Volume 45 Number 1 March 2021

ISSN 0350-5596

Informatica

An International Journal of Computing and Informatics

Editorial Boards

Informatica is a journal primarily covering intelligent systems in the European computer science, informatics and cognitive community; scientific and educational as well as technical, commercial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers accepted by at least two referees outside the author's country. In addition, it contains information about conferences, opinions, critical examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and information industry are presented through commercial publications as well as through independent evaluations.

Editing and refereeing are distributed. Each editor from the Editorial Board can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the list of referees. Each paper bears the name of the editor who appointed the referees. Each editor can propose new members for the Editorial Board or referees. Editors and referees inactive for a longer period can be automatically replaced. Changes in the Editorial Board are confirmed by the Executive Editors.

The coordination necessary is made through the Executive Editors who examine the reviews, sort the accepted articles and maintain appropriate international distribution. The Executive Board is appointed by the Society Informatika. Informatica is partially supported by the Slovenian Ministry of Higher Education, Science and Technology.

Each author is guaranteed to receive the reviews of his article. When accepted, publication in Informatica is guaranteed in less than one year after the Executive Editors receive the corrected version of the article.

Green Computing Approaches - A Survey

Mahdi Dhaini, Mohammad Jaber, Amin Fakhereldine, Sleiman Hamdan and Ramzi A. Haraty
Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon
E-mail: mahdi.dhaini@lau.edu, mohammad.jaber@lau.edu, amin.fakhereldine@lau.edu, sleiman.hamdan@lau.edu, rharaty@lau.edu.lb

Overview Paper

Keywords: green computing, sustainable software engineering, green cloud computing, mobile development, data center, education

Received: October 30, 2019

The upsurge in global warming and release of greenhouse gases are major issues that intensified over the past years due to the increasing usage of technological resources in our daily routines. That is why a call for going green in the technological field is highly recommended. This paper reviews various approaches of green computing in five main areas - software engineering, cloud computing, mobile computing, data centers, and the educational sector.

Povzetek: V tem članku so predstavljeni različni pristopi zelenega računalništva po petih glavnih področjih - programsko inženirstvo, računalništvo v oblaku, mobilno računalništvo, podatkovni centri in izobraževalni sektor.

1 Introduction

Global warming and climate change are causing the increase of global temperature and the rise of sea levels. The main cause of environmental impact is people and their irresponsible and harmful behavior. As an example, for such behaviors are the huge amounts of CO_2 emissions from the industries and vehicles, cutting trees, and the exhaustive use of resources by technology. Studies have shown that the amount of CO_2 emissions has been increasing in the past few years [1]. Efforts for reducing harm to the environment must start from changing peoples' behaviors. Citizens of planet Earth must start thinking "Green" in all aspects of their lives in order to save and protect their future and the future of their children on this planet. Nowadays, technology has entered people's lives intensively - it reached their jobs, homes, and education. As a contribution to achieve environmental sustainability, people can start by changing the way they deal with technology. Most efforts addressed the hardware perspective of green computing with little attention to the importance of the software perspective. Efficient software reduces the use of hardware resources; therefore, reducing energy consumption. In this study, we examine different green computing approaches in the literature in various domains of software development. In particular, we study green approaches in software engineering models, cloud computing, mobile development, data centers. In addition, we highlight the importance of introducing green computing principles in the educational sector.

The remainder of this paper is organized as follows. Section 2 provides a literature review. Section 3 describes the models for sustainable software engineering. Section 4 addresses green cloud computing. Section 5 overviews green mobile development. In section 6 green data centers are addressed, while section 7 sheds the light on green computing in education. And finally, section 8 concludes the paper.

2 Literature review

Many efforts were done in the literature with the aim of achieving green computing in different domains and reducing the negative impacts of ICT on environmental sustainability. A review of the different software process models commonly used in practice is presented in [2]. Naumann et al. [3] presented a reference model for sustainable software engineering (GREENSOFT) that supports different stakeholders in the whole lifecycle of software production. Berkhout and Hertin [4] defined three levels of ICT impacts on the environment and highlighted the importance of studying their rebound effect in which negative impacts overcompensate positive ones. Mahmoud and Ahmad [5] defined green metrics in the stages of software production and stressed the importance of two stages: requirements definition and testing. The model also discusses the role of software itself in achieving green computing. Capra et al. [6] studied the impact of software on sustainability and proved that achieving a better performance does not guarantee better energy efficiency.

Atrey et al. [7] studied how the cost of the unlimited services of cloud computing leads to overcompensating the benefits and increases energy consumption and CO_2 emissions. Dougherty et al. [8] described a model-driven green technique to avoid over-provisioning of idle virtual resources in cloud servers. The aim of their model is to provide a green auto-scaling technique for allocating VM

configurations that preserves a satisfactory QoS. The problem is solved as a feature selection problem. Gai et al. [9] presented an energy-aware mobile cloud computing model that takes advantage of cloudlets to reduce the energy consumption of wireless communications. Xu et al. [10] described an energy-efficient algorithm for VM scheduling inspired by physical principles. Zhao-Hui and Qin-Ming [11] proposed a virtual machine scheduling algorithm that deploys VMs on data nodes with the least growth of energy consumption. Mukhtar et al. [12] presented a green strategy for determining the least energy-consuming fog device to offload client application modules. Verma S. et al. [13] presented an energy-efficient, yet costly, algorithm that integrates load balancing and data replication for fog-cloud computing. Previous researches acknowledge that using virtualization cloud computing is itself energy-efficient technology [14]. A.J. Younge and his colleagues proposed a green cloud framework that covered virtualization and data center operations [15].

Green computing is responsible for designing, manufacturing, using, and disposing of computers, servers, and its hardware like monitors, printers, storage devices, and networking and communications systems in order to efficiently and effectively consume energy with minimal or no impact on the environment [16]. Mobile devices are becoming an important and irreplaceable resource in our daily life. According to the International Telecommunication Union, the number of registers in the worldwide mobile network operators has already reached more than four billion users [17]. Moreover, based on the International Data Corporation's statistics, 494 million smartphones were sold worldwide in 2011. The sales of smartphones reached an annual growth of 62% from 2010 till 2011, expecting this increase to continue furthermore [18]. With this huge number of mobiles and mobile users, and taking into consideration their effect on the environment, regarding their energy consumption and the toxic wastes, one should think of a solution that preserves the environment and prevents harming it. Many approaches have tried to reduce the burden of mobile devices; probably the most common is trying to execute heavy computational operations on the cloud rather than executing them on mobile devices. For example, CloneCloud [19] is a system that allows partial offloading from smartphones to the phone's clone in the cloud. A similar idea was also investigated by Satyanarayanan et al. [20] and Cuervo et al. [21]. Another example from Chen et al. [22] who introduces a framework allowing heavy tasks on an Android phone to be offloaded to an Android virtual machine in the cloud. Others suggest that mobile devices could be the source of computing power.

When talking about going green in an environment that encounters rapid ongoing changes in the technological fields, the need for a business to "Go Green" is much needed to help reduce the hazardous outcomes that humanity is facing. And for an industry to "Go Green" it must simply maintain an eco-friendly and energy-efficient computing resources. Green data centers offer a great aspect of offering an energy-efficient and eco-friendly computing environment. The burst of data centers began in 1946 where data centers were created by the U.S. army to serve the military [23]. And throughout the years the idea of placing data centers has increased and we began to see data centers placed more often in industries. Data centers are defined simply by information and server storage as well as network infrastructure of the company's huge amount of data. And because of this aspect being spread, and adding to that the rapid booming that the world encountered in the field of technology, the consumption of energy became larger and emission of toxics increased as well. That opened the door for data centers to go green and become more and more eco-friendly. A green data center differs from a normal data center where the mechanical, electrical, and computer infrastructure is designed in a way to obtain maximum energy efficiency and minimum environmental damages [24]. Nada and Elgelany [25] mentioned in their article that a data center consumes a huge amount of energy as the same time it plays a major role in producing large amounts of carbon dioxide because data centers are mainly composed of thousands of servers. Uddin et al. [26] mentioned in their study that a data center is composed of thousands of servers and is equal to the amount of a small city.

The topic of green computing in the education sector has been studied intensively in the literature. Many studies were conducted to assess the awareness and knowledge levels of green computing in educational institutions [27]. In [28], a study was conducted to check the level of awareness of green computing among students at the University of Technology in Mauritius. In [29], German software users were surveyed for a study that addressed the environmental issues of software. The integration of sustainability into computing education was studied in [30], where three different strategies were presented. In [31], different techniques for practicing green computing in universities were proposed.

3 Models for sustainable software engineering

Berkhout and Hertin [4] studied the impacts caused by Information and Communication Technologies (ICTs) on the environment. They presented a summary of the literature on the topic and classified the environmental impacts of ICTs into three categories as follows:

- *First-order impacts:* the most obvious environmental impacts that result directly from the production and use of ICT infrastructure and devices such as resource use and pollution, electricity consumption of ICT hardware, and disposal of electronic waste.
- *Second-order impacts:* indirect environmental impacts of using ICT such as resource and energy conservation caused by dematerialization, demobilization and substitution of information goods for tangible goods.
- *Third-order impacts:* indirect environmental impacts of using ICT that appears in the long term such as changing lifestyles and values

systems. These impacts may overcompensate for the energy savings by ICT (rebound effects).

A major issue considered in relation to the third order effects is the ***rebound effect*** of the use of ICT. It has been shown, through the first and second order effects, that ICTs have the potential to reduce resource usages and energy consumption. However, a critical question is whether or not the long-term consumption of ICTs will over-compensate the conserved resources.

This concept is presented as the ***rebound effect***. Naumann et al. [3] presented a reference model for Green and Sustainable software named *GREENSOFT* and gave definitions for *Green and Sustainable Software* and *Green and Sustainable Software Engineering*. The model also performs a refinement for the environmental impacts of ICT, defined by Berkhout and Hertin [4], to cover human and social sustainability issues instead of limiting them to environmental issues. The effects were identified as:

- effects of ICT supply (representing *first-order* effects)
- effects of ICT usage (representing *second-order* effects)
- systemic effects of ICT (representing *third-order* effects)

Naumann et al. [3] claimed that a sustainable software product should have a low impact on *Sustainable Development*, and that the development process of the software product should be environment-friendly. This is reflected in the definitions of *Green and Sustainable Software* and *Green and Sustainable Software Engineering* that was provided:

Green and Sustainable Software: is software that leaves a small footprint on the environment.

Green and Sustainable Software Engineering: is the art of defining and developing software products in a way, which the negative and positive impacts on sustainable development that result and/or are expected to result from the software product over its entire life cycle are continuously assessed, documented, and used for further optimization of the software product.

The *GREENSOFT* model also supports different stakeholders of a software product (developers, administrators, users) in developing, maintaining, and using it in a sustainable manner. The model, as shown in Figure 1, comprises the following four parts:

- *Life Cycle of Software Products:* Based on the three levels of impacts, it directs stakeholders to take into consideration the impacts on *Sustainable Development* during development, distribution, usage, deactivation, and disposal phases
- *Sustainability Criteria and Metrics:* Defines metrics and criteria helpful for the assessment of a software product's sustainability.

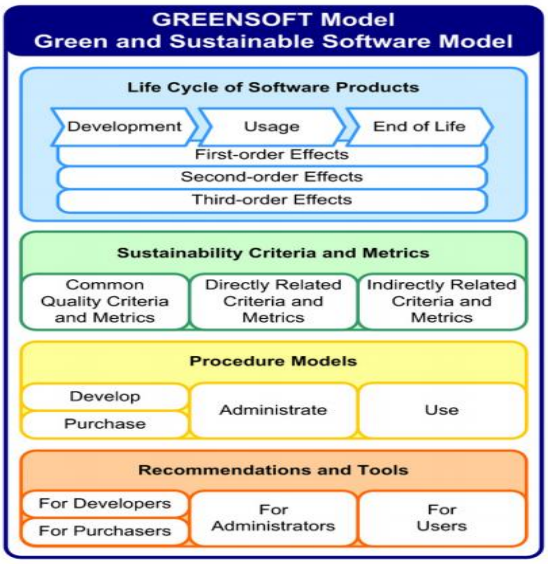


Figure 1: The GREENSOFT model. Reprinted from [3].

- *Procedure Models:* Classifies the software life cycle into four sub-procedure models and proposes a set of activities and processes that are geared towards sustainable development in each model. The four sub-models are: Develop, Purchase, Administrate, and Use.
- *Recommendations and Tools:* This part guides every stakeholder of a software product to comply with green guidelines and procedures while dealing with the software product.

Mahmoud and Ahmad [5] proposed a model for a green software engineering process consisting of two levels. The first level comprises metrics for the green assessment of every stage in the software engineering process according to the ICT *first-order impacts*. The second level of the model addresses the role of software itself in sustainable development and green computing. The model followed the definitions of *Green and Sustainable Software* and *Green and Sustainable Software Engineering* presented by Naumann et al. [3] and their definition of ICT impacts because they consider human and social sustainability issues [5]. They also included an additional definition which is *Green and Sustainable Software Process* because their aim is to provide green instructions for the whole lifecycle of software production.

3.1 First level

Defines a software engineering process to mitigate the negative impacts of ICT on the environment. The process consists of seven stages of the lifecycle of a software product: requirements, design and implementation, testing, green analysis, usage, maintenance, and disposal. The model describes instructions and guidelines that can be used for the green performance of each stage. In the model proposed by Naumann et al. [3], the part of *Life Cycle of Software Products* discusses the impacts of ICT

on sustainable development in the stages of the product's lifecycle. However, they do not include the requirements and testing stages. These two stages were considered in [4] in contrast to many green software models. In addition, a green analysis stage was added to measure the greenness of the output of every stage. We will particularly discuss the importance of including these three additional stages in a green software process:

- *Requirements Stage:* in this stage, the model recommends that requirements engineers will perform requirements functionalities. Therefore, the rate of changes in the software will be reduced, and resources needed for changes will be conserved.
- *Testing Stage:* this stage helps the developing organization to evaluate the compliance of the software product with the customer's requirements. Therefore, it limits the chances of going back to the requirements stage in case the products do not meet the customer's satisfaction. The resources needed for going back to previous stages will thus be conserved.
- *Green Analysis Stage:* the objective of this stage is measuring the "greenness" of the outputs of other stages and allowing for going back to previous stages in order to apply changes that aid green and sustainable development.

3.2 Second level

Another idea missing from the *GREENSOFT* model [3] is the software's role in reducing the negative impacts of ICT and improving sustainable development. The second level of the model presented by Mahmoud and Ahmad [4] describes how software can act as a tool to monitor the efficient use of resources. Several software packages that aid in regulating resource consumption are presented in this level such as operating systems, codes written for energy allocation purposes, and some approaches like *SPAN* [32] that correlate power estimation with source codes, and *GREENTRACKER* [32] that measures energy consumption.

4 Green cloud computing

Cloud computing reduces power consumption by providing cloud applications with virtualized computational resources dynamically upon request such as virtual OS instances. This technique requires keeping idle VM instances in a queue as a standby for any request. As a result, 70-80% of power consumption in data centers is wasted [34-37]. In order to avoid over-provisioning of idle resources, *Auto-scaling* techniques were introduced to improve server utilization of resources and support greener cloud computing by allocating virtualized computational resources, dynamically and accurately, to cloud applications based on their current loads. *Auto* in an *auto-scaling queue* to be provisioned instantly to cloud applications. If no entry in the queue matches the requested configuration, a new VM will be booted and configured then provided to the requesting application. This mechanism is shown in Figure 2. The objective is to maintain the *auto-scaling queue* in a green manner while preserving *QoS*. A preferred *auto-scaling queue* is one that reduces energy and resources consumption by minimizing the number of idle VMs. However, it is very difficult to determine the number of VMs to fill the queue and their configurations. Examples of configuration options provided by Amazon EC2 are Linux vs Windows operating systems, SQL Server vs MySQL databases, Apache HTTP vs IIS/Asp.Net webhosts [8].

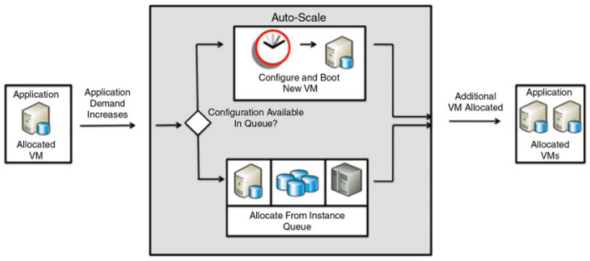


Figure 2: Auto-scaling in a cloud infrastructure. Reprinted from [8].

Dougherty et al. [8] describe a model-driven green technique for sustainable *auto-scaling* cloud computing infrastructures called *Smart Cloud Optimization for Resource Configuration Handling* (SCORCH). The authors mention three main challenges of configuring VMs:

- The need for recognizing the VM configuration options of cloud applications (processors, OS) and their constraints (power consumption).
- The choice of VM configurations to be kept in the auto-scaling queue that can warrant a satisfactory QoS.
- Determining the optimal auto-scaling queue size that minimizes energy consumption.

SCORCH addresses these challenges based on the following functionalities:

- *Feature models* [33] are used to represent VM configurations, implementation details (*e.g.*, whether to use Windows 7 or Redhat 9), and other information about the configurations (*e.g.*, energy consumption, operating costs, etc.).
- Cloud applications are requested to inform *SCORCH* about the VM configurations that it will ask for during its lifetime.
- Feature configuration problems are transformed into constraint satisfaction problems (*CSPs*) and an objective function is defined, to aid in deciding on the appropriate settings of the *auto-scaling queue* taking into consideration several parameters (mainly: expected response time, expected time to fulfill a request, time to boot a new VM instance, and energy consumption of configurations).

- Optimizing the objective function, using a standard constraint solver, yields a combination of VM configurations that minimizes the number of idle VMs to be maintained in the *auto-scaling queue*. This minimizes the operation costs, and respects the satisfactory *QoS* and response time requirements.

Experimental studies were run to assess the contributions of SCORCH in green cloud computing. It was compared to two approaches: the first does not consider auto-scaling, while the second provides auto-scaling without optimizing the queue. The first approach performed worse while SCORCH performed better and reduced cost, power consumption, and CO_2 emissions by 50% [8].

Mobile Cloud Computing (MCC) allows mobile users to have a green experience in using their mobiles in terms of offloading data processing and storage to cloud-based servers. However, the greenness of this approach is linked directly to the stability and efficiency of wireless communications where weak communications cause a waste of energy and resources due to the continuous search for wireless signals. As an approach to solve this problem, Gai et al. [9] introduced a dynamic energy-aware cloudlet-based mobile cloud computing model (*DECM*) that takes advantage of cloudlets to reduce the amount of energy consumed by wireless communications.

The *DECM* mainly consists of mobile devices, cloudlets supported by dynamic searching, and cloud computing as shown in Figure 3. The main objective of the model is to provide green computing on mobile devices without affecting the *QoS* in cloud services. In this model, the nearest cloudlets receive requests from mobiles through the virtual machines corresponding to client applications. A cloudlet may switch a client's connections to another one if it can provide better service based on a more stable network, nearer location, or greener computations. Cloudlets are coupled with dynamic programming algorithms that enable them to find the most convenient cloud servers to connect with. They are named *DCLs* (Dynamic Cloudlets) in this model. This technique guarantees that the mobile devices and the cloud servers will communicate in the most efficient way. The *DECM* algorithm is a minimizing algorithm for the cost of wireless communications in *mobile cloud computing*. The two factors affecting the cost is the energy consumption in a communication route (between mobile, cloudlet, and server), and the service performance provided by a specific route (in order to have a satisfactory *QoS*). This algorithm will ensure that *MCC* will perform with minimum energy consumption. This was proved

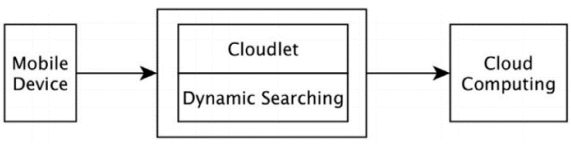


Figure 3: A high-level conceptual model of DECM. Reprinted from [9].

experimentally by comparing the approach with other approaches.

Xu et al. [10] addressed the issue of VM scheduling algorithms that affect the efficient migration of virtual machines between nodes of the cloud. The authors stress that efficient algorithms for scheduling and migration of VMs are crucial because inefficient virtualization of resources will cause unnecessary waste, increase in CO_2 emissions, and overheating inside data centers. In order to solve this problem, Xu et al. [10] presented a VM scheduling algorithm, *VMSAGE*, inspired by the physical gravitational effect. The algorithm is proposed as an improvement of the simple scheduling algorithms that are based generally on placing VMs on nodes with low memory usage (CPU, RAM) without considering other factors such as heat distribution. *VMSAGE* aids in avoiding overheating in data centers, reducing energy consumption, load balancing, and energy efficiency. According to the physical concept of gravitation, the algorithm shuts down data nodes having a low utilization rate and migrates its VMs to other nodes with good heat dissipation in order to avoid overheating. The algorithm also uses another physical concept: "among two objects, having the same acceleration, the one with higher initial velocity will reach its destination faster". Each VM is thus assigned an initial speed of migration in order to enable the system to decide which VMs will migrate before others. VMs with very high utilization rates, having large amounts of resources, or placed in servers with very high temperatures are assigned higher values of initial VMs than others. VMs are migrated to other servers that are selected based on lower costs of migration and heat dissipation. The objective of *VMSAGE* is to reach a point where no server is in need to reschedule its VMs. In order to assess the algorithm, it was compared with two approaches: *Best Fit Heuristic*, and *Dynamic Voltage and Frequency Scaling*. The comparison was based on energy consumption, performance, and heat distribution in the data centers. The experimental results showed that *VMSAGE* reduces energy consumption rates and VM migration times significantly.

The *Cloud of Things* paradigm *(CoT)* was introduced to overcome the problems of limited storage capacities and computational capabilities in *IoT* devices. *CoT* combines cloud services with *IoT* and allows client applications to offload storage or computations to the cloud. However, *CoT* was shown to be inefficient for applications that require high latency, such as applications of healthcare, due to the severity of delays. Consequently, *Fog Computing* was introduced by *Cisco* to support the provisioning of *IoT* applications and services by bringing computations towards the edge of the network. *Fog* can process part of the data collected by *IoT* devices, reserving the cloud capabilities for complex computations and permanent storage [38]. This technique has various benefits such as reduced energy consumption in data centers, and improved latency and network bandwidth. Mukhtar et al. [12] present a green strategy for the allocation of application modules in fog devices. Its objective is to determine the best suitable place for offloading, in the Fog or the Cloud, taking into

consideration energy consumption, CPU capacity, and desired response requirements (or tolerable delays). In the system model, allocated tasks are defined by their type (sensing, processing, etc.) and the workload which defines the needed resources (memory, CPU, energy). This approach was assessed by measuring energy efficiency in a remote patient monitoring system (RPM) and comparing the results with those of two other approaches (default and Cloud-only strategies). The results show that the proposed approach reduces energy consumption by 1.61% and 2.72% with respect to the default and cloud-only strategies, respectively. Energy efficiency in *Fog* devices was improved by 8.27% compared to the default strategy.

5 Green mobile development

In [17] different perspectives to study energy consumption on mobile devices are discussed. The first approach is from the perspective of instructions processed by the Central Processing Unit (CPU). Whenever the amount of code or data the system needs to fetch from the cache increase, the energy consumption will eventually increase as well. Another approach is from the network perspective. For example, using 3G network connections consumes more energy than using 2G network connection [17]. The last approach discussed was from the application perspective. Two factors were addressed in this section: (1) Bluetooth usage and (2) the SMS message size. It was proven that using a mobile with Bluetooth enabled consumes much more energy than using it with Bluetooth turned off as shown in Figure 4. Regarding the SMS message size, sending multiple SMS messages of smaller size will consume more energy than concatenating these messages into fewer SMS messages but of larger size, as shown in Figure 5.

In [18] several actions were recommended to save energy. The first recommendation was for mobile applications to operate in networks that offer the best-cost benefit rate. For applications that are used to send data (e.g. email applications), consider the alternative of delaying the sending of data, so that the maximum number or requests can be triggered at once. Moreover, applications must make use of parallel connections to transfer data, a strategy that would save a lot of energy. In [18] a mobile computing prototype called GMECloud that utilizes energy-efficient mobile devices (e.g., smartphones and tablets) as computing resources is proposed.

Description of Bluetooth State	Consume (mW)
Mobile device with Bluetooth Off	10,4
Mobile device with Bluetooth On	12,52
Mobile device with Bluetooth connected (idle)	62,44
Mobile device performing a search	220,19
Mobile device with Bluetooth receiving data	415,98

Figure 4: Average consumption of Bluetooth technology. Reprinted from [17].

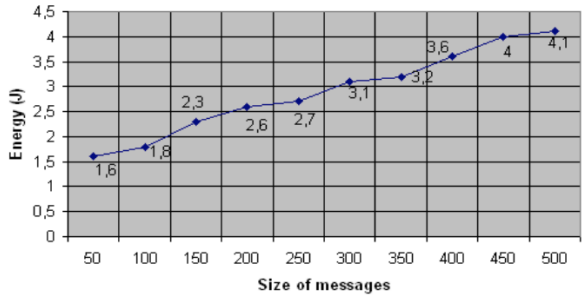


Figure 5: Energy consumption based on the SMS message size. Reprinted from [17].

In the proposed prototype, the clients follow the server-client protocol flow chart described in Figure 6. As shown below, the mobile client's application checks the status of the device: if the device is ready, it connects to the server. The status of the device is defined in terms of different characteristics; for example, the CPU usage, the device battery level, etc.

The server splits the job into smaller tasks; these tasks are then distributed to multiple clients. If the number of active clients is high, the server will assign to each client fewer tasks. This means that the time required by each client to finish the assigned tasks will be less.

In [16], a novel approach is proposed where middleware is coordinating between mobile and cloud computing techniques to achieve green computing for the next generation. The major approaches of Green Computing are product longevity, software and deployment optimization, power management, materials recycling, telecommuting, and low-performance computing. A middleware is a software infrastructure. It binds together the applications, operating systems, network hardware, and network stacks. Its major task in this proposed architecture is to evaluate a data center power, operating system support, power supply, storage, video card, and display. Figure 7 shows the architecture of mobile-cloud computing based middleware for green computing. As the figure shows, a MANET is a set of wireless nodes. The data gathered by these nodes can be delegated to the cloud or middleware of cloud based on the requirements and applications.

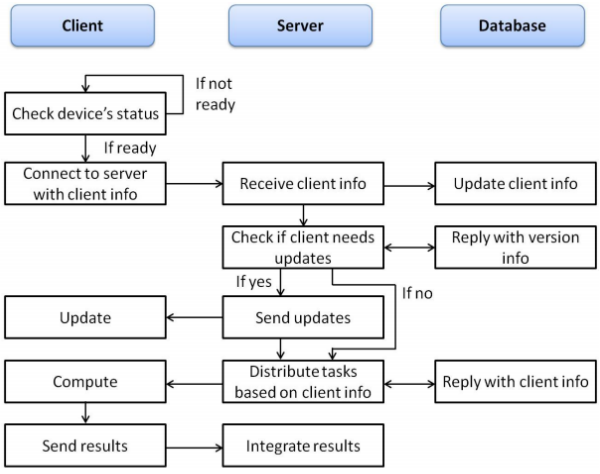


Figure 6: Server-Client Protocol Flow Chart. Reprinted from [18].

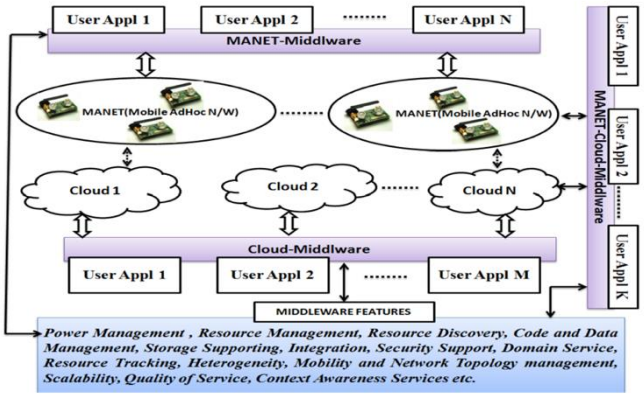


Figure 7: Overview of Mobile-Cloud Computing based Middleware Architecture. Reprinted from [16].

Green cloud computing System Architecture Technologies consists of five core technologies: Scalable Network Architecture, Energy-efficient Cooling, and Power Efficient System, Modular Cloud Computing System, Scalable Virtual Internet Appliance and, Flash Memory Based Cloud Storage System. The purpose of this architecture is to reduce the unnecessary power consumption in a cloud computing environment.

Moving to [39], the main concern is to highlight the energy-related issues as early as possible in the software development life cycle (SDLC) of an application making it more energy-efficient and reducing the cost regarding energy consumption. This work divides the green technology, as mentioned before, on all the stages of the software development life cycle (SDLC) of a given application. It starts with **Green software requirement specification**, which means that there may be additional software requirements to maintain the developed software. Next, **Green software design**; the main concern of the software developers is always the software structure, the modules needed, the software architecture, etc. While energy efficiency is the main part of good software design, it's rarely taken into consideration by software developers at this stage. Moving forward, **Green software implementation** focuses on reducing the application CPU consumption, the number of parameters used, and many other factors that affect energy consumption. Regarding the testing phase, **Green software testing** takes into consideration the number of people and the amount of equipment allocated to test the energy used by the application, based on predefined test cases related to energy consumption. Finally, **Green software maintenance** tends to perform regular maintenance tasks that will keep data transmission at optimal efficiency.

To save energy for mobile-to-mobile communication, an energy-aware adaptive compression approach for battery constrained mobile devices is proposed in [40]. The adaptive approach utilizes a lightweight compression scheme by using the Zip Level 0 compression algorithm whenever compression is required. In the proposed scheme, the signal strength is monitored, and data is read into blocks during the transmission. The signal strength level decides whether blocks will be compressed or not before being sent. Strong signal strength leads to sending the blocks uncompressed, while weak signal strength leads to compressing the blocks using a light compression algorithm. The proposed adaptive scheme was implemented and tested where experimental results show that the adaptive scheme saves energy remarkably compared to the no-compress and always-compress schemes.

6 Green data centers

6.1 How green is a data center

Before taking a risk, an industry must first study the level of energy efficiency, which is the level of energy consumed by the industry. Based on the quantitative results that issued from energy-efficient measures we can decide what suitable techniques we can use in order to turn the data centers to become eco-friendly. In [24], Mata-Toledo and Gupta mentioned two important metrics which are, the Power Usage Effectiveness (PUE) and Data Center Efficiency (DCE). They mentioned too an aid to the following analysis such as tools known as the "The Green Report". However, Siso et al. [41] focused on metrics technique called CoolEmAll that focuses not only on energy consumption but also on heat-aware metrics. They pointed out, that the reason they introduced CoolEmAll metrics is because of the standard metrics such as, CFD, PUE that does not allow any space for predication of energy performance to enhance energy efficiency. CoolEmAll provides analysis tools for data centers' efficiency according to IT equipment. Moreover, Wang, and Khan [42] presented in their study more metrics in order to measure the consumption of energy performance in data centers. The aim of their study is to know how green a data center is through different matrices and measurements and according to that information, possible techniques can be taken for a data center to go green. They pointed out that there are two methods for going green either to involve green requirements into building the infrastructure of the process or to green up the process of a working data center in everyday usage.

6.2 Optimization methodologies

After measuring the efficiency of a data center in an industry, there must be certain measures taken in order to come up with an eco-friendly data center. In their study, Sari and Akkaya [43] mentioned that one of the greatest threats that affect green data centers can be divided into two groups. The first threat consists of the inability to manage the cost crises that are born due to the divergence in determining the efficiency performance technique and calculating the performance of the server. Another threat relates to data centers is the release of carbon dioxide that results from data centers to the atmosphere. The authors presented two methods or techniques for data centers to handle these threats. They first presented a cooling method known as a liquid cooling approach that is put into action. However, its only limitation is that it is geographically dependent, which means it must be located in cold areas so that cold water is formed that will reduce temperature;

and hence, reduce the consumption of energy. Another cooling approach is known as direct cooling that is responsible for reducing energy consumption. It works by implementing cooling coils into the rack to remove waste heat by transferring it into fluids rather than deploying it into the air. In addition, another technique is presented that relates to using energy efficiency servers, which are achieved by using renewable energy resources to power up the data centers.

Another way to produce efficient energy that was provided in this study is either the use of virtualization software through a virtual machine or avoids the rapid change between alternating and direct current. The only limitation that these tests face is the need for a huge amount of tests before implementing the proposed techniques. Furthermore, Ghani, Nikejad, Jeong [23] presented in their paper a series of techniques that enable a data center to go green by saving the consumption of energy. They managed to divide their area of work into four fields by presenting energy-saving techniques for servers, energy-saving techniques for networks, energy-saving techniques for a combined environment of servers and networks, and finally by energy-saving techniques by using renewable energy. As for servers, it is known that a server is the main consumer of energy in a data center. So establishing power saving environment is vital in this area and it is covered through methodologies such as, server virtualization that tends to minimize the number of hardware in use and decrease the amount of functioning servers through making more than one virtual machine on a server. Another technique - known as dynamic power management - handles putting down the computing servers when they are not in use; thus, helping reduce power consumption.

A third technique, known as dynamic voltage scaling, sets the CPU power according to the level of load. Ghani et al. also managed to cover techniques that help to reduce energy consumption in networking fields due to the fact that network infrastructure is the second consumer of energy after servers by utilizing 30% of the energy used for powering data centers. One of the techniques is known as sleep mode that manages to switch off the network resources or putting them to sleep mode whenever they are not in use. Virtual network embedding is a technique that reduces energy consumption by assigning virtual network resources on a small number of physical infrastructures while the idle network resources could be switched off.

Recent research on green computing in cloud data centers includes energy-aware approaches that utilize several techniques to reduce energy consumption [44, 45]. In [44], a hybrid framework was proposed to improve the efficiency of consuming electrical energy in cloud data centers. The proposed energy-efficient framework is based on request scheduling and server consolidation techniques. In this framework, tasks are scheduled after being sorted based on their power and time needs. The framework includes several algorithms: scheduling algorithm to schedule requests, a consolidation algorithm for servers, and a migration algorithm for transferring migrated virtual machines to new servers [44]. In [45], A novel energy-aware technique is introduced to address the green resource management problem in container-based cloud data centers. The proposed approach considers multiple objectives: violation of service level agreements, energy consumption, and the number of both container and VM migrations. The proposed technique is based on a multi-criteria decision-making method where joint container and VM migration are considered in the decision-making process. For that, it utilizes a new joint VM and container consolidation procedure.

7 Green computing in education

Many work and studies have been made in order to improve the green computer awareness and practices in the education sector.

7.1 Awareness on green computing in education sector

The awareness and knowledge of green computing have been addressed by many studies and different surveys where different target groups were surveyed. In [46], Manotas et al. studied the relationship between practitioners and the research community, where they found that practitioners could be more effective in creating efficient applications that consider energy consumption. In 2011, Kogelman conducted a study in which Information and Communication Technology (ICT) managers were surveyed about the use of hardware-focused energy-efficient methods in organizations [47]. In [29], German software users were surveyed for a study that addressed the environmental issues of software. Results showed that although environmental issues seem to be topics of interest to software users, software's environmental issues are not part of these interests [29]. The findings of such studies prove the importance of spreading green computing awareness from the early stages of people's lives (i.e., in the early stages of education). It is important for educational institutions to start investigating the importance of saving the environment at an early stage [29].

Many studies were conducted in the literature to comprehend the level of knowledge and awareness of green computing among university students [28, 48, 31]. In [28], Dookhitran et al. conducted a study to check the level of awareness of green computing among students at the University of Technology in Mauritius. The survey was designed for students of the School of Innovative Technologies and Engineering. The main goal of the study was to determine the levels of awareness and knowledge in green computing and its practices [28]. In [48], Selyamani and Ahmad conducted a study that addressed the student's awareness of green computing issues in higher education institutes focusing mainly on hardware aspects. The survey was undertaken by students from higher education institutes in Malaysia. The study findings indicated that students, mainly non-ICT, lack green computing knowledge. In [31], students and academic staff at Botho College in Botswana were

surveyed in order to check and measure the levels of awareness regarding green computing and the negative

University/Institute	Target group of Students	Focus and Objective	Findings	Year	Source
University of Technology in Mauritius	School of Innovative Technologies and Engineering	Check and analyze the level of awareness of green computing focusing on student's computing-related activities and their computer literacy. The survey questions focused on the hardware aspects.	Students, in the majority, are computer literate. Students lack knowledge of some major green computing practices (e.g. screen savers.)	2012	[28]
Botho College in Botswana	Students and academic employees	Measure and check awareness levels of computer users regarding the negative influence of IT on the environment and with regards to green computing.	The need of green education to reach a green usage technology. Changes behaviors and use of technology and IT can be reached by exalted education	2012	[31]
Higher Education Institute in Malaysia	Higher Education Institute ICT and non–ICT students in Malaysia.	The level of awareness, knowledge and practices of green computing. The study focused on hardware aspects such as the usage of computer and its resources.	Students, mainly non-ICT, lack knowledge about green computing Lack of knowledge among students regarding the benefits of green computing and its practices	2015	[48]

Table 1: An overview of some studies addressing the awareness of green computing issues in higher education institutions.

influences of IT on the environment. The study was also conducted to check if any green computing policies are established in the institution. Interviews with staff of the IT department were also prepared and organized.

The results showed that the level of awareness regarding green computing is low and that no green computing policies are set in the institution. An overview of some studies addressing the awareness of green computing issues in higher education institutions is included in Table 1.

7.2 Approaches to Create and Raise Awareness

The topic of creating and raising awareness in the education sector has been studied in the literature where various ideas have been published. According to [31], creating a website that contains different green computing information, procedures, policies and tips is one solution to create awareness among students in a university. Pang et al. called for extending the aspects of green computing in educational programs [49]. Dookhitram et al. proposed that environmental IT information can be spread by the information channels that are mainly used by the students [28]. Haraty et al. suggested engaging students in educational activities and including awareness campaigns in the educational curriculum [27]. In [50], Suryawanshi presented a number of techniques that raise the awareness of Green ICT such as including an obligatory green ICT program course in all universities, to train learners about the importance of implementing green practices by starting Green Computing Certification course, to present rewards for educational institutions and educators (Green Institute, Green Teacher) that best embrace green practices efficiently as a motivation, effective promotion of green ICT practices and encouraging faculty and students to choose webinars instead of traveling and adopting online education mechanisms in universities that will lead to less carbon footprint [50].

7.3 Green computing techniques to be used by educational institutes

Several studies have been conducted and ideas have been published in the literature about the measures that need to be taken by educational institutes in order to improve the practice of green computing [51, 52, 53]. An overview of some measures is provided in Table 2. Many works and studies have been made in order to improve the green computing awareness and practices in the education sector. However, more green awareness should be raised among students. Educators and educational institutions have a crucial role to play in order to promote and spread green computing awareness among students. Moreover, many measures should be taken and many techniques should be used in order to practice real green computing in educational institutes.

8 Conclusion

Technology has become a major cause of global warming whenever treated inefficiently. A huge urge is needed to save our environment before it's too late. In this paper, we went over different approaches to "GO GREEN". We also

Measure/Technique	Source
Online Learning: Reduce the pollution that results from students and faculty travels by adopting online learning techniques such as video conferencing and web conferencing. **Implementing Green Computing in Administration and in Sharing Information:** Use of online examination systems instead of paper-based exams, using software application to submit student information such as grades and attendance, reduce the use of papers by introducing online applications, forms and petitions, introducing online system for fees payment, use of online brochures; thus, saving papers and conserving power. **Saving electricity:** Educational institutes should consider the huge power consumption that result from the use of computers on its different offices and classes. ENERGY STAR labeled computer equipment should be purchased and should replace energy-inefficient equipment (e.g., LCD monitor instead of CRT one)	[50]
Upgrading Computers: upgrading specific components (CPU, system memory) in the computer in order to prolong the lifecycle computers and improve performance. **Power Saving Modes:** computer power consumption can be managed in an efficient way by using the most "green" and efficient computer power-saving mode. Different modes include sleep mode, hibernate mode, system standby mode, and hard disk sleep mode. The hibernate mode proved to be the most effective among others as it power off the computer completely. **Eliminate Phantom Loads:** by using of power strip devices that powers off in an automatic way powered off devices that are plugged into the strip.	[51]
Virtual Desktop Infrastructure (VDI): exploiting the green benefits of virtualization (operation efficiency, compatibility, ease of management, simplicity of deployment, low carbon emissions, etc.) by implementing VDI in educational institutes. The implementation of VDIs has proved to be power-efficient as it saves power and consumes low energy compared to non-virtual infrastructure.	[52]

Table 2: An overview of some green computing measures to be used by educational institutes.

addressed the software engineering models for green computing, and the four different perspectives for this topic: Green cloud computing, green mobile development, green data centers, and the importance of green computing in the educational sector. Consequently, the usage of green computing by normal people contributes to their harmonic existence in the knowledge society, and this corresponds quite well to the basic objectives of cognitonics [54-57]. To sum up, this paper is intended to be part of the research that has been ongoing to increase the awareness of people towards this topic and presents different approaches that will help whenever applied in software development.

References

[1] New Global CO2 Emissions Numbers Are In. They're Not Good (2018). (n.d.). Retrieved from https://www.wri.org/blog/2018/12/new-global-co2-emissions-numbers-are-they-re-not-good.

[2] Haraty, R. A., Hu, G. (2018). Software process models: a review and analysis. *International Journal of Engineering & Technology*, 7(2.28), pp. 390-397. https://doi.org/10.14419/ijet.v7i2.29.13206

[3] Naumann, S., Dick, M., Kern, E., Johann, T. (2011). The GREENSOFT model: a reference model for green and sustainable software and its engineering, Sustain. *Comp. Inf. Syst.* 1 (4) pp. 294-304. https://doi.org/10.1016/j.suscom.2011.06.004

[4] Berkhout, F. and Hertin, J. (2001). Impacts of Information and Communication Technologies on Environmental Sustainability: Speculations and Evidence, *Report to the OECD*, http://www.oecd.org/dataoecd/4/6/1897156.pdf (accessed 2019-04-03).

[5] Mahmoud, S.S., Ahmad, I. (2013). A green model for sustainable software engineering. *Int. J. Soft. Eng. Appl.* 7(4), pp. 55–74.

[6] Capra, E., Formenti, G., Francalanci, C., Gallazzi, S. (2010). The impact of MIS software on IT energy consumption, in: *18th European Conference on Information Systems*, June 7–9, Pretoria, South Africa, http://web.up.ac.za/ecis/ECIS2010PR/ECIS2010/Content/Papers/0073.R1.pdf (accessed 2010-10-25).

[7] Atrey, A., Jain, N., Iyengar, N. (2013). A study on green cloud computing**,** *Int. J. Grid Distrib. Comp*., pp. 93-102. https://doi.org/10.14257/ijgdc.2013.6.6.08

[8] Dougherty, B.,White J. and Schmidt, C. D. (2012). Model-driven auto-scaling of green cloud computing infrastructure, *Future Generation Computer Systems*, vol. 28, no. 2, pp. 371–378. https://doi.org/10.1016/j.future.2011.05.009

[9] Gai, Keke (2016). Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing, *Journal of Network and Computer Applications*, vol. 59, pp. 46-54. https://doi.org/10.1016/j.jnca.2015.05.016

[10] Xu X., Zhang, Q., Maneas, S., Sotiriadis, S., and Gavan C. (2018)., Simulation modelling practice and theory. VMSAGE: A virtual machine scheduling algorithm based on the gravitational effect for green Cloud computing, *Simul. Model. Pract. Theory*, no. June, pp.1–17, https://doi.org/10.1016/j.simpat.2018.10.006

[11] Zhao-Hui Y., Qin-Ming, J. (2012). Power management of virtualized cloud computing platform, *Chin. J. Comput*. 6 015. https://doi.org/10.3724/sp.j.1016.2012.01262

[12] Mahmoud, M. M. E., Rodrigues, J. J. P. C., Saleem, K., Al-Muhtadi, J., Kumar, N., Korotaev, V. (2018). Towards energy-aware fog-enabled cloud of things for healthcare. *Comput Electr Eng* 2018;67: pp. 58–69. https://doi.org/10.1016/j.compeleceng.2018.02.047

[13] Verma, S., Kumar, Y. A., Motwani, D., Raw, R.S., Singh, H.K. (2016). An efficient data replication and load balancing technique for fog computing environment. In: 3rd international conference on Computing for sustainable global development (INDIACom), New Delhi, India, 16–18 March, pp. 2888–2895.

[14] Salama, A. (2012), Energy-efficient cloud computing application solutions and architectures. http://dx.doi.org/10.18419/opus-2989

[15] Younge, A.J. (2010). Efficient resource management for cloud computing environments. Proceedings of the International Green Computing Conference. https://doi.org/10.1109/greencomp.2010.5598294

[16] Naikodi, C. (2013). Green computing and mobile-cloud-computing inspired middleware for next generation. *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)*. 2.

[17] Siebra, C., Costa, P., Marques, R. Santos, A. and Silva, F. (2011). Towards a green mobile development and certification, *2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Wuhan, 2011, pp. 288-294, doi: 10.1109/WiMOB.2011.6085386. https://doi.org/10.1109/wimob.2011.6085386

[18] Ba, H., Heinzelman, W., Janssen, C., Shi, J. (2013). Mobile computing - A green computing resource. *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC).* https://doi.org/10.1109/wcnc.2013.6555295

[19] Chun. B. -G., Ihm S., Maniatis, P., Naik M., Patti, A. (2011). Clonecloud: elastic execution between mobile device and cloud, in Proceedings of the sixth conference on Computer systems, ser. EuroSys '11. New York, NY, USA: pp. 301–314. [Online]. Available: http://doi.acm.org/10.1145/1966445.1966473.

[20] Satyanarayanan, M., Bahl, P., Caceres R., Davies N. (2009). The case for vm-based cloudlets in mobile computing, *Pervasive Computing*, IEEE, vol. 8, no. 4, pp. 14 –23, oct.-dec. 2009. https://doi.org/10.1109/mprv.2009.82

[21] Cuervo, E., Balasubramanian, A., Cho, D.-k., Wolman A., S. Saroiu, S., Chandra, R., and Bahl P. (2010). Maui: making smartphones last longer with code offload, in *Proceedings of the 8th international conference on Mobile systems, applications, and services, ser. MobiSys '10*. New York, NY, USA: ACM, pp. 49–62. [Online]. Available: http://doi.acm.org/10.1145/1814433.1814441.

[22] Chen, E., Ogata, S. and Horikawa, K. (2012). Offloading Android applications to the cloud without customizing Android, *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, Lugano, 2012, pp. 788-793, https://doi.org/10.1109/percomw.2012.6197619

[23] Ghani, I., Niknejad, N., & Jeong, S. R. (2015). Energy saving in green cloud computing data centers: A review. *Journal of Theoretical & Applied Information Technology*, 74(1).

[24] Mata-Toledo, R., & Gupta, P. (2010). Green data center: how green can we perform, *Journal of Technology Research, Academic and Business Research Institute*, 2(1), pp. 1-8.

[25] Elgelany, A., and Nada, N. (2013). Energy efficiency for data center and cloud computing: A literature review. *Energy*, 3(4).

[26] Uddin, M., Shah, A., Alsaqour, R., & Memon, J. (2013). Measuring efficiency of tier level data centers to implement green energy efficient data centers, *Middle-East Journal of Scientific Research*, 15(2), pp. 200-207.

[27] Haraty, R. and Bitar, G.. (2019). Associating learning technology to sustain the environment through green mobile applications. *Heliyon*. 5. https://doi.org/10.1016/j.heliyon.2019.e01141

[28] Dookhitram, K., Narsoo, J., Sunhaloo, M. S., Sukhoo, A., and Soobron, M. (2012). Green computing: An awareness survey among University of Technology, *Mauritius Students*.

[29] Kern, E. (2018). Green Computing, Green Software, and Its Characteristics: Awareness, Rating, Challenges.In book: From Science to Society, pp.263-273. https://doi.org/10.1007/978-3-319-65687-8_23

[30] Cai, Y. (2010). Integrating sustainability into undergraduate computing education. *In Proceedings of the 41st ACM technical symposium on Computer science education (SIGCSE '10)*. ACM, New York, NY, USA, pp. 524-528. DOI: https://doi.org/10.1145/1734263.1734439.

[31] Batlegang, B. (2012). Green computing: students, campus computing and the environment – a case for Botswana, *Proceedings of the International Seminar on Green Communication, Computing & Management Systems (ISGC2MS-12)*, At Chandigarh, India, Volume: 3. 10.13140/RG.2.1.2227.9528.

[32] Wang, S., Chen, H., and Shi, W. (2011). SPAN: A software power analyzer for multicore computer systems. *Sustainable Computing: Informatics and Systems*, vol. 1, no. 1, pp. 23-34. https://doi.org/10.1016/j.suscom.2010.10.002

[33] Amsel, N., Ibrahim, Z., Malik, A., and Tomlinson, B. (2011). Toward sustainable software engineering: NIER track", *Proceedings of the 2011 IEEE 33rd International Conference on Software Engineering (ICSE)*, pp. 976-979. https://doi.org/10.1145/1985793.1985964

[34] Computer center powernap plan could save 75% of data center energy. (2009). http://www.sciencedaily.com/releases/2009/03/090305164353.htm (accessed 11.04.2019).

[35] Rubin, E., Rao, A., Chen C. (2005). Comparative assessments of fossil fuel power plants with CO2 capture and storage, *in: Proceedings of 7th International Conference on Greenhouse Gas Control Technologies*, vol. 1, pp. 285–294. https://doi.org/10.1016/b978-008044704-9/50029-x

[36] Cassar, C. (2010). Electric power monthly, http://www.eia.doe.gov/cneaf/electricity/epm/epm_sum.html, (accessed 7.07.2020).

[37] Montes, A. F., Cerero, D. F., Abril, L. G., García, J. A. A., Ortega, J. A. (2015). Energy wasting at internet data centers due to fear, *Pattern Recognition*

Letters, Volume 67, Part 1, pp. 59-65,ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2015.06.018.

[38] Shi, Y., Ding, G., Wang, H., Roman H.E., Lu, S. (2015). The fog computing service for healthcare. *In: 2015 2nd international symposium on future information and communication technologies for ubiquitous healthcare* (Ubi-HealthTech), Beijing, China, 28–30 May, pp. 1–5. https://doi.org/10.1109/ubihealthtech.2015.7203325

[39] Kirmani, M. M. (2017). Integrated approach for efficient mobile application development using Cloud Computing and Green SDLC: A Study [PDF]. Srinagar, J&K, India: *Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir*.

[40] Maddah, R., Sharafeddine, S. (2008). Energy-aware adaptive compression scheme for mobile-to-mobile communications, *Proceedings of the 2008 IEEE 10th International Symposium on Spread Spectrum Techniques and Applications*, Bologna, 2008, pp. 688-691, https://doi.org/10.1109/isssta.2008.134

[41] Sisó, L., Salom, J., Jarus, M., Oleksiak, A., & Zilio, T. (2013). Energy and heat-aware metrics for data centers: Metrics analysis in the framework of CoolEmAll project. *Proceedings of the International Conference on Cloud and Green Computing*, pp.428-434. https://doi.org/10.1109/cgc.2013.74

[42] Wang, L., Khan, S. U. (2013). Review of performance metrics for green data centers: A taxonomy study. The Journal of Supercomputing, 63(3), pp. 639-656. https://doi.org/10.1007/s11227-011-0704-3

[43] Sari, A., Akkaya, M. (2015). Security and optimization challenges of green data centers. *International Journal of Communications, Network and System Sciences*, 8(12). https://doi.org/10.4236/ijcns.2015.812044

[44] Alarifi, A. (2020). Energy-efficient hybrid framework for green cloud computing," in *IEEE Access*, vol. 8, pp. 115356-115369, 2020, https://doi.org/10.1109/access.2020.3002184

[45] Gholipour, N., Arianyan, E., Buyya, R. (2020). A novel energy-aware resource management technique using joint VM and container consolidation approach for green computing in cloud data centers, *Simulation Modelling Practice and Theory*, Volume 104, 102127, ISSN 1569-190X, https://doi.org/10.1016/j.simpat.2020.102127

[46] Manotas, I., Bird, C., Zhang, R., Shepherd, D., Jaspan, C., Sadowski, C., Pollock, L., Clause, J. (2016). An empirical study of practitioners' perspectives on green software engineering, Proceedings of the ICSE '16 - 38th International Conference on Software Engineering, May, pp. 237-248. https://doi.org/10.1145/2884781.2884810

[47] Kogelman, C.-A. (2011) CEPIS Green ICT Survey – Examining Green ICT Awareness in Organisations: *Initial Findings. Carol-Ann Kogelman on behalf of the CEPIS Green ICT Task Force*. CEPIS UPGRADE XII(4):6–10.

[48] Selyamani, S., Ahmad, N. (2015) Green Computing: The overview of awareness, practices and responsibility among students in higher education institutes, *Journal of Information Systems Research and Innovation* Zhang C, Hindle A, German DM (2014) The Impact of User Choice on Energy Consumption. Software, IEEE 31(3): pp. 69–75.

[49] Pang, C., Hindle, A., Adams, B., Hassan, A. E. (2015). What Do Programmers Know about Software Energy Consumption? *IEEE Software*. 33. https://doi.org/10.1109/ms.2015.83

[50] Suryawanshi, K. (2018). Green Information and Communication Technology Techniques in Higher Technical Education Institutions for Future Sustainability: i, Volume 2. https://doi.org/10.1007/978-981-13-1274-8_3

[51] Agarwal, S., Basu R. S, Nath, A. (2013). Green Computing and Green Technology based teaching learning and administration in Higher Education Institutions, *International Journal of Advanced Computer Research* (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-3 Number-3 Issue-11 September-2013.

[52] Talebi, M., Way, T. (2009). Methods, metrics and motivation for a green computer science program. ACM *SIGCSE Bulletin*. 41, pp. 362-366. https://doi.org/10.1145/1539024.1508995

[53] Agrawal, S., Biswas, R. and Nath, A. (2014). Virtual desktop infrastructure in higher education institution: Energy efficiency as an application of green computing," *2014 Fourth International Conference on Communication Systems and Network Technologies*, Bhopal, 2014, pp. 601-605, https://doi.org/10.1109/csnt.2014.250

[54] Fomichova, O. S. and Fomichov, V. A. (2009). Cognitonics as an Answer to the Challenge of Time. *In Proceedings of the 12th International Multiconference Information Society - IS 2009, Slovenia, Ljubljana*, 12 – 16 October 2009. The Conference Kognitonika/Cognitonics. Jozef Stefan Institute, Ljubljana, 2009, pp. 431-434.

[55] Fomichov V. A., Fomichova O. S. (2012). A Contribution of Cognitonics to Secure Living in Information Society. Informatica. *An International Journal of Computing and Informatics* (Slovenia). Vol. 36. No. 2.

[56] Fomichov, V.A. and Fomichova, O.S. (2019). The Student-Self Oriented Learning Model as an Effective Paradigm for Education in Knowledge Society. Informatica. An International Journal of Computing and Informatics. Vol. 43. No. 1. P. 95-107. https://doi.org/10.31449/inf.v43i1.2356

[57] Craig P., Roa-Seiler N., Diaz M.M., Lara-Rosano F. (2014). A Cognitonics Approach to Computer Supported Learning in the Mexican State of Oaxaca. Informatica. An International Journal of Computing and Informatics (Slovenia) 38. pp. 241-248.

A Novel Borda Count Based Feature Ranking and Feature Fusion Strategy to Attain Effective Climatic Features for Rice Yield Prediction

Subhadra Mishra
Department of Computer Science and Application, CPGS
Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, India
E-mail: mishra.subhadra@gmail.com

Debahuti Mishra
Department of Computer Science and Engineering,
Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha, India
E-mail: debahutimishra@soa.ac.in

Pradeep Kumar Mallick
School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India
E-mail: pradeep.mallickfcs@kiit.ac.in

Gour Hari Santra
Department of Soil Science and Agricultural Chemistry, IAS
Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, Odisha, India
E-mail: santragh@gmail.com

Sachin Kumar (Corresponding Author)
Department of Computer Science, South Ural State University, Chelyabinsk, Russia
E-mail: sachinagnihotri16@gmail.com

Keywords: rice crop yield prediction, climatic variability, extreme learning machine, feature ranking, feature fusion

Received: July 29, 2020

An attempt has been made in the agricultural field to predict the effect of climatic variability based on rice crop production and climatic features of three coastal regions of Odisha, a state of India. The novelty of this work is Borda Count based fusion strategy on the ranked features obtained from various ranking methodologies. Proposed prediction model works in three phases; in the first phase, three feature ranking approaches such as; Random Forest, Support Vector Regression-Recursive Feature Elimination (SVR-RFE) and F-Test are applied individually on the two datasets of three coastal areas and features are ranked as per the their algorithm. In the second phase; Borda Count as a fusion method has been implemented on those ranked features from the above phase to obtain top five best features. The multiquadratic activation function based Extreme Learning Machine (ELM) has been used to predict the rice crop yield using those ranked features obtained from fusion based raking strategy and the number of varying features are obtained which gives prediction accuracy above 99% in the third phase of experimentation. Finally, the statistical paired T-test has been used to evaluate and validate the significance of proposed fusion based ranking prediction model. This prediction model not only predicts the rice yield per hector but also able to obtain the significant or most affecting features during Rabi and Kharif seasons. From the observations made during experimentation, it has been found that; relative humidity is playing a vital role along with minimum and maximum temperature for rice crop yield during Rabi and Kharif seasons.

Povzetek: Članek opisuje izviren pristop pri iskanju vzorcev vremenske variabilnosti s pomočjo metod za izbiro in združevanjem atributov.

1 Introduction

Agriculture is the major source of livelihood for people in Odisha as well as India, but here it is said that 'Agriculture is the gamble of the monsoon'. Due the climatic changes the production of major yield is reduced in the Kharif. While Kharif rain fall over the country might be increased by 10-15%, but winter rain fall is expected to decrease by 5-25% and seasonal variability would be further compounded [1].

It is highlighted that, due to heavy temperature, including water shortage, distribution of rainy days, maximum loss is expected in Rabi crops and the productivity of Rabi crops is decreased from 10% to 40% by 2100 [2]. Rice yield is expected to decline by 6% for every 10°C rise in

temperature [3]. The scientific and policy personnel have accepted the susceptibility of agriculture crop to climate change and raised question the capability of farmers to adapt because of the direct and strong dependence of crop agriculture on climate [4]. There are different forecasting methodologies available and evaluated by the research workers all over the world in the field of Agriculture. On all India basis, the imitation study developed shows that the yield of rice crop is affected by weather change from 2.5 to 12% [5]. The rice is the main food in eastern India specifically in the states of Odisha, West Bengal, Jharkhand and Bihar. In India green revolution is mainly Wheat as contributed states was mainly Punjab, Haryana and UP. So, Government of India is expecting the 2nd green revolution from eastern India. The amount of data set is very large in Indian agriculture. Earlier, the different model form dataset was done only by manual system, when there was no outset of computer. But with advancement of computer technology, collection of huge data, their classification and storage has been increased. This has established enormous improvement in pattern perception. In this paper, the main focus to develop a user friendly network for farmers which provide the study of rice production on the basis of important climatic parameter.

The current age is the age of data. As we are taking the large dataset for accuracy of the result, so for modeling of the dataset the feature selection technique becomes the prerequisite method [6, 7]. To increase the correctness level of the experiment we have to increase the attributes of the training examples that is the dataset [8, 9, 10]. As the knowledge discovery technique is finding the knowledge from the vast amount of data, so it is dare to do future research for solving the real world troubles. Ranking is a method to find a rank between all the features according to their importance. Selecting a least number of features produce a simple model, this will take less time for computation and can be understood easily. Due to the simpler model fewer resources also required, which can be affordable. Now the question is how we can rank the features or variables [9, 10, 11, 12, 13, 14]. There are so many algorithms in machine learning to find the significant variables. Thus, the concept of feature selection or variable selection arises. It is the selection of the variables or selecting the subset of the variables and this technique does not change the original illustration of the variables.

During the application of the various feature ranking techniques on the dataset, on each iteration small subsets are being generated. For each feature, there is a rank order of the result of each run and then united with the earlier runs to form an ensemble [15, 16]. The Monte Carlo algorithm states that an conclusion can be achieved by the combining random consecutive rough calculation to the same result [17]. This method stimulated the ensemble method.

As agriculture is the backbone bone of Indian economy and rice is the main staple food, so the prediction of rice and the timely advice on variation of climatic condition for the farmers is required. This factor motivate us to prepare a computational model for the farmers and ultimately to the society also. The main aim of this work is to prepare a computational model to find the feature affected most for the rice production. Here we have used three different feature ranking methods such as Random Forest [18, 19, 20, 21, 22, 23], SVR-RFE [24, 25, 26] and F-Test [27, 28] for regression. These are mainly used for ranking of genes in gene expression datasets. The same methods are used here to rank the features of rice crop prediction datasets. Three ranking algorithms gave three different ranks to each feature of the dataset. Then, a feature fusion method has been proposed to evaluate the final rank of each feature and then, these newly ranked features are evaluated by Extreme Learning Machine (ELM) [29, 30, 31, 32] based regressor to measure the importance of each feature. The accuracy of ELM-Regressor has been calculated by decreasing one by one feature from the dataset. Finally, the comparison between proposed fusion based ranking strategy and non fusion based ranking strategy has been made to obtain the number significant features contributing towards the maximum accuracy of regressor. These features decide the importance of climatic parameters in rice crop production both for the Rabi season and Kharif season in the collected districts namely, Balasore, Cuttack and Puri. Thus the important finding of the study is temperature and humidity affect mostly for the crop production in the coastal district of Odisha.

1.1 Study area

In the Figure 1, the rice crop production dataset of three districts such as: Balasore, Puri, Cuttack are shown [33]. The production of rice is mainly in two seasons, such as: Rabi and Kharif. There are different features considered for this production, such as: rainfall, minimum and maximum temperature and relative humidity in the morning and afternoon hour. To avoid the inconsistency in the dataset there are various methods for missing value [36] imputation. In this paper mean value used to solve the missing value problem.

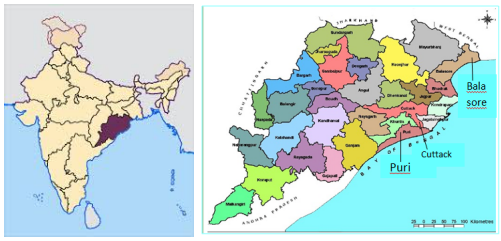


Figure 1: Odisha complete area taken from Google Map an state of India [34]

1.2 Goal

Considering the typical data available in the above mentioned section, the use of data mining or machine learning

strategies should be able to produce a natural decision for crop production based on the important or significant climatic parameters which affects the yield of rice during both the Rabi and Kharif seasons. This paper mainly focuses on the capabilities of ranking and fusion strategies, on two aspects such as; feature ranking and fusion of those ranked features. Specifically, the goal of this study can be outlined as follows:

(a) Collection of climatic data of rice yield for both the Rabi season and the Kharif season of three coastal areas of Odisha, a state of India.

(b) Feature importance evaluation and selection;

  (i) Ranking of features by applying various ranking strategies.

  (ii) Fusion of those ranked features.

(c) Selection important climatic features derived from the ranked and fused features.

(d) Model tuning or searching for appropriate algorithm parameters for better performance.

(e) Model evaluation and validation through performance comparisons and statistical validation.

1.3 Paper layout

The rest of the paper is outlined as follows; the related work in this field is discussed in Section 2. The diagrammatic representation of proposed regressor has been detailed in Section 3. The methodologies such as Random Forest, SVR-RFE, F-Test and ELM regressor and various fusion strategies are discussed in Section 4. The experimentation and model evaluation is discussed in Section 5 and Section 6 discusses the principal findings obtained from this study. Finally Section 7 concludes the paper with future scope of this work.

2 Literature survey

To contextualize the effect of goals set and discussed in Section 1.2 in rice yield modeling, many papers were selected for review which are based on machine learning or data mining techniques be useful for modeling in this serial; (a) ranking of features based on Random Forest, F-Test and SVR-RFE (b) fusion strategies for feature selection and; (c) model evaluation and validation for proper classification. This section explores the various works done on prediction on agricultural field based on random forest, F-Test and SVR-RFE etc. SML Venkata et al. [35] used the dataset consisting of rainfall, precipitation and temperature and applied random forest which is the collection of decision trees, on the two-third of the records and then the resulting decision trees are applied on the remaining records and lastly for the prediction of the crop data, the resultant training sets applied on the test data based on the input attributes. They have used R Studio and they evaluated their results by using other performance measures. Evathia E et al. [18] modified the structure and selection mechanism of the random forest algorithm to improve the prediction performance. Authors have verified all the evaluation measure and basing on the feature selection, clustering etc, they have done the voting procedure. The main objective of their work was the combination of the construction and voting method of random forest algorithm. They found the positive effect on the performance by using 24 datasets. Hari Dahal et al. [36] took six soil variables with crop yield data to find the level of crop productivity. They found some of the soil variables have extremely correlated. So to estimate the potency of the relationship they developed the multiple regression models and applied F-Test to know which variable is most significant and found that total nitrogen, organic matter and phosphorous affect the yield of paddy. J. P. Powell et al. [37] analyses the various weather events on the crop winter wheat taking the data on the farm based and of 334 farms for 12 years. They have used the F-Test to find the significance of weather events in the model. They observed and concluded that, the effect of weather events on yield is time specific and also found that the high temperature and precipitation events significantly decrease yields.

Ke Yan et al. [24] studied both the linear and nonlinear SVM-RFE algorithm. They have analyzed the correlation bias and anticipated a new algorithm such as, SVM-RFE+CBR. They have implemented in the synthetic dataset. Lastly they found the accuracy on their proposed method. Meng-Dar Shieh et al. [25] proposed one method to eliminate the problem of choosing the features subset. Shruti Mishra et al. [26] recommended one extensive deviation of SVM-RFE and SVM-T-RFE. They found the maximum accuracy in case of classification taking the less subset of gene sets and also of high dimensional data. They have also compared with other two methods such as SVM-T-RFE and SVM-RFE and conclude that the projected step by step method is 40% better than SVM-RFE and 25% better than SVM-T-RFE. The ranking strategies adopted by the above mentioned authors have motivated us to carry forward our research on agricultural and climatic datasets.

3 Schematic representation of proposed method

The feature ranking methods are mainly used to rank the features. In this study, a revolutionary effort based on feature ranking methods to find the significant climatic features which affects mostly on the yield of rice of the three coastal districts of Odisha for both the season such as :Rabi and Kharif have been introduced. This empirical study mainly focuses on the selection of significant features through feature ranking and feature fusion based strategies. It works in three important phases, in the first phase known as feature ranking, Random Forest, SVR-FRE and

F-Test based regression methods are explored to rank all the features of the datasets, then in second phase, new ranks have been evaluated by considering all the ranked features from above mentioned ranking techniques and finally, ELM based regressor has been used to empirically evaluate and validate the yield modeling. The Figure 2 illustrates the flow of implementation of proposed ELM based regressor model to obtain the important features that contribute to the yield of rice production in the coastal areas of state of Odisha.

3.1 Data set description

The dataset D is composed of Odisha district of India (Figure 1). Let $d_i \in D \quad \forall i = 1, \cdots, 31$ features that is 31 years of data. where $|d_i| = 25$ features that is represents the attributes of the datasets. Different parameters are, such as $p = \{maxtemperature,\ mintemperature,\ rainfall,\ humidity\}$ that effect the rice production. Since, there are two types of rice production seasons such as; *Rabi* and *Kharif* produced between months 'January–May' and 'June–December', hence p_i is collected over the range of six months each resulting 24 set of attributes and 25^{th} attribute is the production in *hector* of crops for particular year.

The rice production graph for those three coastal areas of Odisha from the year 1983-2014 is shown in **Figure 3(a)** and **Figure 3(b)** for Rabi and Kharif season respectively. The detail description of datasets with standard deviation (Std. Dev.) for three areas is shown in Table 1.

The range and average values of the parameters such as; rainfall in mm/hector, maximum and minimum temperature in °C, mean relative humidity both at 8.30 am and 5.30 pm, of all three datasets with respect to three coastal districts are shown in **Table 2** for Rabi and Kharif seasons.

3.2 Study procedures

This section presents a usable scheme to predict the effect of climatic parameters for rice yield in the coastal areas of a state of India, Odisha, during both the Rabi and the Kharif season. These steps are narrated as follows:

- Collection of the raw data including climatologic characteristics and rice production per hector.
- Calculating the range and average of parameters of those datasets for proper knowledge about the features.
- Defining the attributes affecting the rice yield.
- Redefining the datasets and constructing the database of all tuples according to the selected attributes.
- Dividing the raw data into training and testing datasets.
- Designing the feature ranking models to rank all the features of individual datasets for further processing.
- Designing a feature level fusion model using Borda Count to generate a new set of ranked features by taking the ranked features from all three feature ranking strategies for further analysis.
- Designing an ELM based regressor to classify the datasets with the newly ranked features to measure the importance of each feature.
- The accuracy of ELM regressor has been calculated using by R2 score decreasing one by one feature from the datasets.
- Finally, with respect to maximum accuracy, top 5 ranked features are selected, which decide the importance of climatic parameters in rice crop production both Rabi and Kharif in three different districts.
- Finally, with respect to maximum accuracy, top 5 ranked features are selected, which decide the importance of climatic parameters in rice crop production both Rabi and Kharif in three different districts.

4 Methodologies adopted for experimentation

This section discusses the various methodologies such as random forest; F-Test and SVR-RFE used for feature reduction and ELM for classification are discussed in this section.

4.1 Random forest

Random forest or Random Forest is one of the most important and popular supervised learning algorithm. It can be used both for classification and regression tasks. In this case multiple trees are grown. Then for the classification of a new object based on the attributes, a classification is given by each tree and that is the tree 'votes' for that class. The most votes over all the trees in the forest are chosen for classification and average of outputs by different trees in case of regression. Random forest is one of the ensemble methods of decision trees. Breiman proposed random forest where he adds an extra layer of randomness to bagging [19]. Random forest has a vast number of applications due to its good constancy and simplification [19, 20, 21, 22, 23].

4.2 F-Test for regression

The F-Test for linear regression is one of the methods to know the significance of any variable among the independent variables in a multiple linear regression. How the null hypothesis can be can tested in a multiple regression model with intercept can be described by the F-Test for regression [27, 28].

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0 \tag{1}$$

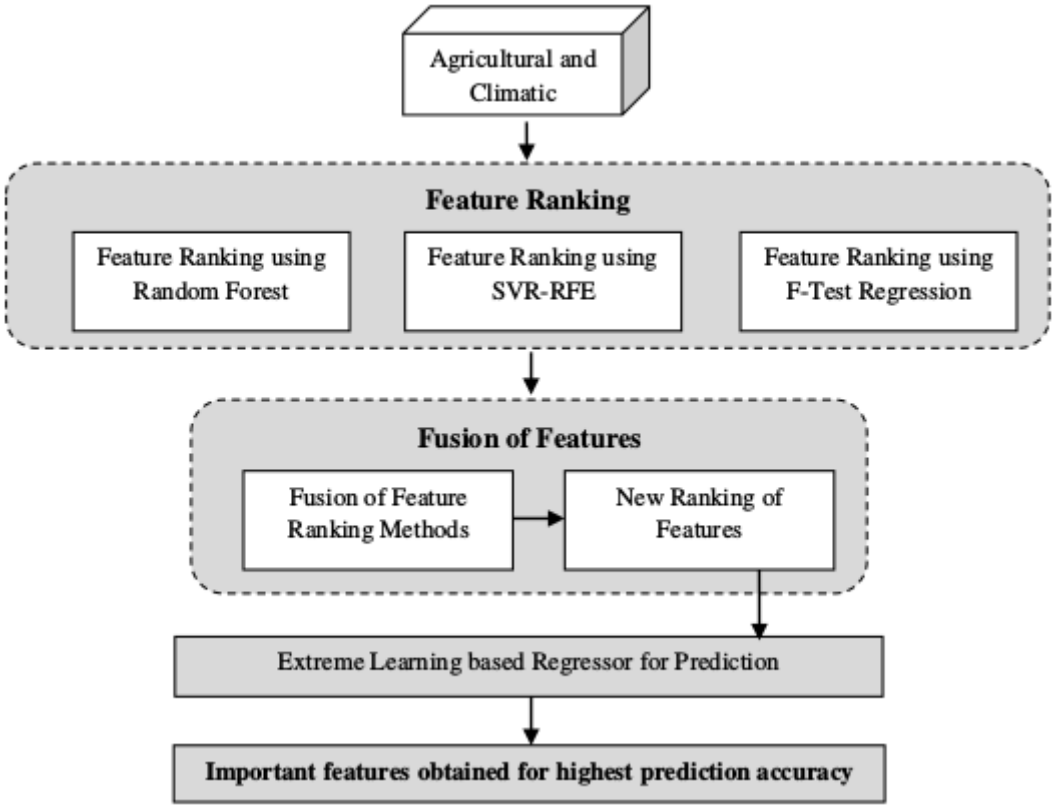


Figure 2: Graphical abstract of proposed model.

Table 1: Description of real datasets collected over period 1983-2014 for Rabi and Kharif production.

Seasons	Rabi			Kharif		
	Dimension	Mean	Std. Dev.	Dimension	Mean	Std. Dev.
Balasore	31×25	47.8386	20.84	31×35	81.6430	43.7791
Cuttack	31×25	44.7391	18.43	31×35	80.6577	50.6339
Puri	31×25	47.6373	25.77	31×35	78.9684	44.2095

$$H_0 : \beta_i \neq 0 \quad \text{for atleast one value of i} \tag{2}$$

Then, assuming the null hypothesis as true we have to test.

$$F = \frac{MSM}{MSE} = \frac{\text{Explained Variance}}{\text{Enexplained Variance}} \tag{3}$$

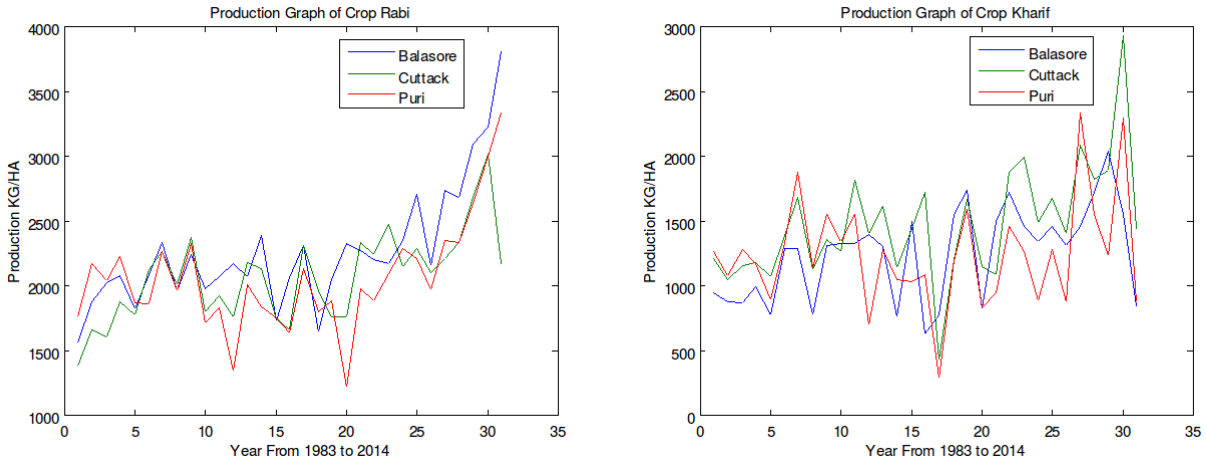
Where, MSM $= \dfrac{\text{SSM}}{\text{DFM}}$ and MSE $= \dfrac{\text{SSE}}{\text{DFE}}$
MSM=Mean Squares for Model
SSM=Corrected Sum of Squares of Models
DFM=Corrected Degrees of Freedom for Models
DFE=Degree of Freedom for Error
Then, using an F-table or statistical software, we have to find confidence interval for degrees of freedom.

4.3 Support vector regressor-recursive feature elimination (SVR-RFE)

SVR-RFE is one of the variable selection or feature selection method. It is an optimization method for finding the best performing feature set. Repeatedly it creates models taking features subset and next with left features and lastly it ranks the features on the basis of order of elimination [24-26]. First the algorithm is trained by SVM with a linear kernel and then the features are detached recursively using the smallest ranking criterion. In order to generate a rank the weight vector needs to be calculated as given in Equation (4).

$$W = \sum_{i=1}^{n} \beta_i x_i y_i \tag{4}$$

Where, i is the number of features ranging from 1to$n$; β_i is the Lagrangian Multiplier estimated from the training set;



(a) (b)

Figure 3: Graphical representation of rice production of three regions for Rabi and Kharif seasons.

Table 2: Range and average values of the parameters in datasets.

Districts	Parameter	Range		Average		Average Rice Production	
		Rabi	Kharif	Rabi	Kharif	Rabi	Kharif
Balasore	Rainfall(mm/hector)	0.0–431.2	0.0–696.5	55.2	280.3	2261.50	1243.8
	Max Temperature (°C)	25-42	13 –37.4	33	32		
	Min Temperature (°C)	9.8 – 32	11.9 - 28	21	25		
	Mean Relative Humidity at 8.30AM (%)	53 - 81	35 - 88	68	79		
	Mean Relative Humidity at 5.30PM (%)	45 - 87	34 - 89	66	78		
Cuttack	Rainfall(mm/hector)	0.0 – 477.8	0 – 752.8	36.42	268.2	2064.71	1472.5
	Max Temperature (°C)	26 – 40	26.8 - 38	31.76	32		
	Min Temperature (°C)	11 - 32	15 - 33	20.92	25		
	Mean Relative Humidity at 8.30AM (%)	58 – 95.5	67 – 95.4	84.33	87		
	Mean Relative Humidity at 5.30PM (%)	29.3 - 89	12 - 90	50.27	73		
Puri	Rainfall (mm/hector)	0.0 – 735.5	0.0 – 826.5	27.13	247	2053	1240
	Max Temperature (°C)	25 – 35.3	20.8 – 40.8	30.43	32		
	Min Temperature (°C)	12 - 29	15.2 - 29	23.49	26		
	Mean Relative Humidity at 8.30AM (%)	70 - 92	66 - 92	80.74	83		
	Mean Relative Humidity at 5.30PM (%)	64 – 90	17 - 91	78.87	81		

x_i is the gene expression vector for sample i and y_i is the class label of $i(y_i \in [-1, +1])$

4.4 Extreme Learning Machine (ELM)

Artificial Neural Network (ANN) is one of the best examples of classification and regression technique which works on back-propagation method. In this case weights are adjusted by trial and error methods. But there are various disadvantages of ANN, such as; *local minima*, *over fitting problem* and *large training time* [38-40]. To overcome the problem of memory requirements, Hung et al. [29] projected new method which is based on the least square algorithm for classification and regression problem, known as ELM. ELM also has unique minimum solution, with both *smallest training error* and *smallest weight norm*, *does not need a stopping methods*.

ELM is a learning neural algorithm, introduced to develop the efficiency of Single Layer Feed Forward Neural Network (SLFN). This section will briefly explain the

Algorithm 1: SVR-RFE [[21, 22, 23]

Input: Initial feature subset, $F = \{1, 2, \cdots, n\}$
Output: Rank list according to smallest weight criterion, R.

1 Set $R = \{\}$
2 Repeat 3 -8 until F is not empty
3 Train the SVM using F.
4 Compute the Weight Vector using (1)
5 Compute the Ranking Criteria, $Rank = W^2$
6 Rank the features as in sorted manner, $New_{Rank} = Sort(Rank)$
7 Update the Feature Rank list $Update \quad R = R + F(New_{Rank})$
8 Eliminate the feature with smallest rank $Update \quad F = F - F(New_{Rank})$

working principle of ELM [30, 31, 32]. N is given as a training sample, where $(X_i, Y_j) \in R^n \times R^m$. Here, $j = 1, 2, \cdots, N$ and the number of hidden nodes is considered as M. Representing the output of SLFN, the equation is formulated in (5).

$$output_k = \sum_{j=1}^{M} \beta_j f(X_k) = \sum_{j=1}^{M} \beta_j f(X_k; a_j, b_j), \quad k = 1, 2, \cdots, N \quad (5)$$

Where, with respect to the input sample, the output vector is $output_k$ and $f(X_k; a_j, b_j)$ is the activation function. a_j and b_j are the randomly generated learning parameter of the k^{th} hidden node and (5) can be compactly written as

$$H \times \beta = Calculated\, Output \quad (6)$$

Here,

$$H = \begin{bmatrix} f(a_1.x_1 + b_1) & \cdots & f(a_M.x_1 + b_M) \\ \vdots & \ddots & \vdots \\ f(a_1.x_N + b_1) & \cdots & f(a_M.x_N + b_M) \end{bmatrix}_{N \times M}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{M \times 1}$$

$$Calculated\, Output = \begin{bmatrix} Output_1^T \\ \vdots \\ Output_N^T \end{bmatrix}_{N \times 1}$$

Where,H is the output matrix, (2) can be linear system by analytically determine the output weights by finding the least square solution, which is defined in (3)

$$\hat{\beta} = inv(H' \times H) \times H' \times trainoutput \quad (7)$$

Where, $trainoutput$ is the output of the training data and the benefit of the ELM is that, the output weight is systematically calculated by using some mathematical transformation, avoiding the lengthy process of training and simultaneously no iterative adjustment of the training parameter is required.

4.5 Fusion strategies

The Borda Count [41, 42] is one of the superior voting system. In this case the voters rank the candidates according to the inclination. Then the points are formed from ranking. The candidates which will gate score one point then ranked last, then score two and next-to-last and so on. Who will secure the more points then declared as winner. There are various other standard voting systems such as: Alternative vote and the single transferable vote, but the advantages of Borda count are, all the MPs have the support of a majority of their votes. The parties nominate the good one. This method is a kind of group consensus functions which maps the inputs of individual rankings to a combined form of ranking which leads to a most appropriate and relevant decision making process. With respect to machine learning, Borda Count is defined as a sum of number of classes ranked below the class by each classifier. The degree of the Borda Count reflects the level of agreement that the input pattern belongs to the considered class. The main advantage of this method is to implement and does not require any training.

4.6 Validation strategies adopted

R^2 is one of the statistical compute to find the fitness of the regression line with the data [43]. Some knowledge regarding the goodness of fit of a model can be defined by this statistic [35, 36]. A linear model explains the proportion of response variable variation and values of R^2 always lie between 0 and 100% or 0 and 1, where; 0% or 0 indicates that the model explains none of the variability of the response data around its mean and 100% or 1 indicates that the model explains all the variability of the response data around its mean and this statistics measure of how well the regression predictions approximate the real data points. An $R^2$of 100% or 1 indicates that the regression predictions perfectly fit the data.

5 Experimentation and model evaluation

5.1 Experimental setup

In this work all the implementations have been carried out using python programming environment in Linux operating system with a minimum hardware configuration of 4GB RAM and 100GB hard disk. First of all, the different activation functions are tested for best suitability to our prob-

lem domain. Then, different feature ranking strategies have been tested with ELM. Finally, the proposed fusion of feature ranking has been tested. The parameters used for experimentation is illustrated in Table 3.

5.2 Parameters used

The **Table 3** gives the details of the parameters used for the implementation.

5.3 Feature ranking methods

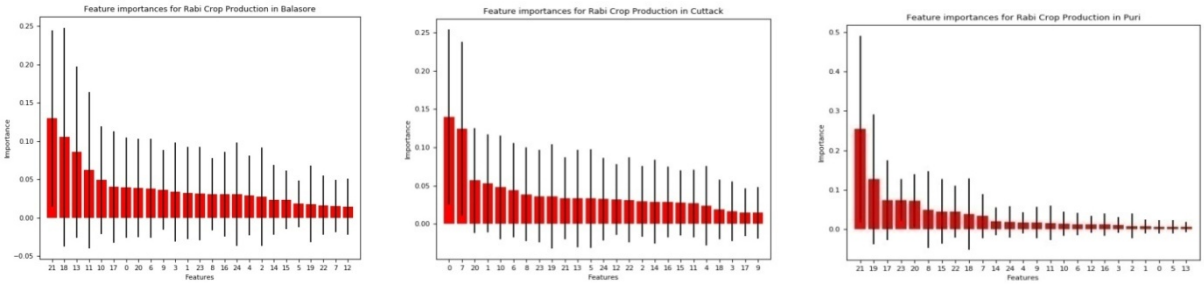
Here three different feature ranking methods such as Random Forest, SVR-RFE and F-Test have been experimented for regression. In literature, it has been found that, these are mainly used for ranking of genes in gene expression datasets and in this study; the same methods are used to rank the features of rice crop prediction datasets. This methodology works in three different steps such as; (a) first, the three ranking algorithms outputs three *different ranks* to each feature of the dataset; (b) secondly, a *feature fusion* method based on *Borda Count* has been used to evaluate the final rank of each feature and; (c) finally, these *newly ranked features are evaluated* by ELM based regressor to measure the *importance of each feature*. The accuracy of ELM regressor has been calculated by decreasing one by one feature from the datasets. Finally, with respect to maximum accuracy, top five ranked features are selected, which decide the importance of climatic parameters in rice crop production both for the Rabi season and the Kharif season in all the districts taken for the analysis. **Figure 4** and **Figure 5** shows the features are arranged in the descending of their R^2 scores measuring the importance of the features after applying the Random Forest feature ranking method on both Rabi and Kharif seasons respectively for Balasore, Cuttack and Puri districts. From **Figure 4** for Rabi season it can be observed that, the features 21, 18, 13 and 11 are having approximate importance scores from 0 to 13, whereas features 7 and 12 are having very less importance scores and rest are in a moderate stage for Balasore district, for Cuttack district, features 0 (first feature) and 7 are having approximate importance scores from 0 to 14, whereas, features 3, 17 and 9 are having very less importance score. Similarly, for Puri district feature 21 has very high importance and 19, 17, 23, 20, 8, 15, 22, 18 and 7 are having moderate scores. Rest others can be ignored due to their very less scores of importance.

Similarly, for Kharif season, from **Figure 5** it can be seen that, the feature 5 is showing highest importance score of 8 and the feature 5 is having the lowest score of importance and rest are lying within the range of 2-6 scores for Balasore district. For Cuttack district, features 1, 24, 8, 9, 23, 14 and 7 are having approximate importance scores from 0 to 7, rest other features are having very less importance scores. Similarly, for Puri district features 8 and are having very high importance with the scores 0 to 16, and 4, 10 and 9 are having moderate scores. Rest others can be ignored due to their very less scores of importance. **Figure 6** and **Figure 7** shows the features with respect to their R^2 scores measuring the importance of the features after applying the SVR-RFE feature ranking method on both Rabi and Kharif seasons respectively for Balasore, Cuttack and Puri districts. From **Figure 6** for Rabi season it can be observed that, the feature 23 is having the 1st rank, then features 15, 9, 21 and 14 are showing better rank and few more are showing moderate rank and feature 4 is having the lowest rank giving rise to non-significant feature. The feature 7 is having the highest rank, and feature 17 is with lowest rank in Cuttack district. Similarly, the feature 19 has very high rank and features 17, 11, 15, 23 are having better rank and feature 4 has less importance in Puri district. Similarly, in **Figure 7**, the feature 27 is experiencing the highest rank, feature 25 and 9 is next to best and feature 0 (first feature) is having less rank with less impact of the feature in Balasore district. For Cuttack district feature 16 is of great importance and feature 34 is of no or less importance, therefore can be ignored. Feature 29 is showing the highest rank and 23, 9, 8 and 20 features are also experiencing better scores, but feature 33 is with the lowest rank in Puri district. The importance of features for both Rabi and Kharif seasons using F-Test for regression has been plotted in **Figure 8** and **Figure 9** respectively. From the experimentation of Rabi season (**Figure 8**), it can be seen that, for Balasore district feature 21 is with the highest score, features 22,24,8,7,5,0 are with lowest scores,4,13,19 are negligible score and rest others are having moderate scores. For Cuttack district feature 6 is with the highest score, 1, 5, 8, 9, 13, 17 and 18 are of no importance and they do not contribute for processing. Similarly for Puri district feature 17 and 21 are having the highest scores, features 1,4,5,13 and 23 are with lowest scores and also it can be seen that rest other features are also not showing better scores. From **Figure 8** for Kharif season, the features 16, 1 and 8 are having the highest importance for Balasore, Cuttack and Puri districts respectively. Features 7, 8, 12, and 31 for Balasore, 2, 5, 6, 10 and 12 for Cuttack and 6, 15, 20, 24, 28 and 31 for Puri datasets are showing scores of least importance.

5.4 Fusion of feature ranking methods

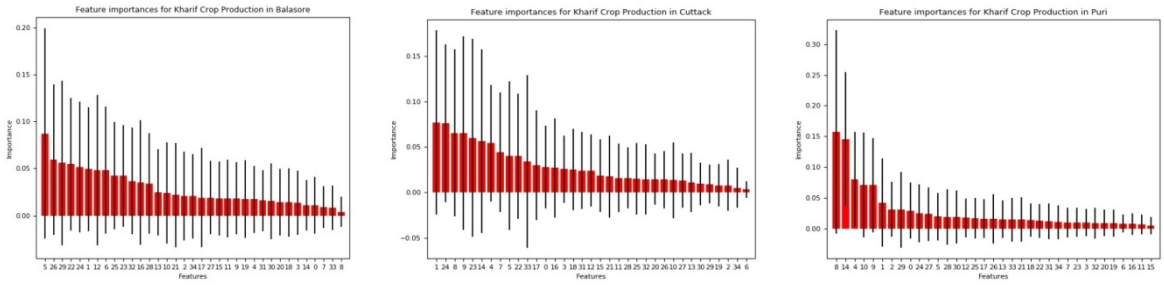
Here, a multiple ranking fusion scheme has been proposed. In this scheme, the individual rankings using different ranking methods have been obtained and then those ranked features are combined to obtain the final rankings of features. The most popular and effective method for fusion used here is Borda count method.

Mathematically, the fusion of features based strategy can be proposed as; let the dataset is defined as $DS = \{x_1, x_2, x_3, \cdots, x_n\}$, where x_1, x_2, x_3, $\cdots$, x_n represents n number of features of the dataset and r_1, r_2 and r_3 are three ranking methods used and the proposed fusion of ranking strategy can be described as shown in **Figure 10**. The importance of features for both Rabi and Kharif seasons using fusion of ranking strategy for regression has



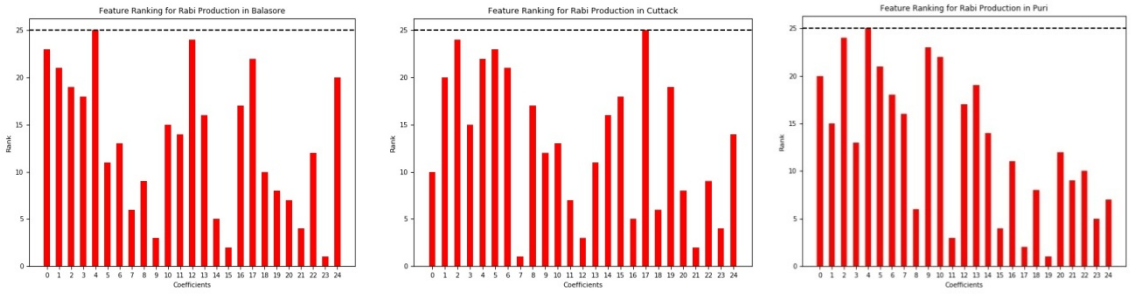
(a) Balesore (b) Cuttack (c) Puri

Figure 4: Feature ranking using Random Forest for Rabi season in three districts.



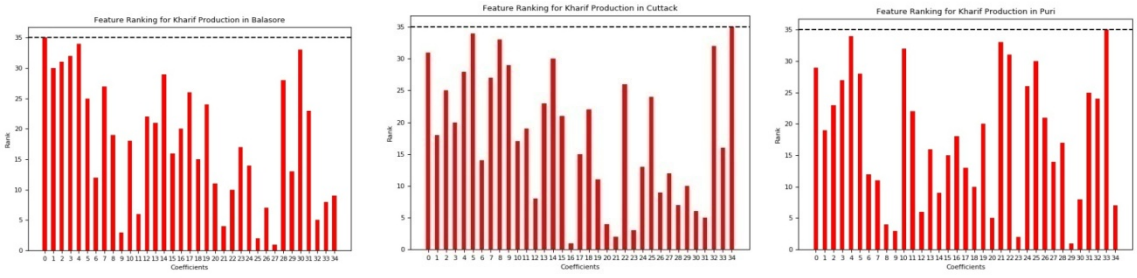
(a) Balesore (b) Cuttack (c) Puri

Figure 5: Feature ranking using Random Forest for Kharif season in three districts.



(a) Balesore (b) Cuttack (c) Puri

Figure 6: Feature ranking using SVR-RFE for Rabi season in three districts.

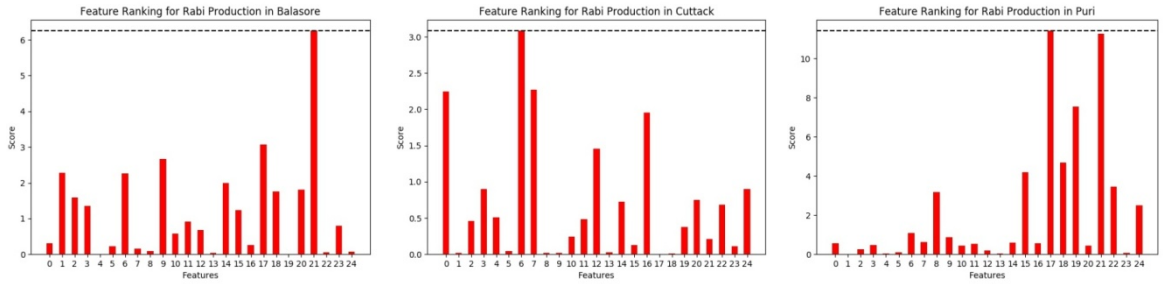


(a) Balesore (b) Cuttack (c) Puri

Figure 7: Feature ranking using SVR-RFE for Kharif season in three districts.

Table 3: Parameter set up for ranking methods.

Techniques	Parameters
Random Forest for feature ranking	No of estimators=1000, criterion=mean square error
SVR-RFE for feature ranking	C=1.0 (Penalty parameter), Base estimator=SVR, kernel=linear, no of features to select=1, step=1
F-Test for feature ranking	Score_function=Ftest, no of features=1
Extreme Learning Machine	No. of hidden layers - 500, Activation function - Multiquadric



(a) Balesore (b) Cuttack (c) Puri

Figure 8: Feature ranking using F-Test for Regression for Rabi season in three districts.

been plotted in **Figure 11** and **Figure 12** respectively and the five top ranked features obtained are listed in **Table 4**.

5.5 Extreme learning machine regressor

In this work, first, all the variants of ELM regressors have been evaluated with different activation functions such as; tanh, sine, tribas, inv-tribas, sigmoid, hardlim, softlim, gaussian, multiquadric, inv-multiquadric etc. Among these functions it has been observed that, tribas, inv-tribas, hardlim, softlim and Gaussian functions gives a negative value of R2 score and score of tanh, sine, sigmoid, multiquadric and inv-multiquadri functions are found to be $\geq$ 98% as detailed in Figure 13 and Figure 14 and also Table 5 and Table 6, shows the graph for R2 score for different activation functions for ELM to predict Rabi and Kharif rice crops respectively. From all those ten activation functions multiquadric is having the highest R2 score while considering all the districts for Rabi and Kharif seasons. Hence, for the experimentation, mutiquadric function has been considered.

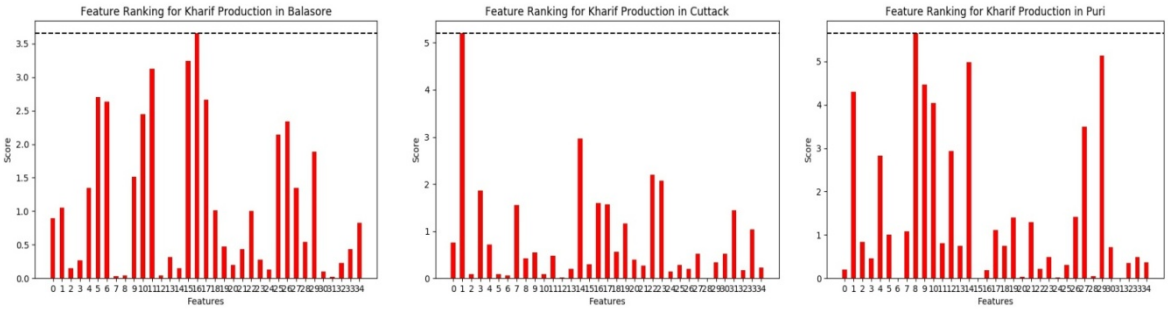
5.6 ELM-Regressor for varying number of features

Once, the newly ranked features are obtained from proposed feature fusion strategy and the activation function (multiquadratic) have been also found to be used by ELM, now the accuracy of ELM Regressor has been calculated by decreasing one by one feature from the datasets as shown in Figure 15 and Figure 16.

Table 7 and Table 8 depicts the accuracy of prediction obtained by multiquadratic based ELM regressor for Rabi and Kharif seasons respectively for all three coastal regions by decreasing the features one by one. The maximum number features those shows above 99% accuracy are coded in red, green and blue colors for Balasore, Cuttack and Puri districts respectively for proper visualization of the readers. From Table 7, it is evident that, while decreasing the number of features from 25 to 20, 15, 14, 13, 12, 8, 6 and even 3 shows above 99% prediction accuracy for Balasore, for Cuttack the number features showing 99% prediction accuracy are 20, 10, 9 and, similarly, for Puri, 18, 15, 11, 10, 6, 3 and 2 number of features are giving maximum prediction accuracy above 99%.
Similarly, from Table 8, it can be observed that, while decreasing the number of features from 35 to 34, 33, 30, 26, 22, 23, 20, 18, 17, 16 and 15 shows above 99% prediction accuracy for Balasore, for Cuttack only 18, 11, 6, 4, 3, 2, and 1 number features are below 99% prediction accuracy and rest are giving above 99%, and, similarly, for Puri, 33, 30, 27, 26, 24, 23, 18, 16, 11, 10, 9, 8, 7, 4, 3, 2 and 1 number of features are giving below 99% prediction accuracy. From those two table and figures this, it can be accomplished that, to predict the crop yield for Rabi season less number of features are working better in comparison to Kharif seasons.

5.7 Result analysis

After obtaining the top five ranked features and the varying number features which give above 99% prediction ac-



(a) Balesore (b) Cuttack (c) Puri

Figure 9: Feature ranking using F-Test for Regression for Kharif season in three districts.

Table 4: Five top ranked features extracted using feature ranking based on Borda Count feature fusion strategy of three districts of Rabi and Kharif season.

Seasons	Balesore		Cuttack		Puri	
	Feature No	Feature Name	Feature No	Feature Name	Feature No	Feature Name
Rabi	23	May RH-8:30 AM	12	Mar Min Temp	22	Apr RH-5:30 PM
	15	Jan RH-8:30 AM	21	Apr RH-8:30 AM	24	May RH-5:30 PM
	9	May Max Temp	7	Mar Max Temp	11	Feb Min Temp
	11	Feb Min Temp	23	May RH- 8:30 AM	15	Jan RH 8:30 AM
	14	May Min Temp	16	Jan RH-5:30 PM	23	May RH-8:30 AM
Kharif	27	Sep RH-8:30 AM	31	Nov RH-8:30 AM	12	Nov Max Temp
	25	Aug RH-8:30 AM	21	June RH-8:30 AM	19	Nov Min Temp
	9	Aug Max Temp	23	July RH-8:30 AM	9	Aug Max Temp
	21	June RH-8:30 AM	20	Dec Min Temp	25	Aug RH-8:30 AM
	26	Aug RH-5.30 AM	16	Aug Min Temp	20	Dec Min Temp

Table 5: R^2 score of all activation functions of ELM for Rabi seasons.

ELM Activation Functions	R^2 score for Rabi Season		
	Balesore	Cuttack	Puri
TANH	0.998093743193	0.994257107884	0.9989356101
SINE	0.996717695318	0.99942453749	0.983896079504
SIGMOID	0.987233114958	0.998330563403	0.999426924698
MULTIQUADRIC	**0.999957522834**	**0.999818219303**	**0.999726152755**
INV-MULTIQUADRIC	0.958613787069	0.935259681129	0.966708028068
TRIBAS	-12.4064777143	-11.4144046567	-9.03257377773
INV-TRIBAS	0.0	-2.22044604925e-16	-2.22044604925e-16
HARDLIM	0.0	-2.22044604925e-16	-2.22044604925e-16
SOFTLIM	0.0	-2.22044604925e-16	-2.22044604925e-16
GAUSSIAN	-1.13177301381	-0.59754939441	-0.0727264050254

curacy for both the seasons, in this section an attempt has been made to validate proposed fusion of feature ranking based strategy with Random Forest, SVR-RFE and F-Test with multiquadratic based ELM to find the impact of fusion based strategy with non-fusion based ranking strategies for the maximum number features that contribute to achieve

Table 6: R^2 score of all activation functions of ELM for Kharif seasons.

ELM Activation Functions	R^2 score for Kharif Season		
	Balesore	Cuttack	Puri
TANH	0.999802124998	0.900092462207	0.941367554859
SINE	0.981838265602	0.983905092261	0.854629459493
SIGMOID	0.993512967504	0.936873558516	0.964438947941
MULTIQUADRIC	**0.999993905565**	**0.999624110222**	**0.991070886794**
INV-MULTIQUADRIC	0.979667648088	0.999512861069	0.960984673885
TRIBAS	-21.7579913615	-37.3921891299	-8.74341923183
INV-TRIBAS	0.103557054691	-4.4408920985e-16	0.0669379515163
HARDLIM	0.0	-4.4408920985e-16	-8.881784197e-16
SOFTLIM	0.103557054691	-4.4408920985e-16	0.0669379515163
GAUSSIAN	-0.308828188795	-1.77193827293	0.549504155107

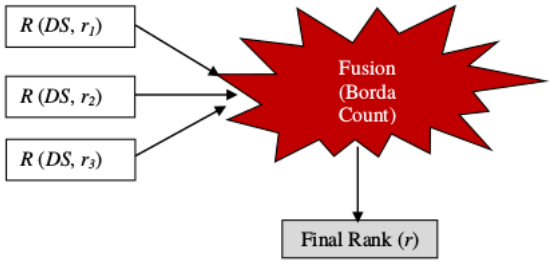


Figure 10: Fusion of feature ranking strategy.

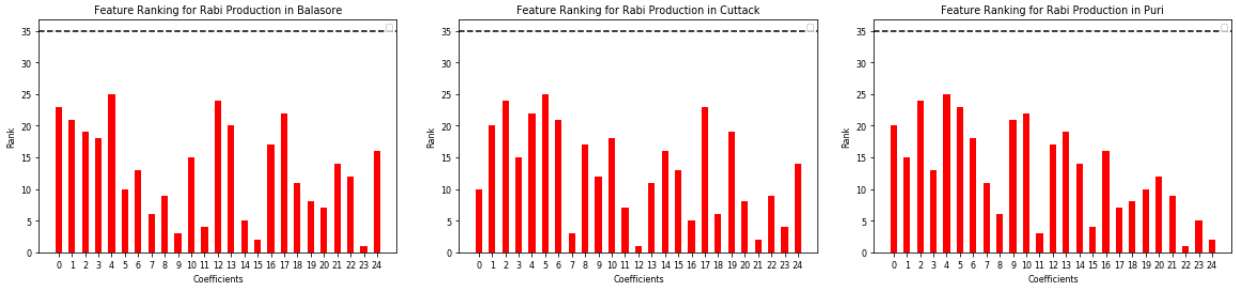
99% prediction accuracy as shown in Table 9 and Table 10 for Rabi and Kharif season crops.

For Rabi season crop from **Table 9**, it can be seen that, proposed fusion based ranking strategy when compared with non fusion based strategies, the maximum number of features that contribute predictive accuracy above 99% for ELM with Random Forest is 7, 10, 6; ELM with SVM-RFE is 5, 9, 4 and similarly ELM with F-Test needs 9, 11, 8 numbers of features to give 99% and above predictive accuracy. While with a very less number of features such as; 3, 5 and 2 can predict above 99% accuracy for Balasore, Cuttack and Puri districts respectively. From Table 4, where the top five ranked features extracted from fusion strategy, it can be concluded that he crop yield for Balasore district in Rabi season can be accurately predicted if we consider only three features out of RH at 5.30 PM of March, April, May, RH of February 8.30 AM and 5.30 PM, because they are affecting the rice crop yield maximum. The five features that affect the rice yield during Rabi season for Cuttack district are; RH of March, April and May and also the minimum and maximum temperature of May month. Similarly, the two features that affect the crop yield of Puri district during Rabi season are out of five features such as; RH of March and May months and minimum temperature of March and May months. From this observation, it can be said that the features containing RH in 8.30 AM and 5.30PM are the mostly affecting rice crop yield in all the three districts for the Rabi season crop.

Table 7: Performance of ELM with varying number of features for Rabi crop prediction.

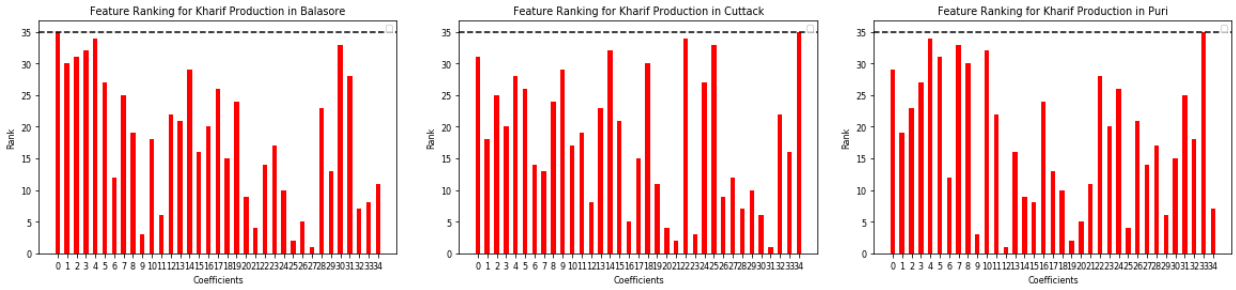
No. of Features	Balesore	Cuttack	Puri
25	0.9721452943	0.983009018	0.8918569778
24	0.8820921735	0.9211651663	0.899522749
23	0.9723284733	0.9717225517	0.8644802695
22	0.9844701984	0.9800205232	0.9897965576
21	0.9668404406	0.9551503977	0.9665234947
20	0.9996177026	0.9999348622	0.9869710443
19	0.9592805399	0.9127841794	0.9398241394
18	0.8356003816	0.9081500374	0.9942342785
17	0.9511241577	0.9780000307	0.9288099884
16	0.9354752886	0.9358388115	0.9363188192
15	0.9930751632	0.9172893274	0.9928662122
14	0.9901838183	0.9617978239	0.9512236619
13	0.9999946834	0.9896162607	0.9585303661
12	0.9934721511	0.9027066465	0.9372090401
11	0.9594424161	0.9510466357	0.9919344792
10	0.8894488099	0.9943953688	0.992055051
9	0.9765177632	0.999323231	0.971105448
8	0.9990643405	0.9784069021	0.9623709905
7	0.9735100076	0.9850878134	0.9978247397
6	0.9968633499	0.9728457206	0.9757838604
5	0.9135013909	0.9969514165	0.9785948706
4	0.9815795296	0.836152001	0.9720037388
3	0.9992149872	0.9126616196	0.998992344
2	0.9183391785	0.9892897973	0.9945608993
1	0.3091946087	0.7773622128	0.5590549638

Similarly, while analyzing the Table 10 for Kharif season for all the district datasets, the observation says, Kharif season crops needs more parameters or features to be considered in comparison to Rabi season crops which is evident from Table 8 and Table 10. The top 15, 5 and 5 ranked features are need to accurately predict the rice yield during this season for Balasore, Cuttack and Puri districts respectively. Observing from Table 4, it can be established that, for Balasore district 15 numbers of features are affecting



(a) (b) (c)

Figure 11: Feature ranking based on Borda Count based feature fusion strategy for Rabi season in three districts.



(a) (b) (c)

Figure 12: Feature ranking based on Borda Count based feature fusion strategy for Kharif season in three districts.

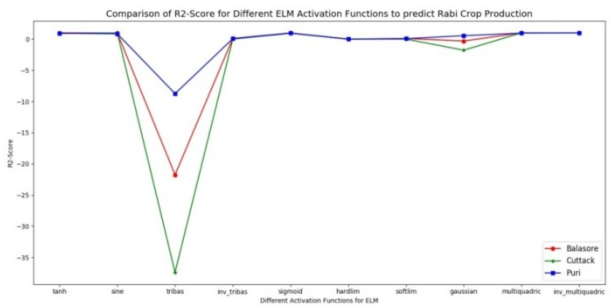


Figure 13: Performance comparison of different activation functions for ELM for Rabi Crop prediction in three different districts.

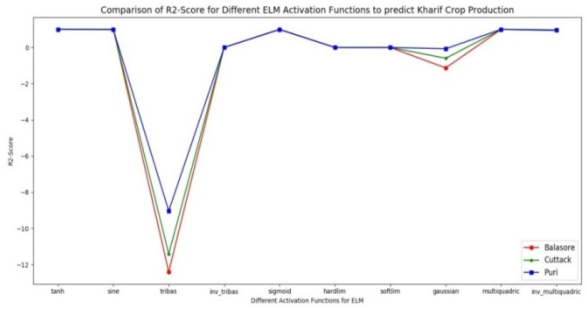


Figure 14: Performance comparison of different activation functions for ELM for Kharif crop prediction in three different districts.

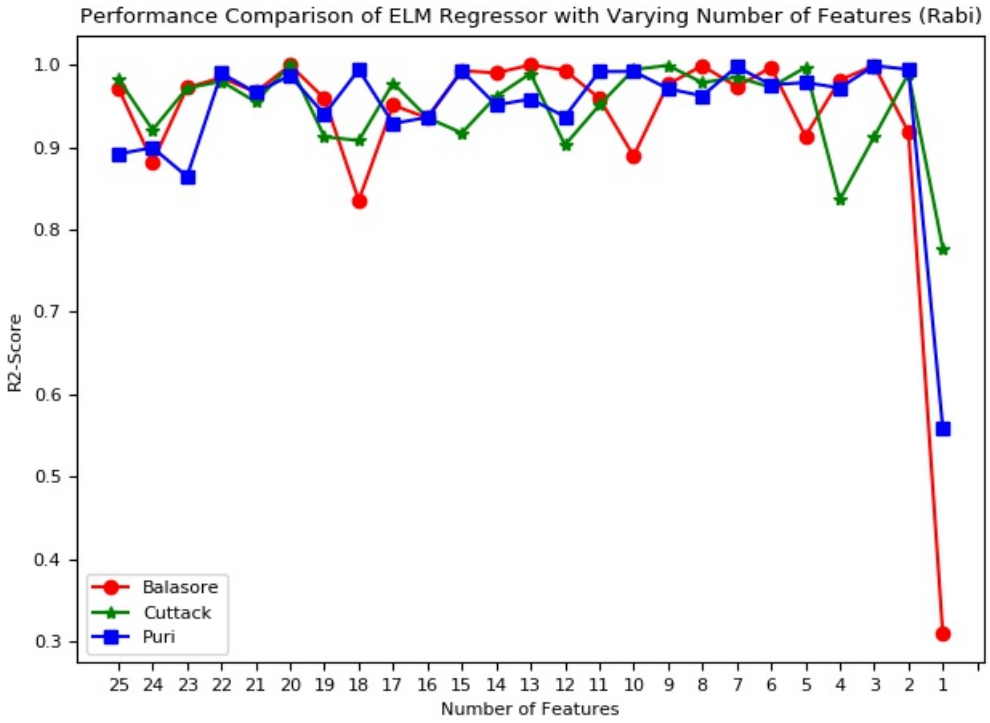


Figure 15: Performance comparison of ELM based Regressor for rice crop prediction (Rabi season) with varying number of features.

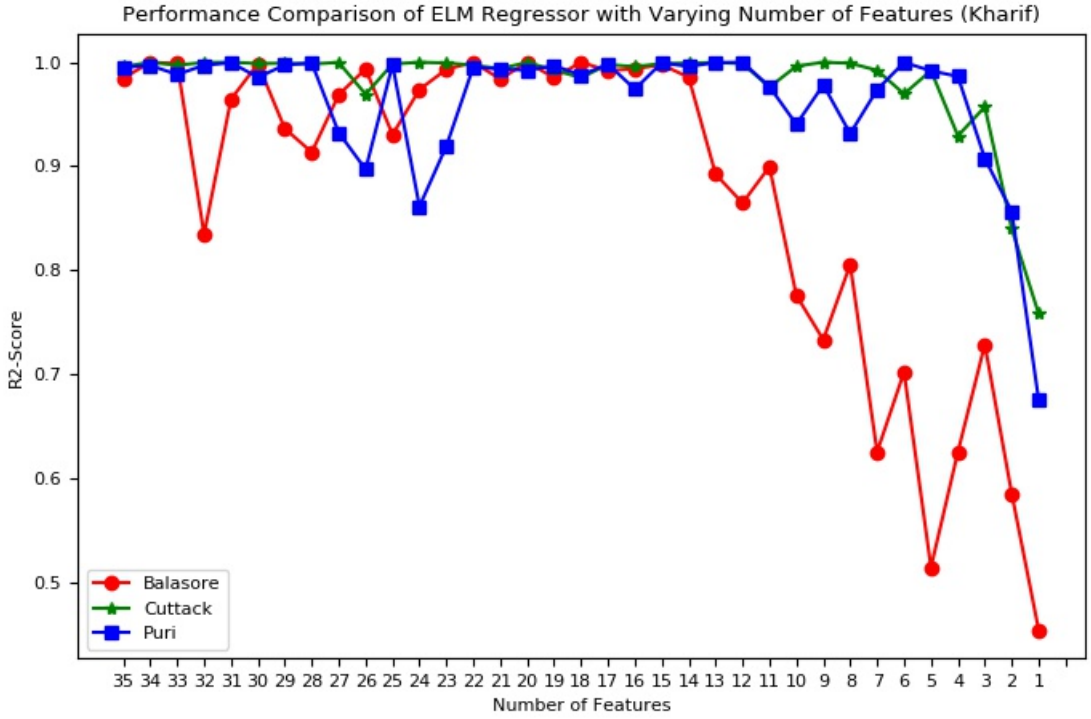


Figure 16: Performance comparison of ELM based Regressor for rice crop prediction (Kharif season) with varying number of features.

Table 8: Performance of ELM with varying number of features for Kharif crop prediction.

No. of Features	Balesore	Cuttack	Puri
35	0.9846787896	0.9957370161	0.9951327585
34	0.9998355712	0.9999942814	0.9968069715
33	0.9988511032	0.9971118552	0.9882099882
32	0.834652697	0.9997670449	0.9970224703
31	0.9644203951	0.9999974004	0.9995567779
30	0.9985549641	0.998758502	0.9855062635
29	0.9361953426	0.9990206928	0.9973169474
28	0.9134833525	0.9982784062	0.9993592848
27	0.9687297599	0.9997739351	0.9323478922
26	0.9935929806	0.9691971505	0.8970090564
25	0.9303826554	0.9983009945	0.9976403045
24	0.9738004793	0.9999393614	0.8613974874
23	0.9931366576	0.9993454567	0.9194678778
22	0.9998339948	0.9970860103	0.9953840259
21	0.9838072021	0.9938437106	0.9934978926
20	0.9998996937	0.9997478388	0.9923063525
19	0.985075577	0.9925039823	0.9966797975
18	0.9999812113	0.9850548875	0.9875049342
17	0.991885116	0.9977728244	0.9982482335
16	0.9937114339	0.9956994688	0.9718,40716746
15	0.9975411687	0.9987271376	0.9994908157
14	0.9855317118	0.9991787244	0.9958694953
13	0.8924205418	0.9997424156	0.9996766792
12	0.8646878928	0.9986409272	0.9999242476
11	0.8996188387	0.9760167761	0.9766086922
10	0.7759922185	0.9961673342	0.9402894268
9	0.7333665426	0.9999887705	0.9782569996
8	0.8055817301	0.9990313839	0.9314026123
7	0.6255842427	0.9922110063	0.9736563656
6	0.7024204135	0.9695098805	0.9997970624
5	0.5146431718	0.9919755462	0.9917328613
4	0.6254793909	0.9292500719	0.9863601838
3	0.7288521573	0.9570653424	0.9075062814
2	0.5854590008	0.83996043	0.8564995075
1	0.4545559232	0.7583238157	0.67564097

the crop yield out of while top five features such as; RH of October, November, December during 8.30 AM and 5.30 PM are shown due to less space. The features affecting the Cuttack district rice yield are RH of July, Sept and October during 8.30 AM and 5.30 PM and also the minimum temperature during September and November months; for the Puri district, the 5 features that affects the rice yield are RH of June, August, September and December mostly 5.30 PM and only 8.30AM in December and also the minimum temperature during October months. From this, it can be concluded that, the features affecting mostly for rice yield are RH during 8.30 AM and 5.30 PM during Kharif season for all three districts as similar to Rabi season.

5.8 Statistical validation

Paired T-test is one of the methods, to assess the consequence of the proposed fusion of feature ranking approach. The outcome produced by ELM-SVR-RFE was compared with proposed approach for five independent runs considering top five ranked features. Here, only ELM-SVR-RFE for statistical validation has been considered for paired test, as it gives better result than the other basic feature ranking based methods. There is no difference found between the outcomes of the two methods that the null hypothesis was the case. The outcomes shown both for the Rabi and the Kharif seasons respectively in the Table 11 and Table 12. From the below tables we can see that, the null hypothesis is rejected and average p-value is 0.0023, 0.0021, 0.0044 for the taken three districts such as: Balasore, Puri and Cuttack of Rabi season and 0.0335, 0.0221 and 0.0450 for Kharif season of all three districts such as: Balasore, Puri and Cuttack. We can observe that the values are closer to zero and for this reason the arguments are strengthened and the projected fusion of feature ranking approach has improved performance than the other only feature ranking based methods.

6 Discussion on principal findings

The principal aim of the present study is to discover the features those have important role or affects mostly in rice crop production both for the Rabi and Kharif seasons of Balasore, Cuttack and Puri. To obtain our desired result, a fusion based strategy based of feature ranking methods has been proposed and explored. This methodology works in three computational phases and not only finds the most significant features contributing towards rice yield but also shows 99% and above prediction accuracy. According to the results obtained the following are few observations made on this study:

- First, the raw data including climatologic characteristics and rice production per hector are collected for three districts and two seasons and the range and average of parameters of those datasets are computed to have a greater insight about the features for proper understanding.
- The importance of features have been evaluated and those features are selected for prediction of rice yield using, ranking of features by applying Random Forest, SVR-RFE and F-Test ranking strategies. These feature ranking models, rank all the features of individual datasets for further processing.
- A feature level fusion model using Borda Count has been explored to generate a new set of ranked features by taking the ranked features from all three feature ranking strategies for further analysis. From this, top five ranked features contributing mostly for rice yield have been listed in **Table 4**.
- Multiquadratic activation has been confirmed from ten activations functions based on R2 score to be used by the ELM regressor to obtain the rice yield prediction above 99% predictive accuracy by decreasing the features one by one for two seasons and three district datasets and results are shown in **Table 7** and **Table 8**.

Table 9: Performance comparison of proposed feature ranking based fusion strategy with feature ranking based methods for Rabi crop prediction.

Districts	Number of top ranked features required to achieve a threshold accuracy of 99%			
	ELM with Random Forest	ELM with SVR-RFE	ELM with F-Test	ELM with Proposed Fusion Strategy
Balasore	7	5	9	3
Cuttack	10	9	11	5
Puri	6	4	8	2

Table 10: Performance comparison of proposed feature ranking based fusion strategy with feature ranking based methods for Kharif crop prediction.

Districts	Number of top ranked features required to achieve a threshold accuracy of 99%			
	ELM with Random Forest	ELM with SVR-RFE	ELM with F-Test	ELM with Proposed Fusion Strategy
Balasore	21	15	24	15
Cuttack	17	10	21	5
Puri	17	12	22	5

Table 11: Paired T-test of Rabi season datasets (all three districts) for the ELM-SVR-RFE approach and proposed Fusion based feature ranking strategy.

Runs	Balasore District Dataset		Puri District Dataset		Cuttack District Dataset	
	Hypothesis Test	p-Value	Hypothesis Test	p-Value	Hypothesis Test	p-Value
1	1	0.002374351	1	0.002145848	1	0.00442727
2	1	0.002376581	1	0.002763544	1	0.00423645
3	1	0.002432856	1	0.002658974	1	0.00445726
4	1	0.002743567	1	0.002738465	1	0.00465187
5	1	0.002267655	1	0.002748983	1	0.00435478

Table 12: Paired T-test of Kharif season datasets (all three districts) for the ELM-SVR-RFE approach and proposed Fusion based feature ranking strategy.

Runs	Balasore District Dataset		Puri District Dataset		Cuttack District Dataset	
	Hypothesis Test	p-Value	Hypothesis Test	p-Value	Hypothesis Test	p-Value
1	1	0.03316396	1	0.022158879	1	0.045061108
2	1	0.03426353	1	0.022165374	1	0.045182873
3	1	0.03326354	1	0.022263667	1	0.044762783
4	1	0.03387623	1	0.021773664	1	0.045002388
5	1	0.03316538	1	0.022377488	1	0.045288384

- Again, the performance comparison of proposed feature ranking based fusion strategy with feature ranking based methods for Rabi and Kharif seasons crop prediction are done to obtain the minimum number of features contributing towards rice crop yield and shown in **Table 9** and **Table 10**. From those tables, it can be concluded that, the features affecting mostly for rice yield are RH during 8.30 AM and 5.30

PM for all three districts taken during both the Rabi and Kharif season and also the minimum temperature plays a vital role.

- The paired T-test was used to calculate the importance of proposed fusion of feature ranking approach. The outcomes found by ELM-SVR-RFE were compared with proposed approach for five independent runs considering top five ranked features. Here, only ELM-SVR-RFE for statistical validation has been considered for paired test, as it gives healthier result than other basic feature ranking based methods.
- It can be observed from Table 11 and Table 12 that, the null hypothesis is rejected in case of Rabi season for all the three districts such as: Balasore, Puri and Cuttack and for three districts of Kharif season, as the values are closer to zero, which strengthens the argument that, proposed fusion of feature ranking approach has improved performance than the other only feature ranking based methods.

7 Conclusion and future scope

In this study an attempt has been made to obtain the climatic effect on rice yield of coastal areas of Odisha. The fusion based strategy is the novelty of this work. This prediction model not only predicts the rice yield per hector but also able to obtain the significant or most affecting features during Rabi and Kharif seasons. This methodology works in three phases, in the first phase, three feature ranking approaches such as; Random Forest, SVR-RFE and F-Test has been applied on the three two datasets of three coastal areas and features are ranked as per the their algorithm. In the second phase, Borda Count as a fusion method has been implemented on those ranked features from the above phase to obtain top five best features. Then in the third phase, multiquadratic based ELM has been used to predict the rice crop yield using those ranked features obtained from fusion based raking strategy of second phase. After applying ELM with fusion strategy, it is seen that by taking at least 3 features for Balasore, 5 features for Cuttack and 2 features for Puri we can get the accuracy of 99% where as in each individual ranking method with ELM we have to take more features. Finally, the statistical paired T-test has been used to evaluate and validate the significance of proposed fusion based ranking prediction model. From the observations made during experimentation, it has been found that; relative humidity and in some case temperature also is playing a vital role for rice crop production both for the Rabi season and the Kharif season. However, in future, the not linked or inconsequential factors can be later dealt with by working on optimized strategies.

Acknowledgement

This work is financially supported by the Ministry of Science and Higher Education of the Russian Federation (Government Order FENU-2020-0022).

References

[1] Central Soil and water Conservation Research & Training Institute (CSWCR & TI), Vision 2030, http://www.cswcrtiweb.org/. (Accessed on 17/10/2014).

[2] Venkateswarlu, B. (2010). The 21st Dr. SP Raychaudhuri Memorial Lecture-Climate change: Adaptation and mitigation strategies in rainfed agriculture. *Journal of the Indian Society of Soil Science*, 58, S27-S35.

[3] Saseendran, S. A., Singh, K. K., Rathore, L. S., Singh, S. V., & Sinha, S. K. (2000). Effects of climate change on rice production in the tropical humid climate of Kerala, India. *Climatic Change*, 44(4), 495-514.

[4] Sarker, M. A. R., Alam, K., & Gow, J. (2012). Exploring the relationship between climate change and rice yield in Bangladesh: An analysis of time series data. *Agricultural Systems*, 112, 11-16.

[5] Soora, N. K., Aggarwal, P. K., Saxena, R., Rani, S., Jain, S., & Chauhan, N. (2013). An assessment of regional vulnerability of rice to climate change in India. *Climatic Change*, 118(3-4), 683-699.

[6] Bocca, F. F., & Rodrigues, L. H. A. (2016). The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and electronics in agriculture*, 128, 67-76.

[7] Gilbertson, J. K., & Van Niekerk, A. (2017). Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Computers and Electronics in Agriculture*, 142, 50-58.

[8] Ma, C., Zhang, H. H., & Wang, X. (2014). Machine learning for Big Data analytics in plants. *Trends in plant science*, 19(12), 798-808.

[9] Hancer, E., Xue, B., & Zhang, M. (2018). Differential evolution for filter feature selection based on information theory and feature ranking. *Knowledge-Based Systems*, 140, 103-119.

[10] Razmjoo, A., Xanthopoulos, P., & Zheng, Q. P. (2017). Online feature importance ranking based on sensitivity analysis. *Expert Systems with Applications*, 85, 397-406.

[11] Teisseyre, P. (2016). Feature ranking for multi-label classification using Markov networks. *Neurocomputing*, 205, 439-454.

[12] Lee, J., & Kim, D. W. (2015). Fast multi-label feature selection based on information-theoretic feature ranking. *Pattern Recognition*, 48(9), 2761-2771.

[13] Fakhraei, S., Soltanian-Zadeh, H., & Fotouhi, F. (2014). Bias and stability of single variable classifiers for feature ranking and selection. *Expert systems with applications*, 41(15), 6945-6958.

[14] Hall, M. A., & Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data engineering*, 15(6), 1437-1447.

[15] Wei, C. C. (2013). Soft computing techniques in ensemble precipitation nowcast. *Applied Soft Computing*, 13(2), 793-805.

[16] Cruz, R. M., Sabourin, R., & Cavalcanti, G. D. (2017). META-DES. Oracle: Meta-learning and feature selection for dynamic ensemble selection. *Information fusion*, 38, 84-103.

[17] Dramiński, M., Rada-Iglesias, A., Enroth, S., Wadelius, C., Koronacki, J., & Komorowski, J. (2008). Monte Carlo feature selection for supervised classification. *Bioinformatics*, 24(1), 110-117.

[18] Tripoliti, E. E., Fotiadis, D. I., & Manis, G. (2013). Modifications of the construction and voting mechanisms of the random forests algorithm. *Data & Knowledge Engineering*, 87, 41-65.

[19] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

[20] Zhang, H. R., & Min, F. (2016). Three-way recommender systems based on random forests. *Knowledge-Based Systems*, 91, 275-286.

[21] Wu, Q., Ye, Y., Zhang, H., Ng, M. K., & Ho, S. S. (2014). ForesTexter: an efficient random forest algorithm for imbalanced text categorization. *Knowledge-Based Systems*, 67, 105-116.

[22] Yeh, C. C., Lin, F., & Hsu, C. Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems*, 33, 166-172.

[23] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.

[24] Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, 212, 353-363.

[25] Shieh, M. D., & Yang, C. C. (2008). Multiclass SVM-RFE for product form feature selection. *Expert Systems with Applications*, 35(1-2), 531-541.

[26] Mishra, S., & Mishra, D. (2015). SVM-BT-RFE: An improved gene selection framework using Bayesian T-test embedded in support vector machine (recursive feature elimination) algorithm. *Karbala International Journal of Modern Science*, 1(2), 86-96.

[27] Xu, Q., Kamel, M., & Salama, M. M. (2004, September). Significance test for feature subset selection on image recognition. *In International Conference Image Analysis and Recognition (pp. 244-252)*. Springer, Berlin, Heidelberg.

[28] Golugula, A., Lee, G., & Madabhushi, A. (2011, August). Evaluating feature selection strategies for high dimensional, small sample size datasets.*In 2011 Annual International conference of the IEEE engineering in medicine and biology society (pp. 949-952)*. IEEE.

[29] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489-501.

[30] Das, S. R., Mishra, D., & Rout, M. (2019). A hybridized ELM using self-adaptive multi-population-based Jaya algorithm for currency exchange prediction: an empirical assessment. *Neural Computing and Applications*, 31(11), 7071-7094.

[31] Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., Min,H. & Deng, X. (2016). Empirical analysis: stock market prediction via extreme learning machine. *Neural Computing and Applications*, 27(1), 67-78.

[32] Balasundaram, S., & Gupta, D. (2016). Knowledge-based extreme learning machines. *Neural Computing and Applications*, 27(6), 1629-1641.

[33] Orissa Agricultural Statistics Year Book, (1983-2013). *Directorate of Agriculture and Food Production, Govt. of Odisha, Bhubaneswar*.

[34] `https://www.google.co.in/images`

[35] Narasimhamurthy, V., & Kumar, P. (2017). Rice Crop Yield Forecasting Using Random Forest Algorithm. *Int. J. Res. Appl. Sci. Eng. Technol*. IJRASET, 5, 1220-1225.

[36] Dahal, H., & Routray, J. K. (2011). Identifying associations between soil and production variables using linear multiple regression models. *Journal of Agriculture and Environment*, 12, 27-37.

[37] Powell, J. P., & Reinhard, S. (2016). Measuring the effects of extreme weather events on yields. *Weather and Climate extremes*, 12, 69-79.

[38] Yusof, M. F., Azamathulla, H. M., & Abdullah, R. (2014). Prediction of soil erodibility factor for Peninsular Malaysia soil series using ANN. *Neural Computing and Applications*, 24(2), 383-389.

[39] Erdil, A., & Arcaklioglu, E. (2013). The prediction of meteorological variables using artificial neural network. *Neural Computing and Applications*, 22(7-8), 1677-1683.

[40] Anitha, A., & Acharjya, D. P. (2018). Crop suitability prediction in Vellore District using rough set on fuzzy approximation space and neural network. *Neural Computing and Applications*, 30(12), 3633-3650.

[41] Zahid, M. A., & De Swart, H. (2015). The borda majority count. *Information Sciences*, 295, 429-440.

[42] García-Lapresta, J. L., Martínez-Panero, M., & Meneses, L. C. (2009). Defining the Borda count in a linguistic decision making context. *Information Sciences*, 179(14), 2309-2316.

[43] `https://www.casact.org/pubs/forum/98wforum/98wf055.pdf`

[44] Hirai, GI., Chiyo, H., Tanka, O., Hikano, T., & Oanotri, M. (1993). Studies on the effect of relative humidity of atmosphere on growth and physiology of rice plants. VIII effect of ambient humidity on dry matter production and nitrogen absorption at various temperatures, *Japanese Journal of Crop Science*, 62(3), 395-400.

[45] Sunil, K. M. (2000). Crops weather relationship in rice *(Doctoral dissertation, Department of Agricultural Meteorology, College of Horticulture, Vellanikkara)*.

[46] Vijayakumar, CM. (1996). Hybrid rice seed production technology- theory and practice. *Directorate of rice research, Hyderabad*, 52-55.

[47] Gridyal, B. P., & Jana, R. K. (1997). Agrometerologycal environmental affecting rice yield. *Agronomy Journal*, 59, 286-287.

[48] Narayanan, A. L. (2004). Relative influence of weather parameters on rice hybrid and variety and validation of CERES- Rice model for staggered weeks of transplanting. *PhD Thesis, Tamilnadu Agricultural University, Coimbatore*.

[49] Shi, C. H., & Shen, Z. T. (1990). Effect of high humidity and low temperature on spikelet fertility in indica rice. *International Rice Research Newsletter*, 15(3), 10-11.

[50] Morita, S., Wada, H., & Matsue, Y. (2016). Countermeasures for heat damage in rice grain quality under climate change. *Plant Production Science*, 19(1), 1-11.

A Generative Model Based Adversarial Security of Deep Learning and Linear Classifier Models

Samed Sivaslioglu
Tubitak Bilgem, Kocaeli, Turkey
E-mail: samedsivaslioglu@gmail.com

Ferhat Ozgur Catak
Simula Research Laboratory, Fornebu, Norway
E-mail: ozgur@simula.no

Kevser Şahinbaş
Department of Management Information System, Istanbul Medipol University, Istanbul, Turkey
E-mail: ksahinbas@medipol.edu.tr

Keywords: adversarial machine learning, generative models, autoencoders

Received: July 13, 2020

In recent years, machine learning algorithms have been applied widely in various fields such as health, transportation, and the autonomous car. With the rapid developments of deep learning techniques, it is critical to take the security concern into account for the application of the algorithms. While machine learning offers significant advantages in terms of the application of algorithms, the issue of security is ignored. Since it has many applications in the real world, security is a vital part of the algorithms. In this paper, we have proposed a mitigation method for adversarial attacks against machine learning models with an autoencoder model that is one of the generative ones. The main idea behind adversarial attacks against machine learning models is to produce erroneous results by manipulating trained models. We have also presented the performance of autoencoder models to various attack methods from deep neural networks to traditional algorithms by using different methods such as non-targeted and targeted attacks to multi-class logistic regression, a fast gradient sign method, a targeted fast gradient sign method and a basic iterative method attack to neural networks for the MNIST dataset.

Povzetek: S pomočjo globokega učenja je analizirana varnost pri sistemih strojnega učenja.

1 Introduction

With the help of artificial intelligence technology, machine learning has been widely used in classification, decision making, voice and face recognition, games, financial assessment, and other fields [9, 12, 44, 45, 48]. The machine learning methods consider player's choices in the animation industry for games and analyze diseases to contribute to the decision-making mechanism [2, 6, 7, 15, 34, 46]. With the successful implementations of machine learning, attacks on the machine learning process and counter-attack methods and incrementing robustness of learning have become hot research topics in recent years [24, 27, 31, 37, 51]. The presence of negative data samples or an attack on the model can lead to producing incorrect results in the predictions and classifications even in the advanced models.

It is more challenging to recognize the attack because of using big data in machine learning applications compared to other cybersecurity fields. Therefore, it is essential to create components for machine learning that are resistant to this type of attack. In contrast, recent works have conducted in this area and demonstrated that the resistance is not very robust to attacks [10, 11]. These methods have shown success against a specific set of attack methods and have generally failed to provide complete and generic protection[43].

Machine learning models already used in functional forms could be vulnerable to these kinds of attacks. For instance, by putting some tiny stickers on the ground in a junction, researchers confirmed that they could provoke an autonomous car to make an unnatural decision and drive into the opposite lane [16]. In another study, the researchers have pointed out that making hidden modifications to an input image can fool a medical imaging operation into labelling a benign mole as malignant with 100% confidence [17].

Previous methods have shown success against a specific set of attack methods and have generally failed to provide complete and generic protection [14]. This field has been spreading rapidly, and, in this field, lots of dangers have attracted increasing attention from escaping the filters of unwanted and phishing e-mails, to poisoning the sensor data of a car or aircraft that drives itself [4, 41]. Disaster scenarios can occur if any precautions are not taken in these systems [30].

The main contribution of this work is to explore the autoencoder based generative models against adversarial machine learning attacks to the models. Adversarial Machine Learning has been used to study these attacks and reduce their effects [8, 32]. Previous works point out the fundamental equilibrium to design the algorithms and to create new algorithms and methods that are resistant and robust against attacks that will negatively affect this balance. However, most of these works have been implemented successfully for specific situations. In Section 3, we present some applications of these works.

This work aims to propose a method that not only presents a generic resistance to specific attack methods but also provides robustness to machine learning models in general. Our goal is to find an effective method that can be used by model trainers. For this purpose, we have processed the data with autoencoder before reaching to the machine learning model.

We have used non-targeted and targeted attacks to multiclass logistic regression machine learning models for observing the change and difference between attack methods as well as various attack methods to neural networks such as fast gradient sign method (FGSM), targeted fast gradient sign method (T-FGSM) and basic iterative method (BIM). We have selected MNIST dataset that consists of numbers from people's handwriting to provide people to understand and see changes in the data. In our previous works [3, 38], we applied the generative models both for data and model poisoning attacks with limited datasets.

The study is organized as follows. In Section 2, we first present the related works. In Section 3, we introduce several adversarial attack types, environments, and autoencoder. In Section 4, we present selection of autoencoder model, activation function and tuning parameters. In Section 5, we provide some observation on the robustness of autoencoder for adversarial machine learning with different machine learning algorithms and models. In Section 8, we conclude this study.

2 Related Work

In recent years, with the increase of the machine learning attacks, various studies have been proposed to create defensive measures against these attacks. Data sterility and learning endurance are recommended as countermeasures in defining a machine learning process [32]. They provide a model for classifying attacks against online machine learning algorithms. Most of the studies in these fields have been focused on specific adversarial attacks and generally, presented the theoretical discussion of adversarial machine learning area [23, 25].

Bo Li and Yevgeniy Vorobeychik present binary domains and classifications. In their work, the approach starts with mixed-integer linear programming (MILP) with constraint generation and gives suggestions on top of this. They also use the Stackelberg game multi-adversary model algorithm and the other algorithm that feeds back the generated adversarial examples to the training model, which is called as RAD (Retraining with Adversarial Examples) [28]. Their approach can scale thousands of features with RAD that showed robustness to several model erroneous specifications. On the other hand, their work is particular and works only in specific methods, even though it is presented as a general protection method. They have proposed a method that implements successful results. Similarly, Xiao et al. provide a method to increase the speed of resistance training against the rectified linear unit (RELU) [36]. They provide that optimizing weight sparseness enables us to turn computationally demanding validation problems into solvable problems. They showed that improving ReLU stability leads to 4-13x faster validation times. They use weight sparsity and RELU stability for robust verification. It can be said that their methodology does not provide a general approach.

Yu et al. propose a study that can evaluate the neural network's features under hostile attacks. In their study, the connection between the input space and hostile examples is presented. Also, the connection between the network strength and the decision surface geometry as an indicator of the hostile strength of the neural network is shown. By extending the loss surface to decision surface and other various methods, they provide adversarial robustness by decision surface. The geometry of the decision surface cannot be demonstrated most of the time, and there is no explicit decision boundary between correct or wrong prediction. Robustness can be increased by constructing a good model, but it can change with attack intensity [50]. Their method can increase network's intrinsic adversarial robustness against several adversarial attacks without involving adversarial training.

Mardy et al. investigate artificial neural networks resistant with adversity and increase accuracy rates with different methods, mainly with optimization and prove that there can be more robust machine learning models [43].

Pinto et al. provide a method to solve this problem with the supported learning method. In their study, they formulate learning as a zero-sum, minimax objective function. They present machine learning models that are more resistant to disturbances are hard to model during the training and are better affected by changes in training and test conditions. They generalize reinforced learning on machine learning models. They propose a "Robust Adversarial Reinforced Learning" (RARL), where they train an agent to operate in the presence of a destabilizing adversary that applies disturbance forces to the system. They presented that their method increased training stability, was robust to differences in training and testing conditions, and outperformed basically even in the absence of the adversary. However, in their work, Robust Adversarial Reinforced Learning may overfit itself, and sometimes it can miss predicting without any adversarial being in presence [39].

Carlini and Wagner propose a model that the self-logic and the strength of the machine learning model with a

strong attack can be affected. They prove that these types of attacks can often be used to evaluate the effectiveness of potential defenses. They propose defensive distillation as a general-purpose procedure to increase robustness [11].

Harding et al. similarly investigate the effects of hostile samples produced from targeted and non-targeted attacks in decision making. They observed that non-targeted samples interfered more with human perception and classification decisions than targeted samples [22].

Bai et al. present a convolutional autoencoder model with the adversarial decoders to automate the generation of adversarial samples. They produce adversary examples by a convolutional autoencoder model. They use pooling computations and sampling tricks to achieve these results. After this process, an adversarial decoder automates the generation of adversarial samples. Adversarial sampling is useful, but it cannot provide adversarial robustness on its own, and sampling tricks are too specific [5]. They gain a net performance improvement over the normal CNN.

Sahay et al. propose FGSM attack and use an autoencoder to denoise the test data. They have also used an autoencoder to denoise the test data, which is trained with both corrupted and healthy data. Then they reduce the dimension of the denoised data. These autoencoders are specifically designed to compress data effectively and reduce dimensions. Hence, it may not be wholly generalized, and training with corrupted data requires a lot of adjustments to get better test results [33]. Their model provide that when test data is preprocessed using this cascading, the tested deep neural network classifier provides much higher accuracy, thus mitigating the effect of the adversarial perturbation.

I-Ting Chen et al. also provide with FGSM attack on denoising autoencoders. They analyze the attacks from the perspective that attacks can be applied stealthily. They use autoencoders to filter data before applied to the model and compare it with the model without an autoencoder filter. They use autoencoders mainly focused on the stealth aspect of these attacks and used them specifically against FGSM with specific parameters [13]. They enhance the classification accuracy from 2.92% to 75.52% for the neural network classifier on the 10 digits and from 4.12% to 93.57% for the logistic regression classifier on digit 3s and 7s.

Gondim-Ribeiro et al. propose autoencoders attacks. In their work, they attack 3 types of autoencoders: Simple variational autoencoders, convolutional variational autoencoders, and DRAW (Deep Recurrent AttentiveWriter). They propose to scheme an attack on autoencoders. As they accept that "No attack can both convincingly reconstruct the target while keeping the distortions on the input imperceptible.". They enable both DRAW's recurrence and attention mechanism to lead to better resistance. Automatic encoders are recommended to compress data and more attention should be given to adversarial attacks on them. This method cannot be used to achieve robustness against adversarial attacks [40].

Table 2 shows the strength and the weakness of the each paper.

3 Preliminaries

In this section, we consider attack types, data poisoning attacks, model attacks, attack environments, and autoencoder.

3.1 Attack Types

Machine Learning attacks can be categorized into data poisoning attacks and model attacks. The difference between the two attacks lies in the influencing type. Data poisoning attacks mainly focus on influencing the data, while model evasion attacks influencing the model for desired attack outcomes. Both attacks aim to disrupt the machine learning structure, evasion from filters, causing wrong predictions, misdirection, and other problems for the machine learning process. In this paper, we mainly focus on machine learning model attacks.

3.1.1 Data Poisoning Attacks

According to machine learning methods, algorithms are trained and tested with datasets. Data poisoning in machine learning algorithms has a significant impact on a dataset and can cause problems for algorithm and confusion for developers. With poisoning the data, adversaries can compromise the whole machine learning process. Hence, data poisoning can cause problems in machine learning algorithms.

3.1.2 Model Attacks

Machine learning model attacks have been applied mostly in adversarial attacks, and evasion attacks being have been used most extensively in this category. For spam emails, phishing attacks, and executing malware code, adversaries apply model evasion attacks. There are also some benefits to adversaries in misclassification and misdirection. In this type of attack, the attacker does not change training data but disrupts or changes its data and diverse this data from the training dataset or make this data seem safe. This study mainly concentrates on model attacks.

3.2 Attack Environments

There are two significant threat models for adversarial attacks: the white-box and black-box models.

3.2.1 White Box Attacks

Under the white-box setting, the internal structure, design, and application of the tested item are accessible to the adversaries. In this model, attacks are based on an analysis of the internal structure. It is also known as open box attacks. Programming knowledge and application knowledge

Table 1: Related Work Summary

Research Study	Strength	Weakness
Adversarial Machine Learning [32]	Introduces the emerging field of Adversarial Machine Learning.	Discusses the countermeasures against attacks without suggesting a method.
Evasion-Robust Classification on Binary Domains [28]	Demonstrates some methods that can be used on Binary Domains, which are based on MILP.	Very specific about the robustness, even though it is presented as a general method.
Training for Faster Adversarial Robustness Verification via Inducing ReLU Stability [36]	Using weight sparsity and RELU stability for robust verification.	Does not provide a general approach, or universality as it is suggested in paper.
Interpreting Adversarial Robustness: A View from Decision Surface in Input Space [50]	By extending the loss surface to decision surface and other various methods, they provide adversarial robustness by decision surface.	The geometry of the decision surface cannot be shown most of the times and there is no explicit decision boundary between correct or wrong prediction. Robustness can be increased by constructing a good model but it can change with attack intensity.
Robust Adversarial Reinforcement Learning [39]	They have tried to generalize reinforced learning on machine learning models. They suggested a Robust Adversarial Reinforced Learning (RARL) where they have trained an agent to operate in the presence of a destabilizing adversary that applies disturbance forces to the system.	Robust Adversarial Reinforced Learning may overfit itself and sometimes it may mispredict without any adversarial being in presence.
Alleviating Adversarial Attacks via Convolutional Autoencoder [5]	They have produced adversary examples via a convolutional autoencoder model. Pooling computations and sampling tricks are used. Then an adversarial decoder automate the generation of adversarial samples.	Adversarial sampling is useful but it cannot provide adversarial robustness on its own. Sampling tricks are also too specified.
Combatting Adversarial Attacks through Denoising and Dimensionality Reduction: A Cascaded Autoencoder Approach [33]	They have used an autoencoder to denoise the test data which is trained with both corrupted and normal data. Then they reduce the dimension of the denoised data.	Autoencoders specifically designed to compress data effectively and reduce dimensions. Therefore it may not be completely generalized and training with corrupted data requires a lot of adjustments for test results.
A Comparative Study of Autoencoders against Adversarial Attacks [13]	They have used autoencoders to filter data before applying into the model and compare it with the model without autoencoder filter.	They have used autoencoders mainly focused on the stealth aspect of these attacks and use them specifically against FGSM with specific parameters.
Adversarial Attacks on Variational Autoencoders [40]	They propose a scheme to attack on autoencoders and validate experiments to three autoencoder models: Simple, convolutional and DRAW (Deep Recurrent Attentive Writer).	As they have accepted "No attack can both convincingly reconstruct the target while keeping the distortions on the input imperceptible.". it cannot provide robustness against adversarial attacks.
Understanding Autoencoders with Information Theoretic Concepts [47]	They examine data processing inequality with stacked autoencoders and two types of information planes with autoencoders. They have analyzed DNNs learning from a joint geometric and information theoretic perspective, thus emphasizing the role that pair-wise mutual information plays important role in understanding DNNs with autoencoders.	The accurate and tractable estimation of information quantities from large data seems to be a problem due to Shannon's definition and other information theories are hard to estimate, which severely limits its powers to analyze machine learning algorithms.
Adversarial Attacks and Defences Competition [42]	Google Brain organized NIPS 2017 to accelerate research on adversarial examples and robustness of machine learning classifiers. Alexey Kurakin and Ian Goodfellow et al. present some of the structure and organization of the competition and the solutions developed by several of the top-placing teams.	We experimented with the proposed methods of this competition bu these methods do not provide a generalized solution for the robustness against adversarial machine learning model attacks.
Explaining And Harnessing Adversarial Examples [19]	Ian Goodfellow et al. makes considerable observations about Gradient-based optimization and introduce FGSM.	Models may mislead for the efficiency of optimization. The paper focuses explicitly on identifying similar types of problematic points in the model.

are essential. White-box tests provide a comprehensive assessment of both internal and external vulnerabilities and are the best choice for computational tests.

3.2.2 Black Box Attacks

In the black-box model, internal structure and software testing are secrets to the adversaries. It is also known as behavioral attacks. In these tests, the internal structure does not have to be known by the tester. They provide a comprehensive assessment of errors. Without changing the learning process, black box attacks provide changes to be observed as external effects on the learning process rather than changes in the learning algorithm. In this study, the main reason behind the selection of this method is the observation of the learning process.

3.3 Autoencoder

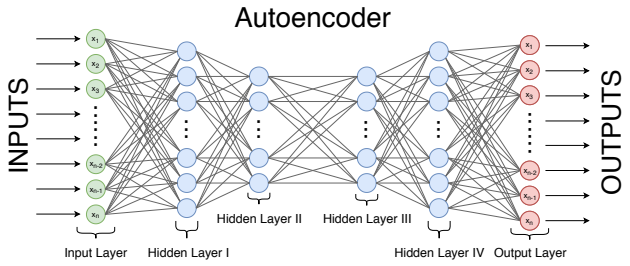


Figure 1: Autoencoder Layer Structure

An autoencoder neural network is an unsupervised learning algorithm that takes inputs and sets target values to be equals of the input values [47]. Autoencoders are generative models that apply backpropagation. They can work without the results of these inputs. While the use of a learning model is in the form of `model.fit(X,Y)`, autoencoders work as `model.fit(X,X)`. The autoencoder works with the ID function to get the output $\mathbf{x}$ that corresponds to $\mathbf{x}$ entries. The identity function seems to be a particularly insignificant function to try to learn; however, there is an interesting structure related to the data, putting restrictions such as limiting the number of hidden units on the network[47]. They are neural networks which work as neural networks with an input layer, hidden layers and an output layer but instead of predicting Y as in `model.fit(X,Y)`, they reconstruct X as in `model.fit(X,X)`. Due to this reconstruction being unsupervised, autoencoders are unsupervised learning models. This structure consists of an encoder and a decoder part. We will define the encoding transition as ϕ and decoding transition as ψ.

$\phi : X \to F$

$\psi : F \to X$

$\phi, \psi = argmin_{\phi,\psi} ||X - (\psi \circ \phi)X||^2$

With one hidden layer, encoder will take the input $x \in \mathbb{R}^d = \chi$ and map it to $h \in \mathbb{R}^p = F$. The h below is referred to as latent variables. σ is an activation function such as ReLU or sigmoid which were used in this study[1, 20]. b is bias vector, W is weight matrix which both are usually initialized randomly then updated iteratively through training[35].

$h = \sigma(Wx + b)$

After the encoder transition is completed, decoder transition maps h to reconstruct x'.

$x' = \sigma'(W'h + b')$ where σ', W', b' of decoder are unrelated to σ, W, b of encoder. Loss of autoencoders are trained to be minimal, showed as L below.

$L(x, x') = ||x - x'||^2 = ||x - \sigma'(W'(\sigma(Wx + b)) + b')||^2$

So the loss function shows the reconstruction errors, which need to be minimal. After some iterations with input training set x is averaged.

In conclusion, autoencoders can be seen as neural networks that reconstruct inputs instead of predicting them. In this paper, we will use them to reconstruct our dataset inputs.

4 System Model

This section presents the selection of autoencoder model, activation function, and tuning parameters.

4.1 Creating Autoencoder Model

In this paper, we have selected the MNIST dataset to observe changes easily. Therefore, the size of the layer structure in the autoencoder model is selected as 28 and multipliers to match the MNIST datasets, which represents the numbers by 28 to 28 matrixes. Figure 2 presents the structure of matrixes. The modified MNIST data with autoencoder is presented in Figure 3. In the training of the model, the encoded data is used instead of using the MNIST datasets directly. As a training method, a multiclass logistic regression method is selected, and attacks are applied to this model. We train autoencoder for 35 epochs. Figure 4 provides the process diagram.

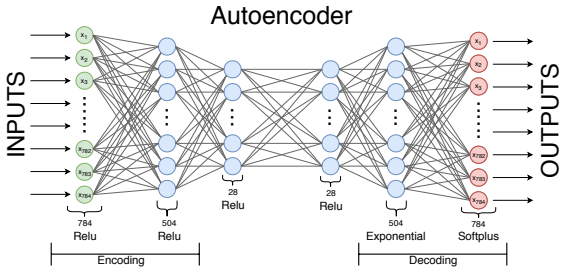


Figure 2: Autoencoder Activation Functions. Note that layer sizes given according to the dataset which is MNIST dataset

4.2 Activation Function Selection

In machine learning and deep learning algorithms, the activation function is used for the computations between

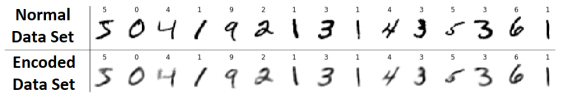


Figure 3: Normal and Encoded Data Set of MNIST

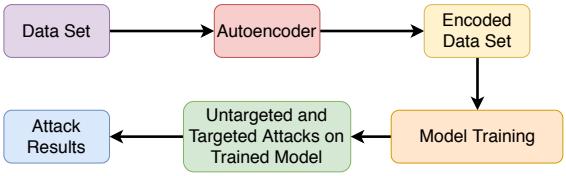
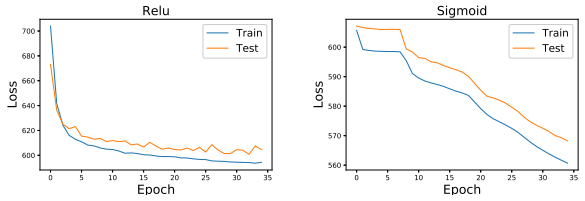


Figure 4: Process Diagram



(a) Relu Loss History (b) Sigmoid Loss History

(c) Softsign Loss History (d) Tanh Loss History

Figure 5: Loss histories of different activation functions

hidden and output layers[18]. The loss values are compared with different activation functions. Figure 5 indicates the comparison results of loss value. `Sigmoid` and `ReLU` have the best performance among these values and gave the best results. `Sigmoid` has more losses at lower epochs than `ReLU`, but it has better results. Therefore, it is aimed to reach the best result of activation function in both layers. The model with the least loss value is to make the coding parts with the `ReLU` function and to use the `exponential` and `softplus` functions in the analysis part respectively. These functions are used in our study. Figure 6 illustrates the result of the loss function, and Figure 2 presents the structure of the model with the activation functions.

4.3 Tuning Parameters

The tuning parameters for autoencoders depend on the dataset we use and what we try to apply. As previously mentioned, `ReLU` and sigmoid function are selected to be activation function for our model [1, 18]. `ReLU` is the activation function through the whole autoencoder while exponential is the softplus being the output layer's activation function which yields the minimal loss. Figure 2 presents

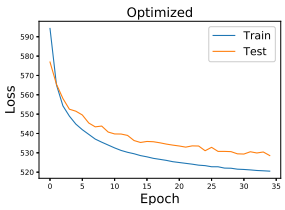


Figure 6: Optimized Relu Loss History

the input size as 784 due to our dataset and MNIST dataset contains 28x28 pixel images[29]. Encoding part for our autoencoder size is $784 \times 504 \times 28$ and decoding size is $28 \times 504 \times 784$.

This structure is selected by the various neural network structures that take the square of the size of the matrix, lower it, and give it to its dimension size lastly. The last hidden layer of the decoding part with the size of 504 uses `exponential` activation function, and an output layer with the size of 784 uses `softplus` activation function [14, 21]. We used `adam` optimizer with categorical crossentropy[26, 49]. We see that a small number is enough for training, so we select epoch number for autoencoder as 35. This is the best epoch value to get meaningful results for both models with autoencoder and without autoencoder to see accuracy. In lower values, models get their accuracy scores too low for us to see the difference between them, even though some models are structurally stronger than others.

5 Experiments with MNIST Dataset

5.1 Introduction

We examine the robustness of autoencoder for adversarial machine learning with different machine learning algorithms and models to see that autoencoding can be a generalized solution and an easy to use defense mechanism for most adversarial attacks. We use various linear machine learning model algorithms and neural network model algorithms against adversarial attacks.

5.2 Autoencoding

In this section, we look at the robustness provided with auto-encoding. We select a linear model and a neural network model to demonstrate this effectiveness. In these models, we also observe the robustness of different attack methods. We also use the MNIST dataset for these examples.

5.2.1 Multi-Class Logistic Regression

In linear machine learning model algorithms, we use mainly two attack methods: Non-Targeted and Targeted Attacks. The non-targeted attack does not concern with how the machine learning model makes its predictions and

Actual Values	Predicted Values									
	0	1	2	3	4	5	6	7	8	9
0	973	0	4	0	1	2	9	1	4	6
1	0	1127	3	0	0	0	3	5	1	2
2	2	6	1016	4	3	0	2	10	4	1
3	0	0	2	1002	0	10	0	5	3	4
4	0	0	3	0	966	0	1	0	0	8
5	0	0	0	1	0	869	3	0	1	2
6	1	1	0	0	1	5	938	0	1	0
7	1	0	4	0	1	1	0	999	3	9
8	3	1	0	3	2	3	1	2	953	3
9	0	0	0	0	8	2	1	6	4	974

Figure 7: Confusion matrix of the model without any attack and without autoencoder

Actual Values	Predicted Values									
	0	1	2	3	4	5	6	7	8	9
0	973	0	4	0	1	2	9	1	4	6
1	0	1127	3	0	0	0	3	5	1	2
2	2	6	1016	4	3	0	2	10	4	1
3	0	0	2	1002	0	10	0	5	3	4
4	0	0	3	0	966	0	1	0	0	8
5	0	0	0	1	0	869	3	0	1	2
6	1	1	0	0	1	5	938	0	1	0
7	1	0	4	0	1	1	0	999	3	9
8	3	1	0	3	2	3	1	2	953	3
9	0	0	0	0	8	2	1	6	4	974

Figure 8: Confusion matrix of the model without any attack and with autoencoder

tries to force the machine learning model into misprediction. On the other hand, targeted attacks focus on leading some correct predictions into mispredictions. We have three methods for targeted attacks: Natural, Non-Natural, and one selected target. Firstly, natural targets are derived from the most common mispredictions made by the machine learning model. For example, guessing number 5 as 8, and number 7 as 1 are common mispredictions. Natural targets take these non-targeted attack results into account and attack directly to these most common mispredictions. So, when number 5 is seen, an attack would try to make it guessed as number 8. Secondly, non-natural targeted attacks are the opposite of natural targeted attacks. It takes the minimum number of mispredictions made by the machine learning model with the feedback provided by non-natural attacks. For example, if number 1 is least mispredicted as 0, the non-natural target for number 1 is 0. Therefore, we can see that how much the attack affects the machine learning model beyond its common mispredictions. Lastly, one targeted attack focuses on some random numbers. The aim is to make the machine learning model mispredict the same number for all numbers. For linear classifications, we select multi-class logistic regression to analyze the attacks. Because we do not interact with these linear classification algorithms aside from calling their defined functions from scikit-learn library, we use a black-box environment for these attacks. In our study, the attack method against multi-class classification models developed in NIPS 2017 is used [42]. An epsilon value is used to determine the severity of the attack, which we select 50 in this study to demonstrate the results better. We apply a non-targeted attack to a multi-class logistic regression trained model which is trained with MNIST dataset without an autoencoder. The confusion matrix of this attack is presented in 9.

The findings from Figure 9 and 10 show that an autoencoder model provides robustness against non-targeted attacks. The accuracy value change with epsilon is presented in Figure 13. Figure 11 illustrates the change and perturbation of the selected attack with epsilon value as 50.

We apply a non-targeted attack on the multi-class logistic regression model with autoencoder and without autoencoder. Figure 13 provides a difference in accuracy metric. The detailed graph of the non-targeted attack on the model

Actual Values	Predicted Values									
	0	1	2	3	4	5	6	7	8	9
0	247	0	17	51	8	73	32	20	8	7
1	0	0	34	8	0	8	0	15	30	13
2	32	18	69	37	181	24	251	288	191	255
3	49	174	222	8	128	106	25	193	489	141
4	4	0	34	49	14	57	59	29	10	231
5	509	58	56	154	43	9	502	110	55	172
6	45	0	93	35	68	109	4	5	25	1
7	23	210	48	22	33	26	43	26	52	1
8	51	678	366	586	31	378	23	141	0	189
9	47	13	60	76	469	60	25	194	137	0

Figure 9: Confusion matrix of non-targeted attack to model without autoencoder

Actual Values	Predicted Values									
	0	1	2	3	4	5	6	7	8	9
0	987	0	7	8	0	13	0	1	5	4
1	0	1137	8	0	1	1	0	4	4	5
2	0	0	958	2	4	2	0	15	0	0
3	0	0	9	886	3	52	1	3	13	9
4	0	0	3	4	923	11	0	10	1	28
5	0	0	0	24	1	643	0	0	0	0
6	5	0	5	2	3	28	962	2	2	0
7	0	0	7	0	1	1	0	932	5	4
8	2	5	31	72	1	116	0	12	944	9
9	1	0	3	14	35	13	0	54	8	931

Figure 10: Confusion matrix of non-targeted attack to model with autoencoder

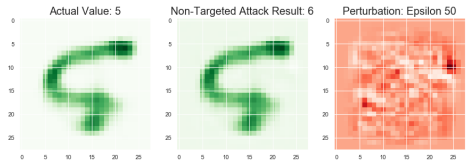


Figure 11: Value change and perturbation of a non-targeted attack on model without autoencoder

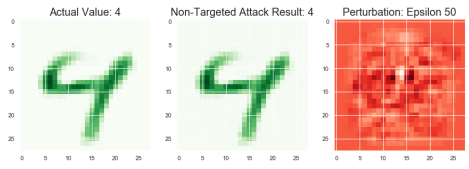


Figure 12: Value change and perturbation of a non-targeted attack on model with autoencoder

with autoencoder is presented in Figure 14. The changes in the MNIST dataset after autoencoder is provided in Fig-

ure 3. The value change and perturbation of an epsilon 50 value on data are indicated in Figure 12.

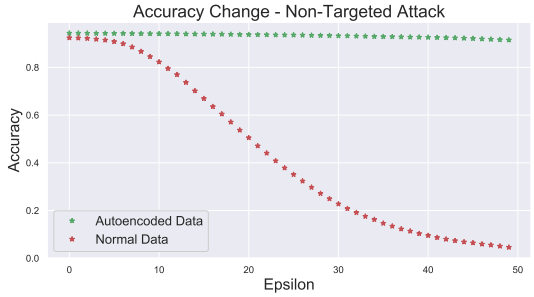


Figure 13: Comparison of accuracy with and without autoencoder for non-targeted attack

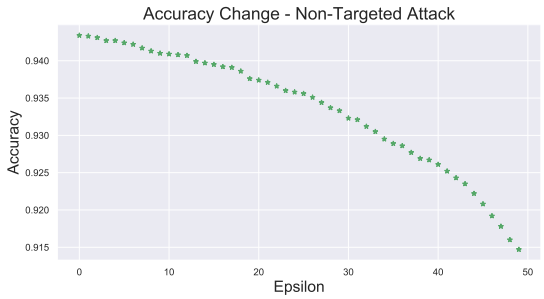


Figure 14: Details of accuracy with autoencoder for non-targeted attack

The following process is presented in Figure 4. In the examples with the autoencoder, data is passed through the autoencoder and then given to the training model, in our current case a classification model with multi-class logistic regression. Multi-class logistic regression uses the encoded dataset for training. Figure 10 provides to see improvement as a confusion matrix. For the targeted attacks, we select three methods to use. The first one is natural targets for MNIST dataset, which is also defined in NIPS 2017 [42]. Natural targets take the non-targeted attack results into account and attack directly to these most common mispredictions. For example, the natural target for number 3 is 8. When we apply the non-targeted attack, we obtain these results. Heat map for these numbers is indicated in Figure 77.

The second method of targeted attacks is non-natural targets which is the opposite of natural targets. We select the least mis predicted numbers as the target. These numbers is indicated as the heat map in Figure 77. The third method is the selection one number and making all numbers predict it. We randomly choose 7 as that target number. Targets for these methods are presented in Figure 16. The confusion matrixes for these methods are presented below.

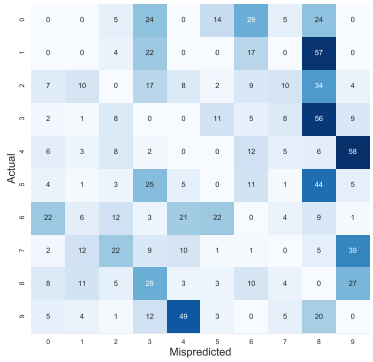


Figure 15: Heatmap of actual numbers and mispredictions

Natural Targets	**Actual Numbers**	0	1	2	3	4	5	6	7	8	9
	Target Numbers	6	8	8	8	9	8	0	9	3	4
Non-Natural Targets	**Actual Numbers**	0	1	2	3	4	5	6	7	8	9
	Target Numbers	1	0	0	1	1	1	1	6	0	6
One Number Targeted	**Actual Numbers**	0	1	2	3	4	5	6	7	8	9
	Target Numbers	7	7	7	7	7	7	7	7	7	7

Figure 16: Actual numbers and their target values for each targeted attack method

		Predicted Values									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Actual Values	**0**	291	0	10	9	1	5	10	16	1	1
	1	0	0	1	0	0	0	0	2	0	0
	2	1	7	70	14	3	1	10	806	25	27
	3	6	10	46	45	7	38	6	17	786	9
	4	9	6	11	10	84	11	13	23	8	920
	5	680	3	22	21	5	49	559	15	29	0
	6	1	0	40	3	8	8	329	2	1	0
	7	0	0	4	1	6	1	3	18	3	0
	8	18	1124	783	917	17	735	26	41	130	17
	9	1	1	12	6	844	2	8	81	14	36

Figure 17: Confusion matrix of **natural targeted attack** to model without autoencoder

		Predicted Values									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Actual Values	**0**	989	0	2	1	0	6	7	1	0	1
	1	0	1105	2	0	0	1	0	0	0	0
	2	0	0	979	4	0	1	0	2	0	0
	3	0	0	0	972	0	12	0	1	4	32
	4	0	0	0	0	889	1	2	1	0	0
	5	0	0	0	0	0	713	0	0	0	0
	6	3	0	3	0	1	8	969	0	1	0
	7	0	0	6	1	0	0	0	943	0	11
	8	3	29	35	46	3	134	2	1	914	6
	9	1	3	1	2	77	2	2	57	44	964

Figure 18: Confusion matrix of **natural targeted attack** to model with autoencoder

5.2.2 Neural Networks

We use neural networks with the same principles as multi-class logistic regressions and make attacks to the machine

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	735	147	281	41	8	36	31	29	694	12
1	3	7	22	565	134	259	105	26	34	39
2	29	88	200	53	107	15	214	170	135	22
3	37	59	96	71	41	95	9	136	59	19
4	3	0	16	8	224	42	53	37	3	362
5	83	0	5	31	1	2	107	14	5	4
6	72	8	99	24	103	110	422	39	28	380
7	5	100	22	6	7	7	6	156	6	0
8	33	741	246	195	30	258	13	104	22	163
9	7	1	12	32	320	26	4	310	11	9

Figure 19: Confusion matrix of **non-natural targeted attack** to model without autoencoder

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	994	0	1	0	0	7	0	0	0	0
1	0	1147	0	2	0	6	0	4	2	1
2	2	1	991	0	0	6	2	2	30	0
3	0	0	4	992	0	71	0	5	2	1
4	0	0	0	0	973	4	0	5	0	1
5	0	0	7	0	1	597	1	1	4	0
6	2	0	3	0	2	32	964	1	0	1
7	0	0	0	0	1	1	0	1001	0	0
8	3	1	5	5	0	170	1	5	917	8
9	0	0	1	3	0	8	0	6	0	992

Figure 20: Confusion matrix of **non-natural targeted attack** to model with autoencoder

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	281	0	17	14	1	27	17	0	1	0
1	0	0	9	0	0	0	0	0	0	0
2	0	0	69	0	1	2	32	0	1	0
3	16	12	330	109	2	132	46	0	96	0
4	1	0	7	4	36	22	16	0	1	1
5	69	0	9	12	0	13	165	0	6	0
6	5	0	38	4	0	27	164	0	3	0
7	612	1114	372	778	828	406	479	1021	731	1005
8	6	25	116	61	0	139	21	0	28	0
9	17	0	32	44	107	82	24	0	130	4

Figure 21: Confusion matrix of **one number targeted attack** to model without autoencoder

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	991	0	3	0	0	8	0	0	0	0
1	0	1139	7	0	0	1	0	0	3	0
2	0	0	955	0	0	0	0	0	0	0
3	0	0	20	991	0	33	1	0	7	0
4	1	0	4	0	947	4	1	0	1	1
5	0	0	0	0	0	775	0	0	0	0
6	0	0	5	0	0	11	960	0	0	0
7	2	3	20	18	25	2	1	1033	19	104
8	0	0	15	0	0	38	0	0	945	0
9	1	0	2	3	0	8	0	0	7	885

Figure 22: Confusion matrix of **one number targeted attack** to model with autoencoder

learning model. We use the same structure, layer, activation functions and epochs for these neural networks as we use in our autoencoder for simplicity. Although this robustness will work with other neural network structures, we will not demonstrate them in this study due to structure designs that can vary for all developers. We also compare the results of these attacks with both the data from the MNIST dataset

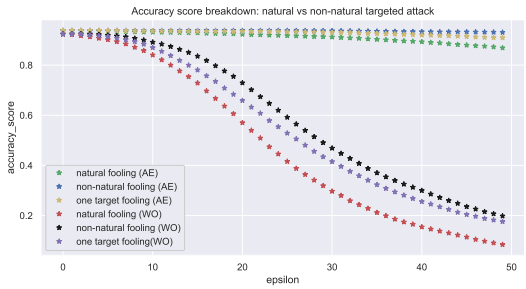


Figure 23: Comparison of accuracy with and without autoencoder for targeted attacks. *AE stands for the models with autoencoder, WO stands for models without autoencoder*

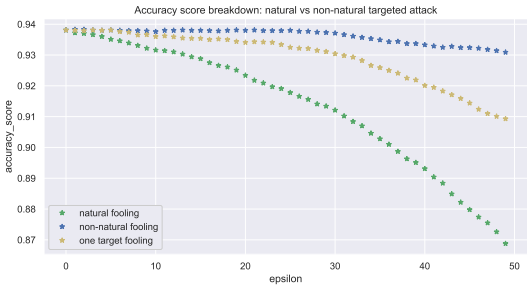


Figure 24: Details of accuracy with autoencoder for targeted attacks

and the encoded data results of the MNIST dataset. As for attack methods, we select three methods: FGSM, T-FGSM and BIM. Cleverhans library is used for providing these attack methods to the neural network, which is from the Keras library.

We examine the differences between the neural network model that has autoencoder and the neural network model that takes data directly from the MNIST dataset with confusion matrixes and classification reports. Firstly, our model without autoencoder gives the following results, as seen in Figure 25 for the confusion matrix and the classification report. The results with the autoencoder are presented in Figure 26. Note that these confusion matrixes and classification reports are indicated before any attack.

Fast Gradient Sign Method:

There is a slight difference between the neural network models with autoencoder and without autoencoder model. We apply the FGSM attack on both methods. The method uses the gradients of the loss accordingly for creating a new image that maximizes the loss. We can say the gradients are generated accordingly to input images. For these reasons, the FGSM causes a wide variety of models to misclassify their input [19].

As we expect due to results from multi-class logistic regression, autoencoder gives robustness to the neural network model too. After the DGSM, the neural network without an autoencoder suffers an immense drop in its accuracy,

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	974	0	5	0	1	2	7	1	5	5
1	0	1125	4	0	0	0	3	4	0	2
2	0	1	1003	1	0	0	2	8	1	0
3	1	4	3	1004	0	7	0	2	0	1
4	0	0	3	0	970	0	7	0	3	14
5	0	0	0	1	0	873	4	0	1	2
6	1	2	0	0	4	5	930	0	0	0
7	2	0	8	0	1	2	0	1005	4	8
8	2	3	6	4	1	2	5	3	957	4
9	0	0	0	0	5	1	0	5	3	973

	Precision	Recall	F1-Score	Support
0	0.99	0.97	0.98	1000
1	0.99	0.99	0.99	1138
2	0.97	0.99	0.98	1016
3	0.99	0.98	0.99	1022
4	0.99	0.97	0.98	997
5	0.98	0.99	0.98	881
6	0.97	0.99	0.98	942
7	0.98	0.98	0.98	1030
8	0.98	0.97	0.98	987
9	0.96	0.99	0.97	987
Micro Avg	0.98	0.98	0.98	10000
Macro Avg	0.98	0.98	0.98	10000
Weighted Avg	0.98	0.98	0.98	10000

Figure 25: Confusion matrix and classification report of the neural network model without autoencoder

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	966	0	4	0	0	2	3	0	3	3
1	0	1122	2	1	1	0	2	6	2	3
2	3	3	1013	7	3	0	2	13	2	2
3	0	0	3	982	0	5	0	8	7	2
4	0	0	1	1	954	1	7	0	4	6
5	1	2	0	10	0	874	4	1	5	6
6	4	4	1	1	2	5	937	0	2	0
7	1	1	4	3	8	2	0	990	2	8
8	3	3	3	3	3	2	3	2	945	7
9	2	0	1	2	11	1	0	8	2	972

	Precision	Recall	F1-Score	Support
0	0.99	0.98	0.99	981
1	0.99	0.99	0.99	1139
2	0.98	0.97	0.97	1048
3	0.97	0.98	0.97	1007
4	0.97	0.98	0.98	974
5	0.98	0.97	0.97	903
6	0.98	0.98	0.98	956
7	0.96	0.97	0.97	1019
8	0.97	0.97	0.97	974
9	0.96	0.97	0.97	999
Micro Avg	0.98	0.98	0.98	10000
Macro Avg	0.98	0.98	0.98	10000
Weighted Avg	0.98	0.98	0.98	10000

Figure 26: Confusion matrix and classification report of the neural network model with autoencoder

and the FGSM works as intended. But the neural network model with autoencoder only suffers a 0.01 percent accuracy drop.

Targeted Fast Gradient Sign Method: There is a directed type of FGSM, called T-FGSM. It uses the same principles to maximize the loss of the target. In this method, a gradient step is computed for giving the same misprediction for different inputs.

In the confusion matrix, the target value for this attack is number 5. The neural network model with the autoencoder is still at the accuracy of 0.98. The individual differences

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	80	1	42	7	16	11	68	7	24	14
1	2	5	125	5	34	6	20	16	30	11
2	177	127	73	120	43	3	47	95	264	5
3	19	13	344	50	7	337	54	504	234	171
4	17	538	35	2	85	1	356	47	18	295
5	68	2	6	351	1	99	177	3	185	63
6	275	8	9	0	32	70	71	0	38	1
7	20	215	177	64	228	7	7	40	48	318
8	109	223	206	303	69	253	154	68	16	105
9	213	3	15	108	467	105	4	248	117	26

	Precision	Recall	F1-Score	Support
0	0.08	0.30	0.13	270
1	0.00	0.02	0.01	254
2	0.07	0.08	0.07	954
3	0.05	0.03	0.04	1733
4	0.09	0.06	0.07	1394
5	0.11	0.10	0.11	955
6	0.07	0.14	0.10	504
7	0.04	0.04	0.04	1124
8	0.02	0.01	0.01	1506
9	0.03	0.02	0.02	1306
Micro Avg	0.05	0.05	0.05	10000
Macro Avg	0.06	0.08	0.06	10000
Weighted Avg	0.05	0.05	0.05	10000

Figure 27: Confusion matrix and classification report of the neural network model without autoencoder after FGSM attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	966	0	5	0	0	2	2	1	3	2
1	0	1122	3	0	3	0	2	5	1	3
2	3	2	1009	8	4	0	2	11	2	3
3	0	0	4	980	0	5	0	9	7	2
4	0	1	1	1	956	2	8	1	4	8
5	1	2	0	11	0	872	3	1	7	5
6	4	4	1	1	2	5	939	0	2	0
7	1	2	4	3	5	2	0	988	2	8
8	3	2	4	4	2	3	2	2	942	8
9	2	0	1	2	10	1	0	10	4	970

	Precision	Recall	F1-Score	Support
0	0.99	0.98	0.99	981
1	0.99	0.99	0.99	1139
2	0.98	0.97	0.97	1044
3	0.97	0.97	0.97	1007
4	0.97	0.97	0.97	982
5	0.98	0.97	0.97	902
6	0.98	0.98	0.98	958
7	0.96	0.97	0.97	1015
8	0.97	0.97	0.97	972
9	0.96	0.97	0.97	1000
Micro Avg	0.97	0.97	0.97	10000
Macro Avg	0.97	0.97	0.97	10000
Weighted Avg	0.97	0.97	0.97	10000

Figure 28: Confusion matrix and classification report of the neural network model with autoencoder after FGSM attack

are presented when compare with Figure 26.

Basic Iterative Method:

BIM is an extension of FGSM to apply it multiple times with iterations. It provides the recalculation of a gradient attack for each iteration.

This is the most damaging attack for the neural network model that takes its inputs directly from the MNIST Dataset without an autoencoder. The findings from Figure 31 show that the accuracy drops between 0.01 and 0.02 percent. The neural network model with autoencoder's ac-

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	3	0
2	0	0	0	0	1	0	0	23	7	39
3	8	6	180	0	59	1	8	7	6	65
4	0	0	0	0	0	0	0	0	0	0
5	972	1119	844	1004	906	890	947	982	956	871
6	0	0	0	0	1	0	0	0	0	12
7	0	1	2	0	5	0	0	0	1	20
8	0	9	6	6	0	0	1	15	1	2
9	0	0	0	0	10	0	2	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	1
1	0.00	0.00	0.00	4
2	0.00	0.00	0.00	70
3	0.00	0.00	0.00	340
4	0.00	0.00	0.00	0
5	1.00	0.09	0.17	9491
6	0.00	0.00	0.00	13
7	0.00	0.00	0.00	29
8	0.00	0.03	0.00	40
9	0.00	0.00	0.00	12
Micro Avg	0.09	0.09	0.09	10000
Macro Avg	0.10	0.01	0.02	10000
Weighted Avg	0.95	0.09	0.16	10000

Figure 29: Confusion matrix and classification report of the neural network model without autoencoder after T-FGSM attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	965	0	4	0	0	2	3	0	3	3
1	0	1123	2	1	1	0	2	7	0	3
2	3	2	1013	7	3	0	1	13	2	2
3	0	0	4	981	0	2	0	7	7	2
4	0	0	1	0	958	2	8	0	4	6
5	1	2	0	14	0	878	7	1	10	6
6	4	4	0	0	2	5	934	0	2	0
7	2	1	3	3	6	1	0	989	2	7
8	3	3	4	3	1	1	3	2	942	6
9	2	0	1	1	11	1	0	9	2	974

	Precision	Recall	F1-Score	Support
0	0.98	0.98	0.98	980
1	0.99	0.99	0.99	1139
2	0.98	0.97	0.97	1046
3	0.97	0.98	0.97	1003
4	0.98	0.98	0.98	979
5	0.98	0.96	0.97	919
6	0.97	0.98	0.98	951
7	0.96	0.98	0.97	1014
8	0.97	0.97	0.97	968
9	0.97	0.97	0.97	1001
Micro Avg	0.98	0.98	0.98	10000
Macro Avg	0.98	0.98	0.98	10000
Weighted Avg	0.98	0.98	0.98	10000

Figure 30: Confusion matrix and classification report of the neural network model with autoencoder after T-FGSM attack

curacy stays as 0.97 percent, losing only 0.1 percent.

Findings indicate that autoencoding before giving dataset as input to linear models and neural network models improve robustness against adversarial attacks significantly. We use vanilla autoencoders. They are the basic autoencoders without modification. In the other sections, we apply the same attacks with the same machine learning models with different autoencoder types.

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	4	1	37	7	21	11	66	7	23	17
1	1	4	125	3	32	3	22	18	31	10
2	201	138	24	132	40	2	51	96	258	4
3	15	12	350	4	8	350	65	492	251	181
4	19	533	42	3	11	2	385	43	19	300
5	48	2	5	342	3	15	160	3	168	58
6	284	8	11	0	47	72	20	0	40	0
7	25	191	184	70	221	7	7	21	45	296
8	136	243	232	323	98	304	178	61	15	119
9	247	3	22	126	501	126	4	287	124	24

	Precision	Recall	F1-Score	Support
0	0.00	0.02	0.01	194
1	0.00	0.02	0.01	249
2	0.02	0.03	0.02	946
3	0.00	0.00	0.00	1728
4	0.01	0.01	0.01	1357
5	0.02	0.02	0.02	804
6	0.02	0.04	0.03	482
7	0.02	0.02	0.02	1067
8	0.02	0.01	0.01	1709
9	0.02	0.02	0.02	1464
Micro Avg	0.01	0.01	0.01	10000
Macro Avg	0.01	0.02	0.01	10000
Weighted Avg	0.02	0.01	0.01	10000

Figure 31: Confusion matrix and classification report of the neural network model without autoencoder after basic iterative method attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	967	0	3	0	0	2	2	1	4	2
1	0	1123	2	0	2	0	2	5	0	3
2	3	1	1008	7	4	0	0	11	3	3
3	0	1	4	983	0	4	0	9	8	2
4	0	1	2	1	959	3	8	2	4	10
5	0	2	0	11	0	872	6	0	7	5
6	4	4	2	0	2	6	936	0	2	0
7	2	1	6	3	3	1	0	989	3	5
8	2	2	4	5	2	3	4	1	938	7
9	2	0	1	0	10	1	0	10	5	972

	Precision	Recall	F1-Score	Support
0	0.99	0.99	0.99	981
1	0.99	0.99	0.99	1137
2	0.98	0.97	0.97	1040
3	0.97	0.97	0.97	1011
4	0.98	0.97	0.97	990
5	0.98	0.97	0.97	903
6	0.98	0.98	0.98	956
7	0.96	0.98	0.97	1013
8	0.96	0.97	0.97	968
9	0.96	0.97	0.97	1001
Micro Avg	0.97	0.97	0.97	10000
Macro Avg	0.97	0.97	0.97	10000
Weighted Avg	0.97	0.97	0.97	10000

Figure 32: Confusion matrix and classification report of the neural network model with autoencoder after basic iterative method attack

5.3 Sparse Autoencoder

Sparse autoencoders present improved performance on classification tasks. It includes more hidden layers than the input layer. The significant part is defining a small number of hidden layers to be active at once to encourage sparsity. This constraint forces the training model to respond uniquely to the characteristics of translation and uses the statistical features of the input data.

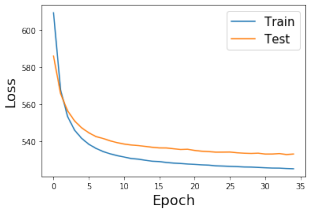


Figure 33: Optimized Relu Loss History for Sparse Autoencoder

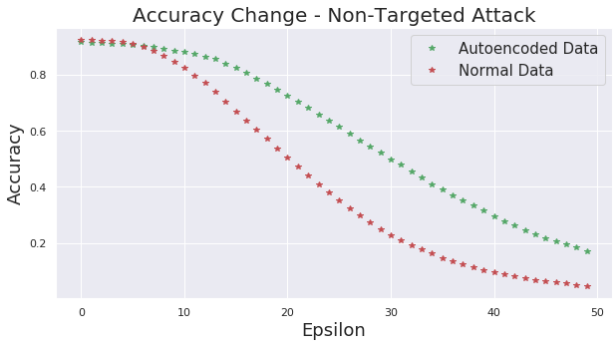


Figure 34: Comparison of accuracy with and without sparse autoencoder for non-targeted attack

Because of this sparse autoencoders involve sparsity penalty $\Omega(h)$ in their training. $L(x, x') + \Omega(h)$

This penalty makes the model to activate specific areas of the network depending on the input data while making all other neurons inactive. We can create this sparsity by relative entropy, also known as Kullback-Leibler divergence.

$\widehat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m}[h_j(x_i)]$ $\widehat{\rho}_j$ is our average activation function of the hidden layer j which is averaged over m training examples. For increasing the sparsity in terms of making the number of active neurons as smaller as it can be, we would want ρ close to zero. The sparsity penalty term $\Omega(h)$ will punish $\widehat{\rho}_j$ for deviating from ρ, which will be basically exploiting Kullback-Leibler divergence. $KL(p||\widehat{\rho}_j)$ is our Kullback-Leibler divergence between a random variable ρ and random variable with mean $\widehat{\rho}_j$.

$\sum_{j=1}^{s} KL(\rho||\widehat{\rho}_j) = \sum_{j=1}^{s}[\rho log\frac{\rho}{\widehat{\rho}_j} + (1-\rho)log\frac{1-\rho}{1-\widehat{\rho}_j}]$

Sparsity can be achieved with other ways, such as applying L1 and L2 regularization terms on the activation of the hidden layer. L is our loss function and λ is our scale parameter.

$L(x, x') + \lambda \sum_i |h_i|$

5.3.1 Multi-Class Logistic Regression of Sparse Autoencoder

This section presents multi-class logistic regressions with sparse autoencoders. The difference from the autoencoder section is the autoencoder type. The findings from Figure 6 and Figure 33 show that loss is higher compared to the autoencoders in sparse autoencoder.

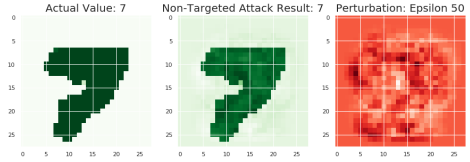


Figure 35: Value change and perturbation of a non-targeted attack on model without sparse autoencoder

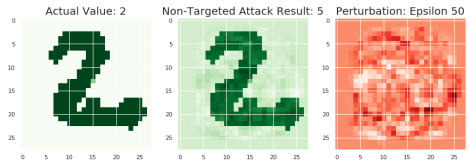


Figure 36: Value change and perturbation of a non-targeted attack on model with sparse autoencoder

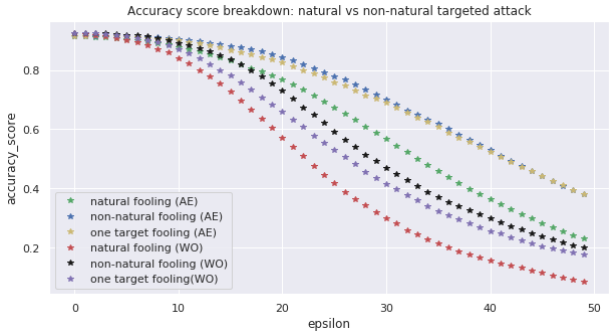


Figure 37: Comparison of accuracy with and without sparse autoencoder for targeted attacks. *AE stands for the models with sparse autoencoder, WO stands for models without autoencoder*

The difference between perturbation is presented in Figure 35 and Figure 36 compared to the perturbation in Figure 11 and Figure 12. The perturbation is sharper in sparse autoencoder.

Figure 37 indicates that sparse autoencoders performs poorly compared to autoencoders in multi-class logistic regression.

5.3.2 Neural Network of Sparse Autoencoder

Sparse autoencoder results for neural networks indicate that vanilla autoencoder seems to be slightly better than sparse autoencoders for neural networks. Sparse autoencoders do not perform as well in linear machine learning models, in our case, multi-class logistic regression.

5.4 Denoising Autoencoder

Denoising autoencoders are used for partially corrupted input and train it to recover the original undistorted input. In this study, the corrupted input is not used. The aim is to achieve a good design by changing the reconstruction prin-

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	972	0	1	0	1	2	5	1	6	4
1	0	1127	2	0	0	0	2	4	0	2
2	3	3	1019	6	4	0	3	10	6	1
3	0	0	0	996	0	10	0	4	0	3
4	0	0	1	0	965	0	3	0	1	10
5	0	0	0	2	0	865	1	0	2	3
6	1	2	0	0	3	6	942	0	0	0
7	1	1	7	2	2	2	0	1008	5	7
8	3	2	2	3	1	5	2	1	952	2
9	0	0	0	1	6	2	0	0	2	977

	Precision	Recall	F1-Score	Support
0	0.99	0.98	0.99	992
1	0.99	0.99	0.99	1137
2	0.99	0.97	0.98	1055
3	0.99	0.98	0.98	1013
4	0.98	0.98	0.98	980
5	0.97	0.99	0.98	873
6	0.98	0.99	0.99	954
7	0.98	0.97	0.98	1035
8	0.98	0.98	0.98	973
9	0.97	0.99	0.98	988
Micro Avg	0.98	0.98	0.98	10000
Macro Avg	0.98	0.98	0.98	10000
Weighted Avg	0.98	0.98	0.98	10000

Figure 38: Confusion matrix and classification report of the neural network model without sparse autoencoder

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	967	0	9	0	2	5	2	1	5	4
1	0	1118	1	0	0	1	3	8	0	2
2	2	4	996	7	2	0	1	14	8	0
3	0	2	10	977	0	18	1	3	8	8
4	0	1	1	0	956	2	5	3	1	12
5	1	0	0	6	0	846	3	0	4	7
6	4	2	1	0	6	6	940	0	1	0
7	1	0	6	5	2	0	0	983	7	12
8	5	7	6	11	3	7	2	4	936	3
9	0	1	2	4	11	7	1	12	4	961

	Precision	Recall	F1-Score	Support
0	0.99	0.97	0.98	995
1	0.99	0.99	0.99	1133
2	0.97	0.96	0.96	1034
3	0.97	0.95	0.96	1027
4	0.97	0.97	0.97	981
5	0.95	0.98	0.96	867
6	0.98	0.98	0.98	960
7	0.96	0.97	0.96	1016
8	0.96	0.95	0.96	984
9	0.95	0.96	0.96	1003
Micro Avg	0.97	0.97	0.97	10000
Macro Avg	0.97	0.97	0.97	10000
Weighted Avg	0.97	0.97	0.97	10000

Figure 39: Confusion matrix and classification report of the neural network model with sparse autoencoder

ciple for using denoising autoencoders. For achieving this denoising properly, the model requires to extract features that capture useful structure in the distribution of the input. Denoising autoencoders apply corrupted data through stochastic mapping. Our input is x and corrupted data is $\widetilde{x}$ and stochastic mapping is $\widetilde{x} \sim q_D(\widetilde{x}|x)$.

As its a standard autoencoder, corrupted data $\widetilde{x}$ is mapped to a hidden layer.

$h = f_\theta(\widetilde{x}) = s(W\widetilde{x} + b)$.

And from this the model reconstructs $z = g'_\theta(h)$.

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	54	0	29	5	2	13	111	12	20	10
1	0	4	114	4	54	8	7	35	38	17
2	369	416	154	295	77	14	222	252	510	14
3	14	12	315	36	5	338	58	297	80	212
4	2	182	19	2	63	1	161	20	10	188
5	41	1	4	329	11	80	185	1	117	47
6	276	9	11	1	48	89	120	0	57	3
7	22	203	183	72	288	7	0	57	80	411
8	108	308	195	188	73	249	89	83	16	88
9	94	0	8	78	361	93	5	271	46	19

	Precision	Recall	F1-Score	Support
0	0.06	0.21	0.09	256
1	0.00	0.01	0.01	281
2	0.15	0.07	0.09	2323
3	0.04	0.03	0.03	1367
4	0.06	0.10	0.08	648
5	0.09	0.10	0.09	816
6	0.13	0.20	0.15	614
7	0.06	0.04	0.05	1323
8	0.02	0.01	0.01	1397
9	0.02	0.02	0.02	975
Micro Avg	0.06	0.06	0.06	10000
Macro Avg	0.06	0.08	0.06	10000
Weighted Avg	0.07	0.06	0.06	10000

Figure 40: Confusion matrix and classification report of the neural network model without sparse autoencoder after FGSM attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	966	0	7	0	2	4	4	2	3	4
1	0	1121	0	1	1	2	3	8	0	4
2	3	7	996	6	2	0	1	15	9	1
3	1	0	10	976	0	17	2	4	10	6
4	0	1	1	0	952	0	7	3	2	15
5	1	0	0	8	0	849	3	0	5	6
6	6	2	2	0	7	6	934	0	2	0
7	0	1	4	3	2	1	0	977	9	15
8	3	3	9	11	4	8	3	6	930	5
9	0	0	3	5	12	5	1	13	4	953

	Precision	Recall	F1-Score	Support
0	0.99	0.97	0.98	992
1	0.99	0.98	0.99	1140
2	0.97	0.96	0.96	1040
3	0.97	0.95	0.96	1026
4	0.97	0.97	0.97	981
5	0.95	0.97	0.96	872
6	0.97	0.97	0.97	959
7	0.95	0.97	0.96	1012
8	0.95	0.95	0.95	982
9	0.94	0.96	0.95	996
Micro Avg	0.97	0.97	0.97	10000
Macro Avg	0.97	0.97	0.97	10000
Weighted Avg	0.97	0.97	0.97	10000

Figure 41: Confusion matrix and classification report of the neural network model with sparse autoencoder after FGSM attack

5.4.1 Multi-Class Logistic Regression of Denoising Autoencoder

In denoising autoencoder for multi-class logistic regression, the loss does not improve for each epoch. Although it starts better at lower epoch values, in the end, vanilla autoencoder seems to be better. Sparse autoencoder's loss is slightly worse.

And just like sparse autoencoder, denoising autoencoder

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	4	1	1	0	0	14	2	9
2	0	5	1	1	2	0	0	34	4	0
3	0	0	0	0	0	0	0	16	1	0
4	0	0	0	0	0	0	0	0	0	0
5	980	1130	1023	1007	976	892	958	961	967	998
6	0	0	0	0	0	0	0	0	0	0
7	0	0	4	1	3	0	0	0	0	2
8	0	0	0	0	0	0	0	3	0	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	31
2	0.00	0.02	0.00	47
3	0.00	0.00	0.00	17
4	0.00	0.00	0.00	0
5	1.00	0.09	0.17	9892
6	0.00	0.00	0.00	0
7	0.00	0.00	0.00	10
8	0.00	0.00	0.00	3
9	0.00	0.00	0.00	0
Micro Avg	0.09	0.09	0.09	10000
Macro Avg	0.10	0.01	0.02	10000
Weighted Avg	0.99	0.09	0.16	10000

Figure 42: Confusion matrix and classification report of the neural network model without sparse autoencoder after T-FGSM attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	966	0	9	0	1	3	4	1	4	4
1	0	1121	1	0	1	1	3	9	0	3
2	2	5	998	7	2	0	1	15	7	0
3	0	1	9	974	0	14	1	3	9	7
4	0	1	1	0	954	0	5	3	1	14
5	1	0	0	9	0	862	6	0	5	11
6	5	2	2	0	7	5	935	0	0	0
7	1	0	5	5	2	0	0	981	7	12
8	5	4	6	11	4	3	2	5	938	4
9	0	1	1	4	11	4	1	11	3	954

	Precision	Recall	F1-Score	Support
0	0.99	0.97	0.98	992
1	0.99	0.98	0.99	1139
2	0.97	0.96	0.96	1037
3	0.96	0.96	0.96	1018
4	0.97	0.97	0.97	979
5	0.97	0.96	0.97	894
6	0.98	0.98	0.98	956
7	0.95	0.97	0.96	1013
8	0.96	0.96	0.96	982
9	0.95	0.96	0.95	990
Micro Avg	0.97	0.97	0.97	10000
Macro Avg	0.97	0.97	0.97	10000
Weighted Avg	0.97	0.97	0.97	10000

Figure 43: Confusion matrix and classification report of the neural network model with sparse autoencoder after T-FGSM attack

also applies a sharp perturbation, which is presented in Figure 48 and Figure 49.

We observe that there is a similarity between accuracy results for denoising autoencoder with multi-class logistic regression and sparse autoencoder results. Natural fooling accuracy drops drastically in denoising autoencoder, but non-targeted and one targeted attack seem to be somewhat like sparse autoencoder, one targeted attack having less accuracy in denoising autoencoder.

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	4	0	35	4	2	13	122	11	22	10
1	0	4	111	4	33	8	6	32	37	9
2	377	360	10	295	65	5	246	205	494	14
3	14	12	398	11	6	337	68	315	98	211
4	2	150	23	2	15	0	226	20	9	199
5	37	0	2	330	6	23	177	2	110	45
6	299	11	15	1	56	103	11	0	59	5
7	19	223	206	72	278	7	1	18	78	392
8	118	374	218	190	89	272	94	95	16	102
9	110	1	14	101	432	124	7	330	51	22

	Precision	Recall	F1-Score	Support
0	0.00	0.02	0.01	223
1	0.00	0.02	0.01	244
2	0.01	0.00	0.01	2071
3	0.01	0.01	0.01	1470
4	0.02	0.02	0.02	646
5	0.03	0.03	0.03	732
6	0.01	0.02	0.01	560
7	0.02	0.01	0.02	1294
8	0.02	0.01	0.01	1568
9	0.02	0.02	0.02	1192
Micro Avg	0.01	0.01	0.01	10000
Macro Avg	0.01	0.02	0.01	10000
Weighted Avg	0.01	0.01	0.01	10000

Figure 44: Confusion matrix and classification report of the neural network model without sparse autoencoder after basic iterative method attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	964	0	6	0	2	4	4	2	3	4
1	0	1119	0	1	1	1	3	8	0	3
2	3	8	998	6	2	0	1	14	8	1
3	1	0	10	972	0	21	2	5	12	7
4	0	1	1	0	955	1	11	2	2	19
5	1	0	0	8	0	844	4	0	5	7
6	7	2	2	0	7	7	929	0	2	0
7	0	1	4	5	2	1	0	974	7	19
8	4	4	9	13	4	9	3	7	931	5
9	0	0	2	5	9	4	1	16	4	944

	Precision	Recall	F1-Score	Support
0	0.98	0.97	0.98	989
1	0.99	0.99	0.99	1136
2	0.97	0.96	0.96	1041
3	0.96	0.94	0.95	1030
4	0.97	0.96	0.97	992
5	0.95	0.97	0.96	869
6	0.97	0.97	0.97	956
7	0.95	0.96	0.95	1013
8	0.96	0.94	0.95	989
9	0.94	0.96	0.95	985
Micro Avg	0.96	0.96	0.96	10000
Macro Avg	0.96	0.96	0.96	10000
Weighted Avg	0.96	0.96	0.96	10000

Figure 45: Confusion matrix and classification report of the neural network model with sparse autoencoder after basic iterative method attack

5.4.2 Neural Network of Denoising Autoencoder

We investigate that neural network accuracy for denoising autoencoder is worse than sparse autoencoder results and vanilla autoencoder results. It is still a useful autoencoder for denoising corrupted data and other purposes; however, it is not the right choice just for robustness against adversarial examples.

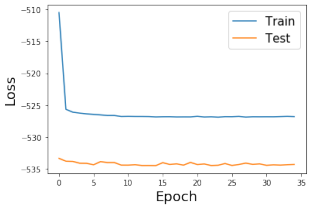


Figure 46: Optimized Relu Loss History for Denoising Autoencoder

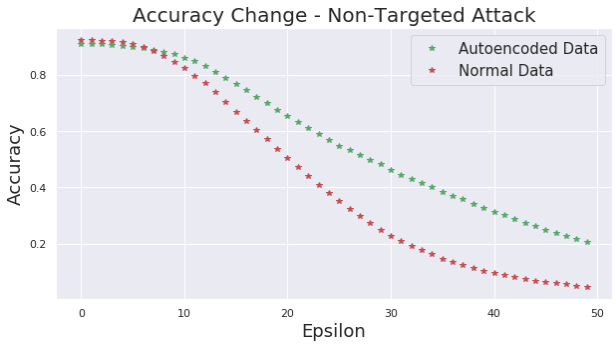


Figure 47: Comparison of accuracy with and without denoising autoencoder for non-targeted attack

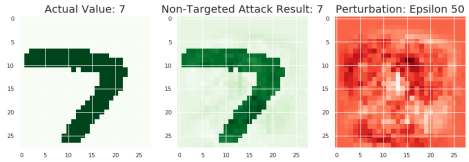


Figure 48: Value change and perturbation of a non-targeted attack on model without denoising autoencoder

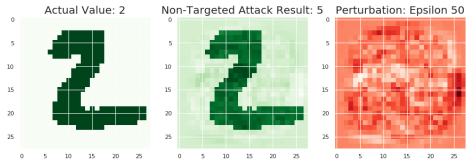


Figure 49: Value change and perturbation of a non-targeted attack on model with denoising autoencoder

5.5 Variational Autoencoder

In this study, we examine variational autoencoders as the final type of autoencoder type. The variational autoencoders have an encoder and a decoder, although their mathematical formulation differs significantly. They are associated with Generative Adversarial Networks due to their architectural similarity. In summary, variational autoencoders are also generative models. Differently, from sparse autoencoders, denoising autoencoders, and vanilla autoencoders, all of which aim discriminative modeling, generative modeling tries to simulate how the data can be generated and to understand the underlying causal relations. It also considers these causal relations when generating new

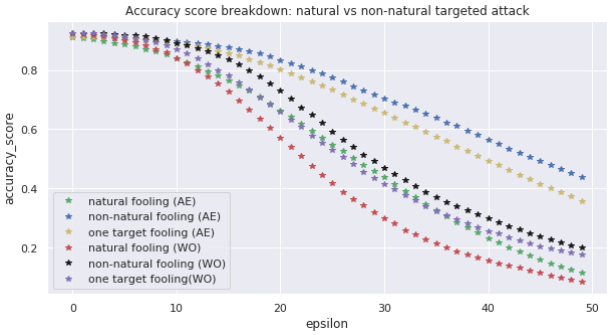


Figure 50: Comparison of accuracy with and without denoising autoencoder for targeted attacks. *AE stands for the models with denoising autoencoder, WO stands for models without autoencoder*

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	974	0	3	0	0	2	7	1	5	3
	1	0	1125	5	0	0	0	3	2	0	3
	2	1	4	1009	4	1	0	2	10	3	0
	3	0	2	0	999	0	5	0	3	1	2
	4	0	0	2	0	973	0	3	0	1	10
	5	0	0	0	3	0	877	1	0	1	4
	6	1	1	0	0	3	4	938	0	0	0
	7	1	1	10	2	0	2	0	1006	4	5
	8	3	2	3	2	1	2	4	2	957	3
	9	0	0	0	0	4	0	0	4	2	979

	Precision	Recall	F1-Score	Support
0	0.99	0.98	0.99	995
1	0.99	0.99	0.99	1138
2	0.98	0.98	0.98	1034
3	0.99	0.99	0.99	1012
4	0.99	0.98	0.99	989
5	0.98	0.99	0.99	886
6	0.98	0.99	0.98	947
7	0.98	0.98	0.98	1031
8	0.98	0.98	0.98	979
9	0.97	0.99	0.98	989
Micro Avg	0.98	0.98	0.98	10000
Macro Avg	0.98	0.98	0.98	10000
Weighted Avg	0.98	0.98	0.98	10000

Figure 51: Confusion matrix and classification report of the neural network model without denoising autoencoder

data.

Variational autoencoders use an estimator algorithm called Stochastic Gradient Variational Bayes for training. This algorithm assumes the data is generated by $p_\theta(x|h)$ which is a directed graphical model and θ being the parameters of decoder, in variational autoencoder's case, the parameters of the generative model. The encoder is learning an approximation of $q_\phi(h|x)$ to a posterior distribution which is showed by $p_\theta(x|h)$ and ϕ being the parameters of the encoder; in variational autoencoder's case, the parameters of recognition model. We will use Kullback-Leibler divergence again, showed as D_{KL}.

$L = (\phi, \theta, x) = D_{KL}(q_\phi(h|x)||p_\theta(h)) - \mathbb{E}_{q_\phi(h|x)}(logp_\theta(x|h))$.

Variational and likelihood distributions' shape is chosen by factorized Gaussians. The encoder outputs are $p(x)$ and $w^2(x)$. The decoder outputs are $\mu(h)$ and $\sigma^2(h)$. The like-

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	961	0	2	0	1	2	7	1	7	3
1	1	1119	7	0	2	0	2	8	2	3
2	5	3	992	7	3	0	1	10	6	0
3	1	1	10	980	1	16	1	8	20	10
4	1	0	1	0	937	3	1	0	5	12
5	3	1	0	7	0	852	5	0	5	4
6	6	5	1	1	6	7	937	0	3	0
7	1	0	11	7	4	0	0	984	3	11
8	1	5	8	5	2	7	3	4	913	7
9	0	1	0	3	26	5	1	13	10	959

	Precision	Recall	F1-Score	Support
0	0.98	0.98	0.98	984
1	0.99	0.98	0.98	1144
2	0.96	0.97	0.96	1027
3	0.97	0.94	0.95	1048
4	0.95	0.98	0.96	960
5	0.96	0.97	0.96	877
6	0.98	0.97	0.97	966
7	0.96	0.96	0.96	1021
8	0.94	0.96	0.95	955
9	0.95	0.94	0.95	1018
Micro Avg	0.96	0.96	0.96	10000
Macro Avg	0.96	0.96	0.96	10000
Weighted Avg	0.96	0.96	0.96	10000

Figure 52: Confusion matrix and classification report of the neural network model with denoising autoencoder

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	82	0	22	7	15	5	71	4	18	7
1	0	8	121	5	85	2	6	42	39	18
2	200	365	86	255	54	15	141	210	304	8
3	9	12	352	54	6	302	47	329	155	114
4	9	272	36	2	69	0	282	50	14	280
5	83	7	8	345	19	92	212	6	232	78
6	303	20	8	1	25	56	77	0	53	1
7	11	206	181	68	278	2	1	56	57	350
8	146	244	213	218	84	352	120	77	16	141
9	137	1	5	55	347	66	1	254	86	12

	Precision	Recall	F1-Score	Support
0	0.08	0.35	0.14	231
1	0.01	0.02	0.01	326
2	0.08	0.05	0.06	1638
3	0.05	0.04	0.05	1380
4	0.07	0.07	0.07	1014
5	0.10	0.09	0.09	1082
6	0.08	0.14	0.10	544
7	0.05	0.05	0.05	1210
8	0.02	0.01	0.01	1611
9	0.01	0.01	0.01	964
Micro Avg	0.06	0.06	0.06	10000
Macro Avg	0.06	0.08	0.06	10000
Weighted Avg	0.06	0.06	0.05	10000

Figure 53: Confusion matrix and classification report of the neural network model without denoising autoencoder after FGSM attack

lihood term of variational objective is defined below.

$q_\phi(h|x) = N(p(x), w^2(x)I)$
$p_\theta(x|h) = N(\mu(h), \sigma^2(h)I)$

5.5.1 Multi-Class Logistic Regression of Variational Autoencoder

The findings from Figure 59 show that variational autoencoder indicates the best loss function result. However, Figure 60 presents that the accuracy is low, especially in low

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	961	0	1	0	2	2	7	1	6	3
1	1	1120	7	1	3	0	2	7	2	4
2	5	3	993	7	2	0	1	10	7	0
3	1	1	11	977	2	16	1	10	18	9
4	1	0	1	0	935	3	1	1	6	11
5	3	1	0	7	0	855	4	0	6	3
6	5	4	2	1	7	5	938	0	3	0
7	1	0	9	8	4	0	0	981	3	11
8	1	6	8	6	3	7	3	4	914	7
9	1	0	0	3	24	4	1	14	9	961

	Precision	Recall	F1-Score	Support
0	0.98	0.98	0.98	983
1	0.99	0.98	0.98	1147
2	0.96	0.97	0.96	1028
3	0.97	0.93	0.95	1046
4	0.95	0.97	0.96	959
5	0.96	0.97	0.97	879
6	0.98	0.97	0.98	965
7	0.95	0.96	0.96	1017
8	0.94	0.95	0.95	959
9	0.95	0.94	0.95	1017
Micro Avg	0.96	0.96	0.96	10000
Macro Avg	0.96	0.96	0.96	10000
Weighted Avg	0.96	0.96	0.96	10000

Figure 54: Confusion matrix and classification report of the neural network model with denoising autoencoder after FGSM attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	3	0	0	0	1	0
2	0	10	0	0	1	0	0	0	2	0
3	0	0	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	0	0	0	0
5	980	1125	1032	1010	976	892	957	1026	971	1009
6	0	0	0	0	2	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	4
2	0.00	0.00	0.00	13
3	0.00	0.00	0.00	2
4	0.00	0.00	0.00	0
5	1.00	0.09	0.16	9978
6	0.00	0.00	0.00	2
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	1
9	0.00	0.00	0.00	0
Micro Avg	0.09	0.09	0.09	10000
Macro Avg	0.10	0.01	0.02	10000
Weighted Avg	1.00	0.09	0.16	10000

Figure 55: Confusion matrix and classification report of the neural network model without denoising autoencoder after T-FGSM attack

epsilon values where even autoencoded data gives worse accuracy than the normal learning process.

Perturbation applied by variational autoencoder is not as sharp in sparse autoencoder and denoising autoencoder. It seems similar to vanilla autoencoder's perturbation.

The variational autoencoder has the worst results. Besides, it presents bad results at the low values of epsilon, making autoencoded data less accurate and only a slight improvement compared to the normal data in high values of epsilon.

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	340	374	1018	773	30	88	6	14	571	28
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	621	274	2	20	269	579	949	0	218	27
7	19	487	12	217	683	225	3	1014	185	954
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	0.99	0.31	0.48	3242
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
5	0.00	0.00	0.00	0
6	0.99	0.32	0.48	2959
7	0.99	0.27	0.42	3799
8	0.00	0.00	0.00	0
9	0.00	0.00	0.00	0
Micro Avg	0.30	0.30	0.30	10000
Macro Avg	0.30	0.09	0.14	10000
Weighted Avg	0.99	0.30	0.46	10000

Figure 56: Confusion matrix and classification report of the neural network model with denoising autoencoder after T-FGSM attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	2	0	23	6	18	5	67	4	18	7
1	0	7	116	3	57	2	7	38	37	12
2	207	323	20	263	34	5	165	190	291	5
3	5	4	384	8	5	310	37	337	163	108
4	10	273	40	1	7	0	304	49	15	285
5	56	1	5	336	14	14	216	5	216	67
6	331	15	8	1	31	64	12	0	58	1
7	12	202	190	76	289	2	0	18	59	346
8	184	308	239	238	120	415	149	84	14	161
9	173	2	7	78	407	75	1	303	103	17

	Precision	Recall	F1-Score	Support
0	0.00	0.01	0.00	150
1	0.01	0.03	0.01	279
2	0.02	0.01	0.02	1503
3	0.01	0.01	0.01	1361
4	0.01	0.01	0.01	984
5	0.02	0.02	0.02	930
6	0.01	0.02	0.02	521
7	0.02	0.02	0.02	1194
8	0.01	0.01	0.01	1912
9	0.02	0.01	0.02	1166
Micro Avg	0.01	0.01	0.01	10000
Macro Avg	0.01	0.01	0.01	10000
Weighted Avg	0.01	0.01	0.01	10000

Figure 57: Confusion matrix and classification report of the neural network model without denoising autoencoder after basic iterative method attack

5.5.2 Neural Network of Variational Autoencoder

Variational autoencoder with neural networks also illustrates the worst results compared to other autoencoder types, where the accuracy for autoencoded data against an attack has around between 0.96 and 0.99 accuracies, variational autoencoder has around between 0.65 and 0.70 accuracies.

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	351	391	1017	773	30	89	6	16	577	28
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	609	273	3	17	261	575	949	0	212	26
7	20	471	12	220	691	228	3	1012	185	955
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	0.99	0.31	0.47	3278
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
5	0.00	0.00	0.00	0
6	0.99	0.32	0.49	2925
7	0.98	0.27	0.42	3797
8	0.00	0.00	0.00	0
9	0.00	0.00	0.00	0
Micro Avg	0.30	0.30	0.30	10000
Macro Avg	0.30	0.09	0.14	10000
Weighted Avg	0.99	0.30	0.46	10000

Figure 58: Confusion matrix and classification report of the neural network model with denoising autoencoder after basic iterative method attack

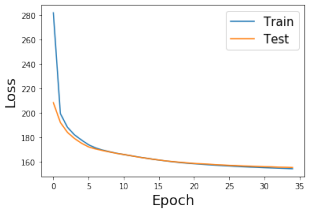


Figure 59: Optimized Relu Loss History for Variational Autoencoder

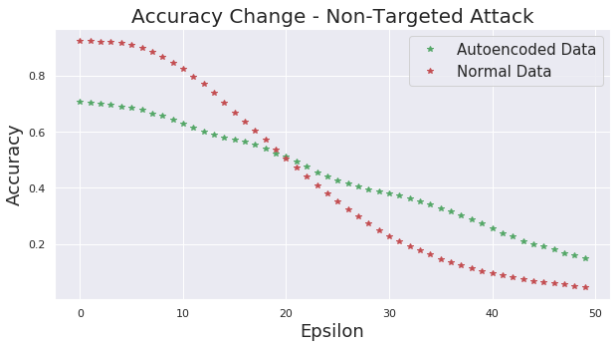


Figure 60: Comparison of accuracy with and without variational autoencoder for non-targeted attack

6 Experiments with Fashion MNIST Dataset

6.1 Introduction

We also used Fashion MNIST dataset. We will briefly show the experiment results for not filling the paper with

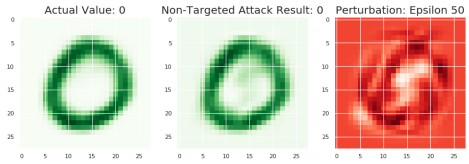


Figure 61: Value change and perturbation of a non-targeted attack on model without variational autoencoder

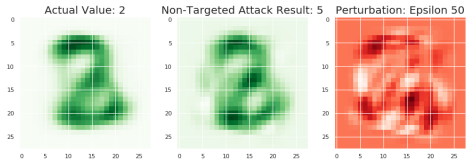


Figure 62: Value change and perturbation of a non-targeted attack on model with variational autoencoder

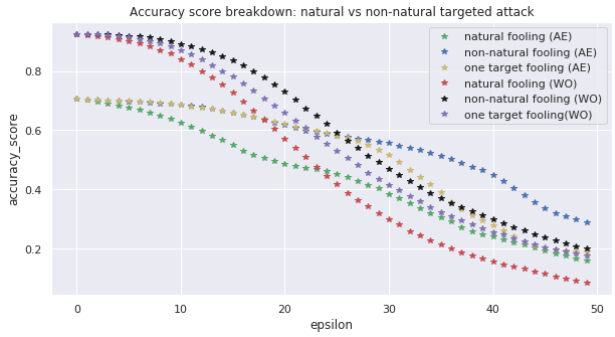


Figure 63: Comparison of accuracy with and without variational autoencoder for targeted attacks. *AE stands for the models with variational autoencoder, WO stands for models without autoencoder*

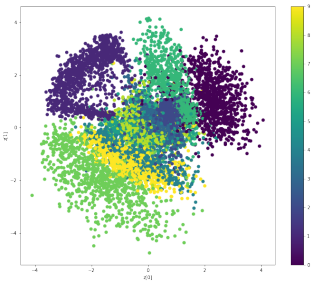


Figure 64: Because of MNIST dataset, our latent space is two-dimensional. One is to look at the neighbourhoods of different classes on the latent 2D plane. Each of these coloured clusters is a type of digit. Close clusters are structurally similar digits, and they are digits that share information in the latent space.

too many images. In these results, we have only changed the imported dataset. All the structure and code for the paper remains the same. Each training and test example is assigned to one of the following labels:

0. T-shirt/top

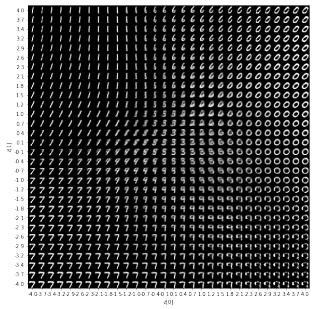


Figure 65: Due to VAE is a generative model, we can also generate new Mnist digits using latent plane, sampling latent points at regular intervals, and generating the corresponding digit for each of these points.

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	976	1	3	0	0	2	6	1	6	7
	1	0	1124	1	0	0	0	3	2	0	2
	2	0	4	1013	6	2	0	1	11	3	0
	3	0	0	1	994	0	4	1	2	0	1
	4	0	0	3	0	970	0	1	0	1	7
	5	0	0	0	2	0	876	5	0	2	3
	6	1	3	0	0	1	5	938	0	1	0
	7	1	2	6	2	1	2	0	1004	3	9
	8	2	1	4	3	2	2	2	2	954	1
	9	0	0	1	3	6	1	1	6	4	979

	Precision	Recall	F1-Score	Support
0	1.00	0.97	0.98	1002
1	0.99	0.99	0.99	1132
2	0.98	0.97	0.98	1040
3	0.98	0.99	0.99	1003
4	0.99	0.99	0.99	982
5	0.98	0.99	0.98	888
6	0.98	0.99	0.98	949
7	0.98	0.97	0.98	1030
8	0.98	0.98	0.98	973
9	0.97	0.98	0.97	1001
Micro Avg	0.98	0.98	0.98	10000
Macro Avg	0.98	0.98	0.98	10000
Weighted Avg	0.98	0.98	0.98	10000

Figure 66: Confusion matrix and classification report of the neural network model without variational autoencoder

1. Trouser
2. Pullover
3. Dress
4. Coat
5. Sandal
6. Shirt
7. Sneaker
8. Bag
9. Ankle boot

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	863	0	3	4	1	7	44	0	2	1
1	0	1095	2	10	2	8	8	11	17	4
2	27	3	810	102	12	49	64	2	37	12
3	8	5	105	673	9	227	5	6	189	12
4	1	0	15	9	688	35	2	66	33	383
5	6	10	22	55	13	190	16	7	108	2
6	69	6	46	3	8	17	815	0	6	4
7	0	0	0	1	2	0	0	745	0	87
8	6	15	27	143	57	350	3	39	578	25
9	0	1	2	10	190	9	1	152	4	479

	Precision	Recall	F1-Score	Support
0	0.88	0.93	0.91	925
1	0.96	0.95	0.96	1157
2	0.78	0.72	0.75	1118
3	0.67	0.54	0.60	1239
4	0.70	0.56	0.62	1232
5	0.21	0.44	0.29	429
6	0.85	0.84	0.84	974
7	0.72	0.89	0.80	835
8	0.59	0.47	0.52	1243
9	0.47	0.56	0.52	848
Micro Avg	0.69	0.69	0.69	10000
Macro Avg	0.69	0.69	0.68	10000
Weighted Avg	0.72	0.69	0.70	10000

Figure 67: Confusion matrix and classification report of the neural network model with variational autoencoder

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	96	4	47	10	20	15	109	8	20	12
1	1	7	106	11	103	3	5	35	42	25
2	243	239	116	210	60	7	143	157	454	10
3	15	129	386	52	7	309	25	396	103	215
4	7	291	38	1	42	0	252	37	13	299
5	48	1	0	310	5	62	178	8	112	53
6	255	14	15	0	35	69	103	0	43	2
7	18	284	131	45	249	7	4	47	46	288
8	47	165	163	172	71	294	134	74	17	85
9	250	1	30	199	390	126	5	266	124	20

	Precision	Recall	F1-Score	Support
0	0.10	0.28	0.15	341
1	0.01	0.02	0.01	338
2	0.11	0.07	0.09	1639
3	0.05	0.03	0.04	1637
4	0.04	0.04	0.04	980
5	0.07	0.08	0.07	777
6	0.11	0.19	0.14	536
7	0.05	0.04	0.04	1119
8	0.02	0.01	0.02	1222
9	0.02	0.01	0.02	1411
Micro Avg	0.06	0.06	0.06	10000
Macro Avg	0.06	0.08	0.06	10000
Weighted Avg	0.06	0.06	0.05	10000

Figure 68: Confusion matrix and classification report of the neural network model without variational autoencoder after FGSM attack

6.2 Autoencoding

6.2.1 Multi-Class Logistic Regression

The process for multi-class logistic regression for fashion MNIST is the same as it was in MNIST Dataset. We apply perturbation to clothes and shoes but it does not matter for the learning model as long as it is labeled correctly. In this particular perturbation example, a shirt is mispredicted as a trouser. We can also observe the line drawn by perturbation

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	858	0	4	3	2	6	47	0	3	1
1	0	1087	3	7	0	12	4	6	17	6
2	25	5	803	109	11	53	70	3	35	11
3	10	4	116	646	10	224	7	11	185	13
4	1	1	13	12	618	33	3	59	42	366
5	5	18	25	85	12	173	10	18	157	3
6	75	4	40	8	8	18	810	0	6	3
7	0	0	0	3	7	0	0	752	1	106
8	6	16	25	129	55	359	6	27	521	24
9	0	0	3	8	259	14	1	152	7	476

	Precision	Recall	F1-Score	Support
0	0.88	0.93	0.90	924
1	0.96	0.95	0.95	1142
2	0.78	0.71	0.74	1125
3	0.64	0.53	0.58	1226
4	0.63	0.54	0.58	1148
5	0.19	0.34	0.25	506
6	0.85	0.83	0.84	972
7	0.73	0.87	0.79	869
8	0.53	0.45	0.49	1168
9	0.47	0.52	0.49	920
Micro Avg	0.67	0.67	0.67	10000
Macro Avg	0.67	0.67	0.66	10000
Weighted Avg	0.69	0.67	0.68	10000

Figure 69: Confusion matrix and classification report of the neural network model with variational autoencoder after FGSM attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	1	0
1	1	1	15	5	37	2	2	26	52	117
2	4	12	1	89	0	0	0	29	17	0
3	1	48	236	1	15	1	0	7	1	0
4	0	0	2	0	0	0	3	0	0	39
5	972	1029	775	914	912	889	948	953	901	853
6	2	1	1	1	8	0	1	0	0	0
7	0	0	0	0	4	0	0	0	1	0
8	0	44	2	0	0	0	2	13	1	0
9	0	0	0	0	6	0	2	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	1
1	0.00	0.00	0.00	258
2	0.00	0.01	0.00	152
3	0.00	0.00	0.00	310
4	0.00	0.00	0.00	44
5	1.00	0.10	0.18	9146
6	0.00	0.07	0.00	14
7	0.00	0.00	0.00	5
8	0.00	0.02	0.00	62
9	0.00	0.00	0.00	8
Micro Avg	0.09	0.09	0.09	10000
Macro Avg	0.10	0.02	0.02	10000
Weighted Avg	0.91	0.09	0.16	10000

Figure 70: Confusion matrix and classification report of the neural network model without variational autoencoder after T-FGSM attack

to the shirt, which surely made it look like a trouser.

When we look at the heat map, we can see ankle boots are mispredicted as sandals most. Sandals mispredicted as ankle boots come second, pullover mispredicted as the coat comes third and trouser as a dress comes forth. In numbers, most mispredicted numbers were 4 as 9 and 8 as 3 and 1. From what we have seen from the perturbation image, clothes can be more deceiving to the human eye than numbers. Now let us see how much it will differ from the

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	858	0	3	4	0	6	40	0	2	1
1	0	1079	2	8	0	4	4	7	12	5
2	23	3	788	90	16	42	57	2	31	12
3	8	4	101	605	8	162	4	6	144	12
4	1	0	12	8	694	28	2	69	33	394
5	18	40	64	195	58	534	36	45	534	21
6	70	6	44	4	9	16	813	0	6	3
7	0	0	0	2	3	0	0	762	0	97
8	2	3	16	85	22	88	1	9	210	12
9	0	0	2	9	172	12	1	128	2	452

	Precision	Recall	F1-Score	Support
0	0.88	0.94	0.91	914
1	0.95	0.96	0.96	1121
2	0.76	0.74	0.75	1064
3	0.60	0.57	0.59	1054
4	0.71	0.56	0.62	1241
5	0.60	0.35	0.44	1545
6	0.85	0.84	0.84	971
7	0.74	0.88	0.81	864
8	0.22	0.47	0.30	448
9	0.45	0.58	0.51	778
Micro Avg	0.68	0.68	0.68	10000
Macro Avg	0.67	0.69	0.67	10000
Weighted Avg	0.70	0.68	0.68	10000

Figure 71: Confusion matrix and classification report of the neural network model with variational autoencoder after T-FGSM attack

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	1	4	52	10	22	15	97	6	20	11
1	1	7	104	5	86	3	7	30	34	18
2	266	213	15	217	50	5	180	143	450	11
3	9	85	397	13	6	296	27	398	110	220
4	7	248	61	1	10	0	290	33	11	305
5	45	1	0	302	5	15	171	9	107	53
6	290	19	18	1	42	76	12	0	43	2
7	18	290	130	47	233	4	6	21	48	272
8	62	266	211	192	86	338	161	79	15	97
9	281	2	44	222	442	140	7	309	136	20

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	238
1	0.01	0.02	0.01	295
2	0.01	0.01	0.01	1550
3	0.01	0.01	0.01	1561
4	0.01	0.01	0.01	966
5	0.02	0.02	0.02	708
6	0.01	0.02	0.02	503
7	0.02	0.02	0.02	1069
8	0.02	0.01	0.01	1507
9	0.02	0.01	0.02	1603
Micro Avg	0.01	0.01	0.01	10000
Macro Avg	0.01	0.01	0.01	10000
Weighted Avg	0.01	0.01	0.01	10000

Figure 72: Confusion matrix and classification report of the neural network model without variational autoencoder after basic iterative method attack

MNIST dataset.

Accuracy scores are below compared to MNIST dataset results, but although our epsilon number increased and therefore our perturbation and attack rate increased, accuracy scores for autoencoded models performed better compared to MNIST dataset. Non-Natural and One Target fooling for models without autoencoder performed better. But with Natural Fooling which we have gathered the targets for mostly mispredicted labels from the heatmap,

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	858	0	8	4	2	6	49	0	3	1
1	0	1068	3	5	0	15	4	3	13	8
2	26	5	780	117	15	53	78	3	35	11
3	9	13	130	638	10	218	3	14	210	11
4	1	1	15	10	575	33	5	73	40	412
5	8	25	36	107	29	266	18	26	336	9
6	74	3	35	8	6	19	796	0	8	3
7	0	0	0	3	9	2	0	750	3	130
8	4	20	22	109	39	266	5	18	316	21
9	0	0	3	9	297	14	0	141	10	403

	Precision	Recall	F1-Score	Support
0	0.88	0.92	0.90	931
1	0.94	0.95	0.95	1119
2	0.76	0.69	0.72	1123
3	0.63	0.51	0.56	1256
4	0.59	0.49	0.54	1165
5	0.30	0.31	0.30	860
6	0.83	0.84	0.83	952
7	0.73	0.84	0.78	897
8	0.32	0.39	0.35	820
9	0.40	0.46	0.43	877
Micro Avg	0.65	0.65	0.65	10000
Macro Avg	0.64	0.64	0.64	10000
Weighted Avg	0.65	0.65	0.65	10000

Figure 73: Confusion matrix and classification report of the neural network model with variational autoencoder after basic iterative method attack

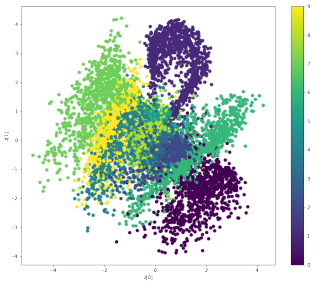


Figure 74: Because of MNIST dataset, our latent space is two-dimensional. One is to look at the neighborhoods of different classes on the latent 2D plane. Each of these colored clusters are a type of digit. Close clusters are digits that are structurally similar, they are digits that share information in the latent space.

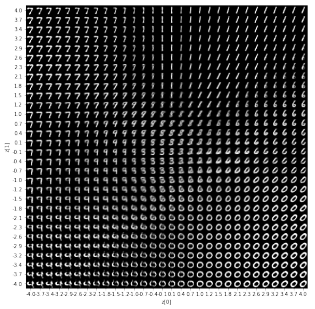


Figure 75: Due to VAE is a generative model, we can also generate new Mnist digits using latent plane, sampling latent points at regular intervals, and generating the corresponding digit for each of these points.

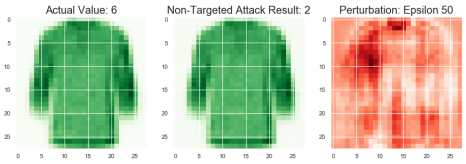


Figure 76: Value change and perturbation of a non-targeted attack on model without autoencoder for Fashion MNIST Dataset

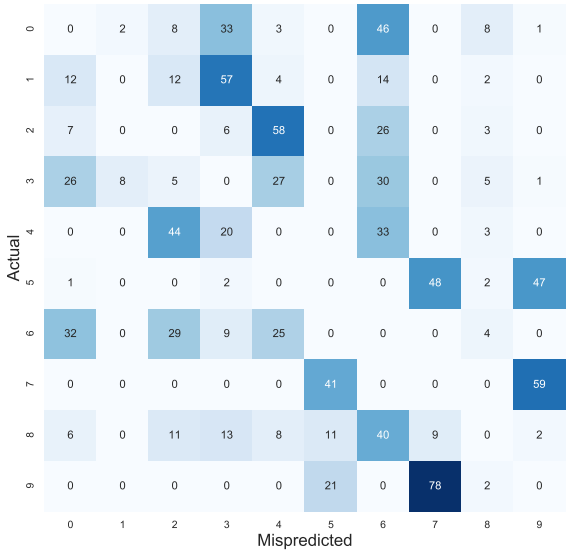


Figure 77: Heatmap of actual numbers and mispredictions for Fashion Mnist

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	60	9	3	435	27	3	71	0	37	1
1	36	872	16	139	21	0	36	0	6	0
2	102	8	10	46	379	10	424	7	157	1
3	376	65	48	186	96	5	111	2	49	2
4	36	8	49	91	3	1	158	2	33	3
5	8	2	39	1	14	2	7	925	81	807
6	331	19	805	92	478	7	1	1	222	10
7	6	3	0	7	6	501	3	15	21	116
8	35	1	55	29	48	101	53	44	391	47
9	4	0	6	8	5	310	1	38	18	36

Figure 78: Confusion matrix of non-targeted attack to model without autoencoder for Fashion MNIST

the model performed poorly. The models with autoencoder performed better as expected, unrelated to the dataset.

6.2.2 Neural Networks

We will do the same attacks with fashion MNIST dataset without changing the code as we did in the previous section.

We can see the changes with FGSM attack in Figure 83. First line is the dataset before the attack and second line is

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	991	0	0	1	1	0	57	0	0	0
1	1	965	0	0	0	0	0	0	0	0
2	0	0	955	1	30	0	58	0	0	0
3	11	2	0	1054	2	0	19	1	0	0
4	0	0	0	0	1060	0	21	0	0	0
5	0	0	0	0	0	923	0	0	0	0
6	0	0	0	0	15	0	735	0	0	0
7	0	0	0	0	0	26	0	1016	0	1
8	1	0	0	1	0	4	2	0	1005	0
9	0	0	0	0	0	25	0	4	0	1012

Figure 79: Confusion matrix of non-targeted attack to model with autoencoder and Fashion MNIST Dataset

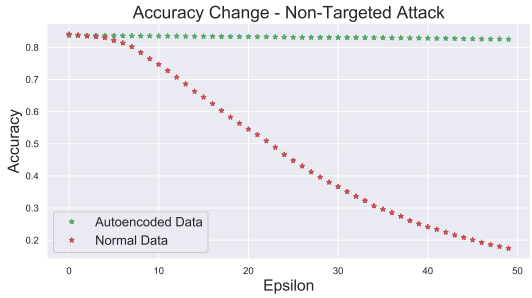


Figure 80: Comparison of accuracy with and without autoencoder for non-targeted attack and Fashion MNIST Dataset

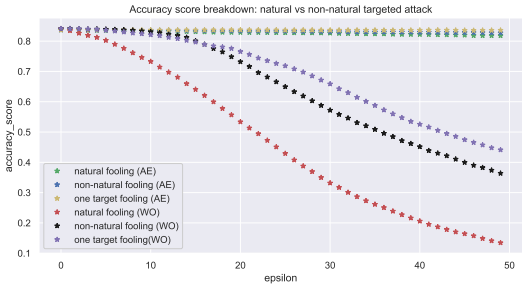


Figure 81: Comparison of accuracy with and without autoencoder for targeted attacks and Fashion MNIST Dataset

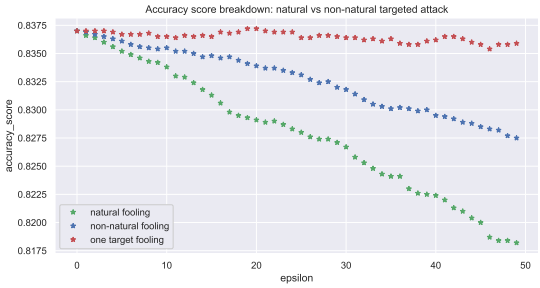


Figure 82: Details of accuracy with autoencoder for targeted attacks and Fashion MNIST Dataset

data after the FGSM attack.

We can also observe the changes with autoencoder in

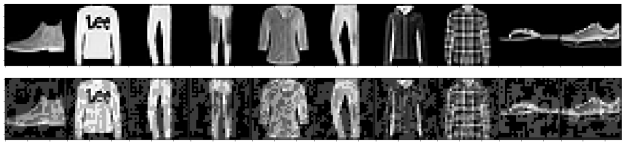


Figure 83: Changes on Fashion MNIST Dataset with FGSM Attack

Figure 84: Changes on Fashion MNIST Dataset with FGSM Attack

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	80	165	17	295	21	45	181	0	97	0
	1	3	32	2	253	3	0	6	0	3	0
	2	35	2	105	51	211	1	177	0	117	0
	3	72	710	6	64	51	2	30	0	71	0
	4	6	12	215	144	100	0	301	0	139	0
	5	5	1	1	0	0	46	1	694	86	359
	6	779	72	621	133	583	24	272	0	410	3
	7	0	0	0	0	0	459	0	57	52	460
	8	20	6	33	60	31	179	32	27	22	133
	9	0	0	0	0	0	244	0	222	3	45

	Precision	Recall	F1-Score	Support
0	0.08	0.09	0.08	901
1	0.03	0.11	0.05	302
2	0.10	0.15	0.12	699
3	0.06	0.06	0.06	1006
4	0.10	0.11	0.10	917
5	0.05	0.04	0.04	1193
6	0.27	0.09	0.14	2897
7	0.06	0.06	0.06	1028
8	0.02	0.04	0.03	543
9	0.04	0.09	0.06	514
Micro Avg	0.08	0.08	0.08	10000
Macro Avg	0.08	0.08	0.08	10000
Weighted Avg	0.12	0.08	0.09	10000

Figure 85: Confusion matrix and classification report of the neural network model without autoencoder after FGSM attack for Fashion MNIST dataset

Figure 84.

The changes in autoencoded data are similar to MNIST Data, more transparent on the edges. We will use the FGSM, T-FGSM, and BIM attacks on models that are using Fashion MNIST Dataset and one is autoencoded, the other is not.

6.3 Sparse Autoencoder

6.3.1 Multi-Class Logistic Regression of Sparse Autoencoder

We will demonstrate the robustness of multi-class logistic regression with sparse autoencoders against attacks with Fashion MNIST dataset. We will give the attack results for easier comparison the process of our experiment is obvious at this point.

The sparse autoencoder is still worse than vanilla autoencoder but as with the MNIST dataset, it will perform as the second-best for the Fashion MNIST dataset.

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	814	3	16	27	2	0	166	0	5	0
	1	2	966	0	14	1	0	3	0	2	0
	2	20	2	769	14	118	0	91	0	8	0
	3	26	18	15	871	41	0	23	0	3	1
	4	5	3	116	43	741	0	86	0	4	0
	5	0	0	0	1	0	942	0	22	1	9
	6	118	5	81	25	94	0	615	0	11	0
	7	0	0	0	0	0	31	0	941	3	39
	8	15	3	3	5	3	3	16	1	962	2
	9	0	0	0	0	0	24	0	36	1	949

	Precision	Recall	F1-Score	Support
0	0.81	0.79	0.80	1033
1	0.97	0.98	0.97	988
2	0.77	0.75	0.76	1022
3	0.87	0.87	0.87	998
4	0.74	0.74	0.74	998
5	0.94	0.97	0.95	975
6	0.61	0.65	0.63	949
7	0.94	0.93	0.93	1014
8	0.96	0.95	0.96	1013
9	0.95	0.94	0.94	1010
Micro Avg	0.86	0.86	0.86	10000
Macro Avg	0.86	0.86	0.86	10000
Weighted Avg	0.86	0.86	0.86	10000

Figure 86: Confusion matrix and classification report of the neural network model with autoencoder after FGSM attack for Fashion MNIST dataset

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	3	0	0	0	0	0	0	0	0	0
	1	0	0	1	5	1	0	1	0	1	0
	2	5	0	1	1	3	0	12	0	0	0
	3	22	3	2	4	32	0	13	0	0	0
	4	0	0	1	0	0	0	0	0	0	0
	5	679	951	779	754	836	1000	913	1000	951	1000
	6	286	45	202	234	54	0	38	0	47	0
	7	0	0	0	0	0	0	0	0	0	0
	8	5	1	14	2	74	0	23	0	1	0
	9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	1.00	0.01	3
1	0.00	0.00	0.00	9
2	0.00	0.05	0.00	22
3	0.00	0.05	0.01	76
4	0.00	0.00	0.00	1
5	1.00	0.11	0.20	8863
6	0.04	0.04	0.04	906
7	0.00	0.00	0.00	0
8	0.00	0.01	0.00	120
9	0.00	0.00	0.00	0
Micro Avg	0.10	0.10	0.10	10000
Macro Avg	0.10	0.13	0.03	10000
Weighted Avg	0.89	0.10	0.18	10000

Figure 87: Confusion matrix and classification report of the neural network model without autoencoder after T-FGSM attack for Fashion MNIST dataset

6.3.2 Neural Network of Sparse Autoencoder

For this section, it is essential to give a confusion matrix and classification report of the neural network model with a sparse autoencoder. Because even without any attack on the model, it labels nearly all of the data as pullovers and sandals.

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	814	5	18	31	1	0	162	0	5	0
1	2	966	0	11	2	0	3	0	2	0
2	20	2	806	12	137	0	109	0	10	0
3	27	17	14	867	36	0	26	0	3	1
4	6	3	92	50	738	0	82	0	4	0
5	1	0	0	2	0	953	0	27	3	11
6	118	5	67	22	83	0	599	0	7	0
7	0	0	0	0	0	26	0	941	3	39
8	12	2	3	5	3	2	19	1	963	2
9	0	0	0	0	0	19	0	31	0	947

	Precision	Recall	F1-Score	Support
0	0.81	0.79	0.80	1036
1	0.97	0.98	0.97	986
2	0.81	0.74	0.77	1096
3	0.87	0.87	0.87	991
4	0.74	0.76	0.75	975
5	0.95	0.96	0.95	997
6	0.60	0.66	0.63	901
7	0.94	0.93	0.94	1009
8	0.96	0.95	0.96	1012
9	0.95	0.95	0.95	997
Micro Avg	0.86	0.86	0.86	10000
Macro Avg	0.86	0.86	0.86	10000
Weighted Avg	0.86	0.86	0.86	10000

Figure 88: Confusion matrix and classification report of the neural network model with autoencoder after T-FGSM attack for Fashion MNIST dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	117	115	10	201	2	9	193	0	28	0
1	8	7	2	246	5	0	5	0	1	2
2	37	3	92	81	256	2	250	0	161	0
3	55	763	14	61	103	0	46	0	65	0
4	4	17	188	231	121	0	232	0	139	0
5	5	5	0	1	0	46	0	610	70	276
6	757	88	685	154	477	3	251	0	397	8
7	0	0	1	0	0	559	0	42	56	591
8	17	2	8	25	36	90	23	24	25	74
9	0	0	0	0	0	291	0	324	58	49

	Precision	Recall	F1-Score	Support
0	0.12	0.17	0.14	675
1	0.01	0.03	0.01	276
2	0.09	0.10	0.10	882
3	0.06	0.06	0.06	1107
4	0.12	0.13	0.13	932
5	0.05	0.05	0.05	1013
6	0.25	0.09	0.13	2820
7	0.04	0.03	0.04	1249
8	0.03	0.08	0.04	324
9	0.05	0.07	0.06	722
Micro Avg	0.08	0.08	0.08	10000
Macro Avg	0.08	0.08	0.07	10000
Weighted Avg	0.12	0.08	0.09	10000

Figure 89: Confusion matrix and classification report of the neural network model without autoencoder after basic iterative method attack for Fashion MNIST dataset

So as it can be seen in from Figure 94 and 95, sparse autoencoder with a neural network is destined to fail from the start.

Sparse autoencoder with fashion MNIST performed surprisingly poorly, even without any attack. Although attack results were unnecessary at this point, we still wanted to show them. Sparse autoencoder's logic to encourage sparsity for classification tasks fails greatly for the Fashion MNIST dataset, which makes sense in a way that clothes

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	805	4	16	28	2	0	162	0	5	0
1	1	964	0	14	1	0	3	0	2	0
2	21	0	770	13	129	1	92	0	7	0
3	23	19	15	869	36	0	23	0	4	1
4	6	3	115	39	734	0	88	0	3	0
5	0	0	0	1	0	941	0	23	1	8
6	129	7	81	30	95	0	613	0	13	0
7	0	0	0	0	0	29	0	937	5	42
8	15	3	3	5	3	5	19	1	960	2
9	0	0	0	1	0	24	0	39	0	947

	Precision	Recall	F1-Score	Support
0	0.81	0.79	0.80	1022
1	0.96	0.98	0.97	985
2	0.77	0.75	0.76	1033
3	0.87	0.88	0.87	990
4	0.73	0.74	0.74	988
5	0.94	0.97	0.95	974
6	0.61	0.63	0.62	968
7	0.94	0.92	0.93	1013
8	0.96	0.94	0.95	1016
9	0.95	0.94	0.94	1011
Micro Avg	0.85	0.85	0.85	10000
Macro Avg	0.85	0.85	0.85	10000
Weighted Avg	0.85	0.85	0.85	10000

Figure 90: Confusion matrix and classification report of the neural network model with autoencoder after basic iterative method attack for Fashion MNIST dataset

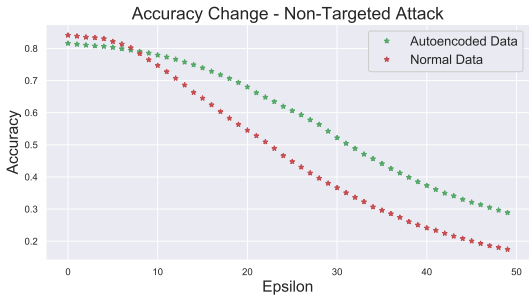


Figure 91: Comparison of accuracy with and without sparse autoencoder for non-targeted attack and Fashion MNIST Dataset

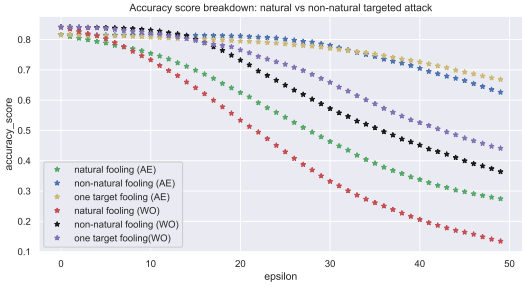


Figure 92: Comparison of accuracy with and without sparse autoencoder for targeted attacks and Fashion MNIST Dataset

on grayscale can be easily confusing. Sparsity turns all the different elements into the same labels. Fashion MNIST dataset after sparse autoencoder is given in Figure 102.

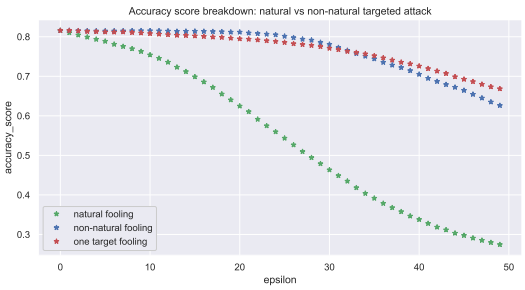


Figure 93: Details of accuracy with sparse autoencoder for sparse targeted attacks and Fashion MNIST Dataset

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	789	1	10	10	1	0	92	0	3	0
	1	0	970	1	4	1	0	2	0	2	0
	2	15	1	760	9	30	0	65	0	1	0
	3	49	20	10	915	36	1	36	0	3	0
	4	4	3	128	30	874	0	84	0	4	0
	5	4	0	0	0	0	962	0	26	2	6
	6	126	4	85	28	56	0	711	0	12	0
	7	0	0	0	0	0	24	0	943	5	40
	8	13	1	6	4	2	1	10	0	968	1
	9	0	0	0	0	0	12	0	31	0	953

	Precision	Recall	F1-Score	Support
0	0.79	0.87	0.83	906
1	0.97	0.99	0.98	980
2	0.76	0.86	0.81	881
3	0.92	0.86	0.88	1070
4	0.87	0.78	0.82	1127
5	0.96	0.96	0.96	1000
6	0.71	0.70	0.70	1022
7	0.94	0.93	0.94	1012
8	0.97	0.96	0.97	1006
9	0.95	0.96	0.95	996
Micro Avg	0.88	0.88	0.88	10000
Macro Avg	0.88	0.89	0.88	10000
Weighted Avg	0.89	0.88	0.88	10000

Figure 94: Confusion matrix and classification report of the neural network model without sparse autoencoder for Fashion MNIST dataset

6.4 Denoising Autoencoder

6.4.1 Multi-Class Logistic Regression of Denoising Autoencoder

Denoising autoencoder with multi-class logistic regression performs as it performed with the MNIST dataset. We do not see too much of a difference with regressions between datasets.

The results for multi-class logistic regression looks kind of similar to sparse autoencoder in the previous section. Let us observe will neural network with denoising autoencoder fails as much as the neural network with sparse autoencoder.

6.4.2 Neural Network of Denoising Autoencoder

The neural network model with denoising autoencoder did not perform as poorly as the neural network model with sparse autoencoder. But as expected, it is still worse than vanilla autoencoder.

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0
	2	996	1000	999	999	999	2	998	0	541	2
	3	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0
	5	4	0	1	1	1	998	2	1000	459	998
	6	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	1.00	0.15	0.27	6536
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
5	1.00	0.29	0.45	3464
6	0.00	0.00	0.00	0
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	0
9	0.00	0.00	0.00	0
Micro Avg	0.20	0.20	0.20	10000
Macro Avg	0.20	0.04	0.07	10000
Weighted Avg	1.00	0.20	0.33	10000

Figure 95: Confusion matrix and classification report of the neural network model with sparse autoencoder for Fashion MNIST dataset

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	129	11	12	244	4	7	177	1	28	0
	1	5	15	1	196	1	0	4	0	1	0
	2	23	0	116	42	166	1	121	0	38	0
	3	81	890	27	69	117	2	95	0	132	1
	4	11	20	189	228	76	0	324	0	188	0
	5	2	0	1	1	0	38	0	714	74	413
	6	729	62	645	171	614	7	251	0	463	7
	7	0	0	0	0	0	481	0	55	33	499
	8	19	2	9	49	22	189	28	9	23	41
	9	1	0	0	0	0	275	0	221	20	39

	Precision	Recall	F1-Score	Support
0	0.13	0.21	0.16	613
1	0.01	0.07	0.02	223
2	0.12	0.23	0.15	507
3	0.07	0.05	0.06	1414
4	0.08	0.07	0.07	1036
5	0.04	0.03	0.03	1243
6	0.25	0.09	0.13	2949
7	0.06	0.05	0.05	1068
8	0.02	0.06	0.03	391
9	0.04	0.07	0.05	556
Micro Avg	0.08	0.08	0.08	10000
Macro Avg	0.08	0.09	0.08	10000
Weighted Avg	0.12	0.08	0.09	10000

Figure 96: Confusion matrix and classification report of the neural network model without sparse autoencoder after FGSM attack for Fashion MNIST dataset

6.5 Variational Autoencoder

6.5.1 Multi-Class Logistic Regression of Variational Autoencoder

The multi-class regression with variational autoencoder performs poorly, as it was in MNIST Dataset which was

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	996	1000	999	999	999	2	998	0	547	2
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	4	0	1	1	1	998	2	1000	453	998
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	1.00	0.15	0.26	6542
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
5	1.00	0.29	0.45	3458
6	0.00	0.00	0.00	0
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	0
9	0.00	0.00	0.00	0
Micro Avg	0.20	0.20	0.20	10000
Macro Avg	0.20	0.04	0.07	10000
Weighted Avg	1.00	0.20	0.33	10000

Figure 97: Confusion matrix and classification report of the neural network model with sparse autoencoder after FGSM attack for Fashion MNIST dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	7	0	0	2	0	0	4	0	0	0
1	0	0	1	1	1	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	31	238	10	10	43	0	64	0	0	0
4	0	0	1	0	0	0	28	0	0	0
5	439	552	558	654	416	1000	792	1000	787	1000
6	521	210	430	333	540	0	111	0	213	0
7	0	0	0	0	0	0	0	0	0	0
8	2	0	0	0	0	0	1	0	0	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.01	0.54	0.01	13
1	0.00	0.00	0.00	3
2	0.00	0.00	0.00	0
3	0.01	0.03	0.01	396
4	0.00	0.00	0.00	29
5	1.00	0.14	0.24	7198
6	0.11	0.05	0.07	2358
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	3
9	0.00	0.00	0.00	0
Micro Avg	0.11	0.11	0.11	10000
Macro Avg	0.11	0.07	0.03	10000
Weighted Avg	0.75	0.11	0.19	10000

Figure 98: Confusion matrix and classification report of the neural network model without sparse autoencoder after T-FGSM attack for Fashion MNIST dataset

expected. Again, due to the variational autoencoder's structure and purpose of use, it is not a fit for defensive measures against attacks.

6.5.2 Neural Network of Variational Autoencoder

We do not observe the problem we have encountered with sparse autoencoder in the neural networks with a variational autoencoder. They still perform poorly and would be

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	996	1000	999	999	999	1	998	0	533	2
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	4	0	1	1	1	999	2	1000	467	998
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	1.00	0.15	0.27	6527
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
5	1.00	0.29	0.45	3473
6	0.00	0.00	0.00	0
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	0
9	0.00	0.00	0.00	0
Micro Avg	0.20	0.20	0.20	10000
Macro Avg	0.20	0.04	0.07	10000
Weighted Avg	1.00	0.20	0.33	10000

Figure 99: Confusion matrix and classification report of the neural network model with sparse autoencoder after T-FGSM attack for Fashion MNIST dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	127	12	9	236	4	5	180	1	29	0
1	6	14	2	215	1	0	4	0	2	0
2	31	1	130	70	197	1	157	0	43	0
3	71	900	22	50	115	0	74	0	119	1
4	13	21	189	252	78	0	315	0	193	0
5	3	0	1	1	0	37	0	720	83	412
6	725	50	637	135	586	5	246	0	459	7
7	1	0	1	0	0	507	0	55	34	508
8	22	2	9	41	19	189	24	8	16	33
9	1	0	0	0	0	256	0	216	22	39

	Precision	Recall	F1-Score	Support
0	0.13	0.21	0.16	603
1	0.01	0.06	0.02	244
2	0.13	0.21	0.16	630
3	0.05	0.04	0.04	1352
4	0.08	0.07	0.08	1061
5	0.04	0.03	0.03	1257
6	0.25	0.09	0.13	2850
7	0.06	0.05	0.05	1106
8	0.02	0.04	0.02	363
9	0.04	0.07	0.05	534
Micro Avg	0.08	0.08	0.08	10000
Macro Avg	0.08	0.09	0.07	10000
Weighted Avg	0.11	0.08	0.08	10000

Figure 100: Confusion matrix and classification report of the neural network model without sparse autoencoder after basic iterative method attack for Fashion MNIST dataset

the last most accurate autoencoder type if sparse autoencoder did not perform so poorly with Fashion MNIST.

We also used the generative aspect of the variational autoencoder for Fashion MNIST Dataset.

We believe that the most significant strength of our model is that the natural practice of implementing an autoencoder between data and machine learning models can provide considerable defense and robustness against attacks and it provide high generalization property, in con-

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	996	1000	999	999	999	1	998	0	548	2
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	4	0	1	1	1	999	2	1000	452	998
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	1.00	0.15	0.26	6542
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
5	1.00	0.29	0.45	3458
6	0.00	0.00	0.00	0
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	0
9	0.00	0.00	0.00	0
Micro Avg	0.20	0.20	0.20	10000
Macro Avg	0.20	0.04	0.07	10000
Weighted Avg	1.00	0.20	0.33	10000

Figure 101: Confusion matrix and classification report of the neural network model with autoencoder after basic iterative method attack for Fashion MNIST dataset

Figure 102: Fashion MNIST Data after being through Sparse Autoencoder

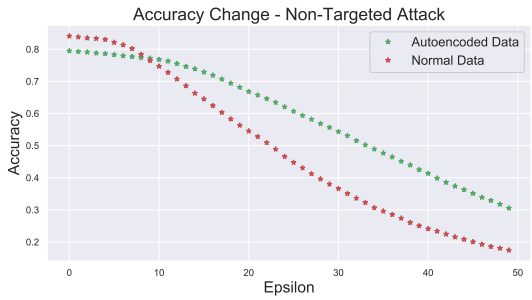


Figure 103: Comparison of accuracy with and without denoising autoencoder for non-targeted attack and Fashion MNIST Dataset

trast to most prior adversarial learning methods. An important feature of our methodology is that not only presents a generic resistance to specific attack methods but also provides robustness to machine learning models in general.

7 Discussion

In linear machine learning model algorithms, we applied non-targeted and targeted attacks to multiclass logistic regression machine learning models to observe the changes and difference between attack methods. Moreover, FGSM,

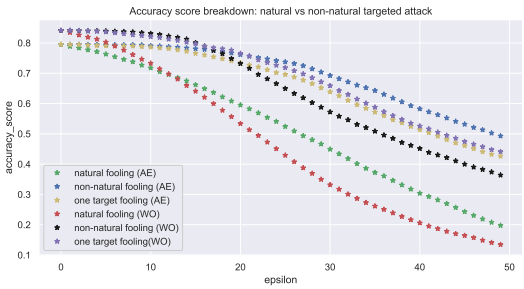


Figure 104: Comparison of accuracy with and without denoising autoencoder for targeted attacks and Fashion MNIST Dataset

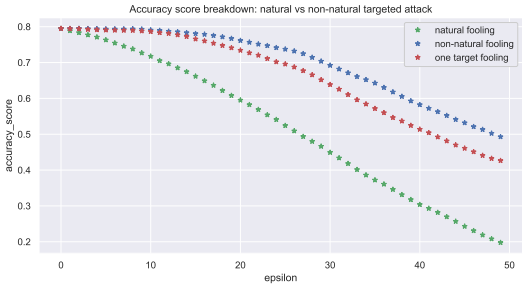


Figure 105: Details of accuracy with autoencoder for denoising targeted attacks and Fashion MNIST Dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	80	165	17	295	21	45	181	0	97	0
1	3	32	2	253	3	0	6	0	3	0
2	35	2	105	51	211	1	177	0	117	0
3	72	710	6	64	51	2	30	0	71	0
4	6	12	215	144	100	0	301	0	139	0
5	5	1	1	0	0	46	1	694	86	359
6	779	72	621	133	583	24	272	0	410	3
7	0	0	0	0	0	459	0	57	52	460
8	20	6	33	60	31	179	32	27	22	133
9	0	0	0	0	0	244	0	222	3	45

	Precision	Recall	F1-Score	Support
0	0.08	0.09	0.08	901
1	0.03	0.11	0.05	302
2	0.10	0.15	0.12	699
3	0.06	0.06	0.06	1006
4	0.10	0.11	0.10	917
5	0.05	0.04	0.04	1193
6	0.27	0.09	0.14	2897
7	0.06	0.06	0.06	1028
8	0.02	0.04	0.03	543
9	0.04	0.09	0.06	514
Micro Avg	0.08	0.08	0.08	10000
Macro Avg	0.08	0.08	0.08	10000
Weighted Avg	0.12	0.08	0.09	10000

Figure 106: Confusion matrix and classification report of the neural network model without denoising autoencoder after FGSM attack for Fashion MNIST dataset

T-FGSM, and BIM attacks have been used for the neural network machine learning model. The effects of these attacks on implementing autoencoder as a filter have been examined for these machine learning models.

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	818	11	17	29	6	0	137	0	5	0
1	4	964	4	7	1	0	3	0	2	0
2	22	2	759	13	115	0	100	0	2	0
3	18	17	8	875	45	0	24	0	4	0
4	2	1	116	20	719	0	71	0	7	0
5	1	0	0	0	0	950	0	28	2	8
6	127	3	94	51	110	0	654	1	12	1
7	0	0	0	0	0	36	0	949	6	33
8	8	2	2	5	4	1	10	0	959	1
9	0	0	0	0	0	13	1	22	1	957

	Precision	Recall	F1-Score	Support
0	0.82	0.80	0.81	1023
1	0.96	0.98	0.97	985
2	0.76	0.75	0.75	1013
3	0.88	0.88	0.88	991
4	0.72	0.77	0.74	936
5	0.95	0.96	0.96	989
6	0.65	0.62	0.64	1053
7	0.95	0.93	0.94	1024
8	0.96	0.97	0.96	992
9	0.96	0.96	0.96	994
Micro Avg	0.86	0.86	0.86	10000
Macro Avg	0.86	0.86	0.86	10000
Weighted Avg	0.86	0.86	0.86	10000

Figure 107: Confusion matrix and classification report of the neural network model with denoising autoencoder after FGSM attack for Fashion MNIST dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	22	3	2	128	2	0	22	0	1	0
1	0	0	0	0	1	0	0	0	0	0
2	0	0	1	0	0	0	2	0	1	0
3	50	14	2	8	5	0	11	0	0	0
4	0	0	0	0	0	0	9	0	0	0
5	695	952	824	840	828	1000	876	1000	939	999
6	213	13	124	5	93	0	25	0	42	0
7	0	0	0	0	0	0	0	0	0	0
8	20	18	46	19	71	0	55	0	1	1
9	0	0	1	0	0	0	0	0	16	0

	Precision	Recall	F1-Score	Support
0	0.02	0.12	0.04	180
1	0.00	0.00	0.00	1
2	0.00	0.25	0.00	4
3	0.01	0.09	0.01	90
4	0.00	0.00	0.00	9
5	1.00	0.11	0.20	8953
6	0.03	0.05	0.03	515
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	231
9	0.00	0.00	0.00	17
Micro Avg	0.11	0.11	0.11	10000
Macro Avg	0.11	0.06	0.03	10000
Weighted Avg	0.90	0.11	0.18	10000

Figure 108: Confusion matrix and classification report of the neural network model without denoising autoencoder after T-FGSM attack for Fashion MNIST dataset

Autoencoders provide robustness against adversarial machine learning attacks to machine learning models for both linear models and neural network models. In this study, we presented that the natural practice of implementing an autoencoder between data and machine learning models can lead significant defense and robustness against attacks.

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	807	12	13	29	3	0	122	0	3	0
1	4	960	4	5	1	0	3	0	1	0
2	21	1	735	14	103	0	89	0	5	0
3	19	19	7	869	46	0	26	0	3	0
4	2	1	137	25	737	0	70	0	6	0
5	1	0	0	0	0	970	1	40	6	12
6	140	5	101	51	106	0	679	0	15	1
7	0	0	0	0	0	21	0	938	6	34
8	6	2	3	7	4	0	8	0	954	0
9	0	0	0	0	0	9	2	22	1	953

	Precision	Recall	F1-Score	Support
0	0.81	0.82	0.81	989
1	0.96	0.98	0.97	978
2	0.73	0.76	0.75	968
3	0.87	0.88	0.87	989
4	0.74	0.75	0.75	978
5	0.97	0.94	0.96	1030
6	0.68	0.62	0.65	1098
7	0.94	0.94	0.94	999
8	0.95	0.97	0.96	984
9	0.95	0.97	0.96	987
Micro Avg	0.86	0.86	0.86	10000
Macro Avg	0.86	0.86	0.86	10000
Weighted Avg	0.86	0.86	0.86	10000

Figure 109: Confusion matrix and classification report of the neural network model with denoising autoencoder after T-FGSM attack for Fashion MNIST dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	72	162	14	290	21	22	176	0	96	0
1	3	15	2	257	4	0	6	0	3	0
2	44	2	109	58	259	2	192	0	151	0
3	70	708	6	57	41	2	26	0	58	0
4	7	15	234	163	104	0	305	0	149	0
5	5	1	1	0	0	32	1	701	89	370
6	779	89	609	119	542	23	268	0	385	3
7	0	0	0	0	0	489	0	57	53	450
8	20	8	25	56	29	161	26	23	10	132
9	0	0	0	0	0	269	0	219	6	45

	Precision	Recall	F1-Score	Support
0	0.07	0.08	0.08	853
1	0.01	0.05	0.02	290
2	0.11	0.13	0.12	817
3	0.06	0.06	0.06	968
4	0.10	0.11	0.11	977
5	0.03	0.03	0.03	1200
6	0.27	0.10	0.14	2817
7	0.06	0.05	0.06	1049
8	0.01	0.02	0.01	490
9	0.04	0.08	0.06	539
Micro Avg	0.08	0.08	0.08	10000
Macro Avg	0.08	0.07	0.07	10000
Weighted Avg	0.12	0.08	0.09	10000

Figure 110: Confusion matrix and classification report of the neural network model without denoising autoencoder after basic iterative method attack for Fashion MNIST dataset

8 Conclusion

In this paper, we have presented the results for pre-filtering the data with an autoencoder before sending it to the machine learning model against adversarial machine learning attacks. We have investigated that the classifier accuracy changes for linear and neural network machine learning models. We have also applied non-targeted and targeted

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	803	11	16	32	10	0	138	0	5	0
1	5	961	4	5	2	0	2	0	1	0
2	25	2	754	14	119	0	95	0	2	0
3	19	21	9	875	48	0	25	0	3	0
4	1	1	116	19	692	0	68	0	7	0
5	1	0	0	0	0	941	1	28	3	9
6	136	3	99	49	124	0	659	1	10	1
7	0	0	0	0	0	41	0	943	5	30
8	10	1	2	6	5	2	11	2	963	1
9	0	0	0	0	0	16	1	26	1	959

	Precision	Recall	F1-Score	Support
0	0.80	0.79	0.80	1015
1	0.96	0.98	0.97	980
2	0.75	0.75	0.75	1011
3	0.88	0.88	0.88	1000
4	0.69	0.77	0.73	904
5	0.94	0.96	0.95	983
6	0.66	0.61	0.63	1082
7	0.94	0.93	0.93	1019
8	0.96	0.96	0.96	1003
9	0.96	0.96	0.96	1003
Micro Avg	0.85	0.85	0.85	10000
Macro Avg	0.86	0.86	0.86	10000
Weighted Avg	0.85	0.85	0.85	10000

Figure 111: Confusion matrix and classification report of the neural network model with denoising autoencoder after basic iterative method attack for Fashion MNIST dataset

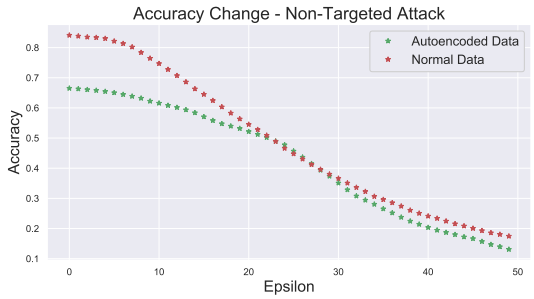


Figure 112: Comparison of accuracy with and without variational autoencoder for non-targeted attack and Fashion MNIST Dataset

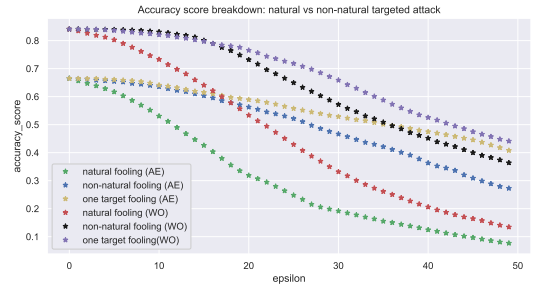


Figure 113: Comparison of accuracy with and without variational autoencoder for targeted attacks and Fashion MNIST Dataset

attacks to multi-class logistic regression. Besides, FGSM, T-FGSM, and BIM attacks have been applied to the neural network machine learning model. The effects of these

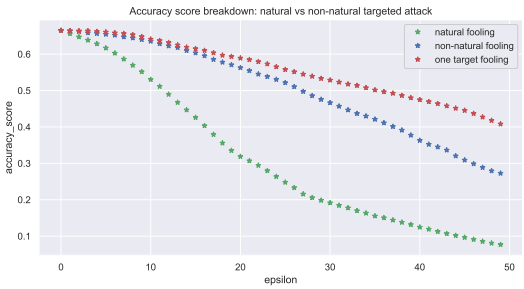


Figure 114: Details of accuracy with autoencoder for variational targeted attacks and Fashion MNIST Dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	124	130	11	205	6	16	194	0	66	0
1	3	10	1	185	1	0	2	0	2	0
2	27	1	124	43	144	23	218	0	82	0
3	81	743	19	61	93	4	83	0	162	1
4	3	44	143	218	143	0	285	0	46	0
5	3	0	1	0	0	35	2	623	58	351
6	739	52	672	234	596	1	194	0	481	3
7	0	0	0	1	0	579	0	32	67	580
8	19	20	29	53	17	107	22	26	24	15
9	1	0	0	0	0	235	0	319	12	50

	Precision	Recall	F1-Score	Support
0	0.12	0.16	0.14	752
1	0.01	0.05	0.02	204
2	0.12	0.19	0.15	662
3	0.06	0.05	0.05	1247
4	0.14	0.16	0.15	882
5	0.04	0.03	0.03	1073
6	0.19	0.07	0.10	2972
7	0.03	0.03	0.03	1259
8	0.02	0.07	0.04	332
9	0.05	0.08	0.06	617
Micro Avg	0.08	0.08	0.08	10000
Macro Avg	0.08	0.09	0.08	10000
Weighted Avg	0.11	0.08	0.08	10000

Figure 115: Confusion matrix and classification report of the neural network model without variational autoencoder after FGSM attack for Fashion MNIST dataset

attacks on implementing autoencoder as a filter have been analyzed for both machine learning models. We have observed that the robustness provided by autoencoder after adversarial attacks can be seen by accuracy drop between 0.1 and 0.2 percent while the models without autoencoder suffered tremendous accuracy drops hitting accuracy score between 0.6 and 0.3 in some cases even 0.1. We have proposed general, generic, and easy to implement protection against adversarial machine learning model attacks. It will be beneficial to remind that all autoencoders in this study were trained with the epoch of 35 with 1024 sized batches, so the results can be improved by increasing the number of epochs. In conclusion, we have discussed that autoencoders provide robustness against adversarial machine learning attacks to machine learning models for both linear models and neural network models. We have examined the other types of autoencoders, which are mostly called vanilla autoencoders, give the best results. The second most accurate autoencoder type is sparse autoencoders, and

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	727	21	10	92	29	0	205	0	21	0
1	1	800	1	157	4	0	1	0	1	0
2	38	7	536	21	369	0	314	0	62	0
3	113	148	6	611	59	0	73	0	34	1
4	25	13	204	52	381	0	135	0	10	0
5	1	0	0	0	0	711	0	190	9	90
6	66	6	173	57	142	2	183	0	35	0
7	1	0	0	0	0	230	0	721	7	76
8	28	5	70	10	16	4	89	5	820	2
9	0	0	0	0	0	53	0	84	1	831

	Precision	Recall	F1-Score	Support
0	0.73	0.66	0.69	1105
1	0.80	0.83	0.81	965
2	0.54	0.40	0.46	1347
3	0.61	0.58	0.60	1045
4	0.38	0.46	0.42	820
5	0.71	0.71	0.71	1001
6	0.18	0.28	0.22	664
7	0.72	0.70	0.71	1035
8	0.82	0.78	0.80	1049
9	0.83	0.86	0.84	969
Micro Avg	0.63	0.63	0.63	10000
Macro Avg	0.63	0.63	0.63	10000
Weighted Avg	0.65	0.63	0.64	10000

Figure 116: Confusion matrix and classification report of the neural network model with variational autoencoder after FGSM attack for Fashion MNIST dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	12	24	1	18	9	0	34	0	11	0
1	2	1	1	57	1	0	2	0	0	0
2	26	0	10	4	15	0	99	0	66	0
3	27	8	2	1	8	0	40	0	0	0
4	3	1	0	6	4	0	19	0	3	0
5	566	935	655	720	497	1000	745	1000	893	1000
6	352	2	310	174	334	0	41	0	26	0
7	0	0	0	0	0	0	0	0	0	0
8	12	29	21	20	132	0	20	0	1	0
9	0	0	0	0	0	0	0	0	0	0

	Precision	Recall	F1-Score	Support
0	0.01	0.11	0.02	109
1	0.00	0.02	0.00	64
2	0.01	0.05	0.02	220
3	0.00	0.01	0.00	86
4	0.00	0.11	0.01	36
5	1.00	0.12	0.22	8011
6	0.04	0.03	0.04	1239
7	0.00	0.00	0.00	0
8	0.00	0.00	0.00	235
9	0.00	0.00	0.00	0
Micro Avg	0.11	0.11	0.11	10000
Macro Avg	0.11	0.05	0.03	10000
Weighted Avg	0.81	0.11	0.18	10000

Figure 117: Confusion matrix and classification report of the neural network model without variational autoencoder after T-FGSM attack for Fashion MNIST dataset

the third most accurate is denoising autoencoders, which gives similar results with the sparse autoencoders. We have observed that the worst autoencoder type for this process is variational autoencoders because variational autoencoders are generative models used in different areas.

In summary, the natural practice of implementing an autoencoder between data and machine learning models can provide considerable defense and robustness against attacks. These autoencoders can be easily implemented with

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	712	16	10	67	17	1	197	0	20	0
1	0	835	1	154	1	0	1	0	1	0
2	43	8	627	30	399	0	371	0	63	0
3	99	108	5	603	59	0	49	0	26	1
4	23	12	198	53	399	0	130	0	9	0
5	2	2	0	1	0	939	1	401	19	179
6	93	14	96	82	106	0	167	0	41	0
7	0	0	0	0	0	35	0	512	2	18
8	28	5	63	10	19	1	84	5	818	1
9	0	0	0	0	0	24	0	82	1	801

	Precision	Recall	F1-Score	Support
0	0.71	0.68	0.70	1040
1	0.83	0.84	0.84	993
2	0.63	0.41	0.49	1541
3	0.60	0.63	0.62	950
4	0.40	0.48	0.44	824
5	0.94	0.61	0.74	1544
6	0.17	0.28	0.21	599
7	0.51	0.90	0.65	567
8	0.82	0.79	0.80	1034
9	0.80	0.88	0.84	908
Micro Avg	0.64	0.64	0.64	10000
Macro Avg	0.64	0.65	0.63	10000
Weighted Avg	0.69	0.64	0.65	10000

Figure 118: Confusion matrix and classification report of the neural network model with variational autoencoder after T-FGSM attack for Fashion MNIST dataset

Actual Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	117	125	8	192	5	14	185	0	66	0
1	4	10	1	187	1	0	2	0	2	0
2	37	1	143	57	181	19	251	0	98	0
3	81	736	21	57	96	3	77	0	150	0
4	6	69	172	247	148	0	271	0	60	0
5	5	0	1	0	0	31	2	638	64	352
6	732	50	643	220	554	1	193	0	453	1
7	0	0	0	0	0	591	0	32	76	584
8	17	9	11	40	15	99	19	26	18	14
9	1	0	0	0	0	242	0	304	13	49

	Precision	Recall	F1-Score	Support
0	0.12	0.16	0.14	712
1	0.01	0.05	0.02	207
2	0.14	0.18	0.16	787
3	0.06	0.05	0.05	1221
4	0.15	0.15	0.15	973
5	0.03	0.03	0.03	1093
6	0.19	0.07	0.10	2847
7	0.03	0.02	0.03	1283
8	0.02	0.07	0.03	268
9	0.05	0.08	0.06	609
Micro Avg	0.08	0.08	0.08	10000
Macro Avg	0.08	0.09	0.08	10000
Weighted Avg	0.11	0.08	0.08	10000

Figure 119: Confusion matrix and classification report of the neural network model without variational autoencoder after basic iterative method attack for Fashion MNIST dataset

libraries such as TensorFlow and Keras. Through the results of this review, it is evident that autoencoders can be used in any machine learning model easily because of their implementation as a separate layer.

		Predicted Values									
		0	1	2	3	4	5	6	7	8	9
Actual Values	0	677	21	11	108	29	0	182	0	18	0
	1	6	719	0	220	5	0	2	0	0	0
	2	36	8	498	17	358	0	288	0	54	0
	3	86	214	5	479	46	0	58	0	33	1
	4	31	17	197	78	374	0	131	0	12	0
	5	2	1	0	1	0	529	0	256	12	107
	6	134	15	214	86	173	2	251	0	45	0
	7	0	0	0	0	0	400	0	642	8	98
	8	28	5	75	11	15	4	88	6	817	2
	9	0	0	0	0	0	65	0	96	1	792

	Precision	Recall	F1-Score	Support
0	0.68	0.65	0.66	1046
1	0.72	0.76	0.74	952
2	0.50	0.40	0.44	1259
3	0.48	0.52	0.50	922
4	0.37	0.45	0.41	840
5	0.53	0.58	0.55	908
6	0.25	0.27	0.26	920
7	0.64	0.56	0.60	1148
8	0.82	0.78	0.80	1051
9	0.79	0.83	0.81	954
Micro Avg	0.58	0.58	0.58	10000
Macro Avg	0.58	0.58	0.58	10000
Weighted Avg	0.58	0.58	0.58	10000

Figure 120: Confusion matrix and classification report of the neural network model with variational autoencoder after basic iterative method attack for Fashion MNIST dataset

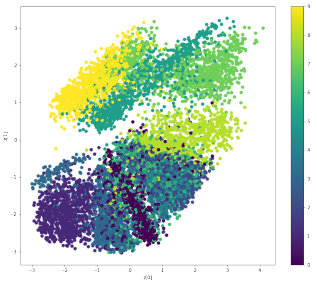


Figure 121: Because of Fashion MNIST dataset, our latent space is two-dimensional. One is to look at the neighborhoods of different classes on the latent 2D plane. Each of these colored clusters are a type of digit. Close clusters are digits that are structurally similar, they are digits that share information in the latent space.

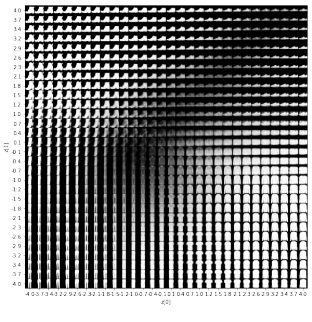


Figure 122: Due to VAE is a generative model, we can also generate new Mnist digits using latent plane, sampling latent points at regular intervals, and generating the corresponding digit for each of these points.

References

[1] A. F. Agarap. Deep learning using rectified linear units (relu). *CoRR*, abs/1803.08375, 2018.

[2] Sharar Ahmadi, Mehran S Fallah, and Massoud Pourmahdian. On the properties of epistemic and temporal epistemic logics of authentication. *Informatica*, 43(2), 2019. doi: 10.31449/inf.v43i2.1617.

[3] M. Aladag, F. O. Catak, and E. Gul. Preventing data poisoning attacks by using generative models. In *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, pages 1–5, 2019. doi: 10.1109/UBMYK48245.2019.8965459.

[4] M. Juuti, B. G. Atli, N. Asokan. Making targeted black-box evasion attacks effective and efficient. *CoRR*, abs/1906.03397, 2019.

[5] W. Bai, C. Quan, and Z. Luo. Alleviating adversarial attacks via convolutional autoencoder. In *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 53–58, 2017. doi: 10.1109/SNPD.2017.8022700.

[6] M. Bakator and D. Radosav. Deep learning and medical diagnosis: A review of literature. *Multimodal Technologies and Interaction*, 2:47, 2018. doi: 10.3390/mti2030047.

[7] Bowen Baker, Ingmar Kanitscheider, Todor Markov, Yi Wu, Glenn Powell, Bob McGrew, and Igor Mordatch. Emergent tool use from multi-agent autocurricula. *arXiv preprint arXiv:1909.07528*, 2019.

[8] X. Yuan, P. He, Q. Zhu, R. R. Bhat and X. Li. Adversarial examples: Attacks and defenses for deep learning. *CoRR*, abs/1712.07107, July 2017.

[9] Financial Stability Board. Artificial intelligence and machine learning in financial services: Market developments and financial stability implications. *Financial Stability Board*, page 45, 2017.

[10] A. Athalye, N. Carlini and D. A. Wagner. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. *CoRR*, abs/1802.00420, February 2018.

[11] N. Carlini and D. A. Wagner. Towards evaluating the robustness of neural networks. *CoRR*, abs/1608.04644, 2016. URL `http://arxiv.org/abs/1608.04644`.

[12] Ferhat Ozgur Catak and Ahmet Fatih Mustacoglu. Distributed denial of service attack detection using autoencoder and deep neural networks. *Journal of Intelligent & Fuzzy Systems*, 37(3):3969–3979, 2019. doi: 10.3233/JIFS-190159.

[13] I. Chen and B. Sirkeci-Mergen. A comparative study of autoencoders against adversarial attacks. *nt'l Conf. IP, Comp. Vision, and Pattern Recognition*, 2018.

[14] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[15] Cherifi Dalila, Boushaba Saddek, Nait-Ali Amine, et al. Feature level fusion of face and voice biometrics systems using artificial neural network for personal recognition. *Informatica*, 44(1), 2020. doi: 10.31449/inf.v44i1.2596.

[16] Murat Dikmen and Catherine M. Burns. Autonomous driving in the real world: Experiences with tesla autopilot and summon. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Automotive'UI 16, page 225–228, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450345330. doi: 10.1145/3003715.3005465.

[17] Samuel G Finlayson, John D Bowers, Joichi Ito, Jonathan L Zittrain, Andrew L Beam, and Isaac S Kohane. Adversarial attacks on medical machine learning. *Science*, 363(6433):1287–1289, 2019.

[18] C. Nwankpa, W. Ijomah, A. Gachagan and S. Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *CoRR*, abs/1811.03378, November 2018.

[19] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[20] Jun Han and Claudio Moraga. The influence of the sigmoid function parameters on the speed of backpropagation learning. In *International Workshop on Artificial Neural Networks*, pages 195–201. Springer, 1995. doi: 10.1007/3-540-59497-3_175.

[21] Hao Zheng, Zhanlei Yang, Wenju Liu, Jizhong Liang, and Yanpeng Li. Improving deep neural networks using softplus units. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–4, 2015. doi: 10.1109/IJCNN.2015.7280459.

[22] Samuel Harding, Prashanth Rajivan, Bennett I Bertenthal, and Cleotilde Gonzalez. Human decisions on targeted and non-targeted adversarial sample. In *CogSci*, 2018.

[23] M. Isakov, V. Gadepally, K. M. Gettings, and M. A. Kinsy. Survey of attacks and defenses on edge-deployed neural networks. In *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–8, 2019.

[24] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru, and B. Li. Manipulating machine learning: Poisoning attacks and countermeasures for regression learning. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 19–35, 2018. doi: 10.1109/SP.2018.00057.

[25] J. Guo, Y. Zhao, X. Han, Y. Jiang and J. Sun. Rnntest: Adversarial testing framework for recurrent neural network systems. *CoRR*, November 2019.

[26] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, December 2014.

[27] K. Auernhammer, R. T. Kolagari and M. Zoppelt. Attacks on machine learning: Lurking danger for accountability. *CoRR*, January 2019.

[28] B. Li and Y. Vorobeychik. Evasion-robust classification on binary domains. *ACM Trans. Knowl. Discov. Data*, 12(4):50:1–50:32, 2018. ISSN 1556-4681. doi: 10.1145/3186282.

[29] F. Chen, N. Chen, H. Mao and H. Hu. Assessing four neural networks on handwritten digit recognition dataset (MNIST). *CoRR*, abs/1811.08278, November 2018.

[30] A. Chernikova, A. Oprea, C. Nita-Rotaru and B. Kim. Are self-driving cars secure? evasion attacks against deep neural networks for steering angle prediction. *CoRR*, abs/1904.07370, April 2019.

[31] Soohwan Park, Hoseok Ryu, Seyoung Lee, Sunmin Lee, and Jehee Lee. Learning predict-and-simulate policies from unorganized human motion data. *ACM Trans. Graph.*, 38(6), November 2019. ISSN 0730-0301. doi: 10.1145/3355089.3356501.

[32] L. Huang, A. D. Joseph, B. Nelson, B.I.P. Rubinstein and J. D. Tygar. Adversarial machine learning. In *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence*, AISec '11, pages 43–58, New York, NY, USA, October 2011. ACM. ISBN 978-1-4503-1003-1. doi: 10.1145/2046684.2046692.

[33] R. Sahay, R. Mahfuz, and A. E. Gamal. Combatting adversarial attacks through denoising and dimensionality reduction: A cascaded autoencoder approach. In *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, 2019. doi: 10.1109/CISS.2019.8692918.

[34] H. Zhang, S. Starke, T. Komura, J. Saito. Mode-adaptive neural networks for quadruped motion control. *ACM Trans. Graph.*, 37(4): 145:1–145:11, July 2018. ISSN 0730-0301. doi: 10.1145/3197517.3201366.

[35] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015. ISSN 0893-6080.

[36] K. Y. Xiao, V. Tjeng, N. M. Shafiullah and A. Madry. Training for faster adversarial robustness verification via inducing relu stability. *CoRR*, abs/1809.03008, September 2018.

[37] A. Siddiqi. Adversarial security attacks and perturbations on machine learning and deep learning methods. *CoRR*, July 2019.

[38] Samed Sivaslıoğlu, Ferhat Ozgur Catak, and Ensar Gül. Incrementing adversarial robustness with autoencoding for machine learning model attacks. In *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2019. doi: 10.1109/SIU.2019.8806432.

[39] L. Pinto, J. Davidson, R. Sukthankar and A. Gupta. Robust adversarial reinforcement learning. *CoRR*, abs/1703.02702, March 2017.

[40] G. Gondim-Ribeiro, P. Tabacof and E. Valle. Adversarial attacks on variational autoencoders. *CoRR*, abs/1806.04646, 2018.

[41] A. Erba, R. Taormina, S. Galelli, M. Pogliani, M. Carminati, S. Zanero, N. O. Tippenhauer. Real-time evasion attacks with physical constraints on deep learning-based anomaly detectors in industrial control systems. *CoRR*, abs/1907.07487, July 2019.

[42] A. Kurakin, I. J. Goodfellow, S. Bengio, Y. Dong, F. Liao, M. Liang, T. Pang, J. Zhu, X. Hu, C. Xie, J. Wang, Z. Zhang, Z. Ren, A. L. Yuille, S. Huang, Y. Zhao, Y. Zhao, Z. Han, J. Long, Y. Berdibekov, T. Akiba, S. Tokui and M. Abe. Adversarial attacks and defences competition. *CoRR*, abs/1804.00097, March 2018.

[43] A. Madry, A. Makelov, L. Schmidt, D. Tsipras and A. Vladu. Towards deep learning models resistant to adversarial attacks. *CoRR*, abs/1706.06083, 2017.

[44] Lev V Utkin and Kirill D Zhuk. Improvement of the deep forest classifier by a set of neural networks. *Informatica*, 44(1), 2020. doi: 10.31449/inf.v44i1.2740.

[45] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019. doi: 10.1038/s41586-019-1724-z.

[46] Ting-Chun Wang, Ming-Yu Liu, Andrew Tao, Guilin Liu, Jan Kautz, and Bryan Catanzaro. Few-shot video-to-video synthesis. *arXiv preprint arXiv:1910.12713*, 2019.

[47] Shujian Yu and José C. Príncipe. Understanding autoencoders with information theoretic concepts. *Neural Networks*, 117:104–123, 2019. ISSN 0893-6080. doi: 10.1016/j.neunet.2019.05.003.

[48] John Paul Tan Yusiong and Prospero Clara Naval. A semi-supervised approach to monocular depth estimation, depth refinement, and semantic segmentation of driving scenes using a siamese triple decoder architecture. *Informatica*, 44(4), 2020. doi: 10.31449/inf.v44i4.3018.

[49] Z. Zhang and M. R. Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *CoRR*, abs/1805.07836, May 2018.

[50] F. Yu, C. Liu, Y. Wang, L. Zhao and X. Chen. Interpreting adversarial robustness: A view from decision surface in input space. *CoRR*, abs/1810.00144, September 2018.

[51] L. Yang, Z. Shi, Y. Zheng and K. Zhou. Dynamic hair modeling from monocular videos using deep neural networks. *ACM Trans. Graph.*, 38(6): 235:1–235:12, November 2019. ISSN 0730-0301. doi: 10.1145/3355089.3356511.

Data Quality Strategy Selection in CRIS: Using a Hybrid Method of SWOT and BMW

Otmane Azeroual (ORCID: 0000-0002-5225-389X)
German Center for Higher Education Research and Science Studies (DZHW), Berlin, Germany
E-mail: azeroual@dzhw.eu

Mohammad Javad Ershadi
Information Technology Department
Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran

Amir Azizi and Melikasadat Banihashemi
Islamic Azad University, Science and Research Branch, Tehran, Iran

Reza Edris Abadi
Islamic Azad University, Central Tehran Branch, Tehran, Iran

Keywords: current research information systems (CRIS), data quality (DQ), knowledge management (KM), SWOT, multi-criteria decision-making (MCDM), best-worst method (BWM)

Received: October 28, 2019

Data quality has been considerably faced with more attention in recent years. While improving the quality of any type of information system needs to apply data quality dimensions, this process is a strategic decision of any organization. Current Research Information System (CRIS) is a state of the art information system which manages different processes for acquisition, indexing, and dissemination of researches funded by research funders. In this paper, quality improvement programs for a CRIS are strategically defined using Strength, Weakness, Opportunity and Threaten (SWOT) approach. According to examined SWOT method, weaknesses (such as failure to evaluate the quality of information contained in the research), strengths (such as the accuracy of information classification), opportunities (such as the presence of university representatives in the process of thesis/dissertation registration) and threats (such as transfer of incorrect information by other systems) are identified and categorized. Besides, data quality dimensions are considered for determining all strategies for improving CRIS. An advanced multi-criteria decision-making method called Best-Worst Method (BWM) is applied for prioritizing obtained strategies. Results of proposed methodology indicated that the development and classification of the appropriate space for recording, controlling, indexing and disseminating the received information is obtained the first rank among the other strategies. Also, the creation of a comprehensive knowledge database for all researches in different universities is another main strategy that is ranked in second priority.

Povzetek: Z metodami multikriterijskega odločanja, kombiniranja SWAT in BMW, je narejeno izbiranje najboljših strategij za CRIS, tj. za informacijske sisteme.

1 Introduction

In today's competitive world, information, equal to capital and human resources, is an influential factor of production and is considered as the most important relative advantage of economic enterprises. One of the features of new organizations is the over-accumulation of data, so increasing the amount of data and consequently obtained information in organizations and the need to use them in organizational decisions over the past two decades has led to the emergence of an approach called knowledge management. This necessitates the planning, organization, leadership, and monitoring of organizational knowledge, as well as the management of the process of access to the right knowledge, in order to be effective. In the current era, organizations have found that they will not survive unless they have a strategy to manage their organizational knowledge. Therefore, strategies and cycles for implementing knowledge management are evolutionally presented. On the other hand, network information systems (NIS) provide new opportunities for data quality management, which can include access to a wider range of data sources, the ability to select and compare information from different sources, to detect and correct errors, and, consequently, an overall improvement on the quality of the data. These contexts provide a wide range of evaluation techniques and data quality improvements for issues such as linkage and background, business rules, and coherent scales. Over time, these techniques evolved to counter the increasing complexity of data in information systems. Given the variability and complexity of these techniques, recent researches focus on different methods

and strategies that help to select, customize and apply evaluation techniques and improve data quality. Recently, a newly developed information system called current research information systems (CRIS) has attracted attention and with this type of information systems, scientific organizations can provide a current scheme for their research activities and results, such as projects, third-party funds, patents, cooperation partners, prices and publications [1]. Furthermore, using CRIS they can manage information about their scientific activities as well as integrate them into websites [1]. Since the lack of proper information in organizations, loss of important information in the knowledge databases and the lack of full access to the important information are the main quality issues in CRIS, data quality plays an important role in the deployment process of CRIS [1]. Studies on research information management have revealed that standardization, coordination, and integration of research information is often required and challenging, but one of the main drivers for the implementation of CRIS is the benefit of integrated data collection on research information. At present, much effort has been put into collecting, integrating and aggregating research information [2]. Since different organizations are constantly requesting reports on research results, so having a uniform data model (or even a standard) for research information could simplify this request [3], [4]. Also, because of the reviewed and recommended data quality techniques (e.g., data cleansing and data profiling by [5] and [6]) that are being used in organizations lately; the application of appropriate data quality strategies is the primary concern in organizations to provide research information for strategic planning to prepare and present in a structured manner.

In this study, we therefore try to develop data quality strategies for a CRIS case using the SWOT approach (strengths, weaknesses, opportunities and threats). Due to the many varieties in the strategies and resource constraints received from organizations for the application of these strategies, an MCDM method (Multi-Criteria Decision-Making) called Best-Worst Method (BWM) is implemented to prioritize the final strategies. This combination was previously used by researchers such as [7], [8] and [9] by integrating methods such as SWOT with BWM or TOPSIS (technique for order preference by similarity to the ideal solution) to achieve the desired results.

The next subsection tries to further introduce the contribution of the current study.

1.1 Contribution of study

Although in some previous studies strategic management has been considered as the main tool for quality improvement in information systems [10], extending this approach to CRIS is a state-of-the-art work. In addition, the combination of this approach according to data quality dimensions and the development of strategies with regard to data quality principles have rarely been investigated [11] [12]. Because of the importance of CRIS as the main category of information systems, this paper therefore provided a SWOT framework to improve the effectiveness of these systems. Due to different data quality dimensions and successively defined strategies, a newly developed MCDM approach is applied to prioritized strategies.

The structure of this paper is as follows. In the next section, background works are introduced. Then in section (3), the methodology of current research is developed. After that, in section (4), the case in which the current research is done is described. In section (5) the main finding of the research is showed. Finally, the discussion and conclusions are provided in section (6).

2 Literature review

Since strategies for improving data quality in the current research information system (CRIS) in the proposed case are inspired by the principles of knowledge management, this section first provides a brief history of the framework for knowledge management. Next, the principles of CRIS and the quality of the data are discussed during the integration of scientific works (such as theses and dissertations), research projects, etc. into CRIS. Through data integration, quality problems can be identified and necessary improvements made. The data quality dimension is then described to provide a framework for the appropriate definition of strategies. Finally, the SWOT method (strengths, weaknesses, opportunities and threats) is presented as a well-known approach to defining strategies.

2.1 Knowledge management

A large number of studies on information systems (IS) have proven the importance of knowledge in the organization [13]. These researches declare that knowledge is more valuable than other assets in the organizations; consequently, it needs to be managed more efficiently. Knowledge Management (KM) has become a prevalent research trend in academia and the business sector [14]. KM is defined as "the process of capturing, storing, sharing, and using knowledge" [15]. Besides, KM is an emerging mechanism that can find particular information more efficiently and organize that information for quick retrieval and reuse [16]. KM can be one of the fundamental approaches of modern institutions as it can lead to the maintenance, growth, success, and innovation of the organization [17]. There are several methods of KM from the perspective of researchers. [18] believes that knowledge management processes include knowledge creation, knowledge transfer, and the application of knowledge. [19] pointed out that KM processes include knowledge gathering, knowledge transfer, and the use of knowledge. [20] and [21] pointed out that KM processes include knowledge acquisition/creation, knowledge sharing/dissemination, and knowledge utilization. [21] has determined that KM processes are working in a continuous cycle, in which, it enables the information systems users to achieve their goals, add a piece of new knowledge and share that knowledge accordingly. One way of evaluating KM processes is whether it is possible to deepen the subject.

Figure 1: Significant events in development of the KM.

The question we have to answer is: How important are databases for scientific production? [22]. Existing databases make explicit knowledge stored and accessible, and since dynamic knowledge is constantly evolving, robust and flexible knowledge management systems are essential to receive frequent updates from all parts of the organization. The access routes must be classified precisely and consciously both as a key word and as key terms for those seeking knowledge [23]. KM processes are considered as the fundamental processes for the successful adoption and implementation of a new IS [24], [25] [26]. Also, IS can be employed to leverage the KM processes of acquiring, storing, sharing, and applying a particular knowledge [27]. The main KM processes are knowledge discovery, knowledge capture, knowledge sharing and knowledge application. Besides, [28] demonstrated that information technologies could serve as a facilitator of KM. Based on this literature, which is briefly shown in Figure 1, it is assumed that KM is mainly related to the support of information system processes and that KM as a scientific area is much more than just supporting the development of IS.

This paper examines in a CRIS which is implemented for the exchange and dissemination of knowledge for all researchers. In this regard, the next subsection is devoted to a brief explanation of CRIS.

2.2 CRIS

Access to information about current research activities and their results across Europe is an essential prerequisite for the success of EU innovation policy [49]. That is why the CRIS was developed and is the most important reporting instrument for research-based funding [50]. CRIS or research information systems (RIS), scientific information systems (SIS), alike enterprise information systems (EIS), should cover interdisciplinary aspects as a dimension that influences significantly research potential and activity of particular scientists. A CRIS is a specialized database or federated information system to collect, manage and provide information on research activities and results [1]. In addition, the CRIS is said to be a useful tool for researchers and research institutions by providing a range of services, such as: For example, simplifying the administrative routines for researchers and widespread reuse of the high-quality data registered in the CRIS [50]. Further literature on the term CRIS can be found on the CRIS Repository website (https://dspacecris.eurocris.org/).

The structure and functionality of CRIS could be divided into three layers according to [1] as follows:

1. The data access layer contains the internal and external data sources, e.g., operational databases (human resources, finance, project management and etc.), open repositories, identifiers (ORCID, DOI, etc.), bibliographic data from the Web of Science, Scopus or PubMed, etc. This layer includes data models for the standardized collection, provision, and exchange of research information, such as the Research Core Dataset (RCD) and the Common European Research Information Format (CERIF). The integration of these data sources into the CRIS takes place via classical Extract, Transform and Load (ETL) processes.
2. The application layer (backend) contains the CRIS and its applications, which merge, manage and analyze the data held at the underlying level.
3. The presentation layer (frontend) shows the target group-specific preparation and presentation of the analysis results for the user, which are made available in the form of reports using business intelligence tools, via portals, websites, etc.

CRIS is becoming increasingly important at European and international universities. Therefore, the special features of CRIS can be explained to make the differences to the other information systems clear. CRIS can combine the university's internal systems such as personnel, student administration, finance and price management systems as

well as a variety of external data sources, including pre-made researcher profiles via Profile Refinement Services, as well as existing data on a single platform. Researchers, administrators, and delegates enter data only once, and staff across the university use the information in CRIS for a variety of purposes. CRIS offers other special features as follows:

- CRIS offers the institution a comprehensive overview of the activities, specialties and achievements of its researchers.
- CRIS can also search external data sources (e.g. Scopus, WoS, PubMed, arXiv, CrossRef, Mendeley, etc.) to determine the results of researchers at their institution. CRIS automatically retrieves the metadata and saves researchers time and effort.
- CRIS makes it easier to create, update and correct researcher profiles by automatically retrieving publication lists from relevant internal and external databases.
- With CRIS, CVs for different requirements can be created at the push of a button and then exported as Word or PDF files or published online.
- CRIS supports universities and their scientists in their search for research opportunities, research sponsors and mentors, etc.
- With the CRIS, universities can find internal and external cooperation partners.
- Much more.

In the next subsection data quality dimensions besides basic strategies for improving data quality are introduced.

2.3 Data quality (DQ)

The research carried out on semi-structured and unstructured events in the DQ domain indicates a strong historical connection between the DQ and the database design [16]. Even complete DQ procedures have bias but their focus is on a set of structured data that provides the most information sources in organizations [12]. Nowadays, switching to semi-structured data and the lack of structure as a corporate information resource is far more common. Also, DQ techniques for semi-structured and unstructured data have recently been investigated. Improving DQ techniques for unstructured and semi-structured data in these domains requires a higher degree of interpersonal communication [29]. The efficacy of scientific data collection and validation processes has always been debated. Traditional approaches are likely to result in poor quality scientific data being recorded [30]. As a result, the scientific results that are mostly based on these data are also of poor quality, and even if the data collection and validation steps are performed correctly, the processes performed are not always qualitatively documented in the scientific paper. This leads to not only a very difficult understanding of scientific literature, but also scientific studies that are difficult to reproduce. This lack of reproduction has been led to a growing concern in various research areas [31], [32], [33]. However, attention to the reproducibility of IS research has so far been limited [34]. Also, the relationship between data quality and process quality is due to the linkage and variety of features of business processes in organizations [35], covering a large part of the research. Different effects of data quality have been investigated in three levels of operational, tactical, and strategic levels in research [16]. Quality of data and its relevance to the quality of services, products, business operations and consumer behavior are widely discussed in the general terms [36], [37]. In these studies, general statements such as “the quality of a company's information is positively related to firm performance" was based on empirical evidence. Also, the issue of how improving information production processes positively affects data and the quality of information has also been analyzed. In the process of improvement, each of the different methods can adopt two general strategies as follows [16]:

1. Data-driven strategy
2. Process-driven strategy

Data-driven strategies improve data quality by directly modifying the value of data. For example, the quality of the data is improved by updating them from another database and replacing them with updated data. Process-oriented strategies also improve quality by redesigning processes that create or modify data. For example, one can redesign a process by including activities that control the data format before storage [38]. Data-centric and process-oriented strategies implement various types of techniques, such as algorithms, intelligent technologies, and knowledge-based activities, aimed at improving data quality. A list of improvement techniques related to strategy-based approaches is as follows:

1. Access to new data, which improves the old data and gets higher quality, and this technique is used instead of methods that cause quality problems in the data.
2. Standardization (or normalization) that replaces or completes non-standard data values with values that conform to the standard. For example, the nickname is replaced by the original name. For example, Bob with Robert, and abbreviations corresponding to the full name are replaced.
3. The history link, which identifies the display of information in two (or several) tables that may refer to the same entity in the real world.
4. Integrating data and designs, which provides an integrated view of the information provided by heterogeneous data sources. The main purpose of integration is to provide a user with access to data stored in heterogeneous data sources and through the integrated view of these data.

According to theory, our data processing consultancy offers a new solution to improve data quality at an early stage, including condition analysis, software design, software implementation and data integration through data consultancy. To construct a framework for the rule-based measurement of IS research data quality, we start from the seven aspects or seven W's of scientific data collection and validation identified by [39] as follows:

1. What explains exactly what is captured in the data.
2. When refers to the time at which the data are collected.

Intrinsic	Accessibility	Contextual	Representation
Accuracy	Accessibility	Relate	Interpretability
Objectivity	Security	Value Added	Ease of understanding
Belief		Up to date	Compatible display
Confidence		Comprehensiveness	
		Data amount	

Table 1: Data quality dimensions [41].

3. Where refers to the location (virtual or real) where the data are collected.
4. How describes the precise process(es) of data collection.
5. Who details the individual(s) involved in the data collection.
6. Which details the instruments or artifacts used in collecting the data.
7. Why provides the set of reasons or goals for collecting the data.

Failure to properly implement each of these seven aspects reduces the quality of the research data [33]. Data quality depends largely on the organization of the information system and how it is processed. Then, measuring and improving data quality in organizations are complex tasks. Hence, to assess the quality of metadata, it is required to employ a standard structure. In this paper, we not only use the four data quality dimensions (completeness, correctness, consistency, and timeliness) in the context of CRIS [40], but also focus on the dimensions used by [41]. These dimensions are accuracy, objectivity, reliability, authenticity, relevance, value-added, update, comprehensiveness, amount of data, interpretability, ease of perception, concise presentation, consistent display, availability, and security. Besides, they are categorized into four categories: intrinsic, contextual, representation, and accessibility. Theses category are shown in the following Table 1. Therefore, identifying appropriate executive policies for improving the data quality of CRIS at hand needs to define suitable KM strategies in the context of DQ dimensions. In the next subsection, a framework called SWOT is introduced for this aim.

2.4 SWOT

SWOT Analysis is a tool used for strategic planning and strategic management in organizations. It can be used effectively to build organizational strategy and competitive strategy. In accordance with the System Approach, organizations are in interaction with their environments and comprised of various sub-systems.

So, an organization communicates with two environments, first its inside and the second its outside. It is a necessity to analyze these environments for strategic management practices. This process of examining the organization and its environment is termed SWOT Analysis. SWOT analysis is a simple but powerful tool for sizing up an organization's resource capabilities and deficiencies, its market opportunities, and the external threats to its future" [42]. The acronym SWOT stands for '*strengths*', '*weakness*', '*opportunities*' and '*threats*'. SWOT analysis is a strategic planning framework used in the evaluation of an organization, a plan or a project. SWOT Analysis has two aspects: internal and external. Internal aspect consists of organizational factors, also strengths and weaknesses, while external aspect consists of environmental factors, also opportunities and threats. Strengths and weaknesses are internal factors and attributes of the organization, opportunities and threats are external factors and attributes of the environment [43]. On the context of CRIS based on some main researches such as [44] and [45] the following general structure of SWOT could be achieved.

The **strengths** of CRIS are:

- Easy access to information regarding research activities
- Research activities are supported and optimized
- It is possible for researchers to manage their own activities
- Administration of the data takes less time
- Information is stored in a system and in a database
- Effective data usage
- Easy retrieval for prospects and cooperation partners of persons and contact information, research activities and services
- Clear presentation of a research profile
- Auxiliary function in the creation of e.g. CVs and publication lists
- Finding and sharing research information
- Research activities can be represented graphically by analysis and visualization function

The **weaknesses** of CRIS

- Introduction of CRIS means high financial and time expenditure
- Furthermore, there are several sources regarding the query
- Several entries necessary
- Scattered information
- Data is publicly available, even if it is only stored in the background

The **opportunities** of CRIS

- As consistent as possible, so comparisons and assessments can be made quickly and easily
- Supporting the design and selection of a CRIS by standardization, so that more benefits arise, such as data exchange
- Collaborations between scientists or departments should be analyzed in order to find out in which areas these cooperation's exist

The **threats** of CRIS

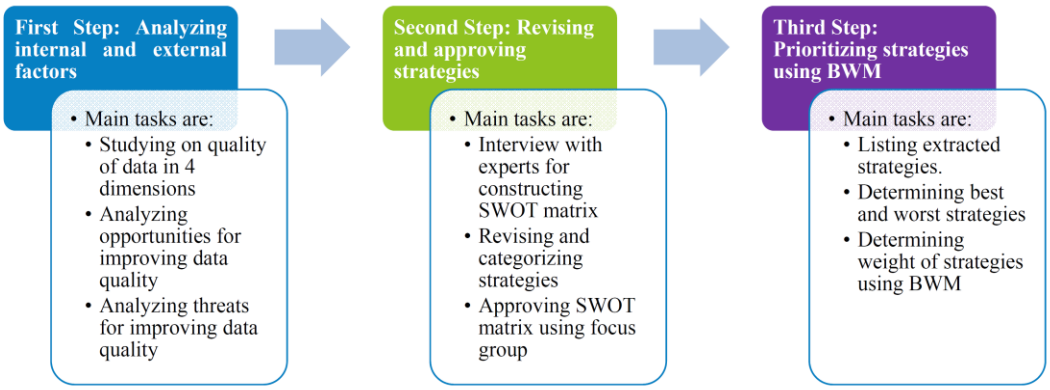


Figure 2: The methodology of the current study.

- Compatibility and interoperability of different CRIS data - different standardizations
- Research institutions and universities are able to develop their own CRIS
- Open source solutions can be used as an alternative to CRIS

After giving an overview of the SWOT analysis in the context of CRIS, the purpose of the present research is to answer the following questions:

1. What are the opportunities and threats and the strengths and weaknesses of the organization in the use of data quality of information systems?
2. What are the effective strategies in the information system of CRIS?

To answer these questions, the brainstorming method was used, which was derived from five experts who were CRIS and data quality professions (these are CRIS managers interviewed). CRIS managers have to do an important balancing act in day-to-day business: It is important to pursue well-founded strategies, which, however, can be adjusted just as dynamically if influencing factors change. In order to master this challenge, SWOT analysis is often used as a means of strategy generation. It is ideal for the development of various strategies. The aim of our SWOT analysis is to derive the appropriate measures for more success from the use of CRIS. The SWOT research has shown that most universities spend a lot of money, time and nerves to eliminate weaknesses and often spend themselves or get bogged down in the process.

The structure of main steps for finding best DQ strategies of CRIS with high priorities is described in the next section.

3 Methodology

This research is executed of the following 3 main steps as are shown in Figure 2. In the first step after studying the quality of data in different four dimensions, complete analysis on inside and outside of the organization is done. In this regard, opportunities for improving the quality of data and threats which have main effects on quality are determined. Then, in the second step, the internal and external factors are revised and approved through an interview with experts. SWOT matrix is constructed and finalized in this step using a focus group. Finally, in the third step, all strategies are ranked using the BWM method. The strategies with high importance are obtained in this step.

According to the research methodology, in this research five experts from the organization which were professions in CRIS and data quality framework have been consulted. Their expertise was data quality, information science, and information system design. For analysis of data in BWM techniques have been and the used software was Lingo[1].

3.1 The Best-Worst Method

In this section, the steps of the BWM method [46] which have been used to gain weight of each criterion is described.

Step 1: Specify the set of criteria: In this step, we consider the criteria {C1, C2, ..., Cn} to be used in decision making.

Step 2: Identify the best (in other words, the most desirable and most important) and the worst (the most unfavorable and the least important) criteria. In this section, the decision-maker generally outlines the best and worst criteria.

Step 3: Determining the performance of the best criterion against other criteria using numbers from 1 to 9. The best criteria for the other criteria may be as follows:

Eq.1

$$A_B = (a_{B_1}, a_{B_2}, \ldots, a_{B_n})$$

Which a_{B_j} specifies the performance of the best B criterion relative to the j criterion. Obviously A_{B_b}= 1. For example, this vector represents the performance of the price benchmark against other criteria.

[1] https://www.lindo.com/index.php/products/lingo-and-optimization-modeling

a_{B_w}	1	2	3	4	5	6	7	8	9
Consistency Index (max ξ)	0.00	0.44	1.00	1.63	2.30	3.00	3.73	4.47	5.23

Table 2: Compatibility rate.

Step 4: Specify the performance of all criteria against the worst-case using numbers *from 1 to 9*. The results of comparisons of criteria to the worst-case criteria can be as follows:

Eq.2

$A_W = (a_{1_w}, a_{2_w}, \dots, a_{n_w})T$

a_{j_w} represents the performance of criterion j versus the worst W criterion. Obviously, the value A_{W_w}=1. For example, our view is that this vector represents the performance of all criteria relative to the appearance criterion.

Step 5: Find the optimal weights ($w_1{}^*, w_2{}^*, \dots, w_n{}^*$)

The optimal values for the unique criteria are for each pair of $W_B/W_J = a_{B_j}$ and $W_J/W_W = a_{j_w}$.

To satisfy these conditions for all j, we need to find a solution that minimizes the magnitude of the maximum difference between $\left|W_B/W_J - a_{B_j}\right|$ and $\left|W_J/W_W - a_{j_w}\right|$.

Given that the weights are non-negative and admissible; the following problem can be expressed in the non-linear model according to formula 3:

Eq.3 Min max $\left\{ \left|W_B/W_J - a_{B_j}\right|, \left|W_J/W_W - a_{j_w}\right| \right\}$

s.t

$\Sigma_j W_j = 1$

$W_j \geq 0$, for all j

The above problem can be expressed in formula 4:

Eq.4

Min ε

s.t

$\left|W_B/W_J - a_{B_j}\right| \leq \varepsilon$, for all j

$\left|W_J/W_W - a_{j_w}\right| \leq \varepsilon$, for all j

$\Sigma_j W_j = 1$

$W_j \geq 0$, for all j

And this is converted into a linear model in formulas 5 which makes it easier to compute:

Eq.5

Min max $\left\{\left|W_B - a_{B_j} W_J\right|, \left|W_J - a_{j_w} W_W\right|\right\}$

The above problem can be expressed using the formula 6 as follows:

Eq.6

Min ξ^L

s.t

$\left|W_B - a_{B_j} W_J\right| \leq \xi^L$, for all j

$\left|W_J - a_{j_w} W_W\right| \leq \xi^L$, for all j

$\Sigma_j W_j = 1$

$W_j \geq 0$, for all j

By solving the above equation, we obtain the optimal values of the weights ($w_1{}^*, w_2{}^*, \dots, w_n{}^*$) and the value of ε^*. Then, using ε^*, we introduce a compatibility rate. It will be clear that larger values for ε^* will result in higher compatibility rates and lower reliability of the comparisons.

3.1.1 Calculation of compatibility rate

In this subsection, a consistency ratio is proposed for the BWM to check the reliability of the comparisons. For each criterion j, a comparison will be perfectly consistent when $a_{Bj} \times a_{jW} = a_{BW}$, where a_{Bj}, a_{jW}, and a_{BW} represent the performance of best criterion related to criterion j, criterion j related to worth criterion, best criterion related to worth criterion respectively [46]. Since the proposed BWM may not be fully compatible with some j we used the compatibility rate to evaluate possible inconsistency. To do this, we compute the lowest compatible value of comparison as follows.

The set a_{ij}={1, … , a_{BW}} indicates that the highest possible value for a_{BW} is 9. The compatibility value decreases when $a_{B_j} \times a_{jw}$ is less or more than a_{BW}, or the equation $a_{Bj} \times a_{jw} \neq a_{BW}$ is established. In other words:

Eq.7

$$\left(a_{B_j} - \varepsilon\right) \times \left(a_{j_w} - \varepsilon\right) = \left(a_{B_w} + \varepsilon\right)$$

As stated above, at least the compatibility is when a_{B_w} = a_{B_j} = a_{j_w}. Thus, we have:

$$\left(a_{B_j} - \varepsilon\right) \times \left(a_{j_w} - \varepsilon\right) = \left(a_{B_w} + \varepsilon\right) \Rightarrow \varepsilon^2 - \left(1 + 2a_{B_w}\right)\varepsilon + \left(a_{B_w}{}^2 - a_{B_w}\right) = 0$$

Solving this equation for ε lead we to the maximum value of ε as indicated in the Table below.

Then we get the compatibility rate value using ε^* from Table 2 and its compatibility index using formula 9. Based on equation 9 if the compatibility rate falls in the appropriate region then the proposed BWM is verified.

Eq.8

$$\text{Compatibility rate} = \frac{\varepsilon^*}{\text{Compatibility Index}}$$

In the rest of this paper, based on case study strategy for improving data quality obtained and using BWM are ranked.

4 Case study

This paper uses the pre-defined methods to determine and evaluate the main data quality strategies of CRIS. CRIS is considered as online dissemination system for Iranian

theses and dissertations (GANJ)[2] and the largest national scientific treasure. In addition, CRIS is the reference of many researchers around the world. GANJ was developed for Iranian metadata of scientific research (such as publications, patents, projects, etc.). It hosts over 10,000 users who perform tens thousands of searches a day (i.e. around 10,000 unique IP-based users).

In this work, the main processes of CRIS are examined and documented. Existing processes in CRIS are divided into three general sections, each referred to as (i) acquisition and registration of scientific document; (ii) indexing and (iii) dissemination.

1. Acquisition and registration process
   Inputs of acquisition and registration of scientific document processes in the CRIS are metadata for scientific works (e.g. theses, dissertations, etc.), research projects and government reports. This process includes quality control operations implemented in all fields of metadata.
2. Indexing process
   The providing information process includes the preparation of documents and records of information. Quality control operations will be implemented in all of these subsections and processes mentioned for ensuring the validity of information in the system.
3. Dissemination process
   Editing metadata received from the indexing process and assigning a unified code to each record is done in this process. Overview of bibliographic information is reviewed too. If there is no problem, the document is approved and disseminated. The main role of this process is storage and dissemination of metadata, but any quality problem is identified that record would be returned to the indexing unit.

In this research using SWOT, while considering data quality dimensions, the main strategies for augmenting the quality of data are defined. Then the proposed strategies are ranked based on the BWM method.

The finding of this research is described in the next section.

5 Findings

Given the importance of data quality in the KM process, it is necessary to review the strategies in accordance with the knowledge transfer hierarchy in the organization. In order to determine the strategies for improving the quality of data in knowledge management, SWOT analysis with BWM has been used. So, according to three main stages of this research which are introduced in section 3, external and internal environment are analyzed and the controllable and uncontrollable sub-factors that affect different dimensions of data quality are identified. To comprehensively implement this stage, as was explained in the section 2, the brainstorming method was employed to do SWOT analysis based on expert's judgments. For better classifying the obtained SWOT, each analysis was done based on data quality dimensions according to Table 1 and the results of are shown in Table 3. Then, based on stage 2 of methodology which was explained in Figure 2 using SWOT sub-factors and finally the SWOT matrix and strategies were formed (see Table 4). The concept of SO strategy is the proper use of opportunities by exploiting the strengths of the organization. The WO strategy seeks to exploit appropriate environmental opportunities in light of the organization's weaknesses. ST strategy is also related to reducing or eliminating the effects of environmental threats through the optimum use of the strengths of the organization. Finally, the WT strategy, taking into account the organization's weaknesses, reduces the effects of environmental threats. The final result after approving experts is shown in Table 4.

In the last sub-step of the stage 2, according to the information gathered from the experts and using focus group method, SWOT components as are shown in Table 4 were verified. In other words, as is demonstrated in Table 4, four basic criteria for formation of data quality strategies on CRIS to respond the first question of our research, are opportunities, threats, weakness and strength.

In the third stage, according to the identification of the organization's strategies, we will rank the strategies, which its result will be shown using the following five steps based on the BWM technique.

Step 1: Determine a set of decision criteria.

Step 2: Determine the best (most desirable, most important) and worst (the most unfavorable, least important) criterion.

In this section, according to an opinion poll from the organization's experts, W_1 and W_9 policies were also evaluated and introduced as the best and worst policy.

Step 3: Determine the importance of the best benchmark against other criteria (see Table 6).

Step 4: Determine the importance of other criteria to the worst criterion (see Table 7).

Step 5: Determine the optimal weight.

Relationship among criteria is constructed based on model (4) as follows:

$$\text{Min } \varepsilon$$

$$\text{s.t}$$

$$|W_1 - 3.6\,W_2| \le \varepsilon \qquad |W_2 - 4.3\,W_9| \le \varepsilon$$
$$|W_1 - 4.2\,W_3| \le \varepsilon \qquad |W_3 - 5.1\,W_9| \le \varepsilon$$
$$|W_1 - 3\,W_4| \le \varepsilon \qquad |W_4 - 4.2\,W_9| \le \varepsilon$$
$$|W_1 - 5.4\,W_5| \le \varepsilon \qquad |W_5 - 6.2W_9| \le \varepsilon$$
$$|W_1 - 4.8\,W_6| \le \varepsilon \qquad |W_6 - 4.7\,W_9| \le \varepsilon$$
$$|W_1 - 6.2\,W_7| \le \varepsilon \qquad |W_7 - 6.2\,W_9| \le \varepsilon$$
$$|W_1 - 4\,W_8| \le \varepsilon \qquad |W_8 - 4\,W_9| \le \varepsilon$$
$$|W_1 - 4.6\,W_9| \le \varepsilon \qquad |W_{10} - 5.2\,W_9| \le \varepsilon$$
$$|W_1 - 4.8\,W_{10}| \le \varepsilon \qquad |W_{11} - 4.7\,W_9| \le \varepsilon$$
$$|W_1 - 4.4\,W_{11}| \le \varepsilon \qquad |W_{12} - 5.2\,W_9| \le \varepsilon$$
$$|W_1 - 4.2\,W_{12}| \le \varepsilon \qquad |W_{13} - 4.7\,W_9| \le \varepsilon$$
$$|W_1 - 4\,W_{13}| \le \varepsilon$$

[2] https://en.irandoc.ac.ir/service-product/94

$$W_1 + W_2 + W_3 + W_4 + W_5 + W_6 + W_7 + W_8 + W_9 + W_{10} + W_{11} + W_{12} + W_{13} = 1$$
$$W_1 + W_2 + W_3 + W_4 + W_5 + W_6 + W_7 + W_8 + W_9 + W_{10} + W_{11} + W_{12} + W_{13} \geq 0$$

Strategies	Identified factors
Opportunities outside the organization	**1. Accuracy** Use Auto-Text Correcting Techniques based on the Deep Learning method The presence of university representatives in the process of registering theses/dissertations
	2. Objectivity Use of knowledge of the other experts out of the organization in the development of the system Study the effect of data quality in future research
	3. Believability Creating a competitive development of knowledge management software The use of data analysis institutes in the development of information quality
	4. Validity Assessing the reputation of external information sources by universities and higher education institutions Identification of successful organizations in the field of data quality
	5. Availability Use valid external available resources
	6. Security The use of modern information protection technologies The use of rival strategy to create internal information security
	7. Relevancy Development of technologies of software provider companies Establishing necessary information infrastructure in the country
	8. Value Added Use of software and data mining analytics for better presentation and dissertation development
	9. Being up to date Knowledge-based development in the evaluation of data quality Use of new technologies in converting transferable data in organizational references
	10. Comprehension Development of operational levels of authoritative scientific and operational references Establishing necessary infrastructure at universities
	11. The amount of data Establishing small scientific bases at educational institutions The motivation of competitors using data quality approaches Create infrastructure to get all useful information
	12. Interpretability Provide training on personnel for augmenting information quality Simplifying information in main resource tanks
	13. Ease of understanding Create new search engines in non-organizational resources Creating ease of access and understanding infrastructures in rival booths
	14. Concise presentation Indexing information on rival information
	15. Compatible display Assessing rival approaches to access information for the audience

External threats of the organization	**1. Accuracy** Transferring the false information by other systems Disruption of the provided information by other resources **2. Objectivity** Improper use of external information **3. Believability** Lack of access to data analysis institutions **4. Validity** Non-conformity between the scientific text and their related resources **5. Availability** Lack of accurate scientific information **6. Security** Using unsupported data in research **7. Relevancy** Non-conformity between the content and purposes of the research **8. Value Added** Disapproval of the value creation in a case study research by an authorized organization **9. Being up to date** Delay in the process of registering theses/dissertations at universities **10. Comprehension** Theses/ dissertations with high similarity **11. The amount of data** Incompatibility of data with research objectives **12. Interpretability** Lack of simplification on information extracted from main sources **13. Ease of understanding** Lack of proper description in the text **14. Concise presentation** Failure to proper indexing of thesis/dissertations on other websites **15. Compatible display** Lack of alignment of texts and resources
Internal organization strengths	**1. Accuracy** Accuracy in information classification Accuracy in the amount of information used Improve the process of registering theses/dissertations **2. Objectivity** Data classification in internal knowledge management In-company software development in referrals **3. Believability** Create the necessary training in data exploitation **4. Validity** Information sampling of resources **5. Availability** Create a new method in the intelligent search **6. Security** Powerful access and plagiarism **7. Relevancy** Development of internal infrastructure for the maintenance **8. Value Added** Verifying of information by experts **9. Being up to date** Use of authoritative references in information classification

Internal organization strengths	**10. Comprehension** Comparisons of interdisciplinary researches **11. The amount of data** Internal alignment of the organization with the content Internal authentication **12. Interpretability** Evaluation of published articles in databases The amount of reference information on different databases **13. Ease of understanding** Research and science assessment by the organization's experts **14. Concise presentation** Indexing in national and accredited libraries **15. Compatible display** Establishing a research plan for researchers to register their research
Weaknesses within the organization	**1. Accuracy** Lack of quality assessment of the information contained in the research Lack of experts in the field of information evaluation **2. Objectivity** Lack of access to all credible library information for research approval Incompatibility of data tanks with new information received **3. Believability** Failure to create research incentives for domestic researchers Lack of financial support from research repositories **4. Validity** Failure to assess the credibility of external sources of information Failure to identify successful organizations in the field of data quality **5. Availability** Not recognizing valid external sources available Failure to classify research data in reservoirs **6. Security** Failure to use security systems to protect the work Creating access to anonymous users **7. Relevancy** The lack of development of software technologies in identifying information Failure to create the necessary information infrastructure in information repositories **8. Value Added** Incompatibility of the organization's policies with industrial relations researches Not information classification **9. Being up to date** Lack of access to new authoritative references for comparison Lack of international cooperation in transferring research achievements **10. Comprehension** Lack of alignment of higher education policies with data quality assessment policies **11. The amount of data** Failure to use decision support systems in the expert system **12. Interpretability** The reluctance of the experts to participate in the research interpretation **13. Ease of understanding** Failure to create research training in the learning environment **14. Concise presentation** Failure to update the information profile on the site **15. Compatible display** Failure to create compatibility tanks for topics with academic disciplines

Table 3: Identification of strategies based on different dimensions of data quality.

SWOT	Strengths (S)	Weaknesses (W)
Opportunities (O)	SO policy: 1. Developing and classifying the appropriate space for recording, controlling, indexing and disseminating received information 2. Development of communication with authoritative databases for the appropriate dissemination of research data 3. Use of experts in research affairs to index the data	WO policy: 1. Establish a comprehensive knowledge repository base of all research in all universities 2. Collaborate with other knowledge databases in the classification process of information 3. Improving the comprehensive thesis/dissertation registration process at universities 4. Establish a comprehensive authentication system
Threats (T)	ST Policy: 1. Development of software for assessing the quality of data using experienced staff 2. Classification of research information 3. The weighting of researches based on the quality of used data	WT policy: 1. Benchmarking of strong information tanks in other countries 2. Establish training courses for key personnel of the organization 3. Survey of users satisfaction

Table 4: Extracted strategies.

Symbol	Strategy	Symbol	Strategy
$\mathbf{W_1}$	Developing and classifying the appropriate space for recording, controlling, indexing and disseminating of the received information	$\mathbf{W_7}$	Comprehensive Authentication / Integration System
$\mathbf{W_2}$	Development of communication with authoritative databases for the proper dissemination of research data	$\mathbf{W_8}$	Development of software for assessing the quality of data using experienced staff
$\mathbf{W_3}$	The use of professors in research affairs to index the data used	$\mathbf{W_9}$	Classification of research information
$\mathbf{W_4}$	Create a comprehensive knowledge base of all research in all universities	$\mathbf{W_{10}}$	Weighing on research based on the quality of data used
$\mathbf{W_5}$	Collaborate with other knowledge bases in the classification of information	$\mathbf{W_{11}}$	Bench Marking has strong information repositories in other countries
$\mathbf{W_6}$	Changes in the comprehensive thesis/dissertation registration process at universities	$\mathbf{W_{12}}$	Creating training courses for key personnel of the organization
		$\mathbf{W_{13}}$	Use surveys of users

Table 5: Categorized strategies for employing BWM.

6 Conclusion

In this research, weaknesses, strengths, opportunities and organizational threats were discussed in the subject areas of data quality and finally 13 key components were identified as main policies for improving data quality of studied CRIS. To achieve these 13 components for augmentation of quality of CRIS, 15 different dimensions of data quality (such as accuracy, objectivity, believability, etc.) and their impact on the studied CRIS were comprehensively studied. Then, using the BWM method, effective strategies were ranked and prioritized. The results of the evaluation showed that the development and classification of the appropriate space for the recording, control, indexing and dissemination of information was given top priority. The second most important component is the creation of a comprehensive knowledge base of all research data at all universities. The organization should use the strategy under investigation to

The average weight of strategy indicators based on expert opinion													
Weight	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_{10}	W_{11}	W_{12}	$\mathbf{W_{13}}$
The most important dimension W_1	1.00	3.6	4.2	3	5.4	4.8	6.2	4	4.6	4.8	4.4	4.2	**4**

Table 6: Paired comparison vector for the best criterion.

The average weight of strategy indicators based on expert opinion													
Weight	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_{10}	W_{11}	W_{12}	$\mathbf{W_{13}}$
The most important dimension W_9	4.2	4.3	5.1	4.2	6.2	4.7	6.2	4	1	5.2	4.7	5.2	**4.7**

Table 7: Paired comparison vector for the worst criterion.

W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_{10}	W_{11}	W_{12}	W_{13}
0.20	**0.08**	**0.07**	**0.10**	**0.06**	**0.06**	**0.05**	**0.07**	**0.02**	**0.06**	**0.07**	**0.07**	**0.07**

Table 8: Calculated weight of research criteria.

Strategy	**The following are the components of the strategy**	**Strategy weight**
SO Policy	$\mathbf{W_1}$	**0.35**
	$\mathbf{W_2}$	
	$\mathbf{W_3}$	
WO Policy	$\mathbf{W_4}$	**0.27**
	$\mathbf{W_5}$	
	$\mathbf{W_6}$	
	$\mathbf{W_7}$	
ST Policy	$\mathbf{W_8}$	**0.15**
	$\mathbf{W_9}$	
	$\mathbf{W_{10}}$	
WT Policy	$\mathbf{W_{11}}$	**0.21**
	$\mathbf{W_{12}}$	
	$\mathbf{W_{13}}$	

Table 9: Ranking Strategies.

drive their data quality goals. Because good data quality leads to an acceleration of digital processes, an increase in productivity and an increase in corporate success. These results are along with [47] who defined three levels of quality, based partially on the extent of documentation provided to downstream users, whether they be partners, aggregators, practitioners, or users. Application profiles, which at the most basic level document the intent of the creator of the metadata, give important clues to those outside the institution or domain of the metadata creators and are increasingly used to provide guidance to specific organizations and communities of practice. [48] defined some strategies based on data quality dimensions for improving the performance of CRIS. Trying to emulate benchmark practices, evaluating the benefits of the implemented system and finally training the users to the system are the most important strategies which are along with the results of current research.

According to the results of this research, in order to complete the obtained results, it is suggested:

1. Using the fuzzy analysis approach in decision making to increase the accuracy in organizing the organization's strategy.
2. Adding other dimensions of KM on data quality and creating a comprehensive framework of organizational strategies in the form of conceptual models.

3. Classification of databases of articles and dissertations in the center of Iran for the classification of strategy for each of the above two main categories.
4. Using the PESTLE method (Political, Economic, Social, Technological, Legal and Environmental) to more accurately assess the opportunities, strengths, weaknesses and organizational threats.

References

[1] Azeroual, O. & Schöpfel, J. (2019). Quality Issues of CRIS data: An exploratory investigation with universities from twelve countries. *Publications*, 7(1), 14. https://doi.org/10.3390/publications7010014

[2] Ershadi, M. J., & Aiassi, R. (2017). A Model for Quality Assurance on Acquisition and Registration, Processing, and Dissemination of Theses and Dissertations Systems. *Journal of Information Technology Management*, 9(2), 167-190. https://www.sid.ir/en/journal/ViewPaper.aspx?id=575896

[3] Waddington, S., Sudlow, A., Walshe, K., Scoble, R., Mitchell, L., Jones, R., & Trowell, S. (2013). Feasibility study into the reporting of research information at a national level within the uk higher education sector. *New review of information networking*, 18(2), 74-105. https://doi.org/10.1080/13614576.2013.841446

[4] Quix, C., & Jarke, M. (2014). Information integration in research information systems. *Procedia Computer Science*, 33, 18-24. https://doi.org/10.1016/j.procs.2014.06.004

[5] Azeroual, O., Saake, G. & Abuosba, M. (2018). Data quality measures and data cleansing for research information systems. *Journal of Digital Information Management*, 16(1), 12–21. https://arxiv.org/abs/1901.06208

[6] Azeroual, O., Saake, G. & Schallehn, E. (2018). Analyzing data quality issues in research information systems via data profiling. *International Journal of Information Management*, 41, 50–56. https://doi.org/10.1016/j.ijinfomgt.2018.02.007

[7] Chitsaz, N., & Azarnivand, A. (2017). Water scarcity management in arid regions based on an extended multiple criteria technique. *Water Resources Management*, 31(1), 233-250. https://doi.org/10.1007/s11269-016-1521-5

[8] Maghsoodi, A. I., Mosavat, M., Hafezalkotob, A., & Hafezalkotob, A. (2019). Hybrid hierarchical fuzzy group decision-making based on information axioms and BWM: Prototype design selection. *Computers & Industrial Engineering*, 127, 788-804. https://doi.org/10.1016/j.cie.2018.11.018

[9] Gupta, H., & Barua, M. K. (2018). A framework to overcome barriers to green innovation in SMEs using BWM and Fuzzy TOPSIS. Science of The Total Environment, 633, 122-139. https://doi.org/10.1016/j.scitotenv.2018.03.173

[10] Cassidy, A. (2016). A practical guide to information systems strategic planning. CRC press. Available at: https://www.routledge.com/A-Practical-Guide-to-Information-Systems-Strategic-Planning/Cassidy/p/book/9780849350733

[11] Dubey, S., Verma, K., Rizvi, M. A., & Ahmad, K. (2018). SWOT Analysis of Cloud Computing Environment. In Big Data Analytics (pp. 727–737). Springer, Singapore. https://doi.org/10.1007/978-981-10-6620-7_71

[12] Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM computing surveys* (CSUR), 41(3), 16. https://doi.org/10.1145/1541880.1541883

[13] Blumenberg, S., Wagner, H. T., & Beimborn, D. (2009). Knowledge transfer processes in IT outsourcing relationships and their impact on shared knowledge and outsourcing performance. *International Journal of Information Management*, 29(5), 342–352. https://doi.org/10.1016/j.ijinfomgt.2008.11.004

[14] Al-Emran, M., Mezhuyev, V., Kamaludin, A., & Shaalan, K. (2018). The impact of knowledge management processes on information systems: A systematic review. *International Journal of Information Management*, 43, 173–187. https://doi.org/10.1016/j.ijinfomgt.2018.08.001

[15] Lee, J. N. (2001). The impact of knowledge sharing, organizational capability and partnership quality on IS outsourcing success. Information & Management, 38(5), 323–335. https://doi.org/10.1016/s0378-7206(00)00074-4

[16] Lee, C. P., Lee, G. G., & Lin, H. F. (2007). The role of organizational capabilities in successful e-business implementation. Business Process Management Journal, 13(5), 677–693. https://doi.org/10.1108/14637150710823156

[17] Lee, J. C., Shiue, Y. C., & Chen, C. Y. (2016). Examining the impacts of organizational culture and top management support of knowledge sharing on the success of software process improvement. *Computers in Human Behavio*r, 54, 462–474. https://doi.org/10.1016/j.chb.2015.08.030

[18] Spender, J. C. (1996). Making knowledge the basis of a dynamic theory of the firm. *Strategic management journal*, 17(S2), 45–62. https://doi.org/10.1002/smj.4250171106

[19] De Long, D. (1997). Building the knowledge-based organization: How culture drives knowledge behaviors. Ernst & Young Center for Business Innovation, Working Paper, Boston. Available at: http://providersedge.com/docs/km_articles/Building_the_Knowledge-Based_Organization.pdf

[20] Soto-Acosta, P., Popa, S., & Palacios-Marqués, D. (2017). Social web knowledge sharing and innovation performance in knowledge-intensive manufacturing SMEs. *The Journal of Technology Transfer*, 42(2), 425–440. https://doi.org/10.1007/s10961-016-9498-z

[21] Tiwana, A. (2000). The knowledge management toolkit: practical techniques for building a knowledge management system. Prentice Hall PTR. Available at:

https://dl.acm.org/doi/book/10.5555/323909

[22] Figueiredo, M. S. N., & Pereira, A. M. (2017). Managing knowledge–the importance of databases in the scientific Production. *Procedia Manufacturing*, 12, 166-173. https://doi.org/10.1016/j.promfg.2017.08.021

[23] Hoegl, M., & Schulze, A. (2005). How to Support Knowledge Creation in New Product Development: An Investigation of Knowledge Management Methods. *European management journal*, 23(3), 263-273. https://doi.org/10.1016/j.emj.2005.04.004

[24] Chong, A. Y. L., Chan, F. T., Goh, M., & Tiwari, M. K. (2013). Do inter-organizational relationships and knowledge-management practices enhance collaborative commerce adoption? *International Journal of Production Research*, 51(7), 2006–2018. https://doi.org/10.1080/00207543.2012.701776

[25] Lin, H. F., & Lee, G. G. (2005). Impact of organizational learning and knowledge management factors on e-business adoption. *Management Decision*, 43(2), 171–188. https://doi.org/10.1108/00251740510581902

[26] Migdadi, M. M., Abu Zaid, M. K. S., Al-Hujran, O. S., & Aloudat, A. M. (2016). An empirical assessment of the antecedents of electronic-business implementation and the resulting organizational performance. *Internet Research*, 26(3), 661–688. https://doi.org/10.1108/intr-08-2014-0203

[27] Turban, E., Sharda, R., & Delen, D. (2010). Decision Support and Business Intelligence Systems (9th Edition). Prentice Hall, Upper Saddle River. Available at: https://www.pearson.com/us/higher-education/product/Turban-Decision-Support-and-Business-Intelligence-Systems-9th-Edition/9780136107293.html

[28] Mitchell, H. J. (2003). Technology and knowledge management: Is technology just an enabler or does it also add value? In Knowledge management: Current issues and challenges (pp. 66–78). IGI Global. https://doi.org/10.4018/978-1-93177-751-3.ch006

[29] Aljumaili, M., Karim, R., & Tretten, P. (2016). Metadata-based data quality assessment. *VINE Journal of Information and Knowledge Management Systems*, 46(2), 232–250. https://doi.org/10.1108/vjikms-11-2015-0059

[30] Ershadi, M. J., Aiasi, R., & Kazemi, S. (2018). Root cause analysis in quality problem solving of research information systems: a case study. *International Journal of Productivity and Quality Management*, 24(2), 284-299. https://doi.org/10.1504/ijpqm.2018.10012949

[31] Collins, F. S., & Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. Nature, 505(7485), 612-613. https://doi.org/10.1038/505612a

[32] Li, F., Hu, J., Xie, K., & He, T. C. (2015). Authentication of experimental materials: A remedy for the reproducibility crisis? *Genes & diseases*, 2(4), 283. https://doi.org/10.1016/j.gendis.2015.07.001

[33] Timmerman, Y., & Bronselaer, A. (2019). Measuring data quality in information systems research. *Decision Support Systems*, 126, 113138. https://doi.org/10.1016/j.dss.2019.113138

[34] Ershadi, M. J., & Forouzandeh, M. (2019). Information Security Risk Management of Research Information Systems: A hybrid approach of Fuzzy FMEA, AHP, TOPSIS and Shannon Entropy. *Journal of Digital Information Management*, 17(6), 321. https://doi.org/10.6025/jdim/2019/17/6/321-336

[35] Ershadi, M. J., Jalalimanesh, A., & Nasiri, J. (2019). Designing a Metadata Quality Model: Case Study of Registration System of Iranian Research Institute for Information Science and Technology. *Iranian Journal of Information processing and Management*, 34(4), 1505-1534. http://jipm.irandoc.ac.ir/article-1-4091-en.html

[36] Mezghani, E., Exposito, E., & Drira, K. (2016). A collaborative methodology for tacit knowledge management: Application to scientific research. *Future Generation Computer Sy*stems, 54, 450–455. https://doi.org/10.1016/j.future.2015.05.007

[37] Ershadi, M. J., Niaki, S. T. A., & Sadeghee, R. (2019). Evaluation and improvement of service quality in information technology department of a detergent production company using the SERVQUAL approach. *International Journal of Services and Operations Management*, 34(2), 228-240. https://doi.org/10.1504/ijsom.2019.10024665

[38] Batini, C., & Scannapieco, M. (2016). Data and information quality. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-24106-7

[39] Marsden, J. R., & Pingry, D. E. (2018). Numerical data quality in IS research and the implications for replication. *Decision Support Systems*, 115, A1-A7. https://doi.org/10.1016/j.dss.2018.10.007

[40] Azeroual, O., Saake, G. & Wastl, J. (2018). Data measurement in research information systems: Metrics for the evaluation of the data quality. *Scientometrics*, 115(3), 1271–1290. https://doi.org/10.1007/s11192-018-2735-5

[41] Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers? *Journal of management information systems*, 12(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

[42] Gürel, E., & Tat, M. (2017). SWOT ANALYSIS: A THEORETICAL REVIEW. *Journal of International Social Research*, 10(51). Available at: https://www.sosyalarastirmalar.com/cilt10/sayi51_pdf/6iksisat_kamu_isletme/gurel_emet.pdf

[43] Phadermrod, B., Crowder, R. M., & Wills, G. B. (2019). Importance-performance analysis based SWOT analysis. *International Journal of Information Management*, 44, 194–203. https://doi.org/10.1016/j.ijinfomgt.2016.03.009

[44] Johansson, Å., & Ottosson, M. O. (2012). A national Current Research Information System for Sweden. In: Jeffery, Keith G; Dvořák, Jan (eds.): E-Infrastructures

for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic), pp. 67-71. Available at: https://dspacecris.eurocris.org/handle/11366/103

[45] Fernandes, S. (2019). Looking deep at current research information systems. *Qualitative and Quantitative Methods in Libraries*, 7(2), 281-291. Available at: http://78.46.229.148/ojs/index.php/qqml/article/view/475

[46] Rezaei, J. (2015). Best-worst multi-criteria decision-making method. Omega, 53, 49–57. https://doi.org/10.1016/j.omega.2014.11.009

[47] Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: defining, expressing, exploiting. ALA editions. Available at: https://ecommons.cornell.edu/handle/1813/7895

[48] Campanella, P., Lovato, E., Marone, C., Fallacara, L., Mancuso, A., Ricciardi, W., & Specchia, M. L. (2015). The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. *The European Journal of Public Health*, 26(1), 60–64. https://doi.org/10.1093/eurpub/ckv122

[49] Zimmerman, E.H. (2002). CRIS-Cross: Research Information Systems at a Crossroads. CRIS2002: 6th International Conference on Current Research Information Systems (Kassel, August 29-31, 2002). Available at: https://dspacecris.eurocris.org/handle/11366/129

[50] Wenaas, L., Karlstrom, N., Vatnan, T. (2012). From a national CRIS along the road to Green Open Access – and back again: Building infrastructure from CRIStin to Institutional Repositories in Norway. CRIS2012: 11th International Conference on Current Research Information Systems (Prague, June 6-9, 2012). Available at: https://dspacecris.eurocris.org/handle/11366/116?mode=full

https://doi.org/10.31449/inf.v45i1.2586 Informatica 45 (2021) 81–91 **81**

An Approach for Automatic Ontology Enrichment from Texts

Nassima Mellal, Tahar Guerram and Faiza Bouhalassa
Department of Mathematics and Computer Science, Larbi Ben Mhidi University, Oum El Bouaghi, Algeria
E-mail: nassima.mellal.univ@gmail.com, tahar.guerram@gmail.com, faiza.bouhalassa@gmail.com

Keywords: ontology, automaticEnrichment, NLP, SVO triplet, wordnet, texts, phytotherapy

Received: November 14, 2018

The automatic ontology enrichment consists of automatic knowledge extraction from texts related to a domain of discourse in the aim to enrich automatically an initial ontology of the same domain. However, the passage, from a plain text to an enriched ontology requires a number of steps. In this paper, we present a three steps ontology enrichment approach. In the first step, we apply natural language processing techniques to obtain tagged sentences. The second step allows us to reduce each extracted sentence to an SVO (Subject, Verb, and Object) sentence, supposed to preserve main information carried by the original sentence(s) from which it is extracted. Finally, in the third step, we proceed to enrich an initial ontology built manually by adding extracted terms in the generated SVO as new concepts or instances of concepts and new relations. To validate our approach, we have used "Phytotherapy" domain because of the availability of related texts on the WWW and also because its usefulness for pharmaceutical industry. The first results obtained, after experiments on a set of different texts, testify the performance of the proposed approach.

Povzetek: Predstavljena je metoda za izboljšave gradnje ontologij iz besedil.

1 Introduction

Ontology allows knowledge representation in graphical and intuitive manner but its construction and management is a hard task and a very time consuming operation. With the apparition of internet and new information and communication technologies, the mass of produced texts relating to different domains becomes huge and almost available for exploitation by interested users.

Hence, it would be very useful if this maintaining operation of ontologies will be done in an automatic or semi-automatic manner. This maintaining operation is sometimes called enrichment, sometimes it is called population as well as, but what is exactly the precise meaning of each one of this words? Ontology population is the process of inserting concept and relation instances into an existing ontology while ontology enrichment is the process of extending ontology, through the addition of new concepts, relations and rules [15]. As a main difference between the two processes is that ontology population preserve the ontology structure but ontology enrichment modifies it. Ontology learning is the process allowing the automatic generation of ontologies from a textual source called corpus. The ontology learning process is composed of several steps which are concept learning, taxonomic relation learning, non-taxonomic relation learning and finally axiom and rule learning.

We will interest, in the context of this paper, to the ontology enrichment process covering the three first steps of the ontology learning process, where we propose an approach for automatic ontology enrichment giving a text relating to a target domain. It is composed of three stages. In the first one, we use natural language processing techniques to extract sentences from text. Each extracted sentence is then annotated with part of speech tags and reduced to one or many binary relations (Subject, Verb, and Object) noted by SVO. The second stage consists of the determination of lexical relations (Hypernyms, hyponymy, synonymy,...) which may exist between the extracted terms (S, V and O) and the ontology concepts. For this purpose, we use an external knowledge source Wordnet. Finally in the third stage, the list of candidate's triplets (SVO) and lexical relations are used to enrich the initial ontology. To validate our work, we have chosen Phytotherapy as domain of discourse and the first results of precision, recall and f-measure metrics obtained are promising.

The remaining of this paper is organized as follows. Section 2 is devoted to the description of similar work, where we give recent and significant work in the field with their advantages and limitations. In Section 3, we

give detailed description of our ontology enrichment approach. Section 4 allows us to discuss the results obtained. Finally, we conclude our work and we give some perspectives in section 5.

2 Related work

Ontology is an explicit, formal specification of a shared conceptualization of a domain of interest [1]. New methods and tools are developed for reducing time and effort in the ontology construction process. The latter is called the ontology learning process. It is defined as the application of a set of methods and techniques in order to develop ontology from scratch or by enriching an

existing ontology using different types of data: unstructured, semi-structured, and fully structured. In our context, we are interested in unstructured data, we speak about textual information.

Ontology learning from text is the process of identifying terms, concepts, relations, and optionally, axioms from textual information and using them to construct and maintain ontology. Techniques from established fields, such as information retrieval, data mining, and natural language processing, have been fundamental in the development of ontology learning systems. The ontology learning process is detailed in [15] (see figure 1)

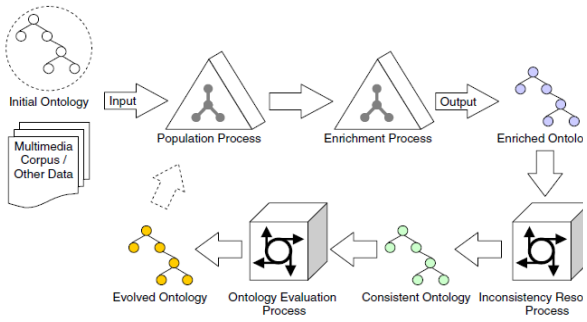


Figure 1: Ontology Learning Process [15].

According to ontology learning process, ontology enrichment is one of its important objectives. It consists of adding automatically new concepts and new relations to an initial ontology constructed manually using a basic knowledge relating to a given domain. Concepts and relations have to be placed in the relevant place in the initial ontology. However, numerous approaches and applications focus only on constructing taxonomic relationships (is-a-related concept hierarchies) rather than full-fledged formal ontologies[5]. For that, we are interesting, in our work, to develop an approach for the ontology enrichment taking in account both taxonomic and non-taxonomic relationships between concepts.

Generally, the process of enrichment attempts to facilitate text understanding and automatic processing of textual resources, moving from words to concepts and relationships. It can be divided into two main phases: the search for new concepts and relationships and the placement of these concepts and relationships within the ontology. According to [15], the process starts with the Concept Identification, then a taxonomy of concepts is constructed, the semantic relation extraction is the last step in enriching the initial ontology (see figure 2).

In [20], the process of enrichment is summarized within three main phases. The first is the **Extraction of representative terms in a specific domain.** It is the most important and difficult task. Several approaches (statistical and linguistic) are proposed for this aim. The second step concerns the **Identification of lexical relations between the terms.** Works in literature have focused on the identification of lexical relationships of hyperonymy, hyponymy, meronymy, synonymy and other more specific relationships that we call "transverse relations"[16],[22],[23]. The last phase aims to add the new terms as concepts/relations in relevant place in the ontology.

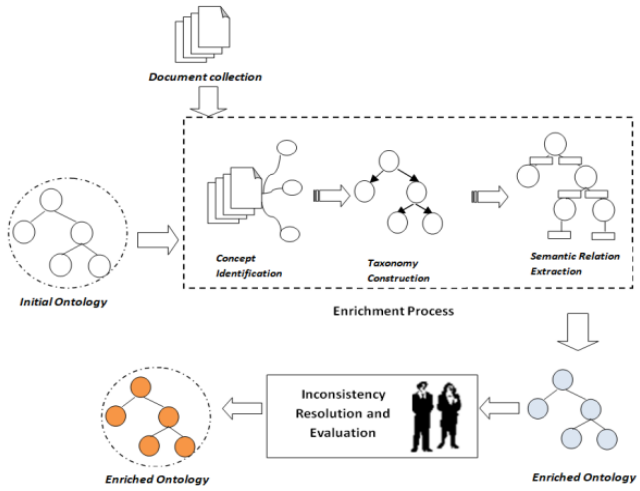


Figure 2: Ontology Enrichment Process [15].

In literature, different works of term extraction from textual corpus use two main approaches: **statistic analysis** and **linguistic analysis** approaches [17], [27], [28], [29]. The first one bases on statistic techniques of measures to facilitate the detection of new concepts and relations between them. Linguistic analysis uses linguistic techniques basing, generally, on detecting morphologic/ syntactic structures from the text in order to measure relativeness. Other works couple these two approaches and constitute an approach said « hybrid approach».

2.1 Statistical methods

They are often performed on large corpora. They are based on the co-occurrences, *TF*IDF*, *C/NC-value*, *T score*, *Dice Factor*, *Church Mutual Information*, and *word frequency* in the text in order to extract relevant terms to the target domain [2]. They base on the idea that is if two words coexist often in the same contexts, then they may be grouped together. This idea has been successfully realized in several works.

Drymonas employed C/NC-value to extract multi-word concepts [3]. Their proposed method “OntoGain”, aims to learn ontology from multi-word concept terms extracted from plain text. This method takes as input a corpus and produces a list of candidate multi-word terms, ordered by the likelihood of being valid terms, namely their C-Value measure [25]. NC-Value provides a method for the extraction of term context words (words that tend to appear with terms) and incorporates this information (from term context words) into the term extraction process [26]. OntoGain is applied on two separate data sources (a medical and computer corpus) the authors have evaluated 150 extracted terms with the help of domain experts. For computer science corpus, they obtained 86.67% precision and 89.6% recall, whereas for medical corpus, 89.7% precision and 91.4% recall were obtained.

Another example, of the statistical method, is the work of Mazari and his colleagues [4]. Their goal is to build ontology from a corpus of domain "Arabic linguistics". The process uses two statistical methods; the first is the “repeated segment”. It aims to identify the relevant terms that denote the concepts associated with the domain. The second is the “co-occurrence” method. It links these new extracted concepts to the ontology by hierarchical or non hierarchical relations. The first method performs an index of all words in the text by assigning a code corresponding to their positions in the corpus. Then it identifies all repeated segments in limiting itself to the same sentence. All of these segments are then filtered to remove unwanted segments and retain only those who are selected as candidate terms. The second method is based on the extraction of binary co-occurrents that meet one of the other more frequently than by chance and these two terms were included in the list found in the previous phase (detection of repeated segments). The co-occurrents will be selected with a frequency exceeding a statistically significant frequency due to chance. Then they will be compared with the labels of the ontology concepts. Terms may be added as new concepts, sub concepts or super concepts in the ontology and linked by Is-a or Part_of relation type. However, this approach is limited to Hyponymy and Meronymy relationships between concepts and the case, when both terms in the pair do not belong to the ontology labels, is not treated.

Therefore, these methods require human intervention for the positioning of the concepts in the ontology, or do not always identify the semantics of the relation, which influences on their accuracy.

2.2 Linguistic approaches

They use filtering techniques to manage text and to extract pieces of relevant information to the target domain. Works like those of Buitelaar and his colleagues [8] proposed a method mainly based on linguistics. It defines linguistic rules that extract concepts and relationships from collections of texts linguistically annotated. It is an approach that integrates linguistic analysis into ontological engineering. It supports the semi-automatic and interactive acquisition of ontologies from texts but also extension of existing ontologies. This methodology is associated with an OntoLT Protected plug-in [9] which uses predefined matching rules that automatically extract classes and candidate relationships from texts. For example, it maps the subject to a class, the predicate to a relation, the object's complement to a class, and creates the corresponding associative relationship between the two classes. If a rule is satisfied, the corresponding operators are enabled to create classes, relationships, or even instances that will later be validated and integrated into the ontology. The extracted ontology is integrated and can be explored in the Protégé development environment [9], which facilitates the management and sharing of the resulting ontologies. This approach has been used to build ontology in the field of neurology.

Other work in [6], aim at enriching an ontology from textual documents by relying on the linguistic analyzer "Insight Discoverer Extractor (IDE)". The analyzer outputs a tagged conceptual tree where each node carries a semantic tag attributed to the extracted textual unit based on the domain being processed. This approach presents a semi-automatic ontology population platform from textual documents. This platform provides an environment for matching linguistic extractions with the domain ontology of the client application using knowledge acquisition rules. These rules are applied, for each relevant linguistic label, to a concept, to one of its attributes or to a semantic relation between several concepts. They trigger the instantiation of these concepts, attributes, and relationships in the domain ontology knowledge base.

In [7], a linguistic method has been proposed in order to build domain ontology from Russian Text Resources. It uses a pipeline of linguistic methods (grafematic, morphological, syntactic and semantic analysis). *Grafematic analysis* is the initial analysis of the text on NL. It presents the input text data in a convenient format for further analysis (separation of input text into words, delimiters etc). *Morphological analysis* aims in construction of morphological interpretation of words of the input text (lemma, morphological part of speech…). *Syntactic analysis* is used for construction of syntactic tree from extracted syntactic groups consisting of sentences. *Semantic analysis* is used for building the semantic structure of one sentence. An algorithm of translating a syntactic tree into a semantic one applying a set of rules is proposed. As a result, the domain ontology can be built from the semantic trees extracted from text resources.

Linguistic approaches defining language rules (expressed as regular expressions) can identify specific terms associated with certain types of concepts in a domain.

The main limitations of rule-based approaches are that implementation requires a good knowledge of the field and requires manual work that is usually complex. In addition, rules are often defined for a specific domain or application and their application in other areas remains problematic.

2.3 Hybrid approaches

Combining linguistic information and statistical information is more commonly used to create term extraction modules. These hybrid systems use, first, linguistic filters to identify candidate terms, then statistical filters to distinguish terms from non-terms. In [10], an iterative method for semi-automatic acquisition of ontology and for enrichment of existing ontologies is proposed. It consists of a set of algorithms organized into modules aiming to extract concepts, relationships from texts. For the extraction of terms, a method based on statistical measures is applied to N-grams. A clustering method is then used to group these terms within concepts. The method proposes an algorithm for discovering Non-Taxonomic conceptual relations. It uses

shallow text processing methods to identify linguistically related pairs of words, which are mapped to concepts using the domain lexicon. The algorithm analyzes statistical information about the linguistic output. Thereby, it uses the background knowledge from the taxonomy in order to propose relations at the appropriate level of abstraction. In this method, the conceptualization is automatic; it allows generating ontology automatically; the latter can then be refined and enriched with the help of an expert (adding new relevant concepts, removing irrelevant concepts).

A methodology implemented in the OntoLearn tool [13] provides different techniques for extracting ontological knowledge from texts. For the extraction of relevant terms from a domain, linguistic and statistical tools are combined to determine their distribution in the corpus. It also uses glossaries available on the Web. Lexical-syntactic patterns described by regular expressions are used to discover the subsumption relations between concepts. The internal structure of multiword terms is also used to extract this type of relationship, as in [8]. Using the WordNet lexical database also makes it possible to extract synonyms and other types of relationships.

In [11], another approach is developed to support the semi-automatic enrichment of ontologies from unstructured texts. It combines NLP and machine learning methods to extract new ontological elements, such as concepts and relations, from text. The method starts by identifying important parts of text and assigning them a set of basic ontological concepts from a given ontology. Then, it extracts new ontological concepts from these revealed pieces of text. Further, it determines hierarchical dependencies between these concepts by assigning them taxonomic relations. Finally, it creates ontological instances for the given ontology. These instances will be represented by concrete occurrences of some ontological concepts in a text document and will be linked by non-taxonomic relations. This method achieves F-measure up to 71% for concepts extraction and up to 68% for relations extraction.

In [14], automatic process for ontology population, from a corpus of texts, is proposed. It is independent from the domain of discourse and aims to enrich the initial ontology with non-taxonomic relations and ontology class properties instances. This process is composed of three phases: identification of candidate instances, construction of a classifier and classification of the candidate instances in the ontology. The first phase applies natural language processing techniques to identify instances of non-taxonomic relationships and properties of an ontology by annotating the inputted corpus. The second phase applies information extraction techniques to build a classifier based on a set of linguistic rules from ontology and queries on a lexical database. This phase has a corpus and an ontology as inputs and outputs a classifier used in the "Classification of Instances" phase to associate the extracted instances with ontology classes. Using this classifier, an annotated corpus and the initial ontology, the third phase aims to the classification of these instances, produces a populated ontology. Implementation of this process applied to the legal domain shows results of 90% as precision 89.50% as Recall and 89.74% as F-measure. Authors conducted others experiments of their process on the touristic domain and obtained the results of 76.50% as precision 77.50% as Recall and 76.90% as F-measure.

In [30] a process of ontology extension is proposed for a selected domain of interest which is defined by keywords and a glossary of relevant terms with descriptions. The methodology is semiautomatic, aggregating the elements of text mining and user-dialog approaches for ontology extension. Authors aimed to the analysis of business news by the means of semantic technologies. The methodology is used for inserting the new financial knowledge into Cyc [31], which maintains one of the most extensive common-sense knowledge bases worldwide.

In [33], a framework for enriching textual data is developed. It is based on natural language information extraction to include more structure and semantics. Authors implemented the proposed framework in a system, named Enrycher, which offers a user-friendly way to qualitatively enhance text from unstructured documents to semi-structured graphs with additional annotations. Since the system offers a full text enrichment stack, it makes the system simpler to use than having the user to implement and configure several processing steps that are usually required in knowledge extraction tasks.

According to the presented approaches, hybrid ones are the most adopted in the domain ontology learning process from texts. These different methods can be chained one by one to lead to better results [32]. But, the main drawback is that the majority of the methods, presented in this state of the art, do not take into consideration an important and preliminary step which can save time and resources. We speak about the automatic simplification / reduction of texts to be processed [20]. Developing a method, that led to reduce texts complexity and upgrades both readability and understandability by removing that which may be less important from texts, could improve and facilitate the enrichment ontology process.

3 Proposed approach

An important task of ontology learning is to enrich the vocabulary for domain ontologies using different sources of information. We propose an approach for automatic ontology enrichment giving a text relating to a target domain. First, a basic knowledge related to this target domain is predefined and represented in an initial ontology through a set of concepts and relationships between those concepts. The objective is to enrich this ontology by the content of texts relating to the same target domain through semantic analysis. As seen in the precedent section, generally, the essential steps in enrichment process are: *Extraction of terms*, *Identification of lexical relationships between terms* and

placing the extracted terms as Concepts/Relations in the existing ontology(see figure 3)

3.1 Extraction of terms

One of our contributions, in this work, is the simplification of text in order to reduce its complexity. The majority of the proposed simplification methods rely on a set of manually defined transformation rules to be applied to sentences. In our approach, the proposed transformation rules are based on the segmentation of text into sentences and each sentence into tokens, each having its own POS (Part Of Speech). Then, based on these POS, we simplify, reduce and transform each sentence into a triplet SVO: (Subject, Verb, Object) supposed to carry the information of the sentences from which they are extracted. For this purpose, we use NLP techniques [18], [19]. The first step is divided into two sub-steps, we start first with parsing the text, and then we extract the significant terms.

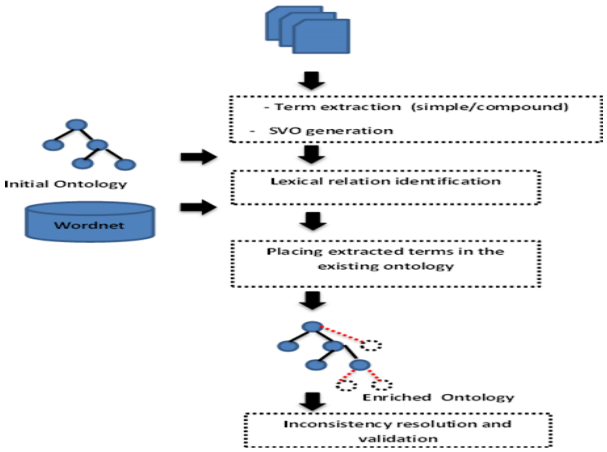


Figure 3: The proposed enrichment Process.

3.1.1 Syntactic analysis:

Syntactic Analysis or what we call preprocessing of texts. We aim, in this phase, to detect the type of words (verb, noun and adjective etc.), by *segmenting the text* into sentences. For each sentence, we extract its tokens having its own POS (Part Of Speech). These tokens may be simple or compound. In this last case, to make easy the detection of compound terms, we have proposed set of rules using English grammar [24] to define all possible compound terms (see the table bellow Table 1).

After term extraction, to simplify the sentences, stop-words will be removed from sentences. The stop words can be defined as words that don't have any remarkable importance. For example, *of, also, here, more, so, very,now.....*

The types of compound words	Description
NN + NNS	noun, common, singular or mass + noun, common, plural
NN +NN	noun, common, singular or mass + noun, common, singular or mass
NNS + NNS	noun, common, plural + noun, common, plural
NNP + NNP	noun, proper, singular+ noun, proper, singular
NN + NNP	noun, common, singular or mass + noun, proper, singular
JJ + NN	adjective or numeral, ordinal + noun, common, singular or mass
JJ + NNS	adjective or numeral, ordinal + noun, common, plural
NNP + NN	noun, proper, singular+ noun, common, singular or mass
NNP + NNS	noun, proper, singular+ noun, common, plural
NN + NN + NN	noun, common, singular or mass+ noun, common, singular or mass+ noun, common, singular or mass
NNPS + NNS	noun, proper, plural + noun, common, plural
NNPS + NN	noun, proper, plural +noun, common, singular or mass+ noun, common, singular or mass
VB + RB	verb, base form + Adverb
VB + RP	verb base form + Particle (up, off……etc)
MD + VB	modal auxiliary (could, should) + verb, base form

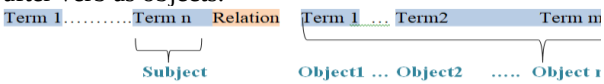
Table 1: Set of rules defining compound words.

3.1.2 Extraction and generation of SVO:

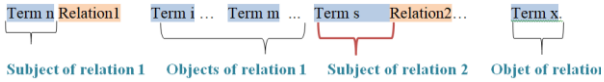
This step consists of simplifying, reducing and transforming each sentence of the text to a set of representative terms in the form of a triplet SVO: Subject, Verb, and Object. Here, we have to analyze all sentences obtained by text segmentation to a set of sentences. First, each sentence is annotated with POS tags and then the three parts of each sentence are delimitated: The subject part, the verbal part and the object part.

We have based, essentially, on the position of each term T (simple or compound) in the sentence S. To extract the relation in S, we test if the grammatical category of T is VB or VB + RB or VB + RP or MD + VB or VB+ADj (ADj: adjective situated directly after the verb), then T is the verbal part of the triplet. For example, in the sentence « The seed is rich in essential amino acidsand is used as cattle or poultry feed.” System detects two verbs *is_rich* and *use*. To extract subject and object parts, we distinguish the following cases:

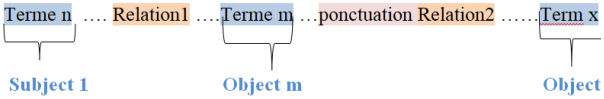
- If the sentence contains one verb, we select the nearest term before the verb as subject, and all terms after verb as objects.



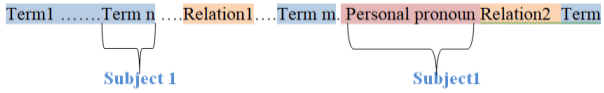
- in complex sentence containing more than one verb, for example according the next example, the subject of the second verb is term s



- For example, in « *Fresh Allspice berries when crushed can be mixed with a few drops of oil and massage onto the affected area to alleviate pain associated with rheumatism and arthritis.* » two subjects (*Allspice berries and pain)* for two verbs *(mixed* and *associated)*.
- If the sentence contains more than one verb and no term is before the verb 2, the subject of this latter is the same of the precedent verb.



- In the case where the personal pronoun is the subject of the sentence. We replace this pronoun by the subject of the precedent sentence.



- For example in “Asparagus is a climbing undershrub with widespread applications. It is useful in nervous disorders, dyspepsia, venereal diseases.” The pronoun “it” is replaced by Asparagus.

Some kinds of sentences are not treated in the present work. For example, in the case of incomplete sentences those do not contain an object or subject. Also, in the case of negative sentences which are in negative form. At the end of this phase, we have a list of candidate’s triplets (SVO) for enrichment.

3.2 Identification of lexical relationships between terms

In this step, we determine the relations which may exist between the extracted terms and the ontology concepts. For this purpose, we use an external knowledge source Wordnet [12]. Terms in WordNet are organized into synonym sets, called synsets, representing a concept by a set of words with similar meanings. Hypernyms, or the IS-A relation, is the main relation type in WordNet. Other types of relations are hyponymy, meronymy, synonymy, equivalence.

Several methods and applications focus on constructing taxonomic relationships rather than full-fledged formal ontologies. For that, our second contribution, in this work, is to develop an approach for the ontology enrichment taking in account taxonomic and non-taxonomic relationships between concepts. The achievement of this step depends on each candidate triplet SVO generated in the previous phase, and the set of concepts in the initial ontology. For each triplet and for each term T of this latter, we identify sets; each one is composed of words Wordnet having a lexical relation with the term T (hypernymy, hyponymy, synonymy). Subsequently, for each concept in the input ontology, we detect the lexical relation between this last and the term T. At the end, we have the types of lexical relations between the terms S, V, and O, and the concepts of ontology. How to place these terms in the ontology? This will be the subject of the next step.

3.3 Placing extracted terms in the initial ontology.

For each of the terms identified in step 1, we first check if it does not appear as a concept in the original ontology. In this case, our algorithm verifies possible approximations of meaning with the concepts of the ontology. The proposed enrichment process is illustrated in the algorithm below. It aims to add new concepts/relationships in the initial ontology. This must take into account the semantic links between concepts such as hyperonymy and hyponymy. The WordNet ontology is used for this purpose. For each triplet SVO, if the extracted term T exists in the initial ontology (IO), then no modification will be realized else, the following cases are distinguished:

Case 1: if the term T is a Subject or Object in the SVO, **then *if*** T is similar to an instance in the initial ontology (IO), so adding T as instance, ***else***, the following possibilities are distinguished (see figure 4),

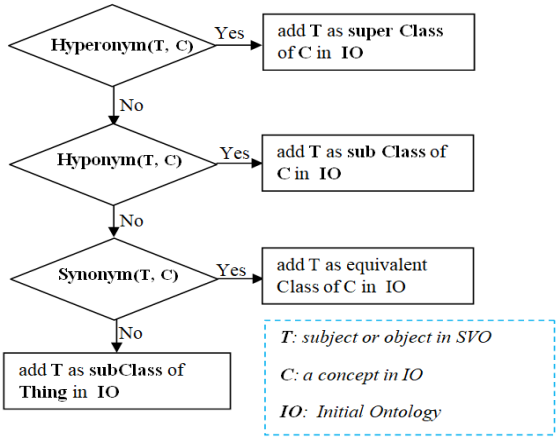


Figure 4: T is Subject or Object in IO.

Case 2: if T appears as a verb in the selected SVO, then as shown in figure 5, the following cases are distinguished.

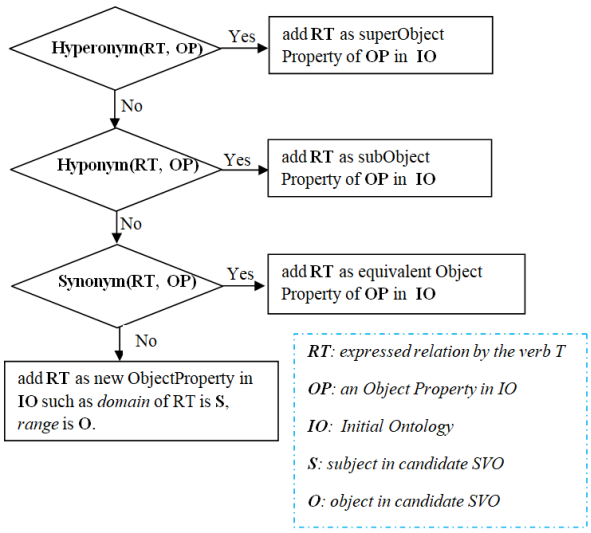


Figure 5: T is a verb in SVO.

Case 3: if the term T is a verb + Adjective, then the following alternatives are distinguished as shown in figure 6)

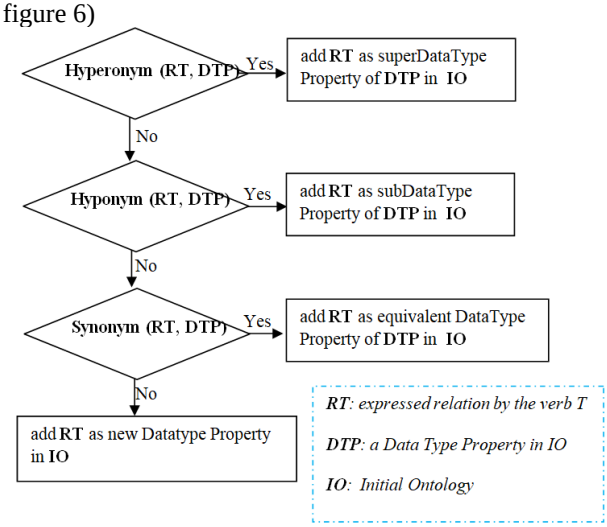


Figure 6: T is verb+adjective.

Indeed, enrichment here consists of adding concepts, instances, axioms and relationships. The following algorithm describes in detail the enrichment process.

Algorithm Enrichment Process

Input: 1- SVO : list of triplet SVO
2- IO: Initial ontology

BEGIN
Semantic_Relation ←Ø ;
For each triplet in SVO **DO**
For each term T in triplet **DO**
LH←Hyperonyms list of T;
LHy←Hyponyms list of T;
LSy←Synonyms list of T;
For each Entity E in IO **DO**
IF T exists in IO **THEN**
No modification
ELSE IF T (Subject or Object) AND (E is a Class)
THEN IF (E ∈ LH) **THEN**
Semantic_Relation← Hyperonym (E,T) ;
IO←IO∪ (T as a sub_ Class of E) ;
ELSE IF (E ∈LHy) **THEN**
Semantic__Relation← Hyponym (E,T) ;
IO←IO∪ (T as a Super_Class of E) ;
ELSE IF (E ∈LSy) **THEN**
Semantic__Relation← Synonym (E,T) ;
IO←IO∪ (T as Equivalent_ Class of E) ;
ELSE
IO←IO∪ (Concept as Class) ;
ENDIF
ENDIF
ENDIF
ENDIF
IF T(Subject or Object) AND(E is Instance)
THEN IF (E ∈LSy) **THEN**
IO←IO∪ (T as Instance) ;
ENDIF
ENDIF
IF T (Verb) AND (E is Object_ property) **THEN**
IF (E ∈ LH) **THEN**
IO←IO∪ (T as Sub_Object_property) ;
ELSE IF (E ∈LHy) **THEN**
IO←IO∪ (T as Super_Object_property) ;
ELSE IF (E ∈LSy) **THEN**
IO←IO∪(T as Equivalant_Object_property);
ELSE /* non-taxonomic relation*/
IO←IO∪ (Relation asObject_property) ;
ENDIF
ENDIF
ENDIF
ENDIF
IF T (Verb + Adj.) AND (E is Data_Type_property) **THEN**
IF (E ∈ LH) **THEN**
IO ←IO∪(T as Sub_Data_Type_Property) ;
ELSE IF (E ∈LHy) **THEN**
IO ←IO∪(T as Super_Data_Type_Property) ;
ELSE IF (E ∈LSy) **THEN**
IO ←IO∪(T as Equivalent_Data_Type_Property) ;
ELSE
/* non-taxonomic relation*/
IO←IO∪(Relation as Data_Type_property) ;
ENDIF ENDIF ENDIF ENDIF
ENDFOR ENDFOR ENDFOR
END
OUTPUT :enriched ontology ;

4 Experiments and results

In this paper, we attempt to evaluate the performance of our proposed automatic ontology enrichment approach. We use **Phytotherapy** which consists of the use of plant derived medications in the treatment and prevention of disease. The World Health Organization (WHO) encourages the integration of the Phytotherapy in the health system [21]. However, the informal nature of its

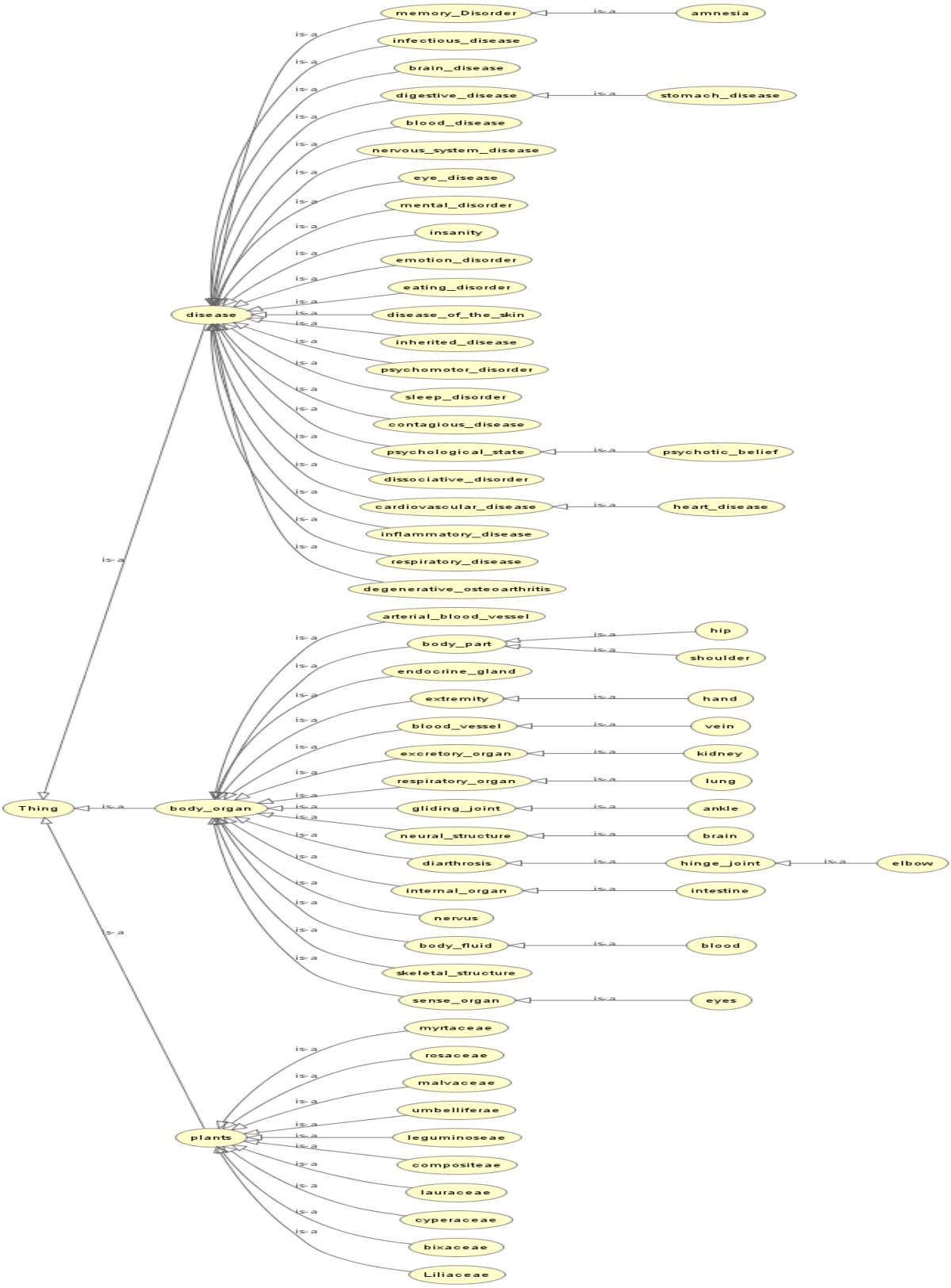


Figure 7: Initial Ontology of Phytotherapy Domain.

content makes difficult its use and practice. Our objective is not only formalizing the content of this medicine by means of ontology but also managing this latter automatically enriched from domain texts. This allows the end-users to be permanently informed about medicinal plants and their natural remedies against different diseases.

The initial ontology developed for the domain Phytotherapy describes some diseases; each disease belongs to a particular organ of the human body. It also

lists the different plants that can cure diseases. For this purpose, we have defined three main classes in the OWL ontology: **Plant**, **Disease** and **Human_Organ.** (see figure 7).

To begin the enrichment process, we have used 25 texts, including 17 075 words speaking about three plants: Ginger, Aloe and Strophanthus. After applying the enrichment process, we obtain the following results (see table 2)

25 texts		
Expert SVO	2475	
Extracted SVO by system	1875 are true	375 are false

Table 2: Extracted SVO by Expert/System.

For example, in the segment of text : "Ginger also shows promise for fighting cancer, diabetes, non-alcoholic fatty liver disease, asthma, bacterial and fungal infections, and it is one of the best natural remedies available for motion sickness or nausea." The generated SVO, are the following:

1. ginger_fighting_cancer
2. ginger_fighting_diabetes
3. ginger_fighting_non-alcoholic fatty liver
4. ginger_fighting_disease
5. ginger_fighting_asthma
6. ginger_fighting_fungal infections

The system takes these SVO one by one and enriches the initial ontology as following:

- *ginger* is added as an individual (instance) of the concept *Umbelliferae* (appearing as a class in the original ontology),
- *non-alcoholic fatty liver*, *fungal infections* are added as subclasses of *disease* Class,
- *asthma* is an existing individual (instance) of *respiratory disease*.
- The verb *fighting* is added as a relation between *Umbelliferae* concept and (*disease*, *fungal infections*, *asthma* and *non alcoholoic fatty liver*, *cancer* and *diabetes*) concepts, see the following figures (figure 8, figure 9 and figure 10).

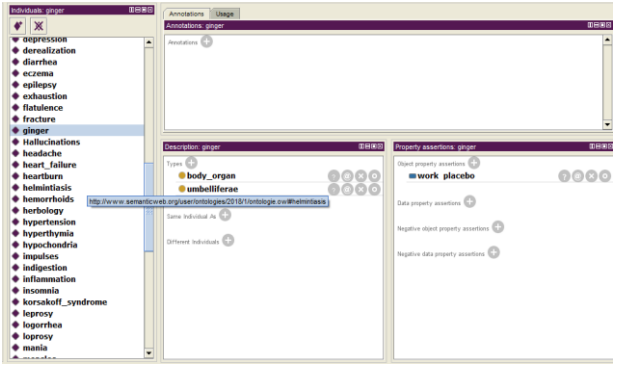


Figure 8: Creation of instance (ginger) from SVO to initial ontology (Protégé Window).

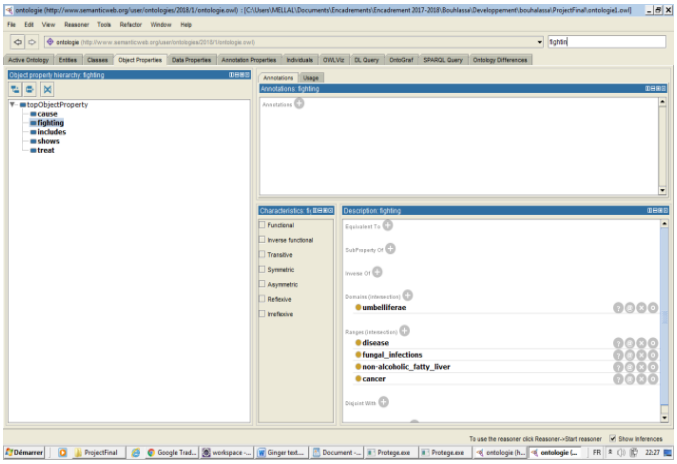


Figure 9: Placing Concepts/ relation from SVO to initial ontology (Protégé Window).

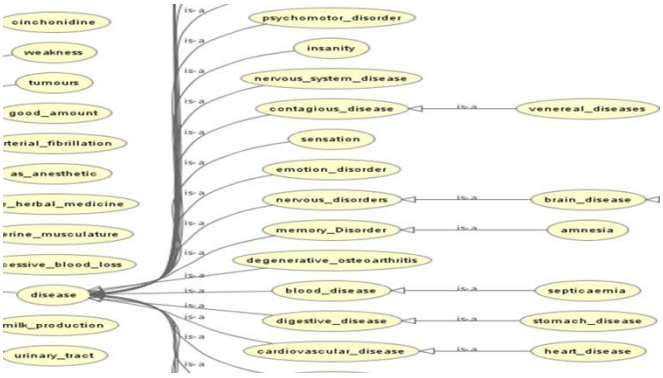


Figure 10: Part of enriched Ontology.

The consistency test and validation of the enriched ontology is done using Fact++ tool basing on the class/properties description. Here, the context of our work is only limited to first order relations.

To evaluate the performance of the proposed enrichment process, we use the precision, recall and F1 measure as follow:

Precision = (Number of generated SVO placed in the correct place in ontology by system/ Number of the generated SVO by system)

Recall = (Number of generated SVO placed in the correct place in ontology by system / Number of correct generated SVO by expert)

F-mesure = 2*Precision*Recall /Precision + Recall

Among the test set of 25 texts, extracted SVO by human experts agreed upon 2475 SVO. But after testing, the system gives 2250 SVO. In these 2250 SVO, 1875 SVO are correct and their terms are inserted in relevant places in the initial ontology. The implementation of the process shows results 83% as precision 75% as Recall and 78% as F-measure.

We have remarked that the proposed approach performs better with texts more than others. This is due to the type of sentences composing these texts. In fact, system gives best results in the case of verbal sentences containing a verb as a main part.

5 Conclusion and future work

In this paper, we have proposed an approach for automatic enrichment of a basic ontology composed of three stages. The first stage consists of applying natural language processing techniques to obtain tagged sentences. In the second stage, we reduce each sentence to a verbal one, called SVO (Subject, Verb, Object) sentence. Finally, in the third stage, we proceed to enrich an initial ontology built manually by adding new concepts, new relations and/or instances of concepts. We have distinguished three different approaches for automatic ontology enrichment: statistic based approach, natural language processing based approach and hybrid approach, which combines the two first approaches. The common problem of these approaches is that they don't reduce compound and complex sentences to their simplified forms before ontology enriching operation, which affect negatively their performance. Our approach is based on natural language processing techniques but augmented by a heuristic algorithm allows reducing extracted sentences to SVO (Subject, Verb, and Object) simple ones. This reducing step is very important because it allows improving the enrichment process performance. Another advantage of our approach is that it takes into account all types of relations, taxonomic and non-taxonomic, which allows us to have a good ontology enrichment rate.

To implement our approach, we have used a set of technologies proposed by the Semantic Web community (OWL, OWL-API, Wordnet ...) and the domain of natural language processing (Stanford Core NLP...). We have used Phytotherapy as domain of expertise since it is very important for pharmaceutical industry and as huge quantity of texts speaking about exits on the WWW. The first results obtained of precision, recall and f- measure are very encouraging (83% of precision, 75% of recall and almost 78 % of F-measure).

For this aim, some guidelines are to be taken into account. *First*, a survey of text segmentation and tagging algorithms must be done in the aim to use the most efficient ones. *Second*, treat the remaining cases of composed sentences and write the process of reducing texts in SVO in the form of an algorithm and try to optimize it. The *third* and the last guideline concerns the step of identifying SVO relationships with those of existing ontology and the placement of new concepts in it, this plays its preponderant role in the performance of the entire system, which is why a study and comparison of different ontology reasoners is imperative in order to use the most efficient one.

As future work, we plan first to enhance the performance of our approach by evaluating and improving the proposed algorithm. Also, we plan to extend our process using textual corpus to ensure that texts are in the domain we are interested in. Another future work consists of defining other new metrics like for example enrichment rate and enrichment efficiency metrics to measure the utility of our approach.

6 References

[1] T. R. Gruber . A Translation Approach to Portable Ontologies.Knowledge Acquisition, 5(2):199220, 1993 https://doi.org/10.1006/knac.1993.1008

[2] V.T. Nguyen. Méthode d'extraction d'informations géographiques à des fins d'enrichissement d'une ontologie de domaine. Doctoral Dissertation, Pau University (France), 2012.

[3] Drymonas,E., Zervanou,K. and Petrakis,E.G. Unsupervised ontology acquisition from plain texts: the Ontogain system. In: *NLDB*. Springer, Cardiff, United Kingdom. 2010. https://doi.org/10.1007/978-3-642-13881-2_29

[4] A.C. Mazari , H. Aliane and Z. Alimazighi. Automatic construction of ontology from Arabic texts. ICWIT, pp. 193-202. 2012.

[5] N. Astrakhantsev, D. Fedorenko, D. Turdakov. Automatic Enrichment of Informal Ontology by Analyzing Domain-Specific Texts Collection". Materials of International Conference "Dialog", vol. 13, no. 20, pp. 29–42. 2014.

[6] F. Amardeilh, P. Laublet, J. L. Minel, "Document Annotation and Ontology Population from Lingusitic Extraction", Proceedings of Third International Conference on Knowledge Capture. 2005 .https://doi.org/10.1145/1088622.1088651

[7] Yarushkina, N.; Filippov, A.; Moshkin, V.; Egorov, Y. Building a Domain Ontology in the Process of Linguistic Analysis of Text Resources. *Preprints* 2018, 2018. https://doi.org/10.20944/preprints201802.0001.v1

[8] P. Buitelaar, D.Olejnik, M. Sintek. A Protege Plug-in for Ontology Extraction from Text Based on Linguistic Analysis. In Proceedings of the 1st European SemanticWeb Symposium(ESWS). 2004. https://doi.org/10.1007/978-3-540-25956-5_3

[9] H. Knublauch, R.Fergerson,.NF, Noy, M.A Musen. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. McIlraith, S.A., Plexousakis, D., Harmelen, F. van (Eds.), The Semantic Web – ISWC, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 229–243. 2004. https://doi.org/10.1007/978-3-540-30475-3_17

[10] A. Maedche, S. Staab, N.Stojanovic, Y. Sure,R.Studer. Semantic portAL - The SEAL approach. In: Spinning the Semantic Web. MIT Press, pp. 317–359. 2001. https://doi.org/10.1007/3-540-45754-2_1

[11] Ivana Lukšová. Ontology Enrichment Based on Unstructured Text Data. Master Thesis, Prague. 2013

[12] C. Fellbaum. WordNet: An Electronic Lexical Database. MIT Press, 1998.

[13] P. Velardi, P. Fabriani, M.Missikoff. Using Text Processing Techniques toAutomatically Enrich a Domain Ontology. In Proceedings of the InternationalConference on Formal Ontology in Information Systems, FOIS '01.ACM, New York,

NY, USA, pp. 270–284. 2001. https://doi.org/10.1145/505168.505194

[14] C. Faria, I. Serra, and R. Girardi, "A domain-independent process for automatic ontology population from text, Elsevier , Journal of Science of Computer Programming vol.95 pp 26–43. 2014. https://doi.org/10.1016/j.scico.2013.12.005

[15] G. Petasis, V. Karkaletsis, G. Paliouras, A. Krithara and E. Zavitsanos, Ontology Population and Enrichment: State of the Art, Berlin/ Heidelberg, pp 134-166, Springer. 2011 https://doi.org/10.1007/978-3-642-20795-2_6

[16] Z. Sellami. Gestion dynamique d'ontologies à partir de textes par systèmes multi agents adaptatifs. Thesis Paul Sabatier University. 2012.

[17] A. Gomez Pérez, D. Manzano Macho. An overview of methods and tools for ontology learning from texts. The Knowledge Engineering Review, Vol. 19:3, 187–212, Cambridge University Press. 2005. https://doi.org/10.1017/S0269888905000251.

[18] D.Jurafsky, J.H.Martin. The Representation of Sentence Meaning. https://web.stanford.edu/~jurafsky/. 2018

[19] D. Jurafsky, J.H. Martin. Computing with Word Senses, https://web.stanford.edu/~jurafsky/. 2018

[20] M.Shardlow. A Survey of Automated Text Simplification. (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Natural Language Processing. 2014. https://doi.org/10.14569/specialissue.2014.040109

[21] M. Smith, A. Burton, T. Falkenberg. World Health Organization : Traditional Medicine Strategy 2014-2023". 2014.

[22] N. Sheena, M.J Smitha, J. Shelbi. Automatic Extraction of Hypernym & Meronym Relations in English Sentences Using Dependency Parser. 6th International Conference On Advances In Computing & Communications. ICACC 2016, Cochin, India https://doi.org/10.1016/j.procs.2016.07.269

[23] M. Khodak, A. Risteski, C. Fellbaum, S. Arora. Automated WordNet Construction Using Word Embeddings. Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications. pages 12–23, Valencia, Spain. 2017 https://doi.org/10.18653/v1/w17-1902

[24] https://www.ef.com/ca/english-resources/english-grammar

[25] N. Hernandez. Ontologies de domaine pour la modélisation du contexte en recherche d'information. Thèse de Doctorat à l'Université Paul Sabatier France. 2006

[26] F. Rousselot et P. Frath. Terminologie et Intelligence Artificielle. 12èmes rencontres linguistiques, Presses Universitaires de Caen. 2002.

[27] A. Imsombut and J. Kajornrit. Comparing Statistical and Data Mining Techniques for Enrichment Ontology with Instances. Journal of Reviews on Global Economics, 6, 375-379. 2017 https://doi.org/10.6000/1929-7092.2017.06.39

[28] M.N. Asim, M. Wasim, M.U.G Khan, W. Mahmoud and H. Abbasi. A survey of ontology learning techniques and applications. Database, 1–24. 2018. https://doi.org/10.1093/database/bay101

[29] W. Wong. Ontology Learning from Text: A Look Back and into the Future. Article *in* ACM Computing Surveys. Volume 44 issue 4 pp 1- 36. 2012. https://doi.org/10.1145/2333112.2333115

[30] Novalija Inna, Mladenić Dunja. Ontology Extension Towards Analysis of Business News. Informatica Journal. Volume 34, N°4, pp 517—522. 2010.

[31] Cycorp, Inc., http://www.cyc.com

[32] Dunja Mladenić, Marko Grobelnik. Automatic Text Analysis by Artificial Intelligence. Informatica Journal, volume 37, N°1, pp 27—33. 2013.

[33] Tadej Štajner, Delia Rusu, Lorand Dali, Blaž Fortuna, Dunja Mladenić and Marko Grobelnik (2010), « A Service Oriented Framework for Natural Language Text Enrichment", Informatica Journal, Volume 34, N°3, pp 307–313. 2010.

Penalty Variable Neighborhood Search for the Bounded Single-Depot Multiple Traveling Repairmen Problem

Ha-Bang Ban
School of Information and Communication Technology
Hanoi University of Science and Technology, Hanoi, Vietnam
E-mail: BangBH@soict.hust.edu.vn

Keywords: nounded-mTRP, penalty variable neighborhood search, metaheuristic

Received: June 4, 2019

Multiple Traveling Repairmen Problem (mTRP) is a class of NP-hard combinatorial optimization problems with many practical applications. In this paper, a general variant of mTRP, also known as the Bounded Single-Depot Multiple Traveling Repairmen Problem (Bounded-mTRP), is introduced. In the Bounded-mTRP problem, a fleet of identical vehicles is dispatched to serve a set of customers. Each vehicle that starts from the depot is only allowed to visit the number of customers within a predetermined interval, and each customer must be visited exactly once. Such restrictions appear in real-life applications where the purpose is to have a good balance of workloads for the repairmen. The goal is to find the order of customer visits that minimizes the sum of waiting times. In our work, the proposed algorithm is encouraged by the efficiency of the algorithms in [15, 19, 20] that are mainly based on the principles of the VNS [14]. The penalty VNS extends the well-known VNS [14] by including constraint penalization, to solve the Bounded-mTRP effectively. Extensive numerical experiments on benchmark instances show that our algorithm reaches the optimal solutions for the problem with 76 vertices at a reasonable amount of time. Moreover, the new best-known solutions are found in comparison with the state-of-the-art metaheuristic algorithms.

Povzetek: Razvita je nova metoda za preiskovanje grafov - za reševanje naloge serviserja, tipičnega NP-polnega problema.

1 Introduction

1.1 Motivation and definition

The Traveling Repairman Problem (TRP) has been studied in the number of previous work [1, 2, 5, 15, 19, 20]. It is known as the Minimum Latency Problem (MLP), or the Deliveryman Problem (DMP). These problems arise applications, e.g., whenever repairmen or servers have to accommodate a set of requests to minimize their total (or average) waiting times [1, 2, 5, 15, 19, 20]. A direct generalization of the TRP is the Multiple Traveling Repairmen Problem (mTRP) that considers k vehicles simultaneously. Applications of the mTRP can be found in Routing Pizza Deliverymen, or Scheduling Machines to minimize mean flow time for jobs. Several prior studies that we can find in the literature are [10, 12]. In this paper, we study the Bounded Single-Depot Multiple Traveling Repairmen Problem (Bounded-mTRP) by involving the restriction of the number of vertices that a repairman must visit in his tour. The restriction is defined by lower (denoted by K) and upper (denoted by L) bounds regarding the traveled vertices. Therefore, the number of vertices that a repairman can visit lies within a predetermined interval with the aim of obtaining balanced solutions. The requirement of the problem is to find a tour such that the above restriction is satisfied, and the overall cost of visiting all vertices is minimized. Such restriction appears in many real-life applications whose purpose is to have a good balance of workloads for the repairmen.

1.2 Approach and contributions

There are three approaches for solving the Bounded-mTRP: 1) exact algorithms, 2) approximation algorithms, and 3) heuristic algorithms. The exact algorithms find the optimal solution with an exponential time in the worst case. Therefore, the exact algorithm only solves the problem with small sizes. To describe related works, we denote an approximation algorithm as p-approximation when the algorithm finds the solution at most p times worse than the optimal solution. Here p is an approximation ratio with a constant value. In this approach, the best approximation ratio of 16.994 is for the mTRP [10, 12]; however, it is still far from the optimal solution. Heuristic algorithms perform well in practice, and their efficiency can be evaluated through experiments. Our algorithm falls into this approach.

Previously, research on the Bounded-mTRP has not studied much, and this work presents the first metaheuristic approach for this problem. Our algorithm is encouraged by the efficiency of the algorithms in [14, 19, 20] that are mainly based on the principles of the VNS [14]. However,

the difference between the Penalty VNS (P-VNS) and their VNS is that our algorithm builds up penalty value during a search. The proposed algorithm includes two phases. The algorithm is developed based on the GRASP [9] to build an initial solution in the construction phase. In the improvement phase, the P-VNS combined with shaking techniques not only exploits good local solution space but also prevents the search from escaping from local optimal. Moreover, several novel neighborhoods' structure as well as a constant time operation for calculating the cost of each neighboring solution is also introduced. The main problem is that there exists no other metaheuristic reported in the literature for this problem; this is, we found no previous attempts to solve this problem, neither exact nor heuristically, to compare with. Therefore, we adapt the metaheuristic algorithms in [17] to solve the Bounded-mTRP, and choose several state-of-the-art metaheuristic algorithms for the mTRP [8, 18], and Bounded-mTSP [17] as a baseline in our research. Extensive numerical experiments on benchmark instances show that our algorithm reaches the optimal solutions for the problems with up to 76 vertices at a reasonable amount of time. Moreover, the new best-known solutions are found in comparison with the state-of-the-art metaheuristic algorithms.

The rest of this paper is organized as follows. Section 2, and 3 present literature review, and neighborhood structure, respectively. The proposed algorithm is described in Section 4. Computational evaluations and discussions are reported in Section 5. Finally, Section 7 concludes the paper.

2 Literature review

The Bounded-mTRP has, as we know, not been studied much, although it is a natural extension of the mTRP problem. In the literature, several variants of the problem are introduced as follows:

- The mTRP is a popular case since no constraint is considered. Numerous works for the mTRP can be found in [8, 10, 12, 18]. Some metaheuristic algorithms [10, 12] can give good solutions fast for large instances.
- The mTRP with Profits (mTRPP) finds a travel plan for server that maximizes the total revenue. Metaheuristic algorithm [3] produces solutions well.
- Another variant of the mMLP is mMLP with distance constraints [4, 16]. Lou et al. [16] proposed an exact algorithm that reaches the optimal solutions for the instances with up to 50 vertices. Ban et al. [4] then presented a metaheuristic algorithm based on VNS. The experimental results concluded that the algorithm found good-quality solutions for small and medium-size instances.
- mTRPD [6] finds a tour with minimum latency sum in post-disaster road clearance. Unlike mMLP, in disaster situations, travel costs need to be added to debris removal times. Their metaheuristic obtained the optimal or near-optimal solutions on Istanbul data within seconds.
- The TRP is a particular case where there is only a repairman. Numerous metaheuristic algorithms [1, 2, 5, 15, 19, 20] for the problem have proposed in the literature. The experimental results showed that their algorithms obtain good solutions fast for the instances with up to 500 vertices.

These algorithms are the best algorithms for some variants of the Bounded-mTRP problem. However, they do not involve the bounded constraint. Therefore, they cannot be used directly to solve the Bounded-mTRP.

3 Mathematical formulation

The formulation is obtained from the formulation proposed by Christofides et al. [7] for the Capacitated Vehicle Routing Problem (CVRP). Let u_i be a non-negative real variable representing the length of the route from the depot 0 to the vertex *i*. Let x_{ij}^r be the following binary variables:

$$x_{ij}^r = \begin{cases} 1 & \text{if if edge}(i,j) \text{ is used in route } r \\ 0 & \text{otherwise} \end{cases} \quad \min z = \sum_{i=1}^{n} u_i$$

Subject to:

$$\sum_{\substack{i=0 \\ i\neq j}}^{n} \sum_{r=1}^{m} x_{ij}^r = 1; \qquad (j = 1, 2, ..., n) \tag{1}$$

$$\sum_{\substack{i=0 \\ i\neq l}}^{n} x_{il}^r - \sum_{\substack{j=0 \\ j\neq l}}^{n} x_{lj}^r = 0; \qquad (l = 0, 1, ..., n; \; r = 1, 2, ..., m) \tag{2}$$

$$\sum_{j=1}^{n} x_{0j}^r = 1; \qquad (r = 1, 2, ..., m) \tag{3}$$

$$K \leq \sum_{i=0}^{n} \sum_{j=0}^{n} x_{ij}^r + 1 \leq L; \qquad (r = 1, 2, ..., m) \tag{4}$$

$$u_i - u_j + (T + c_{ij}) \times \sum_{r=1}^{m} x_{ij}^r + (T - c_{ji}) \times \sum_{r=1}^{m} x_{ji}^r \leq T; \quad (i, j = 1, ..., n; i \neq j) \tag{5}$$

$$u_i \geq c_{0i} \times \sum_{r=1}^{m} x_{0i}^r; \qquad (i = 1, 2, ..., n) \tag{6}$$

$$x_{ij}^r \in \{0, 1\}; \qquad (i, j = 0, 1, ..., n; j \neq i; \quad r = 1, ..., m) \tag{7}$$

$$u_i \geq 0; \qquad (i = 1, ..., n) \tag{8}$$

Constraints (1) show that each vertex is contained in only one route. Constraints (2) indicate that when a vertex is in a route, then it has a predecessor and a successor in that route. Constraints (3) ensure that one vertex is sequenced as first in each route and constraints (4) guarantee that and the number of vertices visited of each repairman must less than *L* and more than *K*. Constraints (5) demonstrate that $u_j = u_i + c_{ij}$ when $\sum_{r=1}^{m} x_{ji}^r = 1$ and they are redundant when $\sum_{r=1}^{m} x_{ji}^r = 0$. Constraints (6) initialize the latency of the first vertex in each route. Finally, constraints (7) and (8) establish the nature of the variables.

4 Neighborhood structure

Seven neighborhoods investigated are divided into two categories: intro-route and intra-route. Now, let $T = (R_1, R_2, ..., R_l, ..., R_k)$ $(l = 1, ..., k)$ be a tour, we introduce a novel neighborhoods' structure and complexity of their exploration. Note that: in [15, 20], the complexity of some neighborhoods is already mentioned. Therefore, we only introduce the complexity of new ones.

For Intro-route: Intro-route is used to optimize on a single route. Assume that, R and m $(m < n)$ are a route and its length, respectively. We then introduce five neighborhoods' structure in turn.

Remove-insert neighborhood considers each vertex v_i in the route at the end of it. This neighborhood of R is defined as a set $N_1(R) = \{R_i = (v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_m, v_i) : i = 2, 3, ..., m-1\}$. Obviously, the size of $N_1(R)$ is $O(m)$.

Property 1. The time complexity of exploring $N_1(R)$ is $O(m^2)$.

Swap adjacent neighborhood attempts to swap each pair of adjacent vertices in the route. This neighborhood of R is defined as a set $N_2(R) = \{R_i = (v_1, v_2, ..., v_{i-2}, v_i, v_{i-1}, v_{i+1}, ..., v_m) : i = 3, 4, ..., m-1\}$. The size of the neighborhood is $O(m)$.

Property 2. The time complexity of exploring $N_2(R)$ is $O(m)$.

Swap neighborhood attempts to swap the positions of each pair of vertices in the route. This neighborhood of R is defined as a set $N_3(R) = \{R_{ij} = (v_1, v_2, ..., v_{i-1}, v_j, v_{i+1}, ..., v_{j-1}, v_i, v_{j+1}, ..., v_m) : i = 2, 3, ..., m-3; j = i+3, ..., m\}$. The size of the neighborhood is $O(m^2)$.

Property 3. The complexity of exploring $N_3(R)$ is $O(m^2)$.

3-opt neighborhood attempts to reallocate three adjacent vertices to another position of the route. This neighborhood of R is defined as a set $N_4(R) = \{R_i = (v_1, v_2, ..., v_{i-1}, v_i, v_{j+1}, ..., v_k, v_{i+1}, ..., v_j, v_{k+1}...., v_m) : i = 2, 3, ..., m-5, j = 4, ..., m-3, k = 6, ..., m-1\}$. The size of the neighborhood is $O(m^3)$.

Property 4. The complexity of exploring $N_4(R)$ is $O(m^3)$.

2-opt neighborhood removes each pair of edges from the solution and reconnects the vertices. This neighborhood of T is defined as a set $N_5(T) = \{T_{ij} = (v_1, v_2, ..., v_i, v_j, v_{j-1}, ..., v_{i+2}, v_{i+1}, v_{j+1}, ..., v_m) : i = 1, ..., n-4; j = i+4, ..., m\}$. The size of the neighborhood is $O(m^2)$.

Property 5. The complexity of exploring $N_5(T)$ is $O(m^2)$.

It is realized that the calculation of a neighboring solution's cost by using the known cost of the current solution can be done in constant time [15, 20]. As a result, the algorithm spends $O(m^3)$ operations for a full neighborhood search.

For intra-route: Let R_l, R_h, ml, and mh be two different routes and their sizes in T, respectively. Intra-route is used to exchange vertices between two different routes or remove vertices from a route and then insert them to another as followings:

The swap-intra-routes neighborhood tries to exchange the positions of each pair of vertices in R_l and R_h in turn. The neighborhood of R_l and R_h is defined as a set $N_8(T) = \{T_i = (R_1, ..., R_2, ..., R_l = (v_{1l}, v_{2l}, ..., v_{ih}, v_{il+1}, ..., v_{ml}), ..., R_h = (v_{1h}, v_{2h}, ..., v_{il}, v_{ih+1}, ..., v_{mh}), ..., R_k) : il = 2, 3, ..., ml-1, ih = 2, 3, ..., mh-1\}$. The size of the neighborhood is $O(ml \times mh)$.

Property 6. The complexity of exploring $N_6(T)$ is $O(ml \times mh)$.

Proof. For a tour $T \in N_6(R)$, we have

$$\begin{aligned} L(T_i) \quad = \quad & L(T) - (ml - i + 1)c(v_{il-1}, v_{il}) \\ & -(ml - i)c(v_{il}, v_{il+1}) \\ & -(mh - i - 1)c(v_{ih-1}, v_{ih}) \\ & -(mh - i)c(v_{ih}, v_{ih+1}) \\ & +(ml - i + 1)c(v_{il-1}, v_{ih}) \\ & +(ml - i)c(v_{ih}, v_{il+1}) \\ & +(mh - i + 1)c(v_{ih-1}, v_{il}) \\ & +(mh - i)c(v_{il}, v_{ih+1}). \end{aligned} \tag{9}$$

□

Hence, we can calculate $L(T_i)$ by the formulation (10) in $O(1)$ time. Therefore, the complexity of exploring $N_6(T)$ is $O(ml \times mh)$.

The insert-intra-routes neighborhood considers each vertex v_i in R_l and insert it into each position in R_h. The neighborhood of R_l and R_h is defined as a set $N_7(T) = \{T_i = (R_1, ..., R_l = (v_{1l}, v_{2l}, ..., v_{ih-1}, v_{ih}, v_{il+1}, ..., v_{ml}), ..., R_h = (v_{1h}, v_{2h}, ..., v_{ih-1}, v_{ih+1}, ..., v_{mh}), ..., R_k) : il = 2, 3, ..., ml-$

$1, ih = 2, 3, ..., mh - 1\}$. The size of the neighborhood is $O(ml \times mh)$.

Property 7. The complexity of exploring $N_7(T)$ is $O(ml \times mh)$.

Proof. For a tour $T \in N_7(R)$, we have

$$\begin{aligned} L(T_i) &= L(T) - (ml - i)c(v_{il}, v_{il+1}) \\ &\quad -(mh - i + 1)c(v_{ih-1}, v_{ih}) \\ &\quad -(mh - i)c(v_{ih}, v_{ih+1}) - \sum_{k=1h}^{ih-1} c(v_k, v_{k+1}) \\ &\quad +(ml - i)c(v_{il}, v_{ih}) \\ &\quad +(mh - i - 1)c(v_{ih}, v_{il+1}) \\ &\quad +(mh - i + 1)c(v_{ih-1}, v_{ih+1}) \\ &\quad + \sum_{k=1l}^{il-1} c(v_k, v_{k+1}). \end{aligned} \tag{10}$$

□

Hence, we can calculate $L(T_i)$ by the formulation (11) in $O(max(mh, ml))$ time. Therefore, the complexity of exploring $N_7(T)$ is $O(max(mh, ml) \times ml \times mh)$.

5 Algorithmic design

The proposed algorithm includes two phases as follows: In the construction phase, the algorithm [9] allows a controlled amount of randomness to overcome the behavior of a purely greedy heuristic. It is used to build an initial solution for our algorithm. In the improvement phase, the penalty VNS [14] is combined with shaking operators to escape from local optima. The proposed algorithm is repeated a number of times, and the best solution found is reported. An outline of the algorithm is shown in Algorithm 1. In Step 1, the algorithm starts with an initial solution. In Step 2, it is explored switches between different neighborhoods. To explore new promising solution spaces, a diversification step is added in Step 3. In the remaining of this section, more details about the three steps of our algorithm are given.

5.1 Feasible solution space

Penalty method is a technique to solve optimization problem with constraints. It adds a penalty value to the original objective function. The advantage of penalty technique is simple to implement. It is used to solve successfully many problem [21]. All infeasible solutions are penalized by a value. With a tour T, let $V(T)$ be the violation. The violation value $V(T)$ is computed as follows:

$$V(T) = \sum_{l=1}^{k} \max\{|R_l| - L, 0\} + \sum_{l=1}^{k} \max\{K - |R_l|, 0\}.$$

Solutions are then evaluated according to the weighted fitness function $L^{'}(T) = L(T) + \rho * V(T)$, where ρ is the penalty parameter

Algorithm 1 The Proposed Algorithm

Input: $v_1, V, N_i(T)(i = 1, ..., 7), level$ are a starting vertex, the set of vertices in K_n, the set of neighborhoods and the number of swap, respectively.
Output: The best solution T^*.
Step 1 (the construction phase):
{Initially, T is an empty tour}
repeat
$T = \phi$;
for $(l = 1; l < k; l++)$ **do**
$R_l = R_l \cup v_1$; {main depot is v_1}
while (all vertices are not visited) **do**
{Pick a random route that still satisfies the constraint if the new insertion is occurred}
$R = \{R_l | R_l \in T$ and $L - 1 \leq |R_l|$ && $|R_l| \leq K - 1\}$;
if $\exists! R_l$ **then**
R = Choose a random route ($R_l \in T$) with minimum cost; {accept an invalid route}
Create a *RCL* of v_e;{v_e is the last vertex of R_l} Select a randomly vertex $v = \{v_i | v_i \in RCL$ and v_i is not visited $\}$ to add to R_l;
for $(l = 1; l \leq k; l++)$ **do**
$T = T \cup R_l$;{update the tour T}
$LT = LT \cup T$;{LT is stored a list of solutions}
until iter
T = The best feasible solution if any. Otherwise the best infeasible one in *LT*;
while stop criteria not met **do**
Step 2 (the improvement phase):
for $i : 1 \rightarrow 7$ **do**
$T^{'} \leftarrow argmin_{T^{''} \in N_i(T)} L(T^{''})$
if $((L(T^{'}) < L(T))$ or $(L(T^{'}) < L(T^*)))$ **then**
$T \leftarrow T^{'}$
if $(L(T^{'}) < L(T^*))$ and ($T^{'}$ is feasible) **then**
$T^* \leftarrow T^{'}$
else
$i++$
Step 3 (Diversification):
type = rand(2);{Select randomly a number from 1 to 2}.
if type==1 **then**
R_l = Select randomly a route $\in T$;
R_l = shaking-single-route$(R_l, level)$;
else
(R_l, R_h) = Select randomly two routes of T;
(R_l, R_h) = shaking-multi-routes$(R_l, R_h, level)$;
Update T;
return T^*;

Algorithm 2 shaking-single-route$(R_l, level)$

Input: $R_l, level$ are the $l - th$ route, and the number of swap, respectively.
Output: a new solution R_l.
while $(level > 0)$ **do**
select i, j positions from R_l at random
if $(i \neq j)$ **then**
Insert $R_l[i]$ between $R_l[j]$ and $R_l[j + 1]$;
$level \longleftarrow level - 1$;
return R_l;

Algorithm 3 shaking-multi-routes($R_l, R_h, level$)

Input: $R_l, R_h, level$ are the $l-th$, $h-th$ route, and the number of swap, respectively.
Output: a new solution R_l and R_h.
  while $(level > 0)$ **do**
    select $i-th$ and $j-th$ positions from R_l and R_h at random, respectively;
    swap $R_l[i]$ between $R_h[j]$;
    $level \longleftarrow level - 1$;
  return R_l and R_h;

5.2 The construction phase

Our construction phase is developed on the GRASP scheme in [9]. Only one iteration is performed, and one solution is found which is either feasible or not. Its steps is described in Algorithm 1. All routes are initialized with v_1 because it is a starting vertex. Each vertex of K_n is then added to the tour by using a Restricted Candidate List (*RCL*). The *RCL* of each vertex includes a number of vertices that are the closest to it. At an iteration, we find a route R_l that does not violate the constraint if a new insertion occurs. Otherwise, if we cannot find any route, then we accept the infeasibility. That means a route R_l with minimum cost in the tour is picked. Let v_e be the current last vertex of the route R_l. An unvisited vertex v is then picked randomly from the *RCL* of v_e to add to R_l. A solution is generated when all vertices of K_n are routed. The above steps are executed *iter* times to create *iter* solutions. They are stored in a *LT* list. The procedure then returns the feasible solution with minimum cost in the list if any. If it cannot produce any feasible solution, the solution with minimum cost is penalized by adding a value to the objective function.

5.3 The improvement phase

For a given current solution T, the neighborhood explores the neighboring solution space set $N(T)$ of T iteratively and tries to replace T by the best solution $T^{'} \in N(T)$. The main operation in exploring the neighborhood is the calculation of a neighboring solution's cost. In straightforward implementation, this operation requires $Tsol = O(n)$. However, by using the known cost of the current solution, we show that this operation can be done in constant time for considered neighborhoods. Thus, we speed up the running time of exploring these neighborhoods.
In a preliminary study, we realize that the efficiency of VNS algorithm relatively depends on the order in which the neighborhoods are used. Therefore, the neighborhoods are explored in a specific order based on the size of their structure, namely, from the small to large, such as the swap-adjacent, remove-insert, swap, 2-opt, or, swap-intra-route, and insert-intra-route. The time complexity of exploring the neighborhoods is reduced by choosing a random vertex, and then we are only interested in neighborhoods generated from this vertex's moves. The strategy is called "restricted". As a result, the size of the neighborhood is reduced by a factor $O(n)$. Another is "without restricted" strategy when the entire neighborhood is explored without first fixing a random vertex. The reduction of neighborhood size is also used in [19]. The aim of using two strategies is to introduce several options to run the proposed algorithm effectively.

5.4 Diversification

Shaking procedure allows to guide the search towards an unexplored part of the solution space. In this work, two types of shaking are used to give a new solution: shaking in a single route (shaking-single-route) and shaking in two (shaking-multi-routes). In shaking procedure in a single route, it selects the l-th route R_l of T and then swaps randomly several vertices for each other. In the rest, it picks two routes R_l and R_h in a random manner, and after that, exchanges randomly some several vertices in them. We finally return to Step 2 with the new solution. The shaking procedure is described in Algorithm 2 and 3.

The last aspect to discuss is the stop criterium of our algorithm. A balance must be made between computation time and efficiency. Here, the algorithm stops if no improvement is found after the number of loop (*NL*).

5.5 The time complexity

The running time of our algorithm mainly spends on exploring in VNS. In the VNS step, insert-intra-routes neighborhood consumes time, at least as well as the others. Assume that if these neighborhoods are invoked k_1 times, then the complexity of neighborhoods' exploration is $O(k_1 \times max(mh, ml) \times ml \times mh) \sim O(k_1 \times n^3)$ (in the worst case the size of mh or ml is n). It is also the theoretical complexity of our algorithm.

6 Computational results

The experiments are conducted on a personal computer, which is equipped with an Intel Pentium core i7 duo 2.10 Ghz CPU and 4 GB bytes RAM memory.

6.1 Datasets

The numerical analysis is performed on a set of benchmark problems for the mTRP and Bounded-mTSP [8, 17, 18]. As testing our algorithm on all instances would have been computationally too expensive, we implement our numerical analysis of some selected instances. In [17], R. Necula et al. propose the Bounded-mTSP instances based on the TSPLIB benchmark. Specifically, they transform TSPLIB four instances (eil51, berlin52, eil76, and rat99) by setting the number of salesmen to be, by turn, 2, 3, 5, and 7. With the aim of obtaining balanced solutions, they choose to set the bounds (L and K) on the number of vertices in each

route, by running the k-means clustering algorithm. Besides that, we add more real instances by randomly choosing some instances from TSPLIB. We divide the instances into two groups: 1) in group one (G1), the vertices are concentrated; 2) in the other (G2) the vertices are scattered.

6.2 Metrics

To evaluate our algorithm's solution quality, we need to compare it with the other metaheuristics. The main problem is that there exists no other metaheuristic reported in the literature for this problem. That means we found no previous attempts to solve the problem, neither exact nor heuristic (or metaheuristic), to compare. We try to solve exactly several small instances using some state-of-art solvers and use those results to evaluate the performance of the proposed algorithm. However, the approach only solve the problem with small instances, while metaheuristics is a suitable approach for the problem with large sizes. Therefore, we adapt the existing algorithms in [17] to compare with the proposed algorithm for the Bounded-mTRP. We define the improvement of our algorithm with respect to *Best.Sol* (*Best.Sol* is the best solution found by our algorithm) in comparison with the initial solution (*Init.Sol*), upper bound (*UB*) obtained by the GRASP and the adapted algorithm in [17], respectively. $Improv[\%] = \frac{Best.Sol - Init.Sol}{Init.Sol} \times 100\%$, and $Gap[\%] = \frac{Best.Sol - UB}{UB} \times 100\%$. In addition, we choose several state-of-the-art metaheuristic algorithms for the Bounded-mTSP [17] (Bounded Multiple Traveling Salesman Problem) and mTRP (Multiple Traveling Repairmen Problem) in [8, 18] as a baseline in our research.

6.3 Results and discussions

Through preliminary experiments, we observe that the values $iter = 10, \rho = 10, \alpha = 5$, *level*=5, *PF*=100, and *NL* = 100 resulted in a good trade-off between solution quality and run time. In this paper, the neighborhoods' order is as follows: swap adjacent, remove-insert, swap, 2-opt, or-opt, swap-intra-routes, and insert-intro-routes. These settings have thus been used in the following experiments.

In the tables, *Init.Sol*, *Best.Sol*, *Aver.Sol*, and T correspond to the initial, best, and average solution, and the average time in seconds of ten executions obtained by our algorithm, respectively. The column ACS in Tables 1 and 2 describe the best results obtained from the adapted algorithms in [17]. Figure 1 and 2 shows the evolution of the average improvement in two strategies. The values in Figures are extracted from Table 4. In Table 5, kM-ACS, g-ACS, s-ACS, gb-ACS, and sb-ACS [17] are developed on Ant Colony System (ACS) with different strategies. The proposed algorithm is tested by selecting a fixed random vertex that is labeled "restricted". Runs in which the search explores all possible moves are labeled "without restricted".

6.3.1 Experimental results for the Bounded-mTRP

The experimental results in Table 3 are the average values calculated from Table 1 and 2. In Table 3, for all instances, it can be observed that our algorithm is capable of improving the solutions in comparison with *Init.Sol*. The average improvement of our algorithm with the two strategies is about 16.94% and 13.69%, respectively. Obviously, our algorithm can obtain a significant improvement for almost instances and required small-scaled running time. Both strategies seem to work well. The neighborhood implementation with fixed random vertex uses significantly less computing time, combined with a slight loss of solution quality (about 3.25%). However, the strategy proves useful for the larger instances, for which full neighborhood search is too time-consuming. Moreover, in comparison with the algorithms in [17], the proposed algorithm also outperforms for most instances.

From Tables 5 to 6 we can draw some conclusions about the working of our algorithm. Unsurprisingly, the multi-start version of our algorithm (algorithm settings 5 to 8) requires a much larger computation time than the single-start version (settings 1 to 4). However, the quality improvement obtained by this method is relatively small. The perturbations in the GRASP+VNS algorithm (Table 6) seem to help marginally, as the solutions obtained by this algorithm are usually slightly better. This may indicate that the GRASP multi-start is not able to provide enough diversification, and that the perturbation move is useful.

For two strategies, Figure 1 and 2 shows the evolution of the average deviation to the initial solutions with respect to $\overline{improv}$ and $\overline{T}$ during the iterations in some instances. The deviations in two strategies are 14.61% (10.38%), 16.25% (11.63%), 16.48% (11.79%), 16.66% (11.94%), 16.77% (12.03%), 16.94% (13.69%), and 16.94% (13.69%) for the first local optimum, obtained by one, ten, twenty, thirty, fifty, one-hundred, and two-hundred iterations, respectively. A major part of the descent obtained by from fifty to one-hundred iterations. As can be observed, additional iterations give a minor improvement with the large running time. Hence, the first way to reduce the large running time is to use no more than one-hundred iterations, and the improvement of the proposed algorithm is about 16.94% (13.69%) for two strategies, respectively. A much faster option is to run the initial construction phase then improve it by using a single iteration, which obtains an average deviation of 14.61% (10.38%) and an average time of 0.28 (0.21) seconds.

6.3.2 Experimental results for some variants

To the best of our knowledge, most algorithms are developed for a specific variant that is not applicable to other variants. Our algorithm can be applicable to the Bounded-mTSP, although it was not designed for solving them. In comparison with the state of the art algorithms for the Bounded-mTSP, and mTRP in [8, 17, 18], our algorithm's solutions are better than the other algorithms. Specifically,

Instances	*Init.Sol*	ACS	GRASP+VNS				
			Best.Sol	*Aver.Sol*	*Improv*[%]	*Gap*[%]	*Time*
eil51_2_23_27	6214.70	5163.14	**5088.79**	5088.79	18.12	-1.44	2.50
eil51_3_15_20	4009.22	3699.57	**3466.89**	3466.89	13.53	-6.29	1.16
eil51_5_7_12	3175.37	2670.60	**2639.06**	2639.06	16.89	-1.18	0.33
eil51_7_5_10	2525.52	2828.36	**2153.82**	2153.82	14.72	-23.85	0.29
eil76_2_36_39	11832.39	10249.57	**9734.12**	9734.12	17.73	-5.03	5.08
eil76_3_21_30	8194.47	6611.05	**6937.95**	6937.95	15.33	4.94	3.71
eil76_5_12_17	5043.09	5708.21	**4252.55**	4252.55	15.68	-25.50	1.83
eil76_7_7_15	5209.80	5040.89	**4266.03**	4266.03	18.12	-15.37	1.89
eil101_2_45_57	18303.24	15942.12	**14917.5**	14917.5	18.50	-6.43	10.54
eil101_3_23_50	13618.56	11090.67	11213.95	11213.95	17.66	1.11	7.79
eil101_5_16_30	9027.06	7268.68	**7268.68**	7268.68	19.48	0.00	5.47
eil101_7_12_25	6455.90	5499.07	**5272.63**	5272.63	18.33	-4.12	5.47
Aver					17.01	-6.93	3.84
KroA100_2_49_52	686248.62	553885.22	558433.17	558433.17	18.63	0.82	13.95
KroA100_3_24_48	487994.45	417822.65	**404763.8**	404763.8	17.06	-3.13	7.44
KroA100_5_16_25	312355.55	241252.38	253177.95	253177.95	18.95	4.94	3.99
KroA100_7_11_18	255989.80	266654.12	**209079.8**	209079.8	18.32	-21.59	2.23
KroB100_2_42_59	667895.70	593219.44	**549901.5**	549901.5	17.67	-7.30	11.62
KroB100_3_27_45	458771.17	432131.45	**396202.6**	396202.6	13.64	-8.31	12.44
KroB100_5_15_27	291305.15	333112.25	**244323.25**	244323.25	16.13	-26.65	4.31
KroB100_7_11_19	244886.31	272217.45	**204165.17**	204165.17	16.63	-25.00	2.64
KroC100_2_49_52	663971.01	523174.65	556508.51	556508.51	16.18	6.37	11.95
KroC100_3_25_49	567620.77	410810.89	466018.69	466018.69	17.90	13.44	9.38
KroC100_5_12_27	360747.36	309966.89	**309055.39**	309055.39	14.33	-0.29	4.38
KroC100_7_11_22	251277.09	277550.32	**207410.95**	207410.95	17.46	-25.27	3.69
KroD100_2_46_55	709352.94	592904.12	**579919.41**	579919.41	18.25	-2.19	14.01
KroD100_5_14_29	341858.76	277429.24	277429.24	277429.24	18.85	0.00	4.94
KroD100_7_9_20	261961.85	218211.38	**217485.09**	217485.09	16.98	-0.33	1.75
KroD100_3_30_37	493176.77	395740.11	**395740.11**	395740.11	19.76	0.00	5.18
berlin52_2_10_41	87767.00	70651.94	**70651.94**	70651.94	19.50	0.00	0.86
berlin52_3_10_27	60961.19	63532.70	**50604.73**	50604.73	16.99	-20.35	1.21
berlin52_5_6_17	40292.05	34479.08	**34479.08**	34479.08	14.43	0.00	0.32
berlin52_7_4_17	36237.36	29728.2	**29728.2**	29728.2	17.96	0.00	0.47
rat99_2_46_52	35642.60	33120.70	**31375.95**	31375.95	11.97	-5.27	12.37
rat99_3_27_36	36575.94	27953.98	**22781.51**	22781.51	37.71	-18.50	8.09
rat99_5_13_30	28373.58	22458.14	**16740.8**	16740.8	41.00	-25.46	3.80
rat99_7_9_22	20659.25	14071.49	**14071.49**	14071.49	31.89	0.00	2.24
pr107_2_48_57	1247880.30	1226071.31	**1222156.88**	1222156.88	2.06	-0.32	7.57
pr107_3_25_54	1163463.77	1168230.35	**990129.24**	990129.24	14.90	-15.25	1.63
pr107_5_18_30	824379.43	793875.99	**793875.99**	793875.99	3.70	0.00	1.55
pr107_7_12_19	832903.62	887050.87	**782543.35**	782543.35	6.05	-11.78	0.41
pr124_2_44_81	1843470.97	2129804.89	**1791978.36**	1791978.36	2.79	-15.86	7.51
pr124_3_22_61	1706257.60	1580239.34	**1531978.87**	1531978.87	10.21	-3.05	6.35
pr124_5_18_42	1163970.02	1059279.15	**1017313.3**	1017313.3	12.60	-3.96	3.85
pr124_7_10_31	1139285.94	1102460.45	**919782.83**	919782.83	19.27	-16.57	2.64
Aver					16.87	-7.21	5.46

Table 1: The experimental results for Bounded-mTRP without restricted neighborhood.

Instances	*Init.Sol*	ACS	GRASP+VNS				
			Best.Sol	*Aver.Sol*	*Improv*[%]	*Gap*[%]	*Time*
eil51_2_23_27	6214.70	5163.14	**5320.66**	5320.66	14.39	3.05	2.22
eil51_3_15_20	4009.22	3699.57	**3622.07**	3622.07	9.66	-2.09	1.02
eil51_5_7_12	3175.37	2670.60	**2765.06**	2765.06	12.92	3.54	0.23
eil51_7_5_10	2525.52	2828.36	**2250.86**	2250.86	10.88	-20.42	0.12
eil76_2_36_39	11832.39	10249.57	**10165.73**	10165.73	14.09	-0.82	5.13
eil76_3_21_30	8194.47	6611.05	**7259.52**	7259.52	11.41	9.81	1.52
eil76_5_12_17	5043.09	5708.21	**4422.10**	4422.10	12.31	-22.53	0.58
eil76_7_7_15	5209.80	5040.89	**4463.30**	4463.30	14.33	-11.46	0.58
eil101_2_45_57	18303.24	15942.12	**14917.50**	14917.50	18.50	-6.43	12.02
eil101_3_23_50	13618.56	11090.67	11673.10	11673.10	14.29	5.25	3.27
eil101_5_16_30	9027.06	7268.68	**7439.82**	7439.82	17.58	2.35	1.52
eil101_7_12_25	6455.90	5499.07	**5524.72**	5524.72	14.42	0.47	1.17
Aver					13.73	-3.27	2.45
KroA100_2_49_52	686248.62	553885.22	584354.04	584354.04	14.85	5.50	11.03
KroA100_3_24_48	487994.45	417822.65	420014.20	420014.20	13.93	0.52	6.87
KroA100_5_16_25	312355.55	241252.38	**264637.54**	264637.54	15.28	9.69	3.64
KroA100_7_11_18	255989.80	266654.12	**219045.30**	219045.30	14.43	-17.85	2.14
KroB100_2_42_59	667895.70	593219.44	**574151.78**	574151.78	14.04	-3.21	10.24
KroB100_3_27_45	458771.17	432131.45	**411341.84**	411341.84	10.34	-4.81	10.64
KroB100_5_15_27	291305.15	333112.25	**254446.94**	254446.94	12.65	-23.62	3.73
KroB100_7_11_19	244886.31	272217.45	**214091.24**	214091.24	12.58	-21.35	2.21
KroC100_2_49_52	663971.01	523174.65	**584134.98**	584134.98	12.02	11.65	10.35
KroC100_3_25_49	567620.77	410810.89	487197.00	487197.00	14.17	18.59	8.68
KroC100_5_12_27	360747.36	309966.89	315991.34	315991.34	12.41	1.94	2.71
KroC100_7_11_22	251277.09	277550.32	**214251.18**	214251.18	14.74	-22.81	2.73
KroD100_2_46_55	709352.94	592904.12	602080.08	602080.08	15.12	1.55	10.35
KroD100_5_14_29	341858.76	277429.24	288911.56	288911.56	15.49	4.14	3.73
KroD100_7_9_20	261961.85	218211.38	**224589.25**	224589.25	14.27	2.92	1.14
KroD100_3_30_37	493176.77	395740.11	408633.90	408633.90	17.14	3.26	4.41
berlin52_2_10_41	87767.00	70651.94	**74100.51**	74100.51	15.57	4.88	0.65
berlin52_3_10_27	60961.19	63532.70	**51355.45**	51355.45	15.76	-19.17	1.03
berlin52_5_6_17	40292.05	34479.08	**35874.04**	35874.04	10.96	4.05	0.33
berlin52_7_4_17	36237.36	29728.2	**31158.58**	31158.58	14.02	4.81	0.36
rat99_2_46_52	35642.60	33120.70	**32886.17**	32886.17	7.73	-0.71	10.47
rat99_3_27_36	36575.94	27953.98	**23857.93**	23857.93	34.77	-14.65	6.74
rat99_5_13_30	28373.58	22458.14	**17512.63**	17512.63	38.28	-22.02	3.22
rat99_7_9_22	20659.25	14071.49	**14757.88**	14757.88	28.57	4.88	1.31
pr107_2_48_57	1247880.30	1226071.31	**1222156.88**	1222156.88	2.06	-0.32	5.06
pr107_3_25_54	1163463.77	1168230.35	**1037508.36**	1037508.36	10.83	-11.19	1.28
pr107_5_18_30	824379.43	793875.99	**828969.83**	828969.83	0.56	4.42	1.12
pr107_7_12_19	832903.62	887050.87	**814688.42**	814688.42	2.19	-8.16	0.25
pr124_2_44_81	1843470.97	2129804.89	**1869437.07**	1869437.07	1.41	-12.22	5.45
pr124_3_22_61	1706257.60	1580239.34	1592631.47	1592631.47	6.66	0.78	5.06
pr124_5_18_42	1163970.02	1059279.15	**1064491.90**	1064491.90	8.55	0.49	3.12
pr124_7_10_31	1139285.94	1102460.45	**965296.64**	965296.64	15.27	-12.44	2.45
Aver					13.64	-3.45	4.46

Table 2: The experimental results for Bounded-mTRP with restricted neighborhood.

Instances	without restricted		restricted	
	Gap[%]	*Time*	*Gap*[%]	*Time*
G1	17.01	3.84	13.73	2.45
G2	16.87	5.46	13.64	4.46
aver	16.94	4.64	13.69	3.45

Table 3: The average results for two schemes.

schemes	Dataset	1 iteration		10 iterations		20 iterations		30 iterations		50 iterations		100 iterations		200 iterations	
		Improv	*T*	*Improv*	*T*	*Improv*	*T*	*Improv*	*T*	*Improv*	*T*	*Improv*	*T*	*Improv*	*T*
without restricted	TSP-G1	13.99	0.23	16.24	0.64	16.57	0.93	16.85	1.55	17.01	3.22	17.01	3.84	17.01	14.15
	TSP-G2	15.30	0.33	16.28	0.93	16.40	1.35	16.48	2.24	16.53	4.47	16.87	5.46	16.87	19.81
	Aver	14.61	0.28	16.25	0.78	16.48	1.14	16.66	1.89	16.77	3.85	16.94	7.65	16.94	16.99
restricted	TSP-G1	11.29	0.15	13.12	0.41	13.37	0.60	13.60	0.99	13.73	2.05	13.73	2.45	13.73	9.03
	TSP-G2	9.34	0.27	9.94	0.76	10.01	1.11	10.05	1.83	10.09	3.65	13.64	4.46	13.64	16.17
	Aver	10.38	0.21	11.63	0.58	11.79	0.85	11.94	1.41	12.03	2.85	13.69	3.45	13.69	12.60

Table 4: Evolution of average deviation to *Init.Sol* without restricted neighborhood.

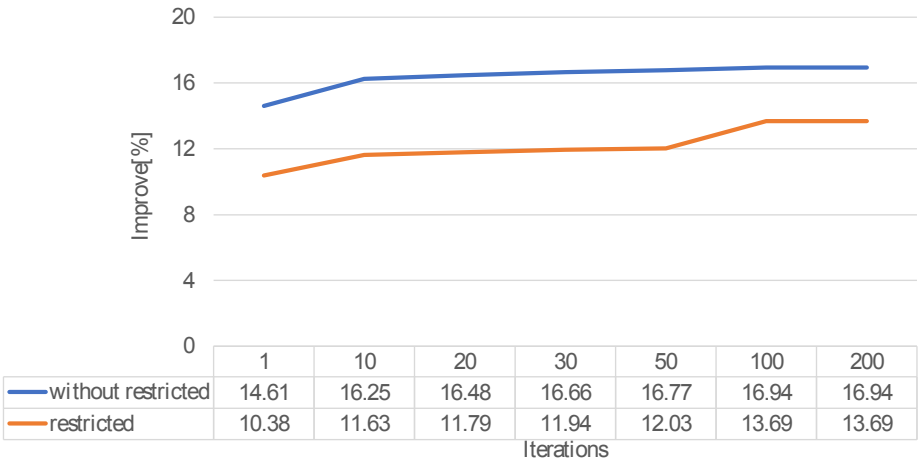


Figure 1: Evolution of average Improve [%].

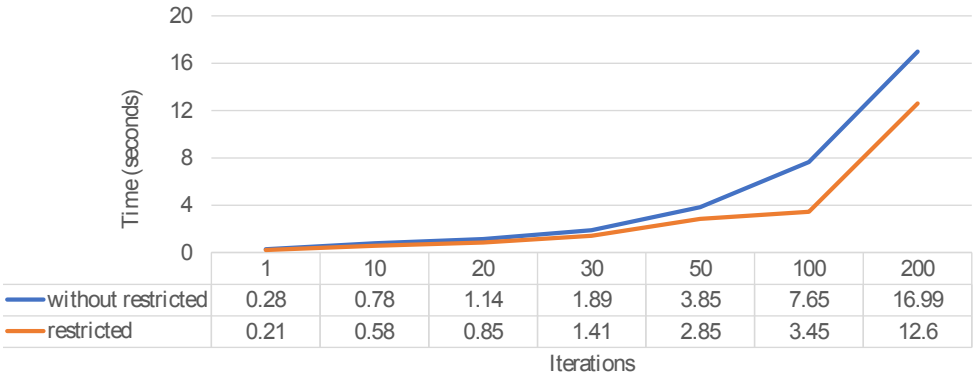


Figure 2: Evolution of average running time.

for the Bounded-mTSP in [17], our algorithm reaches better solutions for 12 out of 16 tested instances at a reasonable computational time. In addition, our algorithm can find the optimal solutions (eil51-2-23-27, eil51-3-15-20) or near-optimal solutions (berlin52-2-10-41, berlin52-3-10-27, berlin52-5-6-17) for the problems with 50 vertices

Instances	*OPT* Or [*LB, UB*]	kM-ACS	g-ACS	s-ACS	gb-ACS	sb-ACS	GRASP+VNS		
		Best.Sol	*Best.Sol*	*Best.Sol*	*Best.Sol*	*Best.Sol*	*Best.Sol*	*Aver.Sol*	*Time*
eil51_2_23_27	442.32	454.3±0.84	452.66±1.77	454.96±2.04	452.22±1.48	453.81±1.63	**442.32**	442.32	2.36
eil51_3_15_20	464.11	500.00±0.24	485.73±3.44	489.64±3.59	479.51±3.37	483.39±3.75	**464.11**	464.11	1.26
eil51_5_7_12	[519.10, 529.70]	563.58±0.52	582.36±3.58	590.63±4.64	585.76±4.34	598.61±5.16	597.17	597.17	0.35
eil51_7_5_10	[584.02, 605.21]	634.47±0.04	674.78±4.32	680.38±3.84	688.26±3.57	699.47±4.34	731.85	731.85	0.33
berlin52_2_10_41	7753.89	8836.80±27.15	8043.92±46.91	8036.08±43.46	8057.38±43.06	8122.44±46.44	**6977.56**	6977.56	0.93
berlin52_3_10_27	8106.85	9009.18±11.83	8653.86±47.04	8806.95±64.40	8795.52±48.13	8839.37±42.63	**7942.45**	7942.45	1.32
berlin52_5_6_17	[8894.50, 9126.33]	10335.03±1.88	10164.58±90.66	10343.52±93.34	10660.46±92.58	10866.66±106.15	**8840.50**	8840.50	0.34
berlin52_7_4_17	[9415.99, 9870.02]	11966.20±2.26	11993.31±137.66	12125.55±121.75	12451.16±104.28	12712.41±97.88	**10191.53**	10191.53	0.51
eil76_2_36_39	558.59	594.21±0.93	580.77±2.91	583.41±3.04	579.68±2.39	578.96±2.81	596.09	596.09	5.52
eil76_3_21_30	579.30	642.89±0.98	622.91±3.41	630.67±5.09	613.76±3.26	619.19±4.12	**592.65**	592.65	4.04
eil76_5_12_17	[623.88, 680.67]	740.35±0.23	747.49±4.75	760.05±5.02	734.61±3.66	744.94±4.39	**685.98**	685.98	2.00
eil76_7_7_15	[675.38, 759.90]	820.35±0.10	873.65±6.61	883.63±5.44	894.70±4.68	911.06±5.00	**693.35**	693.35	2.05
rat99_2_46_52	[1296.35, 1350.73]	1485.56±2.98	1398.01±8.72	1391.89±7.35	1382.05±4.45	1389.08±5.45	1491.65	1491.65	14.55
rat99_3_27_36	[1357.30, 1519.49]	1672.11±3.33	1691.56±11.64	1707.20±12.07	1661.04±8.89	1651.68±11.57	**1449.34**	1449.34	9.30
rat99_5_13_30	[1523.95, 1855.83]	1996.04±1.23	2260.74±14.03	2297.05±16.83	2286.73±16.12	2337.94±10.45	**1810.60**	1810.60	4.38
rat99_7_9_22	[1712.1467, 2291.8207]	2361.55±0.74	2859.98±22.36	2878.97±21.11	3004.37±26.22	2984.42±17.38	**2155.53**	2155.53	2.52

Table 5: Comparisons with the state of the art metaheuristics for Bounded-mTSP without restricted neighborhood.

Instances	IOE		SNG		Our Algorithm	
	Best.Sol	*T*	*Best.Sol*	*T*	*Best.Sol*	*T*
P-n40-k5	-	-	1537.79*	0.25	1580.21	0.32
P-n45-k5	-	-	1912.31*	0.39	**1912.31**	0.37
E-n51-k5	3320	2.25	2209.64*	0.7	2247.83	0.48
P-n50-k7	-	-	1547.89*	0.70	1590.41	0.68
P-n51-k8	-	-	1448.92*	0.67	**1448.92**	0.61
E-n76-k10	4094	1.48	2310.09*	4.2	2419.89	0.91
E-n76-k14	3762	0.5	2005.4*	3.4	**2005.4**	0.83
E-n101-k8	6383	89.4	-	-	4051.47	2.94
E-n101-k14	5048	5.43	-	-	3288.53	2.92

* is the optimal value

Table 6: Comparisons with the state of the art metaheuristics for mTRP.

in several seconds in Table 5. For the mTRP in [8, 18] in Table 6, the quality of our solutions is much better than I. O. Ezzine et al.'s algorithm (IOE) in [8] and every comparable with S. Nucamendi-Guillen et al.'s algorithm (SNG) in [18]. Moreover, our algorithm can find the optimal solutions for the problems for the mTRP instance with up to 76 vertices in several seconds.

6.4 Discussions

Metaheuristic approach is a suitable approach to solve the large sized-problem. The VNS [14] is a popular schemes used widely to solve NP-hard problems. They are very effective for some variants of the mTRP problem [15, 19, 20]. The proposed algorithm is encouraged by the efficiency of the algorithms in [15, 19, 20]. However, the difference between the proposed VNS and their VNS is that our algorithm builds up penalty value during a search. Our main contribution is to adapt the VNS scheme, that extends the well known VNS by including constraint penalization, to solve the Bounded-mTRP effectively.

A good metaheuristic must balance between exploration and exploitation. Exploration is to create diverse solutions on a global space, while exploitation is to focus on the search good current local regions. In the proposed algorithm, the VNS implements exploitation while shaking maintains exploration. In terms of experiments, the proposed algorithm obtains better solutions than the adapted algorithms in [17] in many cases. We also implement two strategies that provide several choices: 1) The first choice is to run the proposed algorithm with one iteration in "restricted" strategy that obtains 10.38% solution quality on average. The running time is very fast; 2) The second is to run the proposed algorithm with one iteration in "without restricted" strategy that obtains an average improvement of 14.61%. The second option trades off solution quality and running time; 3) The last is to run the proposed algorithm no more than 100 iterations. The average improvement is 16.94%. The option is the best in terms of solution quality.

Moreover, the proposed algorithm obtains comparable or better solutions than the algorithms for the mTRP and Bounded-mTSP in [8, 17, 18]. It shows that our algorithm is still effective for various problems.

7 Conclusions

In this paper, we propose the first metaheuristic algorithm, which is mainly based on the VNS and GRASP's principles to solve the problem. Extensive numerical experiments on benchmark instances show that, on average, our algorithm leads to significant improvement. For small instacnes, our algorithm obtains the optimal solutions for the problem with 76 vertices at a reasonable a mount of time. For larger instances, the proposed algorithm reaches better solutions than the state-of-the-art algorithms in many cases.

References

[1] A. Archer, A. Levin, and D. Williamson, "A Faster, Better Approximation Algorithm For The Minimum Latency Problem", J. SIAM, Vol. 37, No. 1, 2007, pp. 1472-1498. https://doi.org/10.1137/07068151x.

[2] F. Afrati, S. Cosmadakis, C. Papadimitriou, G. Papageorgiou, and N. Papakostantinou, "The Complexity Of The Travelling Repairmen Problem", J. Informatique Theorique Et Applications, Vol. 20, pp.79–87. https://doi.org/10.1051/ita/1986200100791

[3] M. Avci, M.G. Avci, 2017, "A GRASP with iterated local search for the Traveling Repairman Problem with Profits", J. Comput. Ind. Eng. 113, PP. 323–332. https://doi.org/10.1016/j.cie.2017.09.032.

[4] Ha-Bang Ban, Duc-Nghia Nguyen, and Kien-Nguyen, "An Effective Metaheuristic for Multiple Traveling Repairman Problem with Distance Constraints", J. CAI, Vol. 38, 2019, pp. 1001-1034. https://doi.org/10.31577/cai_2019_4_883.

[5] A. Blum, P. Chalasani, D. Coppersmith, W. Pulleyblank, P. Raghavan, and M. Sudan, "The Minimum Latency Problem", Proc. STOC, 1994, pp.163-171. https://doi.org/10.1145/195058.195125

[6] M.E. Bruni, P. Beraldi, S. Khodaparasti, "A Fast Heuristic for Routing in Post-Disaster Humanitarian Relief Logistics", J. Computers and Operations Research, Vol. 30, pp. 304-313, 2018. https://doi.org/10.1016/j.trpro.2018.09.033.

[7] N. Christofides, A. Mingozzi, p. Toth, "Exact Algorithms for the Vehicle Routing Problem based on Spanning Tree and Shortest Path Relaxations". j. Math Program, Vol. 20, 1981, pp. 255–282. https://doi.org/10.1007/bf01589353.

[8] I. O. Ezzine, and Sonda Elloumi, "Polynomial Formulation and Heuristic Based Approach for the k-Travelling Repairmen Problem", Int. J. Mathematics In Operational Research, Vol. 4, No. 5, 2012, pp. 503-514. https://doi.org/10.1504/ijmor.2012.048928.

[9] T.A. Feo, and M.G.C. Resende, "Greedy Randomized Adaptive Search Procedures", J. Global Opt., 1995, pp. 109–133.

[10] F. Jittat, C. Harrelson, and S. Rao, "The k-Traveling Repairmen Problem", Proc. ACM-SIAM, 2003, pp.655-664.

[11] D. S. Johnson, and L. A. Mcgeoch, "The Traveling Salesman Problem: A Case Study In Local Optimization In Local Search In Combinatorial Optimization", E. Aarts and J. K. Lenstra, Eds., pp. 215-310. https://doi.org/10.2307/j.ctv346t9c.13

[12] R. Jothi, and B. Raghavachari, "Minimum Latency Tours and The k-Traveling Repairmen Problem", Proc. LATIN, 2004, pp. 423–433. https://doi.org/10.1007/978-3-540-24698-5_46.

[13] O. Martin, S. W. Otto, and E. W. Felten, "Large-Step Markov Chains For The Traveling Salesman Problem", J. Complex Systems, Vol. 5, No. 3, 1991, pp. 299-326.

[14] N. Mladenovic, and P. Hansen, "Variable Neighborhood Search", J. Operations Research, Vol.24, No. 11 24, 1997, pp.1097-1100. https://doi.org/10.1016/s0305-0548(97)00031-2.

[15] N. Mladenovic, D. Urosevi, and S. Hanafi, "Variable Neighborhood Search for the Travelling Deliveryman Problem", J. 4OR, 2012; 11: 1-17. https://doi.org/10.1007/s10288-012-0212-1.

[16] Z. Luo, H. Qin, and A. Lim, "Branch-and-Price-and-Cut for the Multiple Traveling Repairman Problem with Distance Constraints", J. Operations Research, Vol. 234, No. 1, 2013, pp. 49-60. https://doi.org/10.1016/j.ejor.2013.09.014.

[17] R. Necula, M. Breaban, M. Raschip, "Performance Evaluation of Ant Colony Systems for the Single-Depot Multiple Traveling Salesman Problem", Proc. HAIS, vol. 9121, pp. 257-268, 2015. https://doi.org/10.1007/978-3-319-19644-2_22.

[18] S. Nucamendi-Guillén, I. Martínez-Salazar, F. Angel-Bello, and J. M. Moreno-Vega, "A Mixed Integer Formulation and An Efficient Metaheuristic Procedure for the k-Travelling Repairmen Problem", J. JORS, Vol. 67, No. 8, 2016, pp. 1121-1134. https://doi.org/10.1057/jors.2015.113.

[19] A. Salehipour, K. Sorensen, P. Goos, and O.Braysy, "Efficient GRASP+VND and GRASP+VNS metaheuristics for the Traveling Repairman Problem", J. Operations Research, 2011, Vol. 9, No. 2, 189-209. https://doi.org/10.1007/s10288-011-0153-0.

[20] M. Silva, A. Subramanian, T. Vidal, and L. Ochi, "A simple and effective metaheuristic for

the Minimum Latency Problem", J. Operations Research, Vol. 221, No. 3, 2012, pp.513-520. DOI: 10.1016/j.ejor.2012.03.044

[21] A. Piwonska, F. Seredynski, “A Genetic Algorithm with a Penalty Function in the Selective Travelling Salesman Problem on a Road Network. Proc. IPDPS, 2011. pp. 381-387. https://doi.org/10.1109/ipdps.2011.177.

A New Hybrid LGPMBWM-PIV Method for Automotive Material Selection

Saif Wakeel[1,2] and Sedat Bingol[2]
[1]Centre of Advanced Materials, Department of Mechanical Engineering
University of Malaya, Kuala Lumpur, Malaysia
[2]Department of Mechanical Engineering, Dicle Universitesi, Diyarbakir, Turkey
E-mail: saifwakeel@gmail.com, sbingol@dicle.edu.tr

Shafi Ahmad
Department of Mechanical Engineering, Jamia Millia Islamia, New Delhi, India
E-mail: shafiahmad.amu@gmail.com

M. Nasir Bashir
National University of Sciences and Technology, Islamabad, Pakistan
E-mail: nasir@pnec.nust.edu.pk

Mir Seyed Mohammad Mohsen Emamat
Department of Industrial Management, Allameh Tabataba'i University, Tehran, Iran
E-mail: sm.emamat@gmail.com

Zhou Ding
Centre of Advanced Materials, Department of Mechanical Engineering
University of Malaya, Kuala Lumpur, Malaysia

H. Fayaz
Modeling Evolutionary Algorithms Simulation and Artificial Intelligence,
Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam
E-mail: fayaz@tdtu.edu.vn

Keywords: natural fibres, car roof, automotive material selection, linear goal programming model for best-worst method (LGPMBWM), PIV method

Received: July 21, 2020

Efforts are continuously being made by researchers to improve fuel efficiency and to reduce CO_2 emissions from the passenger cars. To achieve these goal, recent trend is to make the cars components light in weight for which manufacturing car roofs using natural fiber reinforced composites (NFCs) is one of the method. Several natural fibers (NFs)are available as alternative reinforcements for the fabrication of NFCs. Different NFs possess different properties and therefore, it is necessary to select the most appropriate natural fiber for fabrication of the composites which in turn will lead to the desired performance of the vehicle. Selection of the optimal natural fiber, amongst the several alternatives, is basically a multi criteria decision making (MCDM) problem as selection is based on the evaluation of several conflicting criteria. In this study, twelve alternative natural fibers (Flax, Hemp, Jute, Kenaf, Ramie, Okra, PALF, Coir, Isora, Cotton, Banana and Sisal) and six evaluation criteria (Tensile strength, Stiffness, Failure strain, Density, Degradation temperature and Moisture gain) are considered and selection of the optimal NF is made using a newly developed hybrid MCDM method i.e. Linear goal programming model for Best-Worst method (LGPMBWM) and Proximity index value method (PIV). Results of the study reveal that among all considered natural fibers, Ramie fiber is the most suitable alternative for the fabrication of composites and coir fiber is the worst candidate for the same. Ranking results were also supported by five other MCDM methods as there was a strong correlation between PIV and other MCDM methods.

Povzetek: V prispevku so opisali izvirno hibridno metodo za iskanje novih delov avtomobilskih motorjev.

1 Introduction

The temperature of earth is increasing due to the emission of global warming gases from various sources such as industries and different modes of transport. Production of transport vehicles needs four times less energy as compared to energy required to drive them during their complete service life [1]. Therefore, emissions from transports contribute major portion of global warming

gases. However, the problem of global warming due to emissions from transports can be minimized by improving their performance in terms of fuel economy which can be achieved by reducing their weight to strength ratio [2]. Consequently, it is suggested to produce light weight vehicles with increased strength. The weight of the vehicular transports can be minimized by using appropriate materials in their production. Thus, it is imperative to select the best material from the existing numerous materials to achieve the objective of making vehicles light in weight.

The car roof is usually encountered with dangerous rollover accidents due to which the occupants get serious head and neck injuries and sometimes the accident is so fatal that they lose their life [3]. Therefore, for safety of passengers, it is necessary to ensure that the car roof is sufficiently strong to withstand the impact in the event of an accident. Toughness of steel sheet is relatively poor which leads to transfer of shocking load to occupants from car roof during accident. Therefore, materials with high value of toughness such as plastic based composite materials are suggested for car roof manufacturing. It has been reported that composites such as Kevlar fiber /Epoxy, carbon fiber/Epoxy and Boron fiber/epoxy possess high tensile and flexural strengths as compared to aluminum and steel [4]. For manufacturing of car roof, thermoset and thermoplastic base fiber reinforced composites have favorable properties such as light in weight, higher toughness and good flexural strength. Generally, composite has high resistance for corrosion as compared to commercial grade of steel which is an additional benefit that restricts deterioration of material of the automotive components such as car roof by corrosion phenomenon which ultimately leads to improved service life of the automotive vehicles.

Selection of a suitable material is based on several desirable conflicting attributes. Therefore, material selection is a multi-criteria decision making problem (MCDM) which needs to be solved by using appropriate MCDM method. Literature reveals that past researchers have used several MCDM methods for solving various problems pertaining to different field of applications. Al-Oqlaand Sapuan [5] employed Analytical Hierarchy Process (AHP) to select the best natural fiber (NF) among coir, date, palm, hemp, flax and sisal for sustainable automotive industry and found that Flax was the best alternative material followed by date palm fiber. Al-Oqlaet al [6] used AHP method for the selection of best polymer based matrix to form flax and date palm reinforced natural fiber composite and found that polypropylene (PP) as the best matrix material. Al-Oqlaet al [7] applied AHP method for the selection of best natural fiber for polypropylene based NFCs and reported that flax fiber was the best reinforcing agent for polypropylene matrix. Maskepatil et al [8] employed AHP method for selection of the most suitable material among wood, steel, aluminium, glass fiber and carbon fiber for designing wind turbine blade and observed that carbon fiber was the best choice for the same. Luqmanet al [9] used AHP method to determine best suitable composite fabrication method to manufacture carbon fiber crank arm of bicycle and observed that compression moulding process was the best choice among all the manufacturing processes. Anojkumar et al [10] employed four different methods: TOPSIS, VIKOR, Electre computational and PROMTHEE computational for the selection of best steel material for pipe manufacturing in sugar industry and found that M 304 steel grade was best suited steel grade among all five steel grades. Anupam et al[11]successfully applied TOPSIS method for selecting optimum material for pulp and paper making industry. Al-Oqla et al [12] successfully investigated the best reinforcing condition for fabrication of NFCs using TOPSIS and AHP method and suggested the importance of NaOH treatment for fabrication of NFCs. Majumdar et al [13] applied AHP for selection of best cotton among Cotton grades (A to H) and suggested Cotton D as the best alternative. Ozturk et al[14] selected light weight fabric from natural cellulose composite employing weighted sum method. Mohammed et al [15] employed TOPSIS method for selecting best glass fiber reinforced epoxy hybrid composite and suggested Glass + epoxy + 5% coal fly ash as the best feasible composite. Jha et al[16] selected optimal biodegradable composite among many composites formed by pine cone with graphite content (0,5,10,15%) /Polycaprolactone using fuzzy TOPSIS method and their result clearly showed that Polycaprolactone/ pine cone + graphite (0%) composite was a suitable choice. Getting motivated from the wide application of AHP and TOPSIS methods, Ahmed [17] developed a Java scripted MCDM weight range method for selection of NFCs material for door panel and suggested that Sisal 30%-PP was suitable NFC. Ishak et al [18] employed fuzzy VIKOR method for selection of NFs for car front hood and the result of their study revealed that kenaf fiber was suitable choice for designing of car front hood. Besides, MCDM methods have also been used in other environments. Chang [19] selected the most suitable public relations personal for tourism industry using hybrid fuzzy Delphi-ANP-TOPSIS method. Chakraborty and Zavadskas [20] successfully used WASPAS method for ranking the alternative involve in the parameter selection of eight machining problem namely, cutting fluid, electrospinning, forging conditions, arc welding process, industrial robot, milling conditions, machinability of materials and process parameters selection for electro-discharge micro machining. Selection of new and right technology for industrial sustainability was done using hybrid fuzzy-ANP and fuzzy- TOPSIS method [21]. Keshavarz et al. [22] applied a EDAS method for solving the multi-criteria inventory classification problem.

From the available literature it is observed that MCDM techniques can be used efficiently for material selection. However, each technique have certain advantages and limitations such as AHP and TOPSIS have rank reversal problem. VIKOR has problem associated with closeness of alternatives from ideal which prompted us to use the recently developed PIV method. PIV method eliminates the rank reversal problem and include short steps. Besides, original BWM has 4n-5 number of constraints (n is number of criteria) however, recently developed LGPMBWM consist of 2n-2

constraints which minimizes the computational complexity and gives better consistency results as compared to original BWM method. To the best of author's knowledge, this kind of technique has never been used for automobile material selection problem. Therefore, recently developed LGPMBWM-PIV method has been applied in this study for ranking the alternative natural fiber to formulate the car roof polymer composite on the basis of various conflicting criteria. As a result of this study, best suited natural fiber was selected among all the alternative materials. Consequently, sensitivity analysis has been performed successfully to verify the results of this MCDM method and their results showed that technique applied is reliable and consistent.

2 Methodology

The methodology adopted for solving the ranking problem of the selected natural fibers is depicted in the research framework shown in Fig. 1.

The details of criteria selection, criteria weights calculation, and different methods used for ranking of alternatives are given in the following sections.

2.1 Selection of criteria for car roof materials

On the basis of car rollover testing, fabrication process temperature, strength to weight ratio and delamination process of NFCs, six criteria (shown in Table.1) are taken in account. As a first criterion, density of reinforcing phase i.e. natural fibres have direct effect on strength to weight ratio of composites used for car roof manufacturing which leads to fuel economy of the vehicles. Therefore, there is a need to minimize the density of the reinforcing phase of the composite to attain better strength to weight ratio. Consequently, density is considered as an important non-beneficial criteria for selection of natural fibre. Second criterion is tensile strength of natural fiber and it is observed that tensile strength of car roof composite (NFC) mainly depends on tensile strength of the reinforcing phase i.e. natural fibres [23]. Therefore, tensile strength of

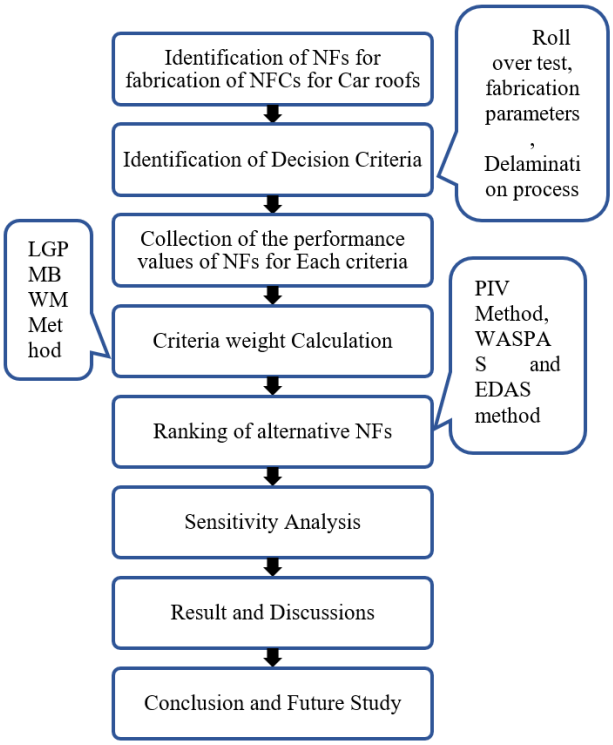


Figure 1: Research Framework.

the natural fibres is considered as an important beneficial criteria. Third criterion is the stiffness of NFCs which is mainly governed by reinforcing phase i.e. natural fibres. The design of lightweight structures such as car roofs requires the use of NFCs having greater stiffness [24]. Therefore, it is necessary to maximize the value of stiffness for natural fibre to fabricate a composite for car roof manufacturing which has greater value of stiffness. Value of Failure strain of composite materials is significantly depends on the reinforcing agent i.e. natural fibres. The composite materials used to manufacture car roofs must have good value for toughness and toughness of composite materials is also depends on failure strain of natural fibres. Therefore, it is necessary to include failure strain of natural fibres in criteria. TGA analysis showed that every natural fibre have different thermal degradation

Alternative NFs	Tensile Strength (TS) in MPa	Stiffness (S) in GPa	Failure Strain (FS) in %	Density (D) in g/cm^3	Degradation Temperature (DT) in ^{0}C	Moisture Gain (M) in %
Goal	(+)	(+)	(-)	(-)	(+)	(-)
Flax(F)	975	32.8	2.35	1.52	250	12
Hemp(H)	781.5	9.93	3.15	1.48	250	12
Jute(J)	583	14.5	1.5	1.375	205	17
Kenaf(K)	423.5	12.75	4.2	1.4	219	17
Ramie(R)	669	76.35	2.5	1.45	300	8.5
Okra(O)	307	9	1.95	0.88	220	13
PALF(P)	170	62.1	3	1.52	220	14
Coir(C)	140.5	6	22.5	1.25	190	13
Isora(I)	580.5	20.5	5.5	1.37	220	1.2
Cotton(CO)	500	8	7	1.55	220	8.5
Banana(B)	750	29.5	2.95	1.3	200	13.5
Sisal(S)	460	15.5	8	1.4	300	14

Table.1 Properties of selected Natural fibres [23-39].

a_{BW}	1	2	3	4	5	6	7	8	9
consistency index	0.00	0.44	1.00	1.63	2.30	3.00	3.73	4.47	5.23

Table 2. Consistency index.

temperature because of the difference in their activation energy which depends on their chemical composition [25]. Generally, Natural fibres degrade at low temperature as compared to synthetic fibres. Available literature revealed that thermoplastic matrix based composites need higher processing temperature during fabrication therefore, it is suggested to improve the thermal degradation temperature of the reinforcing agent i.e. natural fibre by choosing the appropriate natural fibre. Consequently, thermal degradation temperature of natural fibres is necessary to include in the required beneficial criteria of composite which is used to manufacture roofs of the car. Natural fibres have hydrophilic nature due to which they gain moisture contents but the chemical composition of natural fibres are different to each other which leads to differences in percentage of moisture gain by the natural fibres. The interfacial shear strength (IFSS) of the natural fibre composites is significantly influenced by moisture contents of natural fibres [26]. It has been observed that IFSS of the NFCs shows deterioration for moisture gain by natural fibres [27]. In NFCs, it is suggested to minimize or eliminate the moisture contents of reinforcing agents i.e. natural fibres. Consequently, percentage of moisture gain by natural fibres is an important criteria for selections of natural fibres to insure the IFSS of NFCs which is used to manufacture car roofs.

In this study, twelve natural fibers (listed in Table 1) with their different attributes/properties have been considered as available alternatives being used for fabrication of composites for car roof manufacturing. These fibers are eco-friendly, biodegradable and their abundance and light weight make them suitable to be used in automotive, aerospace and sports industries.

In order to rank the natural fibers, it is necessary to evaluate the weights of criteria which can be done by various methods but in our study simple, consistent and reliable LGPMBWM method is employed for the weight calculation. The comprehensive study of LGPMBWM is explained in the following section.

2.2 Linear goal programming model for best worst method (LGPMBWM)

In this study, recently developed linear goal programming model for best-worst method (LGPMBWM) [40] is employed for weight calculation which involves the selection of best (most important criterion) and worst factor (least important criterion) and their comparison with other criteria on the basis of comparison scale from 1 to 9. Thus, this comparison leads to the formation of two pairwise comparison vector i.e. Best to Others (BO) and Others to Worst (OW) vectors. Further, using LGPMBWM the optimal weights of criteria and consistency are calculated. The BWM has certain advantages over the other methods as [41]; (i) It provides minimum total deviation and thereby ensures closer weight ratios, (ii) It provides consistent comparisons, and (iii) In comparison to AHP, it provides minimum violation i.e. better ordinal consistency. Based on the significant advantages of BWM, researchers have used this method for calculating weight of criteria in various applications [42-50]. In addition to these advantages, the LGPMBWM has 2n-2 number of constraints while the original BWM had 4n-5 number of constraints (n is the number of criteria). The LGPMBWM has fewer constraints in comparison with the BWM, which results in improved computational solution and reducing the complexity in original BWM. Further, detail of this method can be found in work done by its developer [40].

However, it is necessary to provide the steps of this method for visualizing the clear picture behind its use in present research.

Step 1: Identify n decision criteria $\{C_1, C_2, \ldots, C_n\}$ for making decision. In decision making problem of natural fibers selection for Car roof, criteria are as follows: Tensile Strength (TS), Stiffness (S), Failure Strain (FS), Density (D), Degradation Temperature (DT) and Moisture Gain (M).

Step 2: Select the best and the worst criteria. In this study best and worst factors have been selected on the basis of academic expert advice.

Step 3: In this step, pairwise comparison is done between best criterion and other criteria by using numbers between 1 to 9 (1: equally important, 2: weakly important, 3: moderately important, 4: moderately plus important, 5: strongly important, 6: strongly plus important, 7: very strongly important, 8: very, very strongly important 9: extremely important) to determine the importance of the best criterion over others which leads to the formation of best to others (BO) vector as:

$$A_B = (a_{B1}, a_{B2}, a_{B3} \ldots . a_{Bn}) \quad (1)$$

Where, a_B is the best to others (BO) vector and $a_{\mathrm{Bj}} =$ Importance of best criterion over the j[th] criterion. It is obvious that $a_{\mathrm{BB}} = 1$.

Step 4: Comparison of all the criteria with the worst criterion is done in same way as in step 3 which leads to the formation of others-to-worst (OW) vector as:

$$A_{\mathrm{w}} = (a_{1w}, a_{2w}, a_{3w} \ldots . a_{nw})^{\mathrm{T}} \quad (2)$$

where a_{jw} = importance of j[th] criterion with respect to the worst criterion. It is evident that $a_{\mathrm{ww}} = 1$.

Step 5: The final step is to calculate the optimal weights ($w_1^*, w_2^*, \ldots, w_n^*$). The amount of inconsistency is reflected to $y_j^+ - y_j^-$ and $z_j^+ - z_j^-$ for indicating the preference of BO and OW. The objective function of LGPMBWM is also about minimizing total deviations. The

LGPMBWM model is presented as Eq. (3).

$$\min z = \sum_j (y_j^+ + y_j^-) + \sum_j (z_j^+ + z_j^-)$$

subject to:

$w_B - a_{Bj} w_j = y_j^+ - y_j^-$, for all j,

$w_j - a_{jw} w_w = z_j^+ - z_j^-$, forall j, (3)

$$\sum_j w_j = 1$$

$w_j, y_j^+, y_j^-, z_j^+, z_j^- \geq 0$, for all j.

Step 6: The consistency ratio can be calculated by Eq. (4) and (5). We also use Table 2 to obtain the consistency index. A value of consistency ratio close to zero indicates a high degree of consistency and vice versa.

$$\xi = \max_j \left\{ y_j^+ - y_j^-, z_j^+ + z_j^- \right\} \quad (4)$$

$$consistency\ ratio = \frac{\xi}{consistency\ index} \quad (5)$$

2.3 Proximity Index Value Method (PIV)

Proximity Index Value (PIV) method has been developed by [51]. This method has advantage of minimizing the rank reversal phenomenon over TOPSIS method. Further, the computational steps involved in PIV method are less as well as simpler than TOPSIS method. Owing to these advantages of PIV method which is a recent method, it has been used in this study. This method involves the following simple steps:

Step 1: Identify the available alternatives A_i (i = 1, 2, ..., m) and decision criteria C_j(j = 1, 2, ..., n)involved in the decision problem.

Step 2: Formulate the decision matrix *Y* by arranging alternatives in rows and criteria in columns as given in Eq.(6).

where, Y_{ij} represents i^{th} alternative performance value on j^{th}criterion, m is the number of alternatives, and n is the number of criteria.

Step 3: Normalized Performance value of each alternative for a given criterion against the performance of remaining alternatives is calculated by taking performance value of one alternative with respect to performances of all the alternatives for the same criterion and mathematically it can be expressed by Eq. (7).

$$y_{ij}^* = \frac{y_{ij}}{\sqrt{\sum_1^m y^2{}_{ij}}} \quad (7)$$

where, $y_{ij}^* =$ Normalized performance of i^{th} alternative ofj^{th} criterion, y_{ij} =performance value of i^{th}alternative on j^{th}criterion.

Step 4: In order to calculate the weighted normalized performance value, weight of the criterion has been calculated using Best-Worst method and then weighted normalized performance value of i^{th}alternative on the j^{th}criterion is calculated by Eq.(8).

$$V_{ij} = y_{ij}^* \times W_j \quad (8)$$

where $V_{ij}$is the weighted normalized value and $W_j$is the weight of criterion

Step 5: Evaluate the Weighted Proximity Index (WPI), u_i using Eq. (9) which is determined by taking the difference of weighted normalized value from its maximum/minimum value (v_{ijmax}, v_{ijmin}) in its range

$$u_i = \begin{Bmatrix} v_{ijmax} - v_{ij}; \text{for beneficial attributes} \\ v_{ij} - v_{ijmin};\ \text{for cost attributes} \end{Bmatrix} \quad (9)$$

Step 6: Overall proximity value (d_i) which determines the closeness of the available alternative w.r.t best alternative and given by taking the overall sum of all the weighted performance index value (u_i) given in Eq.(10).

$$d_i = \sum_{j=1}^{n} u_i \quad (10)$$

Step 7: Ranking of the alternatives on the basis of d_i values. The alternative with least value of d_i represents minimum deviation from the best and therefore, it is ranked first, followed by alternatives with increasing d_i.

3 Results and Discussion

3.1 Results of LGPMBWM-PIV Method

Decision matrix of this problem is shown in Table 1 which reveals twelve alternative natural fibers i.e. flax, hemp, jute, kenaf, ramie, okra, pine apple leaf fibre (PALF), coir, isora, cotton, banana and sisal and six decision criteria viz tensile strength (TS), Stiffness (S), failure strain (FS), density (D), thermal degradation temperature (TD) and percentage moisture gain (M).

In order to evaluate the importance and comparison of criteria with each other, Academic experts from various university and industry are interviewed and they were asked to select the best and worst criterion for Car roof

$$Y = [Y_{ij}]_{m\times n} = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1j} & \dots & Y_{1n} \\ Y_{21} & Y_{22} & \dots & \dots & \dots & Y_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{i1} & \dots & \dots & Y_{ij} & \dots & Y_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{m1} & \dots & \dots & Y_{mj} & \dots & Y_{mn} \end{bmatrix} \quad (6)$$

where, $i = 1, 2, \dots, m; j = 1, 2, \dots, n$

based on their experience related to automotive area. Thus, best and worst criteria identified by these expert are listed in Table 3.

Factors	Identified as 'Best' by Expert No.	Identified as 'Worst' by Expert No.
Tensile strength (TS)	4	5
Stiffness (S)	1,2,5,6	
Failure strain (FS)	3	2
Density (D)		
Degradation temperature (DT)		3,4,6
Moisture gain (M)		1

Expert No	TS	S	FS	D	DT	M	Consistency ratio
1	0.1952	0.3903	0.0781	0.1952	0.1301	0.0112	0.0554
2	0.1206	0.4824	0.0151	0.2412	0.0804	0.0603	0.0663
3	0.1842	0.1228	0.3684	0.1842	0.0175	0.1228	0.0403
4	0.5680	0.1136	0.1420	0.0811	0.0142	0.0811	0.0842
5	0.0239	0.5972	0.0853	0.0746	0.1194	0.0995	0.0731
6	0.1838	0.5234	0.0748	0.1308	0.0218	0.0654	0.0625
Average	0.2126	0.3716	0.1273	0.1512	0.0639	0.0734	0.0636

Table 6: Optimal weights of criteria and consistency ratio.

Alternative NFs	TS	S	FS	D	DT	M
Goal	(+)	(+)	(-)	(-)	(+)	(-)
F	0.4856	0.2883	0.0880	0.3167	0.3066	0.2736
H	0.3892	0.0873	0.1180	0.3084	0.3066	0.2736
J	0.2903	0.1274	0.0562	0.2865	0.2514	0.3875
K	0.2109	0.1121	0.1573	0.2917	0.2686	0.3875
R	0.3332	0.6711	0.0937	0.3021	0.3680	0.1938
O	0.1529	0.0791	0.0730	0.1834	0.2698	0.2964
P	0.0847	0.5458	0.1124	0.3167	0.2698	0.3192
C	0.0700	0.0527	0.8429	0.2605	0.2330	0.2964
I	0.2891	0.1802	0.2060	0.2855	0.2698	0.0274
CO	0.2490	0.0703	0.2622	0.3230	0.2698	0.1938
B	0.3735	0.2593	0.1105	0.2709	0.2453	0.3078
S	0.2291	0.1362	0.2997	0.2917	0.3680	0.3192

Table 7: Normalized decision matrix for Car roof NF's problem.

Table 3: Best and Worst Criteria Identified by Experts from 1 to 6.

Based on the expert interview, identification of best vector and its comparison with other vectors is done using Eq.(1) and pairwise comparison matrix is formulated as depicted in Table 4.

Expert No.	Best	TS	S	FS	D	DT	M
1	S	2	1	5	2	3	9
2	S	4	1	9	2	6	8
3	FS	2	3	1	2	9	3
4	TS	1	5	4	7	9	7
5	S	9	1	7	8	5	6
6	S	2	1	7	4	9	8

Table 4: Best to Others (BO) pairwise comparison matrix.

Similarly, Expert interview revealed the identification of worst factor and the comparison of other criteria with respect to worst factor using Eq.(2) is shown in Table 5.

Expert No.	1	2	3	4	5	6
Worst	M	FS	DT	DT	TS	DT
TS	7	8	8	9	1	8
S	9	9	7	8	9	9
FS	7	1	9	7	2	3
D	8	7	7	3	1	6
DT	7	3	1	1	5	1
M	1	2	6	3	4	2

Table 5: Others to Worst factor (OW) pairwise comparison matrix.

Based on the Pairwise comparison matrices shown in Table 4 and 5, final optimal weights and consistency are calculated using Eq. (3), (4) and (5) presented in Table 6. In the table, last row gives the final average optimal weights of all the criteria whereas last column gives consistency ratio.

Normalized decision matrix is formed by using Eq. (6), (7) and depicted in Table 7. In table (+) represents the beneficial criteria whereas (-) is for non-beneficial criteria.

Ranking of the NF's was done based on the proximity value by employing the Eq. (8), (9)& (10) as given in the Table 8.

Table 8 clearly shows that best available alternative material for the design of Car roof is ramie (R) natural fiber whereas worst choice for the Car roof material among all the NF's is Coir (C) fiber. There are plethora of reasons behind the ranking but some of the highlighted reasons can be (a) Maximum Stiffness (~ 76 GPa) of

	TS (+)	S (+)	FS (-)	D (-)	DT (+)	M (-)	Proximity Value	Rank
F	0.0000	0.1422	0.0041	0.0202	0.0039	0.0181	0.1884	3
H	0.0205	0.2169	0.0079	0.0189	0.0039	0.0181	0.2862	6
J	0.0415	0.2020	0.0000	0.0156	0.0074	0.0264	0.2930	7
K	0.0584	0.2077	0.0129	0.0164	0.0063	0.0264	0.3282	10
R	0.0324	0.0000	0.0048	0.0180	0.0000	0.0122	0.0673	1
O	0.0707	0.2200	0.0021	0.0000	0.0063	0.0197	0.3189	8
P	0.0852	0.0465	0.0072	0.0202	0.0063	0.0214	0.1868	2
C	0.0884	0.2298	0.1001	0.0117	0.0086	0.0197	0.4583	12
I	0.0418	0.1824	0.0191	0.0154	0.0063	0.0000	0.2650	5
CO	0.0503	0.2232	0.0262	0.0211	0.0063	0.0122	0.3394	11
B	0.0238	0.1530	0.0069	0.0132	0.0078	0.0206	0.2254	4
S	0.0545	0.1987	0.0310	0.0164	0.0000	0.0214	0.3221	9

Table 8: Final ranking of NF's obtained through PIV method.

Alternative	PIV	WASPAS	TOPSIS	EDAS	ROV	COPRAS
F	3	2	3	3	2	3
H	6	7	7	7	6	7
J	7	6	6	6	8	6
K	10	10	9	10	10	10
R	1	1	1	1	1	1
O	8	8	8	8	7	8
P	2	4	2	2	4	2
C	12	12	12	12	12	12
I	5	5	5	5	5	5
CO	11	11	11	11	11	11
B	4	3	4	4	3	4
S	9	9	10	9	9	9

Table 9: Ranking of alternative using PIV, WASPAS, EDAS, TOPSIS, ROV and COPRAS methods.

MCDM methods	WASPAS	TOPSIS	EDAS	ROV	COPRAS
Correlation coefficient	0.9860	0.9510	0.9860	0.9720	0.9930

Table 10: Correlation between ranks obtained using PIV method with other MCDM methods.

Ramie fiber whereas, Coir has lowest stiffness (~6GPa), Car roof is a critical part of car which is subjected to high aerodynamics forces which can be major reason behind the bending of car roof. Therefore, it is necessary that sustainable fiber being used for car roof should be highly stiff in order to withstand against all the bending forces acting on Car roof, (b) Tensile strength of Ramie fiber is ~*4 times greater than the tensile strength of Coir fiber which is also one of the major reason and (c) Natural fiber should have minimal moisture gain to obtained good interfacial integrity between NF reinforcement and the polymer matrix for successful fabrication of NFC. In this study, ramie fiber has lowest moisture gain ~8% whereas coir has high moisture gain therefore, ramie fiber is suitable choice for natural fiber reinforced composite being used for car roof. Properties of other fibers is lying in between ramie and coir fiber so they are ranked accordingly.

3.2 Comparison with other MCDM methods

Since, PIV is a newly developed method, the ranking results obtained from PIV method are compared with the ranks obtained with existing methods namely weighted aggregated sum product assessment (WASPAS), technique for order by preference by similarity to ideal solution (TOPSIS), Evaluation based on distance from average solution (EDAS), Range of value (ROV) and Complex proportional assessment (COPRAS) method. The details of these methods can be found in literature [52-57]. The ranking results are shown in Table 9.

It has been observed again from Table 9 that the best available alternative material for the design of car roof is ramie (R) natural fiber. Whereas, worst choice for the car roof material among all the NF's is Coir (C) fiber using all the six methods. Further, there is very few variation in the ranks of other NF's. Subsequently, correlation coefficient which is a statistical parameter used to measure relationship between two measures is computed for the ranks of NFs using different methods. The correlation

Criterion	Normal (0.3274)	Modified weights of all indicators when the weight of S is varied from 0.1 to 0.9								
TS	0.2126	0.3045	0.2707	0.2368	0.2030	0.1692	0.1353	0.1015	0.0677	0.0338
S	0.3716	0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000
FS	0.1273	0.1823	0.1621	0.1418	0.1215	0.1013	0.0810	0.0608	0.0405	0.0203
D	0.1512	0.2165	0.1925	0.1684	0.1444	0.1203	0.0962	0.0722	0.0481	0.0241
DT	0.0639	0.0915	0.0813	0.0712	0.0610	0.0508	0.0407	0.0305	0.0203	0.0102
M	0.0734	0.1051	0.0934	0.0818	0.0701	0.0584	0.0467	0.0350	0.0234	0.0117
Total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 11: Weights of Criteria in sensitivity analysis.

Alternative NF's	Normal	Run 0.1	Run 0.2	Run 03	Run 0.4	Run 0.5	Run 0.6	Run 0.7	Run 0.8	Run 0.9
F	3	2	2	2	3	3	3	3	3	3
H	6	4	6	6	6	7	7	8	9	9
J	7	6	7	7	7	6	6	6	6	7
K	10	10	10	10	10	10	9	9	8	8
R	1	1	1	1	1	1	1	1	1	1
O	8	7	8	8	8	9	10	10	10	10
P	2	8	5	3	2	2	2	2	2	2
C	12	12	12	12	12	12	12	12	12	12
I	5	5	4	5	5	5	5	5	5	5
CO	11	9	11	11	11	11	11	11	11	11
B	4	3	3	4	4	4	4	4	4	4
S	9	11	9	9	9	8	8	7	7	6

Table 12: Ranking of alternative materials after weight modifications.

coefficient of ranks obtained using PIV method with the rank obtained using other method is shown in Table 10.

It is observed from Table 10 there is a strong correlation (correlation coefficient is nearly 1) between the ranks obtained using PIV method and other MCDM methods. Hence, it can be concluded that the results obtained using PIV method are similar to that of other five MCDM methods. This also support the consistency and reliability of PIV method.

3.3 Sensitivity analysis

Sensitivity analysis has been done to ensure that the obtained results do not show any biasness and also to purge the effect of the highest weight criterion on other criteria considered in the present study. A methodology available in the literature [52-54] to carry out sensitivity analysis has been used in the present research where weights of all criteria have been varied in proportion to the weight of the highest ranked criterion. In the present study since the top ranked criterion is stiffness as its weight is maximum i.e. 0.3274 and therefore, its weight has been varied from 0.1 to 0.9 and the weights of all other criteria have been calculated as presented in Table 11.

In sensitivity analysis, the effect of changing weight on the ranking of alternative materials are observed as presented in Table 12.

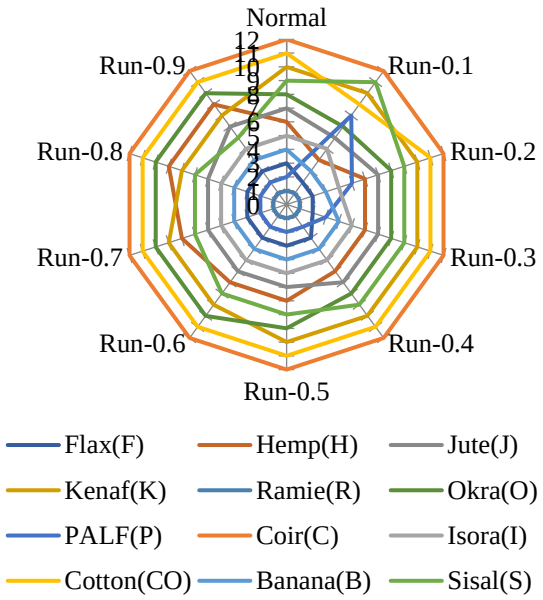


Figure 2: Ranking of alternatives in Sensitivity Analysis.

Table 12 clearly shows the ranking of alternative NF's by changing the weight of S, it can be clearly observed that on changing weight of S from 0.1 to 0.9 ramie acquired first rank and Coir is placed at last which shows the reliability of the LGPMBWM-PIV method. Thus, it is verified that ramie is best alternative NF for composite fabrication of car roof whereas coir is worst choice for the same . Further, ranking of alternatives on varying the

weight of criterion are shown in Figure 2 which clearly represent the variation of ranking of different alternatives.

4 Conclusion

In this study, selection of appropriate natural fibre for Car roof was successfully done by employing LGPMBWM-PIV method. The ranking sequences obtained from PIV method suggest Ramie fiber as the most suitable alternative NFs for fabrication of NFCs used in manufacturing of car roofs whereas coir fiber being worst alternative NF. This ranking sequence was also supported by ranking obtained through PIV, WASPAS, EDAS, TOPSIS, ROV and COPRAS methods. Sensitivity Analysis results show that for any weight modification of criteria, the ranking of alternatives obtained from hybrid LGPMBWM-PIV method remains same which support the consistency and reliability of the method.

5 Future work

Automotive industries are engaged with the manufacturing of several complex components and each components involves the material selection. Thus, as a future research direction, hybrid LGPMBWM-PIV MCDM method can also be applied to the selection of other automotive components which will minimize the cost and time involve in their production.

6 Acknowledgement

Authors want to thanks to all the academic expert chose from Jamia milia Islamia, Delhi Technical University, Indian Institute of Technology-Delhi, University of Malaya, Altus Muhedisliki for their cooperation, understanding and consideration of questionnaire.

7 References

[1] L. Chapman (1997). Transport and climate change: a review, Int. J. Environ. Pollut.,7(3), 327-342.

[2] H. Helms, U. Lambrecht (2017). The potential contribution of light-weighting to reduce transport energy consumption, Int. J. Life Cycle Assess. 12(1), 58-64. http://dx.doi.org/10.1065/lca2006.07.258

[3] J.P Howell (1996). The side load distribution on a Rover 800 saloon car under crosswind conditions. J. Wind Eng. and Ind. Aero. 60, 139-153. https://doi.org/10.1016/0167-6105(96)00029-3

[4] P.M. Samuel, K. Robert D. Christopher, M.B (2010). Eileen Mortality and injury patterns associated with roof crush in rollover crashes. Accident Analysis and Prevention, 42. https://doi.org/10.1016/j.aap.2010.02.013

[5] F.M. Al-Oqla, S.M. Sapuan (2014). Natural fiber reinforced polymer composites in industrial applications: feasibility of date palm fibers for sustainable automotive industry. J. Cleaner Prod., 66, 347-354. https://doi.org/10.1016/j.jclepro.2013.10.050

[6] F.M. AL-Oqla, S.M. Sapuan, M.R. Ishak, A.A. Nuraini (2015). A model for evaluating and determining the most appropriate polymer matrix type for natural fiber composites. Int. J. Polym. Anal. and Charact., 20(3), 191-205. https://doi.org/10.1080/1023666X.2015.990184

[7] F.M. Al-Oqla, S.M Sapuan, M.R Ishak, A.A. Nuraini (2016). A decision-making model for selecting the most appropriate natural fiber–Polypropylene-based composites for automotive applications. J. Compos. Mater., 50(4), 543-556. https://doi.org/10.1177/0021998315577233

[8] L.P. Maskepatil, A.U. Gandigude, S.A. Kale (2014). Selection of material for wind turbine blade by analytic hierarchy process (AHP) method. In Appl. Mech. Mater., 612,145-150. https://doi.org/10.4028/www.scientific.net/AMM.612.145

[9] M. Luqman, M.U. Rosli, C.Y. Khor, S. Zambree, H. Jahidi (2018). Manufacturing Process Selection of Composite Bicycle's Crank Arm using Analytical Hierarchy Process (AHP). In IOP Conf. Ser. Mater. Sci. Eng., 318 (1), 012058. https://doi.org/10.1088/1757-899X/318/1/012058

[10] L. Anojkumar, M. Ilangkumaran, V. Sasirekha (2014). Comparative analysis of MCDM methods for pipe material selection in sugar industry, Expert Systems with Applications 41(6), 2964-2980. https://doi.org/10.1016/j.eswa.2013.10.028

[11] K. Anupam, P.S. Lal, V. Bist, A.K. Sharma, V. Swaroop (2014). Raw material selection for pulping and papermaking using TOPSIS multiple criteria decision making design. Environ. Prog. Sustainable Energy, 33(3), 1034-1041. https://doi.org/10.1002/ep.11851

[12] F.M. Al-Oqla, S.M. Sapuan, M.R. Ishak, A.A. Nuraini (2015). Decision making model for optimal reinforcement condition of natural fiber composites. FiberPolym., 16(1), 153-163. https://doi.org/10.1007/s12221-015-0153-3

[13] Majumdar, B. Sarkar, P.K. Majumdar (2004). Application of analytic hierarchy process for the selection of cotton fibers. FiberPolym., 5(4), 297-302. https://doi.org/10.1007/BF02875528

[14] M.K. Ozturk, O.B. Berkalp, B. Nergis (2017). Design of a light weight fabric from natural cellulosic fibers with improved moisture related properties. In IOP Conf. Ser.: Mater. Sci. Eng. 254 (18), 182005. https://doi.org/10.1088/1757-899X/254/18/182005

[15] R. Mohammed, B.R. Reddy, S. Kakarla, B.B. Krishna, M.P. Khan (2017). Mechanical Characterization & TOPSIS Ranking of Glass Fiber Reinforced particulate filled Epoxy based Hybrid Composites. J. Chem. Pharm. Sci., Special, (2), 311-317.

[16] K. Jha, R. Kumar, K. Verma, B. Chaudhary, Y.K. Tyagi, S. Singh (2018). Application of modified TOPSIS technique in deciding optimal combination for bio-degradable composite. Vacuum, 157, 259-267. https://doi.org/10.1016/j.vacuum.2018.08.063

[17] S. Ahmad Fadli (2017). An integrated software quality model in a fuzzy analytical hierarchy process-based evaluation framework for e-learning software/Ahmad Fadli Saad(Doctoral dissertation, University of Malaya). http://studentsrepo.um.edu.my/id/eprint/8129

[18] N.M. Ishak, S.D. Malingam, M.R. Mansor (2016). Selection of natural fibre reinforced composites using fuzzy VIKOR for car front hood. Int. J. Mate. Product Technol., 53(3-4), 267-285.

[19] Chang, K. L. (2015). The use of a hybrid MCDM model for public relations personnel selection. Informatica, 26(3), 389-406.

[20] Chakraborty, S., Zavadskas, E. K. (2014). Applications of WASPAS method in manufacturing decision making. Informatica, 25(1), 1-20.

[21] Aliakbari Nouri, F., Khalili Esbouei, S., Antucheviciene, J. (2015). A hybrid MCDM approach based on fuzzy ANP and fuzzy TOPSIS for technology selection. Informatica, 26(3), 369-388.

[22] Keshavarz Ghorabaee, M., Zavadskas, E. K., Olfat, L., Turskis, Z. (2015). Multi-criteria inventory classification using a new method of evaluation based on distance from average solution (EDAS). Informatica, 26(3), 435-451.

[23] Ku, H., Wang, H., Pattarachaiyakoop, N., Trada, M. (2011). A review on the tensile properties of natural fiber reinforced polymer composites. Composites Part B: Engineering, 42(4), 856-873. https://doi.org/10.1016/j.compositesb.2011.01.010

[24] Saheb, D. N., Jog, J. P. (1999). Natural fiber polymer composites: a review. Advances in Polymer Technology: Journal of the Polymer Processing Institute, 18(4), 351-363. https://doi.org/10.1002/(SICI)1098-2329(199924)18:4<351::AID-ADV6>3.0.CO;2-X

[25] F. Yao, Q. Wu, Y. Lei, W. Guo, Y. Xu (2008). Thermal decomposition kinetics of natural fibers: activation energy with dynamic thermogravimetric analysis. Polym. Degrad. Stab. 93(1): 90-98. https://doi.org/10.1016/j.polymdegradstab.2007.10.012.

[26] S. Tsai (2018). Introduction to composite materials. Routledge.

[27] R. Latif, S. Wakeel, N.Z. Khan, A.N. Siddiquee, S.L. Verma Z.A. Khan (2018). Surface treatments of plant fibers and effects on mechanical properties of fiber-reinforced composites, J. Reinf. Plast. Compos. https://doi.org/10.1177/0731684418802022

[28] Pickering, K. L., Efendy, M. A., Le, T. M. (2016). A review of recent developments in natural fibre composites and their mechanical performance. Composites Part A: Applied Science and Manufacturing, 83, 98-112. https://doi.org/10.1016/j.compositesa.2015.08.038

[29] Akil, H., Omar, M. F., Mazuki, A. A. M., Safiee, S. Z. A. M., Ishak, Z. M., Bakar, A. A. (2011). Kenaf fiber reinforced composites: A review. Materials & Design, 32(8-9), 4107-4121. https://doi.org/10.1016/j.matdes.2011.04.008

[30] Thakur, V. K., Thakur, M. K., Gupta, R. K. (2014). raw natural fiber–based polymer composites. International Journal of Polymer Analysis and Characterization, 19(3), 256-271. https://doi.org/10.1080/1023666X.2014.880016

[31] Wang, W., Sain, M., Cooper, P. A. (2006). Study of moisture absorption in natural fiber plastic composites. Composites science and technology, 66(3-4), 379-386. https://doi.org/10.1016/j.compscitech.2005.07.027

[32] Athijayamani, A., Thiruchitrambalam, M., Natarajan, U., Pazhanivel, B. (2009). Effect of moisture absorption on the mechanical properties of randomly oriented natural fibers/polyester hybrid composite. Materials Science and Engineering: A, 517(1-2), 344-353. https://doi.org/10.1016/j.msea.2009.04.027

[33] Robertson, N. L. M., Nychka, J. A., Alemaskin, K., Wolodko, J. D. (2013). Mechanical performance and moisture absorption of various natural fiber reinforced thermoplastic composites. Journal of applied polymer science, 130(2), 969-980. https://doi.org/10.1002/app.39237

[34] R. Kumar, Ray (2015). A Selection of material under conflicting situation using simple ratio optimization technique. In: Proceedings of fourth international conference on soft computing for problem solving, advances in intelligent systems and computing, 335, 513–519. https://doi.org/10.1007/978-81-322-2217-0_42

[35] Komuraiah, N.S. Kumar, B.D. Prasad (2014). Chemical composition of natural fibers and its influence on their mechanical properties. Mech. Compos. Mater., 50(3), 359-376. https://doi.org/10.1007/s11029-014-9422-2

[36] S.P.S. Tita, R. Mederios, J.R. Tarpani (2018). Chemical modification of sugarcane baggase and sisal fibers using hydroxylmethylated lignin: Influence on Impact strength and water absorption of phenolic composites J. Compos. Mater. 52 (20), 2743-2753. https://doi.org/10.1177/0021998317753886

[37] M. Kracka, W.K.M. Brauers, E.K. Zavadskas (2010). Ranking heating losses in a building by applying the Multimoora. Eng. Econ. 21(4), 352–359.

[38] B.W. Rosen (1973). Stiffness of fibre composite materials. Composites, 4(1), 16-25. https://doi.org/10.1016/0010-4361(73)90291-7

[39] C. Scarponi, C.S. Pizzinelli (2009). Interface and mechanical properties of natural fibres reinforced composites: a review. International J. Mater. Product Technol., 36(1-4), 278-303.

[40] Amiri, M., Emamat, M. S. M. M. (2020). A Goal Programming Model for BWM. Informatica, 31(1), 21-34. https://doi.org/10.15388/20-INFOR389

[41] Rezaei J (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. Omega; 64, 126-130. https://doi.org/10.1016/j.omega.2014.11.009

[42] Abadi F., Sahebi I., Arab A., Alavi A., Karachi H (2018). Application of best-worst method in evaluation of medical tourism development strategy. Decision Science Letters; 7(1), 77-86. 10.5267/j.dsl.2017.4.002

[43] Shojaei P., Haeri S. A. S., Mohammadi S (2018). Airports evaluation and ranking model using Taguchi loss function, best-worst method and VIKOR technique. Journal of Air Transport Management; 68, 4-13. https://doi.org/10.1016/j.jairtraman.2017.05.006

[44] Salimi N., Rezaei J (2018). Evaluating firms' R&D performance using best worst method. Evaluation and program planning; 66, 147-155. https://doi.org/10.1016/j.evalprogplan.2017.10.002

[45] Gupta H (2018). Evaluating service quality of airline industry using hybrid best worst method and VIKOR. Journal of Air Transport Management; 68, 35-47. https://doi.org/10.1016/j.jairtraman.2017.06.001

[46] Ahmad W. N. K. W., Rezaei J., Sadaghiani S., Tavasszy L. A . (2017). Evaluation of the external forces affecting the sustainability of oil and gas supply chain using Best Worst Method. J. Cleaner Prod.; 153, 242-252. https://doi.org/10.1016/j.jclepro.2017.03.166

[47] Gupta H., Barua M. K. (2017). Supplier selection among SMEs on the basis of their green innovation ability using BWM and fuzzy TOPSIS. J. Cleaner Prod; 152, 242-258. https://doi.org/10.1016/j.jclepro.2017.03.125

[48] van de Kaa G., Kamp L., Rezaei, J (2017). Selection of biomass thermochemical conversion technology in the Netherlands: A best worst method approach. J. Cleaner Prod.; 166, 32-39. https://doi.org/10.1016/j.jclepro.2017.07.052

[49] Rezaei J., Hemmes A., Tavasszy L (2017). Multi-criteria decision-making for complex bundling configurations in surface transportation of air freight. Journal of Air Transport Management; 61, 95-105. https://doi.org/10.1016/j.jairtraman.2016.02.006

[50] Rezaei, J (2015). Best-worst multi-criteria decision-making method. Omega; 53, 49-57. https://doi.org/10.1016/j.omega.2014.11.009

[51] Mufazzal S., Muzakkir S. M. (2018). A new multi-criterion decision making (MCDM) method based on proximity indexed value for minimizing rank reversals. Computers & Industrial Engineering, 119, 427-438.https://doi.org/10.1016/j.cie.2018.03.045.

[52] Prakash C., Barua M. K (2015). Integration of AHP-TOPSIS method for prioritizing the solutions of reverse logistics adoption to overcome its barriers under fuzzy environment. Journal of Manufacturing Systems; 37, 599-615. https://doi.org/10.1016/j.jmsy.2015.03.001

[53] Mangla S. K., Kumar P., Barua M. K (2015). Risk analysis in green supply chain using fuzzy AHP approach: a case study. Resour., Conserv. Recycl.; 104, 375-390. https://doi.org/10.1016/j.resconrec.2015.01.001

[54] Triantaphyllou E (2000). A Sensitivity Analysis Approach for MCDM Methods. In Multi-criteria Decision Making Methods: A Comparative Study (pp. 131-175), Springer; Boston, MA. https://doi.org/10.1007/978-1-4757-3157-6_8

[55] Wakeel, S., Bingol, S., Bashir, M. N., & Ahmad, S. (2020). Selection of sustainable material for the manufacturing of complex automotive products using a new hybrid Goal Programming Model for Best Worst Method–Proximity Indexed Value method. Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications, 1464420720966347. https://doi.org/10.1177/1464420720966347

[56] Wakeel, S., Ahmad, S., Bingol, S., Bashir, M. N., Paçal, T. C., & Khan, Z. A. (2020, August). Supplier Selection for High Temperature Die Attach by hybrid Entropy-Range of Value MCDM Technique: A Semiconductor Industry. In 2020 21st International Conference on Electronic Packaging Technology (ICEPT) (pp. 1-5). IEEE. 10.1109/ICEPT50128.2020.9202994.

[57] Ahmad, S , Bingöl, S , Wakeel, S . (2020). A hybrid multi-criteria decision making method for robot selection in flexible manufacturing system. Middle East Journal of Science , 6 (2) , 68-77 . https://doi.org/10.23884/mejs.2020.6.2.03

https://doi.org/10.31449/inf.v45i1.3111 Informatica **45** (2021) 117–125 **117**

A Comparative Analysis of Machine Learning Algorithms to Build a Predictive Model for Detecting Diabetes Complications

Ali A. Abaker
Department of Accounting Information Systems, Faculty of Computer Science
Al-Neelain University, Khartoum, Sudan
E-mail: aliabdallah@neelain.edu.sd

Fakhreldeen A. Saeed
Department of Software Engineering, Faculty of Computer Science, Al-Neelain University, Khartoum, Sudan
E-mail: fasaeed@neelain.edu.sd

Keywords: diabetes complications, machine learning, logistic regression, electronic health records, feature selection.

Received: April 4, 2020

Diabetes complications have a significant impact on patients' quality of life. The objective of this study was to predict which patients were more likely to be in a complicated health condition at the time of admission to allow for the early introduction of medical interventions. The data were 644 electronic health records from Alsukari Hospital collected from January 2018 to April 2019. We used the following machine learning methods: logistic regression, random forest, and k-nearest neighbor (KNN). The logistic regression algorithm performed better than the other algorithms achieving an accuracy of 81%, recall of 81%, and F1 score of 75%. Also, attributes such as infection years, swelling, diabetic ketoacidosis, and diabetic septic foot were significant in predicting diabetes complications. This model can be useful for the identification of patients requiring additional care to limit the complications and help practitioners in making decisions on whether the patient should be hospitalized or sent home. Furthermore, we used the sequential feature selection (SFS) algorithm which reduced the features to six, which is fewer than any model built before to predict diabetes complications. The primary goal of this study was achieved. The model had fewer attributes which means we have a simple and understandable model in addition to, it has a better performance.

Povzetek: Podana je analiza metod strojnega učenja za napovedovanje komplikacij pri sladkorni bolezni.

1 Introduction

Diabetes is a chronic disease and considered a serious global health challenge[1]. Diabetes Complications happen when diabetes is uncontrolled. That leads to serious health problems; the patient could suffer from a diabetic coma or die from a heart attack or stroke. The number of people with diabetes has risen to 422 million in 2014 which was about 17% of the world population[2]. More than 68% of diabetes-related deaths are caused by diabetes complications[3]. About 25% of people with diabetes are undiagnosed in the United States[4]. To address diabetic complications, doctors need to be allowed to identify and monitor patients at risk of complications[5]; [6]. Early discovery can prevent or delay diabetes-related complications and allow for effective intervention at both the individual and population levels, which are desperately needed to slow the diabetes epidemic[7].

The objective of this study was to predict which patients were more likely to be in a complicated health condition at the time of admission, which helps for the early introduction of medical interventions. We used the following machine learning methods: logistic regression, random forest, and KNN. The dataset was collected from Alsukari Hospital for building the machine learning model. The final model used the most six significant attributes and 644 records. This model is useful for predicting the likelihood that the patient should be hospitalized or sent home.

The rest of this paper is organized into Sections detailed as follows: section 2 literature review, data collection in section 3, feature selection in section 4, sequential feature selection section 5, model design in section 6, evaluation metrics section 7, results in section 8, section 9 conclusion, study limitations in section 10, last future work section 11.

2 Literature review

Various traditional methods, based on the physical and chemical tests are available for diagnosing diabetes. However, methods based on data mining techniques can be effectively implemented[8]. The authors agreed on the importance of developing machine learning algorithms to learn patterns and decision rules from data[9]. Although, many studies were conducted to assess the main causes of diabetes mellitus. But, a few were directed to discover the clinical risk factors[10]. A machine learning model was built to predict wound complications and mortality[11].

Electronic health records were used for many studies related to diabetes[12]; [13]. A method that enables risk assessment from electronic health records(EHR) on a large population, they also added administrative claims and pharmacy records[14]. Another study proposed a model that predicts the severity as a ratio interpreted as the impact of diabetes on different organs of the human body, the algorithm estimated the severity on different parts of the body like the heart and kidney[15]. A rapid model for glucose identification and prediction based on the idea of model migration[16]. Despite the data was collected from different sources and places but, some attributes were used in many studies such as age, gender, body mass index(BMI), glucose, blood pressure, time of diagnosis, and smoking [9]; [17]; [18]; [11].

Algorithms	2015 - 2019
ANN	[8],[19],[2],[20],[21],[11]
K-means	[22]
Logistic	[22],[9],[2],[20],[23],[24],[11]
Decision Tree	[1],[20],[3],[25],[24], [10],[26]
SVM	[27],[2],[1],[20],[3],[21],[24]
KNN	[1],[21],[24]
Random Forest	[1],[24]
Naive Bayes	[3],[23],[25],[24]
Designed algorithms	[28],[13],[5],[29],[9]

Table 1: Shows the trends of used algorithms in previous literature.

Table 1 above is the review of the popular algorithms used in diabetes studies related to machine learning from 20015 to 2019. And the figure below demonstrates the trend.

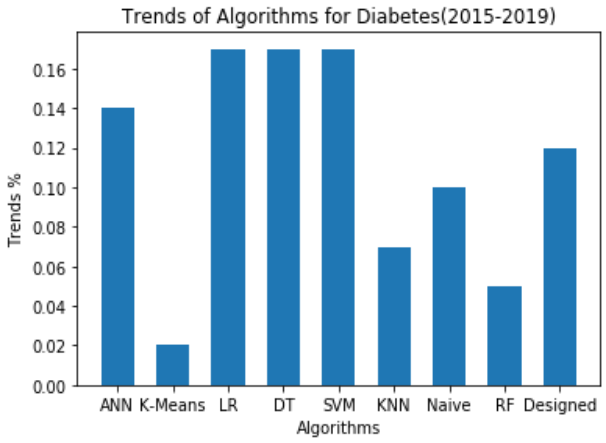


Figure 1: Shows the trend of algorithms used in the last five years.

According to the figure above, logistic regression (LR), decision tree (DT), and supervised machine learning (SVM) are the most popular algorithms used for diabetes studies among the research community with about 17%. Followed by, the artificial neural network (ANN) with 14%. According to, literature review researchers designed about 12% out of all used algorithms as new methods for diabetes problems.

We built a machine learning model to predict diabetes complications, using six attributes. This model introduced new attributes such as diabetic ketoacidosis, swelling, infection years, and diabetic septic foot were found to be significant. These attributes were not included in the previously mentioned studies[30]; [13]; [31]; [9]. Also, in this paper, we investigated five performance metrics such as F1 and recall.

A logistic regression model was used to assess the factors associated with glycemic control. The model indicated that patients older than 65 years old were more likely to have complications compared to the younger[32]. Demographic and treatment data were collected and logistic regression was used to predict complications[33]. Another study was conducted to predict 30-day complication rate using random forest and logistic regression, the analysis showed that age is the most significant attribute in predicting complications[34]. Random forest and simple logistic regression methods showed the best performance compared to the evaluated algorithms[35]. Diabetes complications prediction model was based on similarity measure. first, they assessed the similarity between textual medical records after data cleaning, then topic mining is conducted, and last building the model[36].

3 Data collection

The dataset was collected from Alsukari Hospital. Ethical approval to use the data for research was obtained both from the Ministry of Health (MOH) and the hospital. The dataset contained 29 attributes and 644 records of diagnosed diabetes patients who were admitted to the hospital in the period from January 2018 to April 2019.

4 Feature selection

Classification problems usually have a big number of features in the dataset, but not all of them are significant for classification. Irrelevant features may reduce the performance and even complicate the model. Feature selection aims to select a small number of relevant features to get similar or better classification performance than using a larger number of features[39].

Thus, it is usual to apply a preprocessing step to remove irrelevant features and reduce the dimensionality of the data[40]. The selection of the features can lead to an improvement of the learning algorithm, either in terms of learning speed, generalization, or simplicity of the model. Furthermore, there are other advantages associated with a reduced feature: low cost, clear model, and a better understanding of the domain knowledge[41].

The figure below illustrates the five approaches for feature selection.

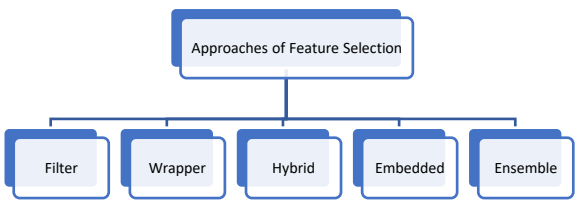


Figure 2: Shows feature selection approaches.

Attribute	Type	Scale	Description
Infection years	Continuous	0 – 35	Is the period from the patient was diagnosed with Diabetes
Sugar	Continuous	53 - 681	Blood sugar or glucose level in the body is measured in mg/dL or milligrams per deciliter.
Swelling	Category	Binary	Swollen body part.
DKA	Category	Binary	Diabetic ketoacidosis called [22]ketones.
DSF	Category	Binary	Diabetic septic Foot is a foot affected by ulceration of the lower limb in a patient with diabetes[37].
HR	Continuous	27 - 139	Is the speed of the heartbeat measured by the number of contractions (beats) of the heart per minute (bpm)[38].
Class	Binary	Binary	The Target variable.

Table 2: Shows the selected attributes and their descriptions to build the model.

4.1 Filter method

Filter based feature methods evaluate features as an individual assessment. Therefore, these methods first assign a score value for each feature using one of the statistical criteria, and then, all features are sorted according to the scores. Then, they select top-n features with the highest score as the final step[42]. Filter feature selection algorithms are useful due to their simplicity and fast speed. A common filter is to use mutual information to evaluate the relationships between each feature and the class variable[43].

4.2 Wrapper method

Feature selection algorithms are divided into two methods. Firstly, it depends on the outcome of the selection algorithm: whether it returns a subset of significant features or an ordered ranking of all the features, recognized as feature ranking. Secondly, feature selection methods are divided into three approaches based on the relationship between a feature selection algorithm and the learning method, which is used for building the model: filters, which rely on overall characteristics of the dataset and are independent of the learning algorithm; wrappers, which use the prediction of a classifier to estimate subsets of features; and embedded methods, which perform the selection in the process of training and are specific to the given learning algorithm.

Wrapper techniques depend on the classification algorithm, which is used to evaluate the subsets of features, but they are more expensive when it comes to computation[44]. Despite this weakness, they often provide better outcomes; wrappers are used widely in many applications[45]. Especially, in healthcare where we care about the accuracy more than the performance of the algorithm in many situations and allow for implementation in real-time systems when we have fewer attributes[46].

4.3 Hybrid method

Recently, researchers are concentrating on developing novel hybrid feature selection methods as they speed up the removal of irrelevant features and give greater classification accuracy compared to other methods[47]. Though various techniques were developed for selecting the perfect subset of features, these methods faced some problems such as instability, high processing time, and selecting a semi-optimal solution as a final result. In other words, they have not been able to fully extract the effective features. Hybrid methods were introduced as a solution to overcome the weaknesses of using a single algorithm[48].

4.4 Embedded method

Embedded feature selection is related to classification algorithms. This relation in embedded methods is stronger than that in wrapper methods. Embedded methods are a sort of combination of filter and wrapper methods[49]. By, embedding feature selection into the model learning. They return both the learned model and selected features and are frequently used for classification[50]. Inserting the feature selection step into the training process can improve the performance of the model.

4.5 Ensemble method

Ensemble learning is an effective method for machine learning. The objective is to attain better learning accuracy by combining different learning models[41]. Ensemble methods are better than using a single machine learning model. Recently, the development of ensemble feature selection is increasingly getting attention[51].

Combined feature selection aims to find multiple optimal features. The advantage of integration technology would produce a stable and efficient method; especially, with high-dimensional data[52].

5 Sequential feature selection (SFS)

The sequential feature selection(SFS) algorithm begins with a blank set and increases one feature for the first step which gives the best value for the model. On the second step onwards the remaining features are added separately to the existing subset and the new subset is assessed. The new feature is permanently added to the subset if it gives the maximum classification accuracy. The process is repeated until the required number of features are added[53]. SFS method has the advantage of improving the prediction performance of the classifier by excluding any characteristic that reduces the performance[54].

6 Model design

Working with methods for reasoning under uncertainty is now one of the most interesting areas of machine learning [20]; [21]. Machine learning has been used for several decades to tackle a broad range of problems in many fields of applications[57]. The machine learning model was built on the historical data using different algorithms for each model; we evaluated the results, and assessed the model accuracy on the testing data. Six attributes were selected out of 29. The dataset was divided into two parts: training and testing set consisting of 70 and 30 percent respectively. The features were selected to be used in the final model based on the training set which achieved the highest accuracy of 86%, with the six attributes as shown in table 2. Logistic regression, random forest, and KNN were selected because of their simplicity and good predictive capability. However, machine learning models are more accurate than normal statistical methods[58].

6.1 Logistic regression

Logistic regression models are a sort of widespread linear model that is used for datasets where the dependent variable is categorical[59]. The logistic model is used to estimate the probability of the response variable based on one or more predictor variables. Logistic regression is used when the dependent variable is categorical and generates output in terms of probabilities[60]. Logistic regression is an effective prediction algorithm. Its applications are efficient when the dependent variable of a dataset is binary[61].

$$f(x) = \frac{L}{1+e^{-k(x-x_0)}} \quad (1)$$

Where

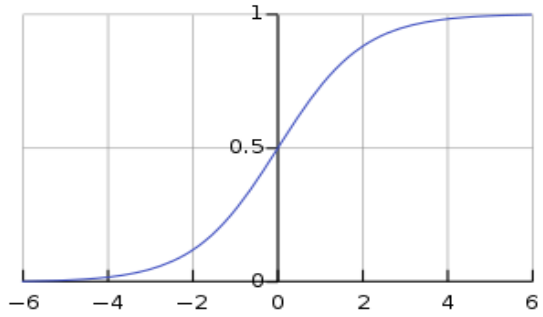


Figure 3: Shows logistic regression curve.

X_0 = the x value of the sigmoid's midpoint.

L = the curve's maximum value.

K = the logistic growth rate or steepness of the curve.

6.2 Random forest

Random Forest was proposed by Dr. Breiman in 2001. It is typically used in classification[62]. It is an algorithm based on statistical learning theory, which uses a bootstrap randomized re-sampling method to extract multiple versions of the sample sets from the original training datasets. Then it builds a decision tree model for each sample set, and finally combines all the results of the decision trees to predict via a voting mechanism[63].

Suppose we have the dataset D = {(x1, y1) … (xn, yn)} and the aim is to find the function f : X →Y where X is the inputs and Y is the produced outputs. Furthermore, let M be the number of inputs. Random forest randomly selects n observations from D with replacement to a bootstrap sample. Each tree is grown using a subset of m features from the overall M features. For regression, it is recommended to set the subset of features at M=3. Then at each node, m features are nominated at random and the best performing split among the M features is selected according to the impurity measure (Gini impurity). The trees are grown to a maximum depth without pruning[64].

6.3 K-nearest neighbor (KNN)

It is a classification algorithm that classifies data based on similarity measure or distance measure[18]. This algorithm can be used in both classification and regression problems[21], [65]. KNN classifies an instance by finding its nearest neighbors[66]. The KNN classifier applies the Euclidean distance or cosine similarity for differentiating the training tuple and test tuples. The same Euclidean distance between tuples Xi and Xt (t = 1,2,3 …n) can be explained as[67]:

$$d(yi, yt) = \sqrt{(y_{i1} - y_{t1})^2 + (x_{i2} - x_{t2})^2 + \cdots + (x_{is} - x_{ts})^2} \quad (2)$$

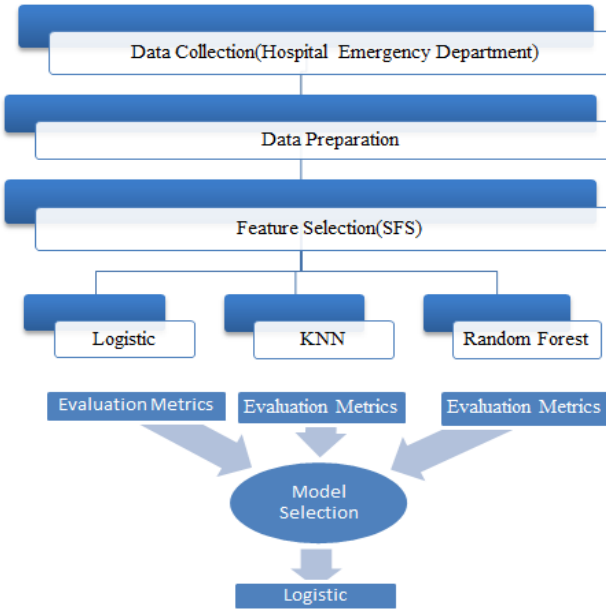


Figure 4: Shows the model selection process.

where yi, n and s be the tuple constant. It can be written as:

$$Dist(y1, y2 = \sqrt{\sum_{i=1}^{n}(y_{1i} - y_{2i})^2} \qquad (3)$$

This equation is prepared based on KNN algorithms, every neighboring point that is closest to the test tuple, which is encapsulated and on the nearby space to the test tuple[68].

7 Evaluation metrics

The three algorithms, Logistic Regression, Random Forest Classifier, and KNN were compared in terms of accuracy, recall, specificity, precision, and F1scores as demonstrated below. The level of efficiency of the classification model is measured with the number of correct and incorrect classifications in each potential value of the variables being classified. From the outcomes gained. The following equations are used to measure the Accuracy, Sensitivity, and Specificity, Precision, and F1 score [69].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (4)$$

Recall or also known as sensitivity refers to the percentage of total relevant results correctly classified by your algorithm.

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (5)$$

Specificity is defined as the proportion of actual negatives, which got predicted as the negative (or true negative).

$$\text{Specificity} = \frac{TN}{TN+FP} \qquad (6)$$

The precision of all the records we predicted positive.

Algorithm/ Metrics	Accuracy	Recall	Specificity	Precision	F1
Logistic Regression	81%	81%	81%	70%	75%
Random Forest Classifier	78%	57%	89%	74%	64%
KNN	76%	62%	84%	68%	65%

Table 3: Shows the three algorithms and their metrics scores using the default threshold.

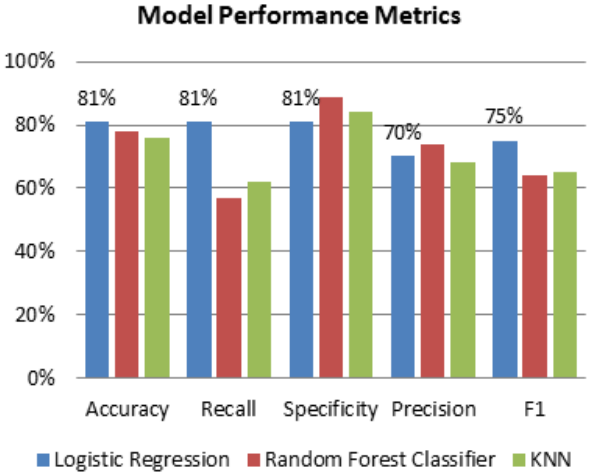


Figure 5: Shows model performance metrics.

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (7)$$

F1-score, which is simply the harmonic mean of precision and recall.

$$\text{F1 Score} = \frac{2TP+TN}{2TP+FP+FN} \qquad (8)$$

From table 3 above the logistic regression achieved the highest accuracy of 81%. According to the medical objective, the model was designed to be more sensitive in predicting true positive which was calculated by recall and F1 of 81 and 79 percent respectively. As this model is used in healthcare our interest in predicting the positive class, which is more important. It is acceptable for the model to fall into a false positive error (type 1 error). But it is very costly for the model to commit the (false negative error) type 2 error, it means that the patient might be in complicated health status and needs special and immediate medical care and intervention, but the model tells us that the patient will not be in a complicated situation. And that could result in more health complications and even the patient's life.

8 Results

New attributes were found to be significant in predicting diabetes complications such as infection years, swelling, diabetic ketoacidosis, diabetic septic foot, which were found to be vital in predicting diabetes complications.

But, blood pressure, body temperature, cholesterol, protein level, oxygen level, and gender were not significant in predicting diabetes complications because they were excluded by the feature selection algorithm. We compared the following three algorithms: logistic regression, random forest classifier, and KNN to find the best algorithm for predicting diabetes complications. Accuracy, recall, specificity, precision, and F1; used as performance metrics. These metrics were used for selecting the best model. The logistic regression algorithm achieved the highest recall score of 81%, followed by KNN and random forest of 62% and 57% respectively, as shown in Table 3 above. This means that the model is more sensitive in predicting the positive class. Also, the model achieved the highest F1score of 75% as shown in Table 3 above. The model was designed to be more sensitive in predicting true positive class which means the diabetes status is complicated. It was calculated by the recall score of 81%. As this model is used in healthcare our interest in predicting the positive class, which is significant.

9 Conclusion

In this paper, we created a dataset from Alsukari Hospital for building our machine learning model. The best model was built using six out of 29 attributes. Three algorithms were compared in selecting the best model as follows: Logistic regression, random forest, and KNN. Furthermore, new attributes were investigated and included in the model. Finally, the best accuracy was obtained using logistic regression. The overall accuracy does not guarantee that the model will perform better and serve the specific domain interest. According to the medical objective, the recall score was used besides general accuracy. The higher recall score indicates that the model is more sensitive in predicting positive cases or medically patients with diabetes complications.

10 Study limitations

The accuracy is not very high as we are working in the medical field higher accuracy is needed. Second, the size of the dataset is small; machine learning models need more data for producing stable and well-trained models.

11 Future work

There are several future research directions. Firstly, for predicting diabetes complications more features could be included which were not included in this study. Secondly, more work should be directed toward identifying the risk factors associated with diabetes complications. Last we are interested in adopting this model to other chronic diseases.

12 Acknowledgement

The authors would like to thank the Ministry of Health (MOH) and Alsukari Hospital for giving us the ethical approval to use the data for this work. Also, we would like to take this opportunity to thank the reviewers for their constructive comments towards improving our manuscript.

13 Conflict of interest

The authors declare no conflict of interest.

References

[1] J. P. Kandhasamy and S. Balamurali, “Performance Analysis of Classifier Models to Predict Diabetes Mellitus,” *Procedia - Procedia Comput. Sci.*, vol. 47, pp. 45–51, 2015. https://doi.org/10.1016/j.procs.2015.03.182

[2] S. Malik, R. Khadgawat, S. Anand, and S. Gupta, “Non-invasive detection of fasting blood glucose level via electrochemical measurement of saliva,” *SpringerPlus*, vol. 5, no. 1. 2016. https://doi.org/10.1186/s40064-016-2339-6

[3] D. Sisodia and D. S. Sisodia, “ScienceDirect Prediction of Diabetes using Classification Algorithms,” *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1578–1585, 2018. https://doi.org/10.1016/j.procs.2018.05.122

[4] A. E. Anderson, W. T. Kerr, A. Thames, T. Li, J. Xiao, and M. S. Cohen, “Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: A cross-sectional, unselected, retrospective study,” *J. Biomed. Inform.*, vol. 60, no. December, pp. 162–168, 2016. https://doi.org/10.1016/j.jbi.2015.12.006

[5] A. Anand and D. Shakti, “Prediction of diabetes based on personal lifestyle indicators,” *Proc. 2015 1st Int. Conf. Next Gener. Comput. Technol. NGCT 2015*, no. September, pp. 673–676, 2016. https://doi.org/10.1109/NGCT.2015.7375206

[6] G. Peddinti *et al.*, “Early metabolic markers identify potential targets for the prevention of type 2 diabetes,” *Diabetologia*, vol. 60, no. 9, pp. 1740–1750, 2017. https://doi.org/10.1007/s00125-017-4325-0

[7] T. P. A. Debray, Y. Vergouwe, H. Koffijberg, D. Nieboer, E. W. Steyerberg, and K. G. M. Moons, “ORIGINAL ARTICLES A new framework to enhance the interpretation of external validation studies of clinical prediction models,” *J. Clin. Epidemiol.*, vol. 68, no. 3, pp. 279–289, 2015. https://doi.org/10.1016/j.jclinepi.2014.06.018

[8] M. Komi, J. Li, Y. Zhai, and Z. Xianguo, “Application of data mining methods in diabetes prediction,” *2017 2nd Int. Conf. Image, Vis. Comput. ICIVC 2017*, no. S Ix, pp. 1006–1010, 2017. https://doi.org/10.1109/ICIVC.2017.7984706.

[9] A. Dagliati *et al.*, “Machine Learning Methods to Predict Diabetes Complications,” 2017. https://doi.org/doi: 10.1177/1932296817706375

[10] Purushottam, K. Saxena, and R. Sharma, “Diabetes mellitus prediction system evaluation using C4.5 rules and partial tree,” *2015 4th Int. Conf. Reliab.*

Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2015, pp. 1–6, 2015.
https://doi.org/10.1109/ICRITO.2015.7359272

[11] J. S. Kim *et al.*, "Examining the Ability of Artificial Neural Networks Machine Learning Models to Accurately Predict Complications Following Posterior Lumbar Spine Fusion," *Spine (Phila. Pa. 1976).*, vol. 43, no. 12, pp. 853–860, 2018.
https://doi.org/10.1097/BRS.0000000000002442

[12] M. Kumar, N. K. Rath, A. Swain, and S. K. Rath, "Feature Selection and Classification of Microarray Data using MapReduce based ANOVA and K-Nearest Neighbor," *Procedia Comput. Sci.*, vol. 54, pp. 301–310, 2015.
https://doi.org/10.1016/j.procs.2015.06.035

[13] B. Liu, Y. Li, Z. Sun, S. Ghosh, and K. Ng, "Early Prediction of Diabetes Complications from Electronic Health Records : A Multi-Task Survival Analysis Approach," pp. 101–108.

[14] N. Razavian, S. Blecker, A. M. Schmidt, A. Smith-mclallen, S. Nigam, and D. Sontag, "Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors," vol. 3, no. 4, 2015.
https://doi.org/10.1089/big.2015.0020

[15] V. R. Balpande and R. D. Wajgi, "Prediction and severity estimation of diabetes using data mining technique," *IEEE Int. Conf. Innov. Mech. Ind. Appl. ICIMIA 2017 - Proc.*, no. Icimia, pp. 576–580, 2017.
https://doi.org/10.1109/ICIMIA.2017.7975526

[16] C. Zhao and C. Yu, "Rapid model identification for online subcutaneous glucose concentration prediction for new subjects with type i diabetes," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, pp. 1333–1344, 2015.
https://doi.org/10.1109/TBME.2014.2387293

[17] O. Geman, I. Chiuchisan, and R. Toderean, "Application of Adaptive Neuro-Fuzzy Inference System for diabetes classification and prediction," *2017 E-Health Bioeng. Conf. EHB 2017*, no. Dm, pp. 639–642, 2017.
https://doi.org/10.1109/EHB.2017.7995505

[18] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," *J. Big Data*, vol. 6, no. 1, 2019.
https://doi.org/10.1186/s40537-019-0175-6

[19] S. Joshi and M. Borse, "Detection and prediction of diabetes mellitus using back-propagation neural network," *Proc. - 2016 Int. Conf. Micro-Electronics Telecommun. Eng. ICMETE 2016*, pp. 110–113, 2016.
https://doi.org/10.1109/ICMETE.2016.11

[20] H. Y. Tsao, P. Y. Chan, and E. C. Y. Su, "Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms," *BMC Bioinformatics*, vol. 19, no. Suppl 9, 2018.
https://doi.org/10.1186/s12859-018-2277-0

[21] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Appl. Comput. Informatics*, no. December, 2019.
https://doi.org/10.1016/j.aci.2018.12.004

[22] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, "Type 2 diabetes mellitus prediction model based on data mining," *Informatics Med. Unlocked*, vol. 10, pp. 100–107, 2018.
https://doi.org/10.1016/j.imu.2017.12.006

[23] B. J. Lee and J. Y. Kim, "Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on Machine Learning," *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 1, pp. 39–46, 2016.
https://doi.org/10.1109/JBHI.2015.2396520

[24] T. Zheng *et al.*, "A Machine Learning-based Framework to Identify Type 2 Diabetes through Electronic Health Records," *Int. J. Med. Inform.*, 2016.
https://doi.org/10.1016/j.ijmedinf.2016.09.014

[25] P. Songthung and K. Sripanidkulchai, "Improving type 2 diabetes mellitus risk prediction using classification," *2016 13th Int. Jt. Conf. Comput. Sci. Softw. Eng. JCSSE 2016*, 2016.
https://doi.org/10.1109/JCSSE.2016.7748866

[26] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," *Procedia Comput. Sci.*, vol. 82, no. March, pp. 115–121, 2016.
https://doi.org/10.1016/j.procs.2016.04.016

[27] J. Zhang *et al.*, "Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images," *Biomed Res. Int.*, vol. 2017, 2017.
https://doi.org/10.1155/2017/7961494

[28] J. Li *et al.*, "Feature selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, 2017.
https://doi.org/10.1145/3136625

[29] K. Zarkogianni, M. Athanasiou, and A. C. Thanopoulou, "Comparison of Machine Learning Approaches Toward Assessing the Risk of Developing Cardiovascular Disease as a Long-Term Diabetes Complication," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 5, pp. 1637–1647, 2018.
https://doi.org/10.1109/JBHI.2017.2765639

[30] L. Liu, "Forecasting Potential Diabetes Complications," 2014.

[31] G. Huzooree, "Glucose Prediction Data Analytics for Diabetic Patients Monitoring," no. i, 2017.
https://doi.org/10.1109/NEXTCOMP.2017.8016197

[32] M. Almetwazi *et al.*, "Factors associated with glycemic control in type 2 diabetic patients in Saudi Arabia," *Saudi Pharm. J.*, vol. 27, no. 3, pp. 384–388, 2019.
https://doi.org/10.1016/j.jsps.2018.12.007

[33] M. YimamAhmed, S. H. Ejigu, A. Z. Zeleke, and M. Y. Hassen, "Glycemic control, diabetes complications and their determinants among ambulatory diabetes mellitus patients in southwest ethiopia: A prospective cross-sectional study,"

Diabetes, Metab. Syndr. Obes. Targets Ther., vol. 13, pp. 1089–1095, 2020. https://doi.org/10.2147/DMSO.S227664

[34] B. N. Armstrong, A. Renson, L. C. Zhao, and M. A. Bjurlin, "Development of novel prognostic models for predicting complications of urethroplasty," *World J. Urol.*, vol. 37, no. 3, pp. 553–559, 2019. https://doi.org/10.1007/s00345-018-2413-5

[35] V. Rodriguez-Romero, R. F. Bergstrom, B. S. Decker, G. Lahu, M. Vakilynejad, and R. R. Bies, "Prediction of Nephropathy in Type 2 Diabetes: An Analysis of the ACCORD Trial Applying Machine Learning Techniques," *Clin. Transl. Sci.*, vol. 12, no. 5, pp. 519–528, 2019. https://doi.org/10.1111/cts.12647

[36] S. Ding, Z. Li, X. Liu, H. Huang, and S. Yang, "Diabetic complication prediction using a similarity-enhanced latent Dirichlet allocation model," *Inf. Sci. (Ny).*, vol. 499, pp. 12–24, 2019. https://doi.org/10.1016/j.ins.2019.05.037

[37] K. Alexiadou and J. Doupis, "Management of diabetic foot ulcers," *Diabetes Ther.*, vol. 3, no. 1, pp. 1–15, 2012. https://doi.org/10.1007/s13300-012-0004-9

[38] Y. J. van de Vegte, B. S. Tegegne, N. Verweij, H. Snieder, and P. van der Harst, "Genetics and the heart rate response to exercise," *Cell. Mol. Life Sci.*, no. 123456789, 2019. https://doi.org/10.1007/s00018-019-03079-4

[39] B. Xue, M. Zhang, S. Member, and W. N. Browne, "Particle Swarm Optimization for Feature Selection in Classification : A Multi-Objective Approach," *Ieee Trans. Cybern.*, pp. 1–16, 2012. https://doi.org/10.1109/TSMCB.2012.2227469

[40] M. A. Sulaiman and J. Labadin, "Feature selection based on mutual information for machine learning prediction of petroleum reservoir properties," *2015 9th Int. Conf. IT Asia Transform. Big Data into Knowledge, CITA 2015 - Proc.*, pp. 2–7, 2015. https://doi.org/10.1109/CITA.2015.7349827

[41] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," *Inf. Fusion*, vol. 52, pp. 1–12, 2019. https://doi.org/10.1016/j.inffus.2018.11.008

[42] R. Cekik and A. K. Uysal, "A novel filter feature selection method using rough set for short text data," *Expert Syst. Appl.*, vol. 160, p. 113691, 2020. https://doi.org/10.1016/j.eswa.2020.113691

[43] E. Hancer, B. Xue, and M. Zhang, "Differential evolution for filter feature selection based on information theory and feature ranking," *Knowledge-Based Syst.*, vol. 140, pp. 103–119, 2018. https://doi.org/10.1016/j.knosys.2017.10.028

[44] M. Monirul Kabir, M. Monirul Islam, and K. Murase, "A new wrapper feature selection approach using neural network," *Neurocomputing*, vol. 73, no. 16–18, pp. 3273–3283, 2010. https://doi.org/10.1016/j.neucom.2010.04.003

[45] V. F. Rodriguez-Galiano, J. A. Luque-Espinar, M. Chica-Olmo, and M. P. Mendes, "Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods," *Sci. Total Environ.*, vol. 624, pp. 661–672, 2018. https://doi.org/10.1016/j.scitotenv.2017.12.152

[46] J. González, J. Ortega, M. Damas, P. Martín-Smith, and J. Q. Gan, "A new multi-objective wrapper method for feature selection – Accuracy and stability analysis for BCI," *Neurocomputing*, vol. 333, pp. 407–418, 2019. https://doi.org/10.1016/j.neucom.2019.01.017

[47] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egypt. Informatics J.*, vol. 19, no. 3, pp. 179–189, 2018. https://doi.org/10.1016/j.eij.2018.03.002

[48] J. Pirgazi, M. Alimoradi, T. Esmaeili Abharian, and M. H. Olyaee, "An Efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets," *Sci. Rep.*, vol. 9, no. 1, pp. 1–15, 2019. https://doi.org/10.1038/s41598-019-54987-1

[49] H. Liu, S. Member, M. Zhou, I. Qing, and G. Liu, "An Embedded Feature Selection Method for Imbalanced Data Classification," *IEEE/CAA J. Autom. Sin.*, vol. PP, pp. 1–13. https://doi.org/10.1109/JAS.2019.1911447

[50] M. Lu, "Embedded feature selection accounting for unknown data heterogeneity," *Expert Syst. Appl.*, vol. 119, pp. 350–361, 2019. https://doi.org/10.1016/j.eswa.2018.11.006

[51] R. Saifan, K. Sharif, M. Abu-Ghazaleh, and M. Abdel-Majeed, "Investigating algorithmic stock market trading using ensemble machine learning methods," *Inform.*, vol. 44, no. 3, pp. 311–325, 2020. https://doi.org/10.31449/INF.V44I3.2904

[52] J. Wang, J. Xu, C. Zhao, Y. Peng, and H. Wang, "An ensemble feature selection method for high-dimensional data based on sort aggregation," *Syst. Sci. Control Eng.*, vol. 7, no. 2, pp. 32–39, 2019. https://doi.org/10.1080/21642583.2019.1620658

[53] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014. https://doi.org/10.1016/j.compeleceng.2013.11.024

[54] J. Lee, D. Park, and C. Lee, "Feature selection algorithm for intrusions detection system using sequential forward search and random forest classifier," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 10, pp. 5132–5148, 2017. https://doi.org/10.3837/tiis.2017.10.024

[55] O. F.Y, A. J.E.T, A. O, H. J. O, O. O, and A. J, "Supervised Machine Learning Algorithms: Classification and Comparison," *Int. J. Comput. Trends Technol.*, vol. 48, no. 3, pp. 128–138, 2017. https://doi.org/10.14445/22312803/ijctt-v48p126

[56] B. J. Frey, S. Member, and N. Jojic, "freyJojicTutorial_pami_sep05.pdf," vol. 27, no. 9, pp. 1392–1416, 2005.

[57] C. M. Bishop, "Model-based machine learning Author for correspondence :," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 371, no. 1984, p. 20120222, 2013.

[58] M. M. Churpek, T. C. Yuen, C. Winslow, D. O. Meltzer, M. W. Kattan, and D. P. Edelson, "Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards," *Crit. Care Med.*, vol. 44, no. 2, pp. 368–374, 2016. https://doi.org/10.1097/CCM.0000000000001571

[59] B. Heung, H. C. Ho, J. Zhang, A. Knudby, C. E. Bulmer, and M. G. Schmidt, "An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping," *Geoderma*, vol. 265, pp. 62–77, 2016. https://doi.org/10.1016/j.geoderma.2015.11.014

[60] M. Maniruzzaman *et al.*, "Accurate Diabetes Risk Stratification Using Machine Learning: Role of Missing Value and Outliers," *J. Med. Syst.*, vol. 42, no. 5, pp. 1–17, 2018. https://doi.org/10.1007/s10916-018-0940-7

[61] C. Zhu, C. U. Idemudia, and W. Feng, "Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques," *Informatics Med. Unlocked*, vol. 17, no. January, p. 100179, 2019. https://doi.org/10.1016/j.imu.2019.100179

[62] M. Jena and S. Dehuri, "Decision tree for classification and regression: A state-of-the art review," *Inform.*, vol. 44, no. 4, pp. 405–420, 2020. https://doi.org/10.31449/INF.V44I4.3023

[63] W. Xu, J. Zhang, Q. Zhang, and X. Wei, "Risk prediction of type II diabetes based on random forest model," 2017. https://doi.org/10.1109/AEEICB.2017.7972337

[64] B. Baba, "Borsa _ Istanbul Review Predicting IPO initial returns using random forest," 2020. https://doi.org/10.1016/j.bir.2019.08.001

[65] D. Panda, S. R. Dash, R. Ray, and S. Parida, "Predicting the causal effect relationship between copd and cardio vascular diseases," *Inform.*, vol. 44, no. 4, pp. 447–457, 2020. https://doi.org/10.31449/INF.V44I4.3088

[66] K. Saxena, Z. Khan, and S. Singh, "Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm," vol. 2, no. 4, pp. 36–43, 2014.

[67] G. G. Várkonyi and A. Gradišek, "Data protection impact assessment case study for a research project using artificial intelligence on patient data," *Inform.*, vol. 44, no. 4, pp. 497–505, 2020. https://doi.org/10.31449/INF.V44I4.3253

[68] S. K. Nayak, M. Panda, and G. Palai, "Realization of optical ADDER circuit using photonic structure and KNN algorithm," *Optik (Stuttg).*, vol. 212, no. March, p. 164675, 2020. https://doi.org/10.1016/j.ijleo.2020.164675

[69] N. Nai-Arun and R. Moungmai, "Comparison of Classifiers for the Risk of Diabetes Prediction," *Procedia Comput. Sci.*, vol. 69, pp. 132–142, 2015. https://doi.org/10.1016/j.procs.2015.10.014

https://doi.org/10.31449/inf.v45i1.3424 Informatica **45** (2021) 127–132 **127**

Research on Emotion Recognition Based on Deep Learning for Mental Health

Xianglan Peng
School of Humanities, Henan Mechanical and Electrical Vocation College, Zhengzhou, Henan 451191, China
E-mail: pe79193@163.com

Keywords: deep learning, artificial intelligence, facial expression, emotion recognition

Received: January 27, 2021

This paper briefly introduced the support vector machine (SVM) based and convolutional neural network (CNN) based healthy emotion recognition method, then improved the traditional CNN by introducing Long Short Term Memory (LSTM), and finally carried out simulation experiments on three emotion recognition models, the SVM, traditional CNN, and improved CNN models, in the self-built face database. The results showed that the CNN model converged faster in training and had a smaller error when it was stable after introducing LSTM; compared with the SVM and traditional CNN models, the improved CNN had a higher recognition accuracy for facial expressions; the time consumed by the improved CNN model was the shortest in both training and testing stages.

Povzetek: Analiziranih je bilo več metod strojnega učenja, tudi globoke mreže, za iskanje čustev v povezavi z mentalnim zdravjem.

1 Introduction

With the progress of science and technology and the improvement of computer performance, artificial intelligence appeared and has been widely used in mechanical operation fields, such as translation, image recognition, and classification, which are not difficult but highly repetitive [1]. The ultimate goal of artificial intelligence is to achieve good human-computer interaction, thus replacing humans to carry out dangerous or repetitive work. However, in the current development of artificial intelligence, although it has been able to realize the recognition and classification of objects, such as images and audio, in human-computer interaction, the perception of human emotions by artificial intelligence is still at a low level [2]. Artificial intelligence needs better emotion recognition ability to achieve better human-computer interaction services [3]. Also, human beings express their emotions in various forms, including actions, language, physiological signals, and facial expressions. These emotions usually reflect their psychological state, especially physiological signals and facial expressions. People's physiological state will directly affect the psychological state, and the psychological state will react to the physiological state. Changes in physiological signals will reflect changes in physiological state, thus indirectly reflecting the psychological state [4]. Facial expression can directly reflect people's emotions, and the changes of emotions also reflect the state of mental health. However, the monitoring of physiological signals needs quite professional equipment, and the collection process is complex, which may delay the judgment of people's mental health. Changes in facial expression are relatively easy to collect as long as a good camera is configured to collect mental health-related images. When mental health is judged by the emotion reflected by the facial expression, the manual observation needs rich clinical experience and has low efficiency. Artificial intelligence has a fast computing speed, and it can extract relevant feature rules from face images more effectively and then judge whether people's emotions are in a healthy state. Atkinson et al. [5] proposed a feature-based emotion recognition model based on an electroencephalogram, which combined the mutual information-based feature selection method with kernel classifier to improve the accuracy of emotion classification tasks. The experimental results verified the effectiveness of the proposed method. Kaya et al. [6] proposed to replace the Deep Neural Network (DNN) and support vector machine (SVM) with the extreme learning machine (ELM) in audio and visual emotion recognition. The results showed that the method could achieve better accuracy in emotion classification in audios and videos. Shojaeilangari et al. [7] proposed a new pose invariant dynamic descriptor to encode the relative motion information of facial landmarks. The results showed that the method could deal with speed changes and continuous head pose changes to realize fast emotion recognition. This paper briefly introduced the emotion recognition method based on SVM and convolutional neural network (CNN), then improved the traditional CNN by introducing Long Short Term Memory (LSTM), and finally carried out simulation experiments on three emotion recognition models, the SVM, traditional CNN, and improved CNN models, in the ORT human face database and self-built face database.

2 Recognition of mental health emotion based on deep learning

Unless specially controlled, the expression of ordinary people is usually rich, and the emotion of the other person can be confirmed by observing the change of expression [8]. Artificial intelligence is difficult to understand the emotions represented by different expressions in images

taken by cameras and judge the mental health level represented by the emotions; therefore, artificial intelligence needs relevant algorithms to improve the experience of human-computer interaction and the accuracy of artificial intelligence in judging the user's mental health.

2.1 Traditional recognition method based on SVM

At present, artificial intelligence needs to recognize emotions through machine learning, and SVM is one of the traditional machine learning methods [9]. The basic principle of SVM for health emotion recognition is to find a hyperplane for space division in the vector space of expression features. The expression on one side of the hyperplane is classified as a healthy emotion, and the other side is classified as one kind of unhealthy emotion. In short, SVM is a classification algorithm, which classifies the expression images collected by cameras to identify whether the emotion is in a healthy psychological state. Since the expressions collected by cameras are generally image data, it is necessary to extract features of expression images to obtain expression features when using SVM for recognition [10]. There are various methods for extracting image features. In this paper, facial expression features are extracted by the LDP (local directional pattern) algorithm. The principle of the LDP algorithm is directional edge statistics. It is assumed that x is a pixel in an image. The gray value of a 3×3 field that centers on pixel x is convoluted with Kirsch template [11] M to obtain the corresponding edge response, $|m_i|$. Then, the edge responses are sorted according to their gradients. The first k edge responses are marked as code 1, and the rest is marked as code 0. The calculation formula of LDP code [12] is as follows:

$$\begin{cases} m_k = kth(M) \\ M = |m_0, m_1, \cdots, m_7| \\ LDP_k(r,c) = \sum_{i=0}^{7} b_i(m_i - m_k) \times 2^i, (1) \\ b_i(m_i - m_k) = \begin{cases} 1 & m_i - m_k \geq 0 \\ 0 & m_i - m_k < 0 \end{cases} \end{cases}$$

where m_k is the k-th edge response, M is Kirsch template, $LDP_R(r,c)$ is the LDP code of central point c, and r is the domain radius, which is set as 3 in this paper. The extraction steps are as follows: ① eight Kirsch templates and equation (1) are combined to convert each pixel in the original face image into LDP code; ② the LDP code image of the human face is constructed according to the LDP code; ③ the LDP code image is divided into $a \times b$ blocks, the histogram of all the blocks is extracted; ④ the histogram of the blocks is connected end to end to get the final feature vector.

After obtaining the expression feature vector, it can be used as a training sample to train SVM to obtain the decision function of SVM. The calculation formula is:

$$\begin{cases} f(x) = sgn(\sum_{i=1}^{l} a_i y_i K(x_i, x_j) + b) \\ \sum_{i=1}^{l} a_i y_i = 0 \quad 0 \leq a_i \leq C \end{cases}, (2)$$

where a is the set of a_i, a_i is the Lagrangian coefficient [13], l is the sample size, $K(x_i, x_j)$ is the kernel function, C is the penalty parameter, y_i is the result of classification, and x_i is the sample data.

2.2 Healthy emotion recognition method based on LSTM-CNN

In addition to SVM, neural network, a kind of deep learning algorithm, has also widely used in artificial intelligence. Neural network realizes machine learning by imitating neural cells of the human brain, which is relatively better in learning effect. CNN is one of the neural networks [14]. Compared with other kinds of neural networks, CNN is more suitable for image recognition. The basic structure of CNN is the input layer, convolution layer, pooling layer, and output layer. The convolution layer and pooling layer are the hidden layers of the neural network, and the number of them depends on the operation requirements. The more the number is, the better the learning effect is, but the lower the efficiency is.

The basic process of emotion recognition by CNN is as follows. An image is inputted into the input layer after preprocessing and then convoluted through convolution kernels in the convolution layer. The image features are extracted. The convoluted image is processed by pooling in the pooling layer (equivalent to compressing the image, including mean-pooling and max-pooling). After repetitive convolution and pooling operations, the results are output in the form of full connection in the output layer according to the transfer formula. The results will be compared with the expected results; if they are not consistent, the weights and bias terms in the hidden layer will be adjusted reversely, and the weights and bias terms in CNN will be adjusted through repetitive training to make the output as close to the expected output as possible.

One of the advantages of CNN in image recognition is that it does not need feature extraction. The convolution operation in the convolution layer plays the role of feature extraction. Moreover, the hidden layer of CNN has an activation function operation, which can transform the linear input data into nonlinear to fit the hidden rules between features better; therefore, it can classify more accurately than the hyperplane in SVM [15].

Although CNN can effectively identify the emotion in expression images, in practical applications, when artificial intelligence recognizes the images taken by cameras, not all of the images are taken under good lighting and from proper angles, and most of the images have incomplete facial expression features. Moreover, the facial expression features have multi-dimensional and multi-scale changes, making it difficult to improve the recognition rate. This study introduces LSTM [16] into CNN to improve its recognition rate of expression and emotion.

The main structure of LSTM includes the input gate, forget gate, and output gate. Parameters to be calculated are input into the input gate, mainly including the current input of cell (x_t), the state of the last hidden layer (h_{t-1}), and the last state of cell (C_{t-1}). A matrix is constructed by these parameters and corresponding weights to determine the number of new information in the cell. The relevant formula is:

$$\begin{cases} i_t = g(\omega_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = tanh(\omega_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \end{cases}, \quad (3)$$

where i_t is the proportion of the new information that can be memorized, $\tilde{C}_t$ is the cell state of the new information added, C_t is the current cell state after the addition of the new information, ω_i and ω_t are the corresponding weights, and b_i and b_C are the corresponding offsets.

The forget gate determines the number of original information to be abandoned, and its formula is:

$$f_t = g(\omega_f \cdot [h_{t-1}, x_t] + b_f), \quad (4)$$

where f_t is the proportion of information that is not forgotten in C_{t-1}, ω_f is the corresponding weight, and b_f is the corresponding offsets.

The output gate is a structure that obtains the output result based on the parameters of the first two structures. The output result can be the final result or the hidden variable when the content is updated next time. The formula is as follows:

$$\begin{cases} o_t = g(\omega_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot tanh(C_t) \end{cases}, \quad (5)$$

where o_t is the weight that determines the final output information quantity and h_t is the final output or the next hidden state.

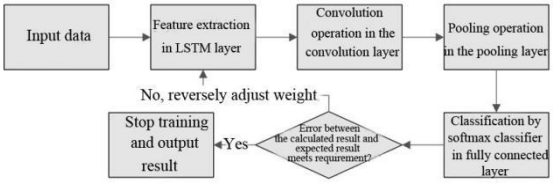


Figure 1: The expression and emotion recognition process based on LSTM-CNN.

After introducing LSTM into CNN, it can effectively associate the changes of expression before and after to obtain the regular features of expression changes. Moreover, the continuously varying features that can reflect emotions in the human face can be more prominent, thus reducing the influence of irrelevant background features and reflect the emotions contained in continuously changing expressions. The training flow of the expression and emotion recognition model based on the LSTM and CNN is shown in Figure 1.

① The data were input, and the relevant parameters were initialized, including convolution kernel, weights in structure layers, offset, etc.

② In the LSTM layer, features were extracted from the image according to equations (3), (4), and (5). The feature map that was needed by the subsequent convolution was constructed according to the extracted h_t .

③ The feature map processed by the LSTM layer was input into the convolution layer for convolution operation by the convolution kernel. The convolution formula is:

$$x_j^l = f\left(\sum_{j \in M} x_i^{l-1} \cdot W_{ij}^l + b_j^l\right), \quad (6)$$

where x_j^l is the output feature map after the activation of the j-th convolution kernel in the l-th convolution layer, x_i^{l-1} is the feature output of the i-th convolution kernel in the last convolution layer after pooling, W_{ij}^l is the weight parameter between the i-th convolution kernel and the j-th convolution kernel, b_j^l is the offset of j convolution kernels of l layers, M is the number of convolution kernels in the l-th convolution layer, and $f(\bullet)$ is the activation function.

④ The convoluted feature map was input into the pooling layer for pooling. The pooling operation included mean-pooling and max-pooling. In this study, the max-pooling operation was adopted. The target box slid on the feature map for some distance, and the largest pixel in the target box was taken as the compression result of the target box.

⑤ The convolution and pooling operations mentioned above were performed many times, depending on the number of convolution layers and pooling layers. After convolution and pooling, the result was output to the fully connected layer. Then the expression images were classified by using a softmax classifier in the fully connected layer.

⑥ The recognition results of CNN were compared with the expected results (the recognition results refer to the results obtained by calculating the input image layer by layer with CNN, and the expected results refer to the corresponding result label of the training sample), and the weights and offset parameters in the calculation formula were adjusted reversely according to the error until the error was within the predetermined range or converged to stability. The calculation formula of error is as follows:

$$E = -\sum_{k=1}^{n} t_k \, log(y_k), \quad (7)$$

where E stands for the error between the calculated output vector and the actual output vector, n is the number of output layer nodes, y_k is the probability of belonging to such kind of label output to the output layer after the forward calculation of the fully connected layer, and t_k is the label of the actual correct solution that is set.

⑦ When the error converged to stability or is within a predetermined range, the training of the recognition model ended, and then the model was tested using the testing set.

3 Simulation experiment

3.1 Experimental environment

The CNN model was simulated and analyzed using MATLAB software [17]. The experiment was carried out on a laboratory server. The server configurations were the Windows7 system, I7 processor, and 16 G memory.

3.2 Experimental data

In this study, a self-built facial expression database was used. Facial expression images came from 100 students randomly selected from Henan Mechanical and Electrical Vocation College after explaining the use of face images to them and obtaining their approval. Since the purpose of this study was to realize the recognition of mental health emotion of human expression fast by artificial intelligence, when collecting the facial expression image data of the volunteers, the corresponding mental health test was carried out, and the corresponding mental health labels were added for facial expression images. To ensure the time correspondence between the expression image and the degree of mental health in the database (i.e., the mental state reflected by the expression image was indeed the psychological state when the expression was collected), the expression data were collected by making the psychological evaluation of the volunteers to judge the mental health status and capturing the volunteers' expressions synchronously during the psychological evaluation [18]. Finally, 30 facial expression images were collected from each volunteer. The results of the psychological evaluation were statistically analyzed, and it was found that 68 volunteers had healthy psychology (2040 expression images), 26 volunteers had sub-healthy psychology (780 expression images), and six volunteers had poor mental psychology (180 expression images). As the number of people with healthy psychology was the largest, followed by people with sub-healthy psychology and people with poor mental psychology, the number of images collected for three mental health states was unbalanced, which would affect the final training result; therefore, the expression images were extended through means such as rotation, extension, and mirroring. The number of images for sub-healthy psychology and poor mental psychology was extended to 2040. The external performance of three kinds of mental health states was described briefly. The volunteers with healthy psychology were relaxed when receiving the psychological counseling test. They smiled unconsciously in the process of communication and showed a bright smile when the communication was smooth. The volunteers with sub-healthy psychology were not relaxed in the process of mental health assessment, but most of them were not tense in facial expression. In the communication process, the expression was relatively flat, and the communication is relatively smooth. Most of the smiles appeared when they talked about the topic of interest. The volunteers with poor mental health were usually tight in facial expression. Although they achieved communication, they gave a sense of tension and anxiety, and some had sweating. Moreover, the atmosphere presented by the dialogue in the process of communication was relatively repressive. When testing the three recognition models, 20 expression images of each volunteer in the database were used as the training set, and the remaining ten images were used as the testing set.

In the simulation test, 60% of the images were taken from every mental health status as the training set, and the remaining 40% was taken as the test set. There were 1224 images in the training set and 816 images in the test set.

3.3 Experimental setup

In this study, the expression recognition model was improved by introducing LSTM to CNN. The structural parameters of CNN are as follows. There were three convolutional layers. Every convolutional layer had 64 convolution kernels in a size of 5×5. Relu function was used as the activation function. There were three pooling layers. In every pooling layer, the size of the pooling box was 2×2, and the moving step length of the pooling box was 2. The size of the image in the input layer was 150×100. In the LSTM layer, there were 64 hidden neurons, weights were initialized using glorot_normal, and the offset was set as 0.

Moreover, to verify the effectiveness and excellence of the improved expression recognition model, it was compared with the SVM model and the traditional CNN model. The comparative experiment was carried out in the same face database. The parameters of the traditional CNN model were consistent with the CNN in the LSTM-CNN model. SVM adopted the sigmoid kernel function, and the penalty parameter was set as 1.

3.4 Experimental results

The SVM model fits the decision function according to the extracted feature vector in training, thus to obtain the hyperplane in the feature vector space for the classification of different expressions; therefore, different from the CNN model, the SVM model needed to be trained repeatedly. Figure 2 shows the convergence curves of the traditional CNN model and the improved CNN model in training. It was seen from Figure 2 that the training error of the two CNN models gradually decreased in the process of training iteration and finally stabilized at a low level. The comparison of the curves showed that the improved CNN model converged to stability faster than the traditional CNN model: the traditional CNN model converged to stability after about 250 times of iterations, and the improved CNN converged to stability about 150 times of iterations; the error of the improved CNN model after convergence to stability was significantly smaller than that of the traditional CNN model.

After training the SVM, traditional CNN, and improved CNN models with the training set in the self-built database, the trained models were tested using the

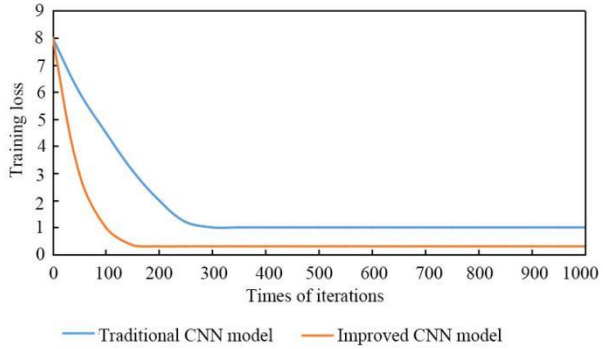


Figure 2: The convergence curves of the traditional and improved CNN models in training.

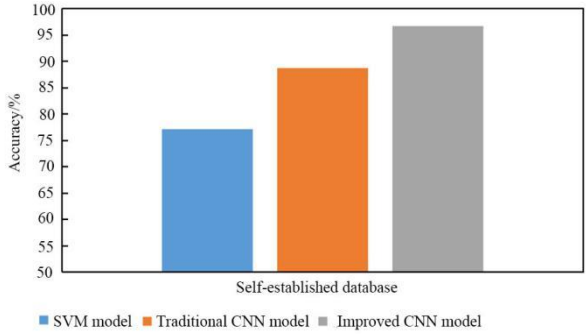


Figure 3: The recognition accuracy of three healthy emotion recognition models in two kinds of face databases.

corresponding testing set, and the results are shown in Figure 3. In the self-built database, the recognition accuracy of the SVM model was 77.1%, that of the traditional CNN model was 88.6%, and that of the improved CNN model was 96.6%. In the same database, the SVM model had the lowest accuracy, that of the traditional CNN model was the second, and that of the improved CNN model was the highest.

For the artificial intelligence that was used for judging emotions, in addition to the high accuracy for emotional judgment, the speed of judgment is also very important. Table 1 shows the time spent in training and testing three healthy emotion recognition models. In the training stage, the training time of the SVM model was 20.2 min, that of the traditional CNN model was 20.4 min, and that of the improved CNN model was 15.3 min; in the testing stage, the SVM model took 835 ms, the traditional CNN model took 621 ms, and the improved CNN model took 378 ms. In the training stage, the SVM model could not perform a parallel operation on the data in the training set but gradually fit it; thus, it needed a long training time. Although the other two CNN models could perform a parallel calculation on the data in the training set, they needed repeated training and gradual adjustment when adjusting the internal weight; thus, they also needed a long time. However, the improved CNN model eliminated the background features that would produce interference from the image as much as possible and highlighted the expression features; therefore, it converged faster and spent less time. In the testing stage, the three models have been trained, and the results could be calculated step by step as long as the data were input; thus, the time consumed was much shorter than that in the training stage.

	The SVM model	The traditional CNN model	The improved CNN model
Training time	20.2 min	20.4 min	15.3 min
Testing time	835 ms	621 ms	378 ms

Table 1: Time consumption of three recognition models in training and testing.

4 Discussion

For the human body, health includes not only physical health but also mental health. However, different from the physiological health state, it is difficult to see the mental health state intuitively. If the physical health status can be directly obtained through various detection instruments, such as blood state, body temperature, etc., then the mental health status needs to be gradually judged by professionals in the process of communication. It not only requires a high professional quality of the tester but also consumes a lot of time. Artificial intelligence has the advantages of fast learning and high work efficiency. Based on the progress of machine vision technology, artificial intelligence has been gradually applied to the judgment of people's emotions. Artificial intelligence combined with machine vision technology can judge mental health through the emotion reflected by the change characteristics of human facial expressions. This paper briefly described two intelligent algorithms, the image LDP features-based SVM algorithm and the LSTM-introduced CNN algorithm. Then, 100 student volunteers were taken to establish the database of facial emotional and mental health. The performance of the SVM, traditional CNN, and improved CNN recognition models was compared. The final experimental results showed that the improved CNN model could identify the mental health state behind the expression more accurately than the other two models, and the training and testing time was shorter. On the one hand, compared with the SVM model, the CNN model did not need to extract image features deliberately as its convolution operation has obtained features; on the other hand, the improved CNN model could associate the images before and after to further extract the core features from the changing expression, which reduced the interference of background features and improved the recognition efficiency and accuracy.

The psychological evaluation on volunteers was carried out by professionals to ensure the accurate correspondence between expression and mental state, and facial expression changes were captured in time in the process of psychological assessment. As a thank to the volunteers, after the psychological assessment, professionals provided guidance and suggestions on the mental health of volunteers according to the evaluation results. After summing up the results of the psychological evaluation, it was concluded that most of the volunteers

had healthy psychology, some volunteers had sub-healthy psychology, and fewer volunteers had unhealthy psychology. The advice to the volunteers with healthy psychology was to keep the current good mood. The reason for the sub-healthy state was mostly related to the heavy academic pressure and the chaotic daily schedule. The advice to the volunteers with a sub-healthy state was to adjust work and rest, be relaxed in the face of study, and attempt to formulate a study schedule. Besides the heavy academic pressure, the reasons for the unhealthy mental state of the volunteers also included introversion, inferiority, and little communication with others. The final suggestion for the volunteers with an unhealthy mental state was to set a good daily routine, walking outside, and starting communication with acquaintances first.

5 Conclusion

This paper briefly introduced the SVM-based and CNN-based healthy emotion recognition methods, then improved the traditional CNN by introducing LSTM, and finally carried out simulation experiments on the SVM, traditional CNN, and improved CNN models through the self-built human face database. The results are as follows: (1) compared with the traditional CNN model, the improved CNN model converged faster and had a smaller error after stabilization; (2) the recognition accuracy of the improved CNN model was the highest, followed by the traditional CNN model and SVM model; (3) the improved CNN model took the least time in the training stage and the shortest time in the testing stage.

References

[1] Jenke R, Peer A, Buss M (2017). Feature Extraction and Selection for Emotion Recognition from EEG. IEEE Transactions on Affective Computing, 5, pp. 327-339. https://doi.org/10.1109/TAFFC.2014.2339834.

[2] Anagnostopoulos C N, Iliou T, Giannoukos I (2015). Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. Artificial Intelligence Review, 43, pp. 155-177. https://doi.org/10.1007/s10462-012-9368-5.

[3] Suja P, Tripathi S (2015). Analysis of emotion recognition from facial expressions using spatial and transform domain methods. International Journal of Advanced Intelligence Paradigms, 7, pp. 57. https://doi.org/10.1504/IJAIP.2015.070349.

[4] Atkinson J, Campos D (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. Expert Systems with Applications, 47, pp. 35-41.

[5] Chen ZB (2018). Facial Expression Recognition Based on Local Features and Monogenic Binary Coding. Informatica, 43, pp. 117-121. https://doi.org/10.31449/inf.v43i1.2716.

[6] Kaya H, Salah A A (2016). Combining modality-specific extreme learning machines for emotion recognition in the wild. Journal on Multimodal User Interfaces, 10, pp. 139-149.

[7] Shojaeilangari S, Yau W Y, Teoh E K (2016). Pose-invariant descriptor for facial emotion recognition. Machine Vision & Applications, 27, pp. 1063-1070.

[8] Liu S, Tong J, Meng J, Yang J, Zhao X, He F, Qi H, Ming D (2018). Study on an effective cross-stimulus emotion recognition model using EEGs based on feature selection and support vector machine. International Journal of Machine Learning and Cybernetics, 9, pp. 721-726. https://doi.org/10.1007/s13042-016-0601-4.

[9] Ghimire D, Lee J (2016). Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. Sensors, 13, pp. 7714-7734. https://doi.org/10.3390/s130607714.

[10] Viet SD, Bao CLT (2018). Effective Deep Multi-source Multi-task Learning Frameworks for Smile Detection, Emotion Recognition and Gender Classification. Informatica, 42, pp. 345-356. https://doi.org/10.31449/inf.v42i3.2301.

[11] Chakraborty S, Singh S K, Chakraborty P (2017). Local directional gradient pattern: a local descriptor for face recognition. Multimedia Tools & Applications, 76, pp. 1201-1216.

[12] Chakraborty S, Singh S K, Chakraborty P (2018). Correction to: Local directional gradient pattern: a local descriptor for face recognition. Multimedia Tools & Applications, pp. 1-1.

[13] Pan H, Xie L, Lv Z, Li J, Wang Z (2020). Hierarchical support vector machine for facial micro-expression recognition. Multimedia Tools and Applications, 79, pp. 1-15. https://doi.org/10.1007/s11042-020-09475-4.

[14] Lopes A T, Aguiar E D, Souza A F D, Oliveira-Santos T (2017). Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order. Pattern Recognition, 61, pp. 610-628. https://doi.org/10.1016/j.patcog.2016.07.026.

[15] Gjoreski M, Gjoreski H, Kulakov A (2014). Machine Learning Approach for Emotion Recognition in Speech. Informatica, 38, pp. 377-384.

[16] Ghimire S, Deo R C, Raj N, Mi J (2019). Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Applied Energy, 253, pp. 113541.1-113541.20. https://doi.org/10.1016/j.apenergy.2019.113541.

[17] Wang F, Wu S, Zhang W, Xu Z, Zhang Y, Wu C, Coleman S (2020). Emotion recognition with convolutional neural network and EEG-based EFDMs. Neuropsychologia, 146, pp. 107506. https://doi.org/10.1016/j.neuropsychologia.2020.107506.

[18] Azam I, Khan SA (2018). Feature Extraction Trends for Intelligent Facial Expression Recognition: A Survey. Informatica, 42, pp. 507-514. https://doi.org/10.31449/inf.v42i4.2037.

Risks Analyzing and Management in Software Project Management Using Fuzzy Cognitive Maps with Reinforcement Learning

Ahmed Tlili and Salim Chikhi
MISC Labs, Abdelhamid Mahri University of Science and Technology, Constantine, Algeria
E-mail: a.tlili@univ-emir.dz, salim.chikhi@univ-constantine2.dz

Keywords: Fuzzy Cognitive maps (FCM), complex system, reinforcement learning, project risk management

Received: March 31, 2020

Many projects fail each year simply because a risk has been misjudged, ignored or unidentified. An essential motivation for analyzing the risk of a project is to inform managers in order to reduce the risk, and therefore the loss of the project. Risk analysis can help identify the best actions that would reduce the risk and assess by how much. In the last decades, the Fuzzy Cognitive Map emerged as a powerful tool for modeling and supervising dynamic interactions in complex systems. There is two ways to construct them, the first way by experts of domain and the second way by learning method based on the historical of data. In this paper, we develop a new learning fuzzy cognitive maps based on a reinforcement learning algorithm so called Q-learning and we propose here a new formulation of kosko causality principle. This connection between fuzzy cognitive maps and reinforcement learning allows us to choose based on the historical of data learning process the best and the most important connections between concepts. In this work, we illustrate the effectiveness of the proposed approach by modeling and studying the analysis of project risk management as an economic decision support system.

Povzetek: Spodbujevalno učenje in metode mehke logike so uporabljene za analizo tveganj pri razvoju programskih sistemov.

1 Introduction

Risks represent a major challenge for organizations and more particularly for organizations developing applications. All activities in general, present risks. The objective of risk management is to better understanding of the factors that contribute to software project risk and to propose an approach to deal them. This approach is no longer reserved for the space or nuclear fields; it has become one of the crucial elements of project management, as well as the management of people, resources, planning and performance. Today, the success of a project is strongly conditioned by the way its leaders know how to recognize the risks. Risk prevention and risk analysis is an important task of the managers that threaten it, to study and overcome them. The information's absorbed by humans; quite complex processes are usually imprecise or approximate [1]. The strategy adopted is usually imprecise in nature with no or partial knowledge of the problem, and generally possible to be expressed in linguistic terms. Thus the main problem with risk estimation is that the input data is imprecise or uncertain in nature and it is difficult to accurately represent them in mathematical models [2]. Usually and naturally, the risk analyst is specified in language terms as high, very high, medium, low… etc., rather than in exact statistical terminology. To this end, the application of the Fuzzy Inference System (FIS) theory to risk analysis seems appropriate because it deals with inaccurate and ambiguous information and the basic idea of this approach is to allow an element to belong to a set with membership degrees within the continuous real interval [0,1], rather than in the set {0,1}.

In risk analysis and management RAM, the most important factors contributing to the risk of failure for any type of socio-economic organization are related to the different criteria as: time constraints, high cost, weak operating resources, poor performance of supervisors…etc., and the identification of the relationships between the risks and the ones that causes' them remains a major challenge for experts in this field because they are in most cases very complex [3].

In this work we propose an approach for risks analysis and management to managing software projects using Koskos' fuzzy cognitive maps FCM improved with reinforcement learning Q-Learning algorithm. This work is implemented and validated on Matlab R2014a platform. The proposed method is summarized by the framework shown on Figure 1.

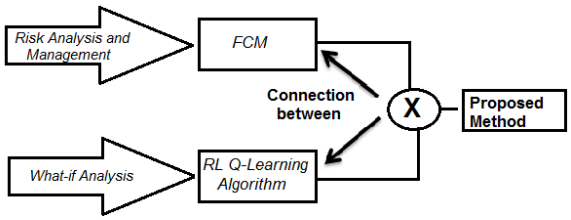


Figure 1: Proposed method's background.

2 Research method

2.1 Literature review

Several methods can be found in literature review for the risks management mainly classified in deterministic and stochastic approaches: what-if analysis, task analysis, Hazard and Operability (HAZOP), Quantitative Risk Assessment (QRA), the Critical Risk and Error Analysis (CREA), Fault Tree Analysis (FTA), the Event Tree Analysis (ETA), Failure Mode and Effects Analysis (FMEA), Probability Distribution of Failure and Reliability (PDEA), Petri networks, Bayesian networks, … etc.

In [4] Samantra et al., explain that the risk associated with a specific risk factor is expressed as a combination of two parameters: the probability of occurrence and the effect. The concept of risk matrix is here to categorize different risk factors at each levels of occurrence to create a plan of actions. A case study of a metropolitan construction project for the construction of an underground metro station was carried out and demonstrated the efficiency of the steps of the procedure for applying the proposed methodology.

Taylan et al. in [5] illustrated risk assessment using AHP and fuzzy TOPSIS where many construction projects were studied according to these main criteria: time, cost, quality, safety, and environmental sustainability. Authors showed that these methods are able of evaluating the overall risk factors of projects and selecting a project with the lowest risk with a relative weight matrix. The results showed that these novel methodologies are able to assess the overall risks of projects, select the project that has the lowest risk with the contribution of relative importance index.

In the work of Dziadosz & Rejment [6] , risk and risk factor are a measurable part of uncertainty and can be estimated from the probability of occurrence. This risk and risk factor represent a deviation from the desired level, which can be positive or negative. Consequently, risk analysis is very important for selecting a win project. the main result of this approach concerns cases in which the schedule, costs and requirements of the project must be defined in the planning phase and deviations will be detected automatically in the progress phase.

The main idea in the paper related by Muriana & Vizzini [7] is that total weight method is used to calculate the current risk level of the project and the risk of the whole project is reduced taken preventive measures.

2.2 Risk analysis and management

Rik is an uncertain event that may have positive or negative impact on project and risk management is the process of identifying and migrating risk. Risk management is more important because it affects all aspects of the project as schedule, budget, delay…etc.

One of the main difficulties of risk management is that it is not "an exact science", in this way:

- It is not possible to predict in the long term without admitting a part of the uncertainty,
- Risks are present at all stages of a project and can take a variety of forms with internal and / or external origins,
- We can reduce the risks of a project, but we cannot eliminate them completely,
- Due to the diversity of the risks and their management, in particular according to the size of the project, the mobilized resources and the sector of activity concerned, there is a difficulty in invariant identifications.

Research in risk analysis and management (RAM) using fuzzy systems [8] have provided several models in recent years. However, to the extent that we have found, there are very few sufficiently representative approaches to be used for complex problems in this area.

Quantifying or assessing risk and its factors consists in measuring the (linguistic) probability of occurrence and the estimated or the staggered by defining a scale of (linguistic) values associated with it as follows:

- Frequent risk with high probabilities of realization, very high.
- Occasional or average risk, can be realized
- Rare, unlikely or low
- Very unlikely or high.

2.3 What-If Analysis method

What–If Analysis is defined as a structured brainstorming method of determining what things can go wrong and estimate the likelihood and consequences of those situations occurring. The answers to these questions are not evident and form the basis for determining a recommended course of action for those risks or risk factor. our proposed method here constitute an automatic alternative to expert review team and can effectively and productively discern major issues concerning a software project or with any other risks project. Lead by an energetic and focused facilitator, each member of the review team participates in assessing what can go wrong based on their past experiences and knowledge of similar situations.After the “What-If” answers are generated by different simulation, the review manager then makes judgments regarding the probability and severity of the risk. If the risk is judged unacceptable then a recommendation is made by the manager for further action. The completed analysis is then summarized as mentioned below:

What-if?	*Answer*	*Likelihood*	*Consequences*	*Recommendations*
What-if the high cost risk immerses?	scheduling process or technological aspects are deficient	possible	Very serious	1 - Include inspection in scheduling procedure. 2- Check the technological aspects in terms of equipment and software plate form.

Table 1: What-If Analysis Form.

3 Theory background

3.1 Fuzzy Cognitive Maps

The cognitive maps were studied by computer scientists from the 80s when Bart Kosko [9] chooses to provide a new formalization of Axelrod's cognitive maps [10]. Kosko notes that Axelrod's cognitive maps applied to fields such as politics, history, international relations, contain concepts and influences between concepts that are by nature fuzzy. He thus formalizes the model of fuzzy cognitive maps using the theory of fuzzy sets [9].

Fuzzy cognitive map is a directed graph in the form $< X,W >$ where $X = [X_1, ...,X_n]$ is the set of the concepts, W is the connection matrix describing weights of the connections, $w_{j,i}$ is the weight of the direct influence between the j-th concept and the i-th concept, taking on the values from the range [−1, 1]. A positive weight of the connection $w_{j,i}$ means X_j causally increases X_i. A negative weight of the connection wj,i means X_j causally decreases X_i and A nul weight of the connection $w_{j,i}$ means there is no causality between X_j and X_i.

Fuzzy cognitive map can be used for modeling behavior of dynamic systems. The state of the FCM model is determined by the values of the concepts at the t-th iteration. The simulation of the FCM behavior requires an initial state vector. Next, the values of the concepts can be calculated according to the selected dynamic model. Simulations show the effect of the changes in the state maps and can be used in a what-if analysis [11].

$$X_i^{k+1} = f(\sum X_j^k . \omega_{ji}) \quad (1)$$

Where $X_i(k)$ is the value of the i-th concept at the *k-th* iteration, i = 1, 2, ..., n, n is the number of concepts. Transformation function $f(x)$ normalizes values of the concepts to a proper range. A logistic function is most often used [12]:

$$f(x) = 1/(1 + e^{-\beta x}) \quad (2)$$

Where $\beta > 0$ is a parameter.

Other alternatives are taking into account the past history of concepts and jointly proposed a popular dynamic model which was used in this work summarized in the following equation [10]:

$$X_i^{k+1} = f(X_i^k + \sum X_j^k . \omega_{ji}) \quad (3)$$

3.2 Reinforcement Learning

Reinforcement Learning (RL) is one effective method in the solution of multi stage decision making problems. For a comprehensive study of the subject, refer [13][143][15].

The Markov Decision Processes (MDP) defines the formal framework of reinforcement learning [13]. More formally, an MDP process is defined by:

- S, a finite set of states. s Є S
- A, a finite set of actions in state *s*. a Є A(s)
- r, a reward function. r(s, a) Є R
- P, the probability of transition from one state to another depending on the selected action. P (s '| s, a) = P_a(s, s').

The problem is to find an optimal policy of actions that achieves the goal by maximizing the rewards, starting from any initial state. At each iteration, the agent being in the state chooses an action, according to these outputs the environment sends either award or a penalty to the agent shown by the following formula: $r_k = h(s_k, a_k, s_{k+1})$.

To find the total cost, which is represented by the formula $\Sigma\, h(s_k, a_k, s_{k+1})$, the costs are accumulated at each iteration of the system. In [15] the expected reward is weighted by the parameter γ and becomes $\Sigma\, \gamma\, h(s_i, a_i, s_{i+1})$ with $0 \le \gamma \le 1$. The RL is to find a policy or an optimal strategy π^*, among the different π possible strategies in the selection of the action. Considering that an optimal policy π exists, and then the Bellman [16] optimality equation is satisfied:

$$V^{\pi^*} = V^*(s_i) = \max\left\{R(s_i, a) + \delta(\sum P(s_i \to s_{i+1}, a) V^*(s_{i+1})\right\}$$
$$\forall s \in S \quad (4)$$

Equation (4) sets the value function of the optimal policy that reinforcement learning will seek to assess:

$$V^{/*}(s) = \max V^{\pi}(s) \quad (5)$$

In Q-Learning algorithm technique [14], the agent, For any policy π and any state $s \in S$, the value of taking action *a* in state *s* under policy π, denoted $Q^{\pi}(s, a)$, is the expected discounted future reward starting in *s*, taking *a*, and henceforth following π. In this case the function (4) can also be expressed for a state-action pair:

$$Q^*(s, a) = \max\ Q^{\pi}(s, a) \quad (6)$$

Q-learning is one of the most popular reinforcement learning methods developed by Watkins [17] in 1989

years and is based on TD (0). It involves finding state-action qualities rather than just state values. Q-Learning algorithm technique is to introduce a quality function Q represents a value for each state-action pair and Q^{π} (s, a) is to strengthen estimate when starting from state s, executing action a by following a policy π: $Q^{\pi}(s, a) = E \Sigma \gamma r_i$ and Q*(s, a) is the optimal state-action pair by following policy π* if $Q^{*}(s, a) = \max Q^{\pi}(s, a)$ and if we reach the $Q^*(s_i, a_i)$ for each pair state-action then we say that the agent can reach the goal starting from any initial state. The value of Q is updated by the following equation:

$$Q^{k+1}(s_i,a_i) = Q^k(s_i,a_i) + \alpha\left[h(s_i,a_i,s_{i+1}) + \gamma \arg\max(Q^k(s_{i+1},a) - Q^k(s_i,a_i)\right] \quad (7)$$

4 Software risks and risk management perceptions

Recent perceptions about risk management from majority of software project organizations contribute to the lack of project stability. in addition to the inherent challenges posed by the nature of software projects. Ibbs and Kwak [18] identified risk management as the least practiced discipline among different project management knowledge areas. Boehm and DeMarco [19] mentioned that "our culture has evolved such that owning up to risks is often confused with defeatism". In many organizations, the tendency to 'shoot the messenger' often discourages people from bringing imminent problems to the attention of management. This attitude is the result of a misunderstanding of risk management. Boehm [20] identified 10 software risk items to be addressed by software development projects:

1. Developing the wrong user interface
2. Personnel shortfalls.
3. Real-time performance shortfalls
4. Unrealistic schedules and budgets.
5. Developing the wrong functions and properties.
6. Gold plating (adding more functionality/features than is necessary).
7. Straining computer-science capabilities.
8. Shortfalls in externally furnished components.
9. Shortfalls in externally performed tasks.
10. Continuing stream of requirements changes.

Jones [21] further presented three key software risk factors and concerns of both executives and software managers. Risk factors always generate a loss, i.e. an event or situation that causes the occurrence of a loss. The risk factor therefore constitutes the origin of a risk or a set of risks.

1. Risks associated with inaccurate estimating and schedule planning.
2. Risks associated with incorrect and optimistic status reporting.
3. Risks associated with external pressures, which damage software projects.

However, most software developers and project managers perceive risk management processes and activities as extra work, not part of their job, and more expense. Risk management tasks are therefore to be removed from project activities when the project schedule is operational. G.F. Jones in [22] mentioned that "complex computer systems can be built with a very low level of control by intelligent and motivated people." Many software development professionals believe that risk management and control prevent creativity.

5 Modeling Software Project Management

In the software project management (SPM), one of the main issues is the consistency of the project in terms of cost, completion time, quality, performance, etc. However, the most significant risk factors (causes) are of external natures that are part of the third point of the risk factors cited by [21]. Among these, there are five main risk factors:

- Bad task scheduling.
- Deficient developers.
- Technological aspect.
- Budget limitation.
- Fuzzy objectives.

In Figure 2, the rectangles are used to represent the risks, the circles to represent the risk factors and the different arcs to represent the different links.

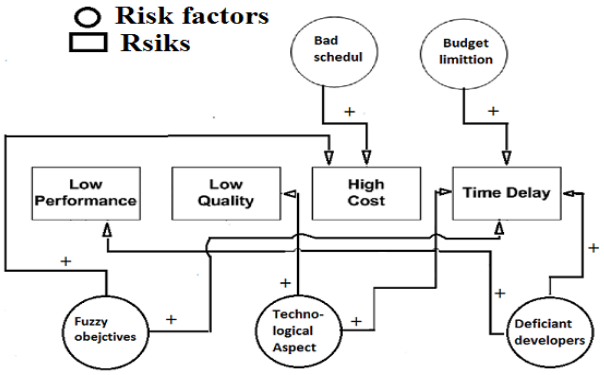


Figure 2: The main different risks, risk factors and influence links of the SPM model [23].

As we can see on Figure 2, we have one link with delay, two conditional links, and two non-linear links. Below, we will discuss in more detail.

- Weighted links with duration: the Technological Aspect risk factor, generally, will not necessarily have an immediate effect on the Time Delay risk concept, but it will affect it after a certain time or duration. Indeed, if, for example, the computers that are used to develop software are old, the immediate effect on the Time Delay concept will not be so obvious, but in the long run, it will certainly cause an increase in the risk of time delay. Note that the same observation also applies to performance, cost and quality.
- Non-linear links: If we increase the risk factor deficient developers, initially it can help meet deadlines, but if we increase more than necessary, it

might not help anymore, and could even lead to the opposite result. Therefore, the relationship here must be non-linear.

- Conditional Links: If initially there is a bad scheduling with a lack of management skills, they will affect the Time Delay risk as well as the High Cost risk. We categorize them as conditional links, because they affect only if they both occur. For example, if the scheduling of tasks is not optimal, but on the other hand, the organization is very experienced in its field to handle this type of frequent situations, the effect would certainly be different.

Once the influences between the risks and the factors are identified, we move on to the second stage, which consists in defining the fuzzy rules by considering the three attributes of the prototype schematized in Figure 2, namely the temporal delay and its conditional links. It remains to be noted here that the construction of fuzzy rules in a general way requires a detailed and complete knowledge of the field studied.

The three fuzzy rules above reflect an influence or a linear link between the time delay risk and the risk factor deficient developers.

- If the risk factor deficient developer is low Then the time delay risk is low.
- If the risk factor deficient developer is Medium Then the time delay risk is medium.
- If the risk factor deficient developer is high Then the time delay risk is high.

For relationships with time weights, we define an additional input delay variable parameter in fuzzy inference rules. For our example application, two fuzzy rules indicating the existence of the delay parameter can be as follows:

- If the technological aspect risk factor is high and the delay is short then the high cost risk is Low.
- If the technological aspect risk factor is high and the delay is long then the high cost risk is high.

Table 2 summarizes the differentiation of concepts into sensory and motor concepts of model associated with SPM.

The without learning FCM that model the SPM system of figure 2 is shown in figure 3.

	C_1	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**
C_1	0	0	0	0	0	0	0	+0.7	0
C_2	0	0	0	0	0	0	0	0	+0.6
C_3	0	0	0	0	0	0	0	+0.4	+0.4
C_4	0	0	0	0	0	0	+0.5	0	+0.8
C_5	0	0	0	0	0	+0.2	0	0	+0.2
C_6	0	0	0	0	0	0	0	0	0
C_7	0	0	0	0	0	0	0	0	0
C_8	0	0	0	0	0	0	0	0	0
C_9	0	0	0	0	0	0	0	0	0

Table 2: SPM sensory and motor concepts model.

As can be seen on Figure 3, the risk factors C_1, C_4 and C_5 that activated the high cost concept C_8 and time delay concept C_9 risks are still active despite the convergence of the non-learning FCM after 46 step. for the organization this implies that the risk remains active.

Among the concepts mentioned in Figure 2, we will discuss the concept of high cost risk and see how based on the proposed approach the organization adapts to its environment by treating this risk.

The High Cost concept is affected by risk factor concepts bad schedule and fuzzy objectives. Adaptation is translated here by the action or actions (decisions) undertaken by the organization to deal with this type of risk. One can imagine that in order to stabilize costs, we must act on the risk factors that directly affect this concept. In other words, either improve the scheduling of tasks, or seek to clarify objectives related to its field or both in parallel. This search is guided by, on the one hand, the values associated with the pairs (state, action) found in the table of the function Q, and on the other hand by the probabilities of the actions as mentioned above.

If the possible or permissible actions are no longer able to meet the needs of the organization, it is called upon to look for other mechanisms that allow it to meet its needs. For example, in our case, the organization can play on the risk factor deficient developers with which

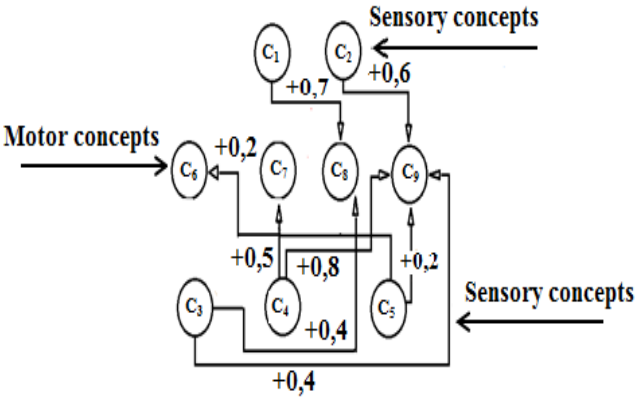


Figure 3: Without learning FCM SPM model.

Concepts	Description	Type
C_1	Bad schedule	Sensory concept
C_2	Budget Limitation	Sensory Concept
C_3	Fuzzy objectives	Sensory Concept
C_4	Technological aspects	Sensory concept
C_5	Deficient developers	Sensory concept
C_6	Low Quality	Motor Concept
C_7	Low Performance	Motor Concept
C_8	High Cost	Motor Concept
C_9	Time Delay	Motor Concept

Table 3: Without learning fuzzy cognitive maps initial matrix.

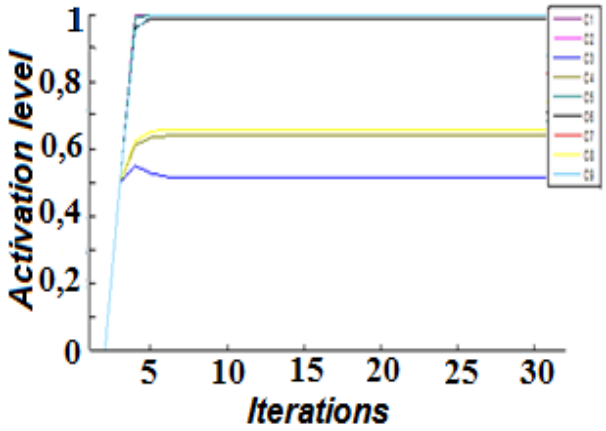


Figure 4: Evolution of activation values of FCM concepts without learning (Matlab R2014a).

Concept	Initial Value	Final Values	Activation Function	Transfert Function	Numb. of Iteration
1. C1	1	1.00	A + A.W	Sigmoid	46
2. C2	0	0,65904607			
3. C3	0	0,65904607			
4. C4	1	1.00			
5. C5	1	1.00			
6. C6	0	0,69586237			
7. C7	0	0,83569675			
8. C8	1	0,72975341			
9. C9	0	0,90204315			

Table 4: Concepts' final values without learning fuzzy cognitive maps.

the concept High Cost has no direct influence link, this action results in the creation of a connection between concept risk high cost and the concept risk factor deficient developers. This last case is represented by figure 4.

The rules that go along with the organization in the search for the optimal actions or decisions allowing it to adapt to the new environmental data in the proposed approach are of the form:

- If High Cost Risk is Active Then // depending on the factor that triggered the risk.

If state *Q (state, a_i)* already visited Then execute action a_i where action a_i is represented here by increase or decrease weight.

Otherwise select the action that has the highest probability or choose any other actions.

The two links (increase, decrease) from C_8 to C_5, figure 4, shematise that in complex systems it is difficult to know if a concept causes or decreases another concept only after several simulation of the model. It also happens that a concept can under certain conditions cause one concept and inhib it in others.

Taking into account this characteristics of complex systems, we give an another equivalent formulation of the Kosko principle of causality mentioned in [9] that is applyed in our cases study.

Defintion 1 : *(C_i causes C_j)* ***OR*** *(C_i causally decreases C_j)* ***Iff*** *($Q_i \subset Q_j$ and $\sim Q_i \subset \sim Q_j$)* ***OR*** *($Q_i \subset \sim Q_j$ and $\sim Q_i \subset Q_j$).*

Were $\subset$ stands for fuzzy set inclusion. The logical operator **OR** is used here with the reward received from the environment, which allows to select the best link, attributed to each applied weight of the two links that connect the concept C_i and C_j and it is defined as follows:

$$r_{\max} = Max(r_{increase}, r_{decrease})$$

therefore, definition 1 is written in our case study as follows:

Defintion 2 : *(C_i causally increases C_j)* ***iff*** *($Q_i \subset Q_j$ and $\sim Q_i \subset \sim Q_j$)* ***and*** $r_{Max} = r_{increase}$

(C_i causally decreases C_j) ***Iff*** *($\sim Q_i \subset Q_j$ and $Q_i \subset \sim Q_j$)* ***and*** $r_{Max} = r_{decrease}$

Based on the theoretical aspects described above, the pseudo code of Algorithm 1 summarizes our aproach [24].

Algorithm 1: Pseudo code of the proposed approach

Step 1: Read the vector $A^{(k)}$ and weight matrix W

Step 2: Calculate the output vector $A^{(k+1)}$:

$$A^{k+1} = f(A^k + \sum A^k W)$$

Step 3: Apply the transfer function f to the output vector $A^{(k+1)}$

Step 4: Among active concepts, choose the one that has the highest value of the Q function, if not the highest in probability.

Step 5: calculate the new output vector (output concepts) $A^{(k+1)}$

Step 6: Depending on the response of the environment:

If r = 1 / / Award

(Updating the probability P_{ij} and the Q value)

$$Q^{k+1}(s_i, a_i) = Q^k(s_i, a_i) + \alpha[1 - Q^k(s_i, a_i)]$$

$$W^{k+1}(C_i, C_j) = W^k(C_i, C_j)$$

$$P^{k+1}(a_i) = P^k(a_i) + \beta[1 - P^k(a_i)]$$

If r = o / / Penalty

(Updating the probability P_{ij}, the weight of the connection and the value of Q)

$$Q^{k+1}(s_i, a_i) = (1 - \alpha) Q^k(s_i, a_i)$$

$$W^{k+1}(C_i, C_j) = W^k(C_i, C_j)$$

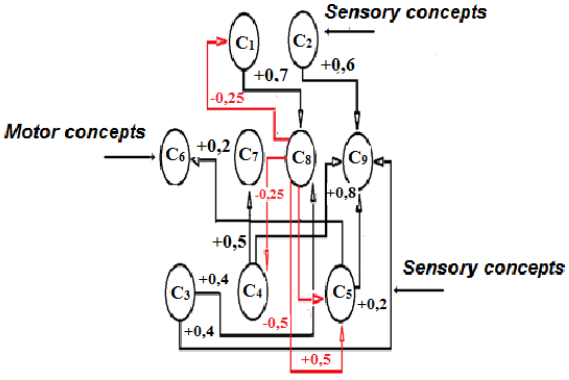


Figure 5: The SPM Reinforcement learning fuzzy cognitive maps model.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
C_1	0	0	0	0	0	0	0	+0.7	0
C_2	0	0	0	0	0	0	0	0	+0.6
C_3	0	0	0	0	0	0	0	+0.4	+0.4
C_4	0	0	0	0	0	0	+0.5	0	+0.8
C_5	0	0	0	0	0	+0.2	0	0	+0.2
C_6	0	0	0	0	0	0	0	0	0
C_7	0	0	0	0	0	0	0	0	0
C_8	-0.25	0	0	-0.25	±0.5	0	0	0	0
C_9	0	0	0	0	0	0	0	0	0

Table 5: Reinforcement learning FCM (RL-FCM) Initial matrix.

$$P^{k+1}(a_i) = (1-\beta)\, P^k(a_i)$$

Step 7: If the termination conditions are realized Stop. Otherwise go to Step 2.

Thereafter, the organization evaluates its actions towards its environment by the feedbacks of the latter (in the form of positive or negative answers) by updating its decision-making policy that allows it to adapt and improve its behavior towards its economic and social partner.

In the next paragraph 6, we discuss the results obtained after simulation of the SPM model in the proposed approach and in the conventional FCMs approach in order to make a comparison between the two approaches to show the effectiveness of the approach proposed in this paper.

6 Experimental results

The simulation of the prototype associated with the SPM model of figure 4 is carried out under MATLAB R2014a. The two scenarios are represented by the results obtained in table 7 in the case where the concept C_8 decreases the concept C_5 and in table 8 where the concept C_8 increases the concept C_5. It can be seen that the best result is obtained in the case where the C_8 concept decreases the C_5 concept.

The following table 5 represents the initial matrix of the reinforcement learning fuzzy cognitive maps RL-FCM that model the software project studied in this paper and summarizes the different weights between the concepts of the map especially the links that express the behavioral adaptation, in particular the concept High cost C_8 and its links with the concepts Bad schedule C_1 with w_{81}*=-0.25*, Technological aspects *C4 with* w_{84}*=-0.25* and deficient developers C_5 with $w_{85incfrease}$=+0.50 in case if C_8 increases C_5 and with w_{85}decrease=–0.50 in case where C_8 decreases C_5.

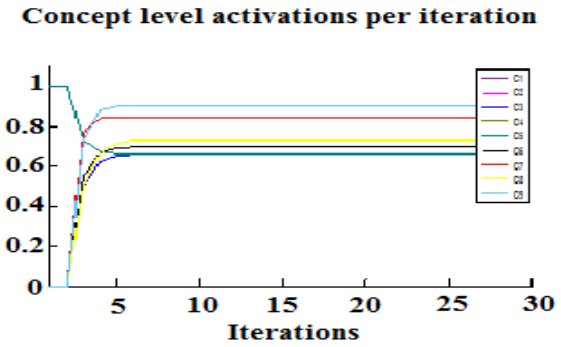


Figure 6: Concept values evolution, the reinforcement learning FCM converge in 27 steps.

Table 6 gives the probability and the function Q quantity values before the simulation, the initial values, and after the simulation, the final values, obtained by application of our algorithms 1 while taking into account the natures of the different weights described above paragraph 5. In the next simulation, figure 6, our simulator will consider the model with the weight that will decrease the concept of deficient developers C5 from high cost concept C8 as being the action taken by the organization to adapt to its environment.

7 Conclusion

In this paper we presented an approach in which there is a connection between reinforcement learning and fuzzy cognitive maps for studying risk analysis and management in software projects. The nature of software projects generates many risks that must be managed carefully to avoid the project's loss.

In this work the What-If scenario analysis technique is automated, used and has been effectively applied to a variety of processes. It can be useful in other processes per example in job shop scheduling with mechanical systems such as production machines. The results of the analysis are immediately available for managers and usually can be applied quickly. On behalf the firm to be able to make an adequate decision, it has to compare the simulation of its SPM model, in our case study, with two links (increase, decrease) from concept C_8 to concept C_5. Similarly, in our proposed approach, another's situations can arise, in which concept influences another concept with two weights (increase, increase) or with two weights (decrease, decr**e**ase). In Also we have presented a new formulation of Kosko causality principle in which one concept increases or decreases another concept according to environmental conditions. The work is realized with MATLAB R2014 platform.

8 Acknowledgments

This paper is the result of a research project titled "Modeling and implementing of complex systems" which was supported by the MISC laboratory of abdelhamid mehri University, Algeria. We thank all the laboratory teams and responsible for their cooperation and supports.

Action a_i	Initial Probability $P(a_i)$	Final Probability $P(a_i)$	$Q(s_i, a_i)$	Initial Value	Final Value
(C_8,C_1)	0,25	0,25	$Q(C_8,C_1)$	0	0,25
(C_8,C_4)	0,25	0,25	$Q(C_8,C_4)$	0	0,25
(C_8,C_5) ***increase***	0,25	0,125	$Q(C_8,C_5)$ ***increase***	0	0
(C_8,C_5) ***decrease***	0,25	0,625	$Q(C_8,C_5)$ ***decrease***	0	0,50

Table 6: Values of the actions probabilities and Q-function with α, β=0.5.

Concept	Initial Values	Final Values	Activation Function	Transfert Function	Number of Itérations
1. C_1	1	0,59937409	A+AW	Sigmoid	24
2. C_2	0	0,65904607			
3. C_3	0	0,65904607			
4. C_4	1	0,59937409			
5. C_5	1	0,69329384			
6. C_6	0	0,74556292			
7. C_7	0	0,76593465			
8. C_8	1	0,78606504			
9. C_9	0	0,65904607			

Table 7: Simulation Results with decrease weight from C_8 to C_5 as the best weight.

Concept	Initial Values	Final Values	Activation Function	Transfert Function	Number of Itérations
1. C_1	1	1	A+AW	Sigmoid	32
2. C_2	0	0,65904607			
3. C_3	0	0,65904607			
4. C_4	1	0,97252654			
5. C_5	1	0,85343066			
6. C_6	0	0,74556292			
7. C_7	0	0,76593465			
8. C_8	1	0,78606504			
9. C_9	0	0,65904607			

Table 8: Simulation Results with increase weight from C_8 to C_5.

References

[1] L. Zadeh, (1965). Fuzzy sets. Inf. Contr., vol. 3, no. 8, pp. 338–353. https://doi.org/10.2307/2272014

[2] Angeline, P.J., Fogel, D.B. (1997). An evolutionary program for the identification of dynamical systems. SPIE Aerosence 97, Symp. On Neural Networks, S.K. Rogers and D. Ruck (eds.), Vol. 3077, 409-417. https://doi.org/10.1117/12.271503

[3] R.T Futrell, L.I Shafer & D.F Shafer (2001). Quality software project management. Prentice Hall PTR. January 24, 2002. ISBN: 0-13-091297-2

[4] C. Samantra, S. Datta., S.S Mahapatra. (2017) Fuzzy based risk assessment module for metropolitan construction project: an empirical study. Eng. Appl. Artif. Intell.; 65:449–464. https://doi.org/10.1016/j.engappai.2017.04.019

[5] O.Taylan, A.O Bafail., R.M Abdulaal., M.R Kabli. (2014) Construction projects selection and risk assessment by fuzzy AHP and fuzzy TOPSIS methodologies. Appl. Soft Comput.; 17:105–116. https://doi.org/10.1016/j.asoc.2014.01.003

[6] A.Dziadosz, M Rejment. (2015) Risk analysis in construction project-chosen methods. Proc. Eng.; 122:258–265. https://doi.org/10.1016/j.proeng.2015.10.034

[7] C. Muriana, G. Vizzini, (2017) Project risk management: a deterministic quantitative technique for assessment and mitigation. Int. J. Proj. Manag.; 35(3):320–340. https://doi.org/10.1016/j.ijproman.2017.01.010

[8] B.W. Boehm and T. DeMarco, Software risk management. IEEE Software 14 (3), (1997), 17–19. https://doi.org/10.1109/ms.1997.589225

Kosko B. (1986), Fuzzy Cognitive Maps, International Journal Man-Machine Studies, 24:65-75. https://doi.org/10.1016/s0020-7373(86)80040-2

[9] Axelrod Robert (1976). Structure of decision. Princeton university press, Princeton, NewJersy. M. Carr, S. Konda, I. Monarch, C. Walker and F. Ulrich, Taxonomy-Based Risk Identification, (1993). https://doi.org/10.21236/ada266992

[10] Maikel Leon1, Ciro Rodriguez1, Maria M. Garcia1, Rafael Bello1 and Koen Vanhoof (2010). Fuzzy Cognitive Maps For Modeling Complex Systems. G. Sidorov et al. (Eds.): MICAI 2010, Part I, LNAI 6437, pp. 166 – 174, © Springer-Verlag Berlin Heidelberg 2010

[11] R. Sutton & A.G. Barto, (2015). Reinforcement Learning: An Introduction. A Bradford Book. Second Edition. 2014-2015 The MIT press Cambridge, Massachusetts London, England.

[12] E.A. Jasmin, T.P. Imthias Ahamed, V.P. Jagathy Raj (2011). Reinforcement Learning approaches to Economic Dispatch problem. Elsevier. https://doi.org/10.1016/j.ijepes.2010.12.008

[13] L.P. Martin (2014). Markov Decision Processes: Discrete Stochastic and Dynamic Programming. Wiley series in probability and statistics. ISBN: 978-1-118-62587-3

[14] Thomas J. Sargent, (2004). Recursive Macroeconomic Theory. Second edition Lars Ljungqvist Stockholm School of Economics. New York University and Hoover Institution. The MIT Press Cambridge, Massachusetts London, England. ISBN: 9780262122740.

[15] C. Watkins (1989). Learning from Delayed Rewards. Ph.D. thesis, King's College, Cambridge, England.

[16] Ibbs, C. W., and Kwak, Y. H. (2000). Assessing project management maturity. Project Management Journal, pp 32–43. https://doi.org/10.1177/875697280003100106

[17] Boehm, B.W., DeMarco, T. (1997). Software risk management. IEEE Software 14 (3), 17–19. https://doi.org/10.1109/ms.1997.589225

[18] Boehm, B.W., 1991. Software risk management principles and practices. In IEEE Software 8 no 1, pp 32–41. Jan. 1991, Doi:10.1109/52.62930.

[19] C. Jones, (1998). Minimizing the risks of software development. Cutter IT Journal 11 (6), 13–21.

[20] Jones, G.F., (2001). What is different about it risks. In: 2001 INCOSE Proceedings of a Symposium on Risk Management..

[21] Beatrice Lazzerini, Member, IEEE, and Lusine Mkrtchyan (2011) Analyzing Risk Impact Factors Using Extended Fuzzy Cognitive Maps. IEEE SYSTEMS JOURNAL, VOL. 5, NO. 2, https://doi.org/10.1109/jsyst.2011.2134730

[22] A. Tlili, S. Chikhi, (2016). Natural Immune System Response as Complexe Adaptive System Using Learning FCMs. IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 4, No. 3, pp 95-104. ISSN: 2252-8938. https://doi.org/10.11591/ijai.v5.i3.pp95-104.

Study of Fuzzy Distance Measure and Its Application to Medical Diagnosis

Taruna and H. D. Arora
Department of Mathematics, Amity University, Noida, India
E-mail: hdarora@amity.edu

Vijay Kumar
Manav Rachna International Institute of Research & Studies, Faridabad, India

Keywords: fuzzy sets, directed divergence, fuzzy relative information measures, multi criteria decision making

Received: June 12, 2020

Ambiguity has an important part in the contrary observations around peripheral world. Entropy is imperative for measuring uncertain information which was first introduced by Shannon (1948) to measure the uncertain degree of randomness in a probability distribution. Fuzzy information measures have been applied widely in the area of decision making. Jensen–Shannon divergence is a useful distance measure in the probability distribution space. The present communication we propose a way of measuring the difference between two fuzzy sets by means of a function, called divergence. In addition, study of their detailed properties for its validity is also discussed. The applications of these newly developed fuzzy divergence measure have been provided to the optimal decision making based on the weights of alternatives. Numerical verification has been illustrated to demonstrate the proposed method for solving optimal decision-making problem under fuzzy environment.

Povzetek: Za probleme medicinske diagnostike je narejena študija nedorečenosti in entropije.

1 Introduction

Information theory advanced out of mathematical studies of the problems linked with communication, storage, and transmission of massages. It originated from the fundamental paper "The Mathematical Theory of Communication" published by Shannon [1]. Shannon developed mathematical schemes for quantitatively defining the ideas of facts and proved several very trendy outcomes with deeper effects. Various generalizations of Shannon entropy were studied by Renyi [2], Arimoto [3], Sharma and Taneja [4], De Luca and Termini [5], Kaufmann [6] and Peerzada et al. [7]. Uncertainty and fuzziness are the primary nature of human wondering and of many real-world objectives. Fuzziness is found in our decision, in our language and inside the way we process information. The fundamental use of information is to get rid of uncertainty and fuzziness.

In reality, we degree data furnished by using the quantity of probabilistic uncertainty eliminated in an experiment and the measure of uncertainty eliminated is also called as a measure of information while degree of fuzziness is the measure of vagueness and ambiguity of uncertainties. The theory of fuzzy sets (FSs) developed by Zadeh [8], as a generalization of classical set theory, for representing vague and indistinct phenomena. This idea serves as an effective tool for know-how of the behaviour of humanistic systems in which human judgment, perceptions and feelings play a critical role. In fuzzy set concept, the entropy is described as a degree of fuzziness which expresses the quantity of ambiguity or problem in we decide whether an element belongs to a set or not. Bhandari and Pal [9] extended the probabilistic exponential entropy idea of Pal and Pal [10] to the fuzzy phenomenon. Kapur [11] discussed fuzzy measures uncertainty due to fuzziness of information.

In fuzzy context, several measures have been proposed to measure the degree of difference between two fuzzy sets. Measure of fuzzy divergence between two fuzzy sets gives the difference between two fuzzy sets and this measure of difference between two fuzzy sets is called the fuzzy divergence measure.

The similarity measure is important tools that can be used in decision-making problem to deal with uncertainty through IFS theory. Various distance measures have been proposed by different researchers. It has been observed that different distance measure produces different values while measuring the distance degree between two IFSs. Also, sometimes existing distance measures are not able to give an appropriate and convenient result for a pair of IFSs. For this reason, it is always necessary to derive advanced measures for better decision making.

To explain the distinction among fuzzy sets, the distance measure was set up and was regarded as dual model of correspondence measure. Many researchers, such as Yager [12], Kosko [13] and Kaufmann [6] had used distance measure to define fuzzy entropy. Several recent methods of fuzzy entropy generated by distance measure and properties of distance measure were extended by Fan et al. [14]. The distances among two fuzzy subsets

on a fuzzy subset of R+ were characterized by Dubois and Prade [15]. Thus, the set of distances between two sets was simplified whereas the shortest distance between two crisp sets was not simplified. The shortest distance among two fuzzy sets as a density function on non-negative reals was described by Rosenfeld [16]. Thus, related to Kullback and Leibler [17] probabilistic measure of divergence, the subsequent measure of fuzzy directed divergence was initiated by Bhandari and Pal [9]. Montes et al. [18] proposed an axiomatic form to measure the difference between fuzzy sets and we study in detail the case of local divergence.

Luo and Zhao [19] gave the algorithms for pattern recognition and use it to solve medical diagnosis problems. Gupta and Tiwari [20] and Datta and Goala [21] proposed cosine similarity measure for intuitionistic and interval-valued intuitionistic fuzzy sets using an advanced distance measure on intuitionistic fuzzy sets.

2 Preliminaries

The model of entropy was initiated to arrange numerical quantity of ambiguity.

Shannon [1] originated a quantity

$$H(P) = -\sum_{i=1}^{n} p_i log p_i \quad (1)$$

for the uncertainty of a probability distribution $(p_1, p_2, p_3 \dots p_n)$ and called it entropy.

A fuzzy set $\tilde{A}$ in a finite Universe of discourse X = $(x_1, x_2, x_3 \dots x_n)$ is given by

$$\tilde{A} = \{\langle x, \mu_{\tilde{A}}(x)\rangle | x \in X\} \quad (2)$$

where $\mu_{\tilde{A}}$: X $\rightarrow [0,1]$ *is the* membership function of $\tilde{A}$.The number

$\mu_{\tilde{A}}(x)$ *describes the degree of belongingness of* $x \in X$ *in* $\tilde{A}$.

De Luca and Termini [5] defined fuzzy entropy for a fuzzy set A corresponding to Shannon Entropy (1948) as

$$\mathrm{H}(\tilde{A}) = -\frac{1}{n}\sum_{i=1}^{n}\left[\mu_{\tilde{A}}(x_i) log\left(\mu_{\tilde{A}}(x_i)\right) + \left(1 - \mu_{\tilde{A}}(x_i)\right) log\left(1 - \mu_{\tilde{A}}(x_i)\right)\right] \quad (3)$$

Motivated by the fundamental properties of directed divergence, Kapur [11] explained the concept of fuzzy directed divergence as follows: The directed divergence of fuzzy set A from the fuzzy set B is a function D (A; B) that should comply with the subsequent requirements which satisfies the following conditions:

1. $D(A;B) \geq 0$
2. $D(A;B) = 0\ iff\ A = B$
3. $D(A;B) \geq 0$ is a convex function in (0,1)
4. $D(A;B) \geq 0$ should not change, when $\mu_A(x_i)$ is changed to $1 - \mu_A(x_i)$ and $\mu_B(x_i)$ is changed to $1 - \mu_B(x_i)$.

Now, corresponding to Kullback – Leibler's [17] measure of divergence, Bhandari and Pal [9] proposed a fuzzy divergence measure A and B given by

$$D(A;B) = \frac{1}{n}\sum_{i=1}^{n}\left[\mu_A(x_i) log \frac{\mu_A(x_i)}{\mu_B(x_i)} + (1 - \mu_A(x_i)) log \frac{(1-\mu_A(x_i))}{(1-\mu_B(x_i))}\right] \quad (4)$$

Later, Shang and Jiang [22] was pointed out that the expression (4) has some limitations, i.e., if $\mu_A(x_i)$ approaches to 0 or 1, then its value tends to ∞. Therefore they proposed a modified version of fuzzy divergence measure (4), given as

$$J(A;B) = \sum_{i=1}^{n}\left[\mu_A(x_i) log \frac{\mu_A(x_i)}{\frac{\mu_A(x_i)+\mu_B(x_i)}{2}} + (1 - \mu_A(x_i)) log \frac{(1-\mu_A(x_i))}{1-(\frac{\mu_A(x_i)+\mu_B(x_i)}{2})}\right] \quad (5)$$

Corresponding to Kerridge [23] inaccuracy measure, Verma and Shrama [24] define a measure of inaccuracy of fuzzy set B with respect to fuzzy set A, as

$$I(A;B) = -\frac{1}{n}\sum_{i=1}^{n}[\mu_A(x_i) log \mu_B(x_i) + (1 - \mu_A(x_i)) log(1 - \mu_B(x_i))] \quad (6)$$

Ohlan [25] proposed a parametric generalized measure of divergence between two fuzzy sets A and B corresponding to Taneja [26] as

$$L_t(A,B) = \sum_{i=1}^{n} \frac{\left(\mu_A(x_i)+\mu_B(x_i)\right)^2}{2^t} \times \left[\frac{\left(\mu_A(x_i)+\mu_B(x_i)\right)^t}{\sqrt{\mu_A(x_i)\mu_B(x_i)}^{t+1}} + \frac{\left(2-\mu_A(x_i)+\mu_B(x_i)\right)^t}{\sqrt{(1-\mu_A(x_i)).(1-\mu_B(x_i))}^{t+1}}\right] \quad (7)$$

$$t = 0,1,2, \dots$$

The generalized measure of fuzzy directed divergence of order α and type β is given by Arora and Dhiman [27] as

$$D_\alpha^\beta(A:B) = \frac{1}{(1-\alpha)\beta}\sum_{i=1}^{n}\left[\left\{\frac{\mu_A(x_i)^{\alpha\mu_A(x_i)}}{\mu_B(x_i)^{\alpha\mu_B(x_i)}} + \frac{(1-\mu_A(x_i))^{\alpha(1-\mu_A(x_i))}}{(1-\mu_B(x_i))^{\alpha(1-\mu_B(x_i))}}\right\}^\beta - 2^\beta\right]$$

$$\text{where } \alpha > 0, \alpha \neq 1, \beta \neq 0. \quad (8)$$

Prakash and Kumar [28] proposed a new fuzzy divergence measure of fuzzy set B with respect to fuzzy set A, as follows:

$$K(A,B) = -log\left(\frac{1+\frac{1}{n}\sum_{i=1}^{n}\left[\sqrt{\mu_A(x_i)\mu_B(x_i)}+\sqrt{(1-\mu_A(x_i)).(1-\mu_B(x_i))}\right]}{2}\right) \quad (9)$$

Kumari et al. [29] proposed Weighted Fuzzy Exponential J-Divergence as

$$H(A;W) = \frac{1}{n(\sqrt{e}-1)}\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} w_{ij}\left[(\mu_A f_{ij})e^{1-\mu_A f_{ij}} + \left(1 - (\mu_A f_{ij})e^{1-\mu_A f_{ij}} - 1\right]\right. \quad (10)$$

where $\mu_A f_{ij}$ is the membership values of the pixels in the image and f_{ij} is the (i, j) th pixel of the image A.

Tiwari and Gupta [30] proposed entropy measures and erived relation between distance, entropy, and similarity measures for IvIFSs.

3 Proposed Fuzzy Distance Measure

Let X = $(x_1, x_2, x_3 \ldots x_n)$be the universe of discourse. Let A= $\{\langle x_i, \mu_A(x_i)\rangle | x_i \in X\}$ and B = $\{\langle x_i, \mu_B(x_i)\rangle | x_i \in X\}$ be two fuzzy sets. Then we propose new distance measure as follows:

$$D(A;B) = \frac{2}{n}\sum_{i=1}^{n} \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}}{1+ sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}} \quad (11)$$

Theorem 3.1. The fuzzy distance measure $D(A;B)$ defined in equation (11) is a valid measure of fuzzy divergence.

Proof. All the necessary four conditions to be a distance measure are satisfied by the new distance measure which are as follows:

(P1) $0 \le D(A;B) \le 1$

(P2) $D(A;B)$ = 0 if and only if $\mu_A(x_i) = \mu_B(x_i)$.

(P3) $D(A;B) = D(B;A)$

(P4) If A, B and C be three fuzzy sets, then the distance measure satisfies the triangular inequality, i.e., $D(A;C) \le D(A;B) + D(B;C)$.

Proof: We will now prove these conditions one by one:

(P1) As we know that, A= $\{\langle x_i, \mu_A(x_i)\rangle | x_i \in X\}$ for degree of membership $0 \le \mu_A(x_i) \le 1$.

That is, for A= $\{\langle x_i, \mu_A(x_i)\rangle | x_i \in X\}$ and B= $\{\langle x_i, \mu_B(x_i)\rangle | x_i \in X\}$

$0 \le |\mu_A(x_i) - \mu_B(x_i)| \le 1 \Rightarrow 0 \le \frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)| \le \frac{\pi}{2}$

$$\Rightarrow 0 \le sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\} \le 1 \quad (12)$$

$$\Rightarrow 0 \le 1 + sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\} \le 2 \quad (13)$$

From (12) and (13), we have

$$0 \le 2.\frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\}}{1 + sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\}} \le 1$$

$$\Rightarrow \ 0 \le \frac{2}{n}\sum_{i=1}^{n} \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}}{1+ sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}} \le 1 \Rightarrow 0 \le D(A;B) \le 1.$$

(P2) $D(A;B)$ = 0

$$\Leftrightarrow \quad \frac{2}{n}\sum_{i=1}^{n} \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\}}{1 + sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\}} = 0$$

$$\Leftrightarrow \quad \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\}}{1 + sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\}} = 0$$

$$\Leftrightarrow sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\} = 0$$

$$\Leftrightarrow |\mu_A(x_i) - \mu_B(x_i)| = 0$$

$$\Leftrightarrow \mu_A(x_i) = \mu_B(x_i) \Leftrightarrow A = B$$

Therefore, $D(A;B)$ = 0 if and only if $\mu_A(x_i) = \mu_B(x_i)$.

(P3) As

$$D(A;B) = \frac{2}{n}\sum_{i=1}^{n} \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\}}{1 + sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\}}$$

$$= \frac{2}{n}\sum_{i=1}^{n} \frac{sin\left\{\frac{\pi}{2}|\mu_B(x_i) - \mu_A(x_i)|\right\}}{1 + sin\left\{\frac{\pi}{2}|\mu_B(x_i) - \mu_A(x_i)|\right\}}$$

$= D(B;A)$

To prove the triangular inequality, i.e., $D(A;C) \le D(A;B) + D(B;C)$, we have to prove that

(P4) In order to prove fourth necessary condition, we must first prove the identity

$\sin(A+B) \le \sin(A) + \sin(B)$, where A and B are acute angles.

or in other words, we have to show that

$$\sin(A) + \sin(B) - \sin(A+B) \ge 0$$

$$\Rightarrow \sin A + \sin B - \sin A.cosB - cosA.sinB \ge 0$$

$$\Rightarrow \sin A(1 - cosB) + \sin B\,(1 - cosA) \ge 0$$

Since A and B are acute angles, therefore, $\sin A, (1 - cosB), \sin B, (1 - cosA)$ are all positive and hence the identity holds good.

Now, consider A = $\{\langle x, \mu_A(x)\rangle | x \in X\}$, B = $\{\langle x, \mu_B(x)\rangle | x \in X\}$ and C = $\{\langle x, \mu_C(x)\rangle | x \in X\}$ be three fuzzy sets.

As,

$|\mu_A(x_i) - \mu_C(x_i)| \le |\mu_A(x_i) - \mu_B(x_i)| + |\mu_B(x_i) - \mu_C(x_i)|$

$(\because Inequality\ of\ real\ numbers)$

$$\Rightarrow sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_C(x_i)|\right\} \le sin\frac{\pi}{2}[\{|\mu_A(x_i) - \mu_B(x_i)| + |\mu_B(x_i) - \mu_C(x_i)|\}]$$

$$\Rightarrow sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_C(x_i)|\right\} \le sin\left\{\frac{\pi}{2}|\mu_A(x_i) - \mu_B(x_i)|\right\} + sin\left\{\frac{\pi}{2}|\mu_B(x_i) - \mu_C(x_i)|\right\}$$

Also.

$$1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}\leq 1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}$$

$$\Rightarrow \frac{1}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}}\geq \frac{1}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}$$

$$\Rightarrow 1-\frac{1}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}}\leq 1-\frac{1}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}$$

$$\Rightarrow \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}}\leq \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}$$

$$\Rightarrow \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}}\leq \frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}+\frac{sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}$$

$$\Rightarrow \frac{2}{n}\sum_{i=1}^{n}\frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_C(x_i)|\right\}}\leq \frac{2}{n}\sum_{i=1}^{n}\frac{sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}+\frac{2}{n}\sum_{i=1}^{n}\frac{sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}{1+sin\left\{\frac{\pi}{2}|\mu_A(x_i)-\mu_B(x_i)|\right\}+sin\left\{\frac{\pi}{2}|\mu_B(x_i)-\mu_C(x_i)|\right\}}$$

$$\Rightarrow D(A;C)\leq D(A;B)+D(B;C).$$

Hence, the proposed distance measure satisfies all the necessary properties.

4 Application of Proposed Fuzzy Measure to Medical Diagnosis

In a classical problem of medical diagnosis, assume that if a doctor needs to diagnose some of patients "$P=\{Alex, Chris,\ James,\ Mike\ and\ Shawn\}$" under some defined diagnosis "$D=\{Viral\ fever\ (VF),\ Malaria(M),\ Typhoid(T),\ Stomach\ problem(SP) and\ Chest\ problem(CP)\}$" and a set of symptom "$S=\{Temperature\ (Temp.),\ Headache(H),\ Stomach\ pain\ (S.Pain),\ Cough(C) and\ Chest\ pain(CP)\}$". The following tables (table 1 and table 2) serve the purpose of the proposed computational application:

In view of the table 3, it is being concluded that "$Alex$" is suffering from "$Malaria$"; "$Chris$", "$James$" and "$Mike$" are suffering from "$Chest\ problem$" and "$Shawn$" is suffering from "$Viral\ fever$".

	Temp.	*H*	*S. Pain*	*C*	*CP*
VF	0.4	0.4	0.3	0.1	0.6
M	0.4	0.2	0.5	0.6.	0.7
T	0.3	0.6	0.1	0.5	0.7
SP	0.2	0.3	0.7	0.4	0.3
CP	0.4	0.6	0.5	0.4	0.6

Table 1: Fuzzy membership values for diseases and their symptoms.

	Temp.	*H*	*S. Pain*	*C*	*CP*
Alex	0.8	0.1	0.7	0.6	0.4
Chris	0.3	0.6	0.8	0.4	0.7
James	0.7	0.7	0.6	0.6	0.4
Mike	0.5	0.4	0.5	0.4	0.6
Shawn	0.5	0.6	0.3	0.2	0.9

Table 2: Fuzzy membership values for patients and related symptoms.

	Alex.	*Chris*	*James*	*Mike*	*Shawn*
VF	0.680965	0.493065	0.634603	0.273332	**0.327418**
M	**0.421368**	0.421368	0.469442	0.296955	0.538941
T	0.689085	0.24253	0.546709	0.444991	0.40803
SP	0.42172	0.381062	0.516246	0.398187	0.670988
CP	0.596858	**0.233026**	**0.42181**	**0.148478**	0.367724

Table 3: Values of Fuzzy relative entropy measure for the patients and the likely diseases.

This is because smaller value of the patient against each distance measure indicates the more probability of having the disease.

5 Comparative Study

Jain and Kumar [31] proposed the intuitionistic fuzzy based trigonometric entropy as:

$$E_{IF}(A) = \frac{1}{n}\sum_{i=1}^{n}\left[cos\frac{\pi}{2}(|{\mu_A}^2(x_i) - {\nu_A}^2(x_i)|)\right]$$

The fuzzy version of the entropy is:

$$E(A) = \frac{1}{n}\sum_{i=1}^{n}\left[cos\frac{\pi}{2}(|{\mu_A}^2(x_i) - {\mu_B}^2(x_i)|)\right]$$

From the table, it is concluded that the larger value in the column is the decision value.

Wei et al. [32] proposed the generalized fuzzy entropy as:

$$H(A) = \frac{1}{n}\sum_{i=1}^{n}\left[\left\{cos\,\pi\left(\frac{\mu_A(x_i) - \nu_A(x_i)}{4}\right) - 1\right\} \times \frac{1}{\sqrt{2}-1}\right]$$

From the table, it is concluded that the smaller value in the column is the decision value.

	P1	*P2*	*P3*	*P4*	*P5*
D1	0.862945	0.909299	0.890581	0.986209	**0.956661**
D2	**0.905346**	0.927966	0.896657	0.980538	0.917587
D3	0.817137	0.907895	0.899024	0.963769	0.940134
D4	0.905073	0.938003	0.898832	0.956289	0.815088
D5	0.882851	**0.958308**	**0.947068**	**0.988228**	0.940318

Table 4: Values of Fuzzy entropy measure for Jain and Kumar (2020).

	P1	*P2*	*P3*	*P4*	*P5*
D1	0.103199	0.058957	0.082638	0.020753	**0.02224**
D2	**0.044361**	0.044368	0.064859	0.014851	0.06058
D3	0.140823	0.072568	0.076627	0.038459	0.031143
D4	0.065938	0.039909	0.069238	0.03408	0.108776
D5	0.078141	**0.016301**	**0.028178**	**0.007425**	0.026691

Table 5: Values of Fuzzy entropy measure for Wei et al. (2012).

6 Conclusions

In this paper, we have proposed a relative distance measure for fuzzy sets. Proof of its validity is also considered through numerical computations. Some of the essential properties of the measure are also studied. It has been observed that this measure is more flexible in terms of their previous derived measures. Application of this measure is also studied in medical diagnosis to check its legitimacy. Also, from the table 4 and 5, it is concluded that the result obtained from the proposed entropy is similar with the results of the existing entropies (shown in table nos.), which validates the fact that the proposed entropy is valid and have applications across disciplines.

References

[1] C.E. Shannon, "A Mathematical Theory of Communication" Bell Syst. Tech. Journal, vol. 27(379-423), pp. 623-656, 1948. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[2] Renyi A: On measure of entropy and information, Proceeding Fourth Berkely Symposium on Mathematical Statistics and probability, University of California Press, 1, pp. 547-561, 1961.

[3] S.C. Arimoto,, "Information -Theoretic Considerations on Estimation Problems". Information and Control, vol. 9, pp. 181-190, 1971. https://doi.org/10.1016/S0019-9958(71) 90065-9

[4] B.D. Sharma, and I.J. Taneja, "Entropies of Type α, β and Other Generalized Measures of Information Theory", Mathematika, vol. 22, pp. 202-215, 1975. https://doi.org/10.1007/BF01899728

[5] A. De Luca and S. Termini, "A Definition of a Non-Probabilistic Entropy in the Setting of fuzzy sets theory", Information and Control, vol. 20, pp. 301-312, 1972. https://doi.org/10.1016/S0019-9958(72)90199-4

[6] A. Kaufmann, "Fuzzy subsets: Fundamental Theoretical Elements", Academic Press, New York, 3, 1980. https://doi.org/10.1109/TSMC.1977.4309751

[7] S. Peerzada, S.M. Sofi and R. Nisa, "A New Generalized Fuzzy Information Measure and its Properties", International Journal of Advance Research in Science and Engineering, vol. 6, no. 12, pp. 1647-1654, 2017.

[8] L.A. Zadeh, "Fuzzy Sets", Information and Control, vol. 8, pp. 338-353, 1965. https://doi.org/10.1016/S0019-9958(65)90241-X

[9] D. Bhandari and N.R. Pal, "Some New Information Measures for Fuzzy Sets", Information Science, vol. 67, pp. 204-228, 1989. https://doi.org/10.1016/0020-0255(93)90073-U

[10] N.R. Pal and S.K. Pal, "Object Background Segmentation Using New Definition of Entropy", Proc. Inst. Elec. Eng., vol. 13, pp. 284-295, 1989. https://doi.org/10.1049/ip-e.1989.0039

[11] J.N. Kapur, "Measures of Fuzzy Information", Mathematical Science Trust Society, vol. 2, no. 2, pp 73-76, 1997.

[12] R.R. Yager, "On the Measure of Fuzziness and Negation, Part I: Membership in the Unit Interval" International Journal of General Systems, vol. 5, no. 4, pp. 221-229, 1979. https://doi.org/10.1080/03081077908547452

[13] B. Kosko, "Fuzziness vs. probability" International Journal of General Systems, vol. 17, pp. 211-240, 1990. https://doi.org/10.1080/03081079008935108

[14] J.L. Fan, Y.L. Ma and W.X. Xie, "On some properties of distance measure", Fuzzy Sets and

Systems, vol. 117, pp. 355-361, 2001. https://doi.org/10.1016/ S0165-0114(98)00387-X
[15] D. Dubois and H. Prade, "On distances between fuzzy points and their use for plausible reasoning", Proc. IEEE Int. Conf. on Cybernatics and Society, Bombay, New Delhi, pp. 300-303, 1993.
[16] A. Rosenfeld, "Distance between fuzzy sets", Pattern Recognition Letters, vol. 3, pp. 229-231, 1985. https://doi.org/10.1016/0167-8655(85)90002-9
[17] S. Kullback and R.A. Leibler, "On Information and Sufficiency", Ann. Math. Stat., vol. 22, pp. 79-86, 1951. https://doi.org/10.1214/aoms/1177729694
[18] S. Montes, I. Couso, P. Gil and C. Bertoluzza, "Divergence measure between fuzzy sets", Int. J. of Approximate Reasoning, vol. 30, pp. 91–105, 2002. https://doi.org/10.1016/S0888-613X(02)00063-4.
[19] M. Luo and R. Zhao, A distance measure between intuitionistic fuzzy sets and its application in medical diagnosis", Artificial Intelligence in Medicine, vol. 89, pp. 34-39, 2018. https://doi.org/10.1016/ j.artmed.2018.05.002.
[20] P. Gupta, and P. Tiwari, "Measures of cosine similarity intended for fuzzy sets, intuitionistic and interval-valued intuitionistic fuzzy sets with application in medical diagnoses", In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1846-1849, 2016.
[21] P. Dutta, and S. Goala, "Fuzzy decision making in medical diagnosis using an advanced distance measure on intuitionistic fuzzy sets". The Open Cybernetics & Systemics Journal, vol. 12, no. 1, pp. 136-149, 2018. https://doi.org/10.2174/ 1874110X01812010136
[22] X. Shang, and G. Jiang, "A note on fuzzy information measures", Pattern Recognition Letters, vol. 18, no. 5, pp. 425-432, 1997. https://doi.org/10.1016/S0167-8655(97)00028-7
[23] D.E.F. Kerridge, "Inaccuracy and inference", J. Royal Statistical Society B, vol. 23, no. 1, pp. 184-194, 1961. https://doi.org/10.1111/j.2517-6161.1961.tb00404.x
[24] R.K. Verma and B.D. Sharma, "A measure of inaccuracy between two fuzzy sets", Cybernetics and Information Technologies, vol. 11, no. 2, pp. 13-23, 2011.
[25] A. Ohlan, "A new generalized fuzzy divergence measure and applications", Fuzzy Inf. Eng., vol. 7, pp. 507-523, 2015. https://doi.org/10.1016/j.fiae.2015.11.007
[26] I.J. Taneja, "Seven means, generalized triangular discrimination and generating divergence measures", Information, vol. 4, no. 2, pp. 198-239, 2013. https://doi.org/10.3390/info4020198
[27] H. D. Arora and A. Dhiman, "On some generalized information measure of fuzzy directed divergence and decision making", Int. J. Computing Sc. and Math., vol. 7, no. 3, pp. 3931-3940, 2016.
[28] O. Parkash and R. Kumar, "Optimal Decision-Making Method Using Interval Valued Intuitionistic Fuzzy Divergence Measure Based on the Weights of Alternatives", Int. J. Engg. Sc. Invention, vol. 7, no. 3, pp. 82-94, 2018.
[29] S. Kumari, P. Tiwari and P. Gupta, "Application of Weighted Fuzzy Exponential J-Divergence Measure in Engiography", Int. J. Applied Engg. Research, vol. 14, no. 13, pp. 2984-2988, 2019.
[30] P. Tiwati and P. Gupta, "Entropy, Distance and Similarity measures under Interval Valued Intuitionistic Fuzzy Environment", Informatica, Vol. 42 (2) pp. 617-628, 2018. https://doi.org/10.31449/inf.v42i4.1303.
[31] S. Jain and V. Kumar, "Trigonometric Entropy on Intuitionistic Fuzzy Sets", Int. J. Adv. Sci. Tech., vol. 29, no. 03, pp. 12234-12243, 2020.
[32] C. Wei, Z. Gao, Z. & T. Guo, "An intuitionistic fuzzy entropy measure based on trigonometric function", Control Decision, Vol. 27, pp. 571–574, 2012. https://doi.org/10.1155/2015/563745.

https://doi.org/10.31449/inf.v45i1.3402 Informatica **45** (2021) 149–156 **149**

An Analysis of Emotional Tendency Under the Network Public Opinion: Deep Learning

Jinze Li, Yizhen Wang and Jun Wang
China People's Police University, School of Intelligence Policing, Langfang, Hebei 065000, China
E-mail: j9z571@163.com

Keywords: network public opinion, emotional orientation, deep learning, OCC model, convolutional neural network

Received: January 5, 2021

Network public opinion refers to the common opinion with tendency and influence formed by the public on certain social events through the Internet. Due to the complexity of interest relations, network public opinion is likely to cause difficulties for individuals, enterprises, or governments. To control the public's emotional tendency to social events, this study designed an OCC sentiment rule system to label the network public opinion case base. The text representation method is Word2Vec in deep learning, and the convolution neural network is used to construct the sentiment tendency analysis model under the network public opinion. Taking the case of Dolce & Gabbana humiliation incident, Xiangshui explosion incident, and baixiangguo girl's murder as the research cases, the accuracy of the model in identifying the above three events was 85.87%, 73.65%, and 85.87%, respectively, under the optimal parameters setting. The experimental results show that the proposed method can improve the accuracy of emotion recognition by 3.00% ~ 8.00% compared with the manual annotation method, i.e., the network public opinion sentiment orientation recognition model constructed in this study has a high recognition accuracy and can be used to assist relevant departments in detecting network public opinion.

Povzetek: Z globokim učenjem je narejena analiza mnenja uporabnikov omrežja o določeni tematiki.

1 Introduction

The dissemination and governance of network public opinion is an important part of government work, and the analysis of network public opinion sentiment tendency is the basic work to eliminate the network environment. At present, scholars have made a detailed analysis on the generation mechanism, evolution mechanism, impact on social economy, and guidance mechanism of network public opinion. Wu et al. used the social network analysis (SNA) framework to analyze the generation mechanism and evolution process of network public opinion [1]. Huang analyzed the changing trend of public policy under the sudden network public opinion event [2]. Song et al. analyzed the social transformation crisis and economic risk generated by network public opinion events [3]. Zhang's team proposed a network public opinion monitoring, tracking, analysis, and guidance system based on extensive investigation of existing network public opinion monitoring research [4]. The academic community has realized the emotional analysis and prediction of network public opinion on a certain field of social events. Li et al. have realized real-time tracking, monitoring and comprehensive evaluation of public opinion information on agricultural product quality and safety in the form of intelligent platform assisted by manual [5]. Based on the authenticity of network information dissemination, Hong's team realized the simulation of the propagation process of food safety network public opinion events [6]. At present, the research on the evolution mechanism of Internet public opinion and the analysis of emotional tendency about social events such as food, environment, and international disputes have been relatively perfect, but there are few studies on the universal model of emotional orientation analysis of network public opinion. Given the above situation, this study relies on the standard emotional rule system to construct the identification model of netizens' emotional orientation to realize a comprehensive and efficient understanding of the emotion contained in the information text.

2 Construction of network public opinion sentiment tendency analysis model

2.1 Design of emotional rules of network public opinion

In this study, the OCC model is used to construct emotion recognition rules. The final output of this model is the emotional tendency of netizens to a certain social event or social phenomenon. The generation process can be divided into four parts: classification, quantification, mapping, and expression [7, 8]. The evaluation criteria of the OCC model are event result, object behavior, and object image. The information receiver will make an emotional judgment on these three components. If the evaluation subject pays more attention to the event results,

the OCC model needs to focus on analyzing the evaluation subject's evaluation of the target; if the evaluation subject pays more attention to the behavior of the object, the model focuses on the behavior criterion of the evaluation subject; if the evaluation subject pays more attention to the image of the object, the model pays more attention to the emotional attitude of the evaluation subject to the evaluation object. The satisfaction degree of the event result, the behavior standardization of the object, and the attitude towards the target object may affect the emotional color of the information receiver. The standard design of emotional orientation recognition rules should not only include common emotional orientation words but also consider the language context of the comment. In this study, the judgment of the information receiver's emotional orientation is mainly based on the semantics of the information receiver's comment events and the logical relationship between sentences and the emotional orientation words.

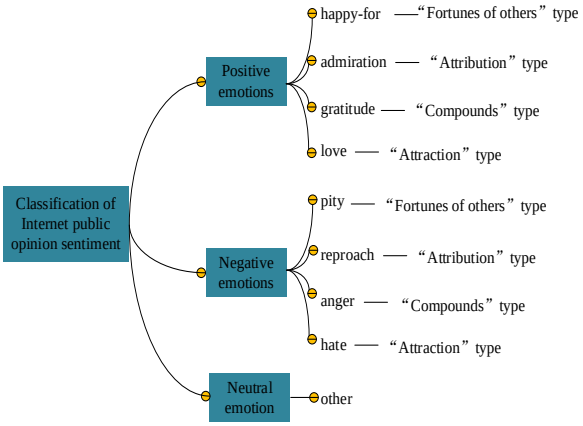


Figure 1: Classification and mapping of the emotional tendency of Internet public opinion.

The emotional orientation recognition model constructed in this study tends to provide a reference for relevant departments to monitor the network environment and control the direction of public opinion; thus, it is more inclined to understand the negative comments of netizens on social events or social phenomena. After understanding the emotional tendency of netizens to the event, the relevant departments can conduct targeted public opinion guidance to prevent large-scale network violence and affect social harmony [9]. Considering the closed-loop principle and the requirements of fine emotional granularity, this study creatively divides the emotional orientation of evaluation subjects into three categories: positive, medium, and negative. Referring to relevant literature, after screening and identifying, eight basic emotions are identified as the basic emotional types of online public opinion, as shown in Figure 1. These eight kinds of emotions are happiness, appreciation, regret, condemnation, anger, hate, love, and thanks for others. Appreciation, happiness for others, love and gratitude are mapped into positive emotions; anger, regret, condemnation, and hate are mapped to negative emotions; the comments that cannot identify emotional tendencies are mapped to neutral emotions [10].

When using the OCC model to identify the emotional orientation of online public opinion, it is necessary to analyze the correspondence between the subjective comment extractor evaluation standard and the emotional rules according to the evaluation criteria published by the evaluation subject. According to the information shown in Figure 1, eight kinds of emotion rules are constructed in this study. These emotional rules can not be used as the input content of the subsequent recognition model directly, and they need to be transformed into functional forms that are easy to implement in text. This paper constructs a 9-dimensional emotional space and assigns emotional variables to each comment text. The calculation formula is:

$$Emotions = [e_{happy-for}, e_{pity}, e_{admiretion}, e_{reproach}, e_{gratitude}, e_{anger}, e_{love}, e_{other}], \quad (1)$$

where $Emotions$ in formula (1) represents the emotional variable of the text and $e^{[0,1]}$ represents the value of each dimension of emotion.

$$\begin{aligned} Emtion(positive) &= Emtion(happy-for) \cup Emtion(admiration) \\ &\cup Emtion(gratitude) \cup Emtion(love) \\ Emtiom(negative) &= Emtion(pity) \cup Emtion(reporach) \\ &\cup Emtion(anger) \cup Emtion(hate) \\ Emtion(neutral) &= 1 - Emtion(positive) - Emtiom(negative) \end{aligned} \quad (2)$$

Equation (2) is the positive, middle, and negative emotional rules constructed in this study.

The emotion of "fortunes for other" is event-driven, and the emotion of the information receiver is based on the result of an event; thus, satisfaction is an important variable of emotion evaluation.

$$\begin{aligned} Emtion(happy-for) &= EventConsequece(txt_i, e) \cap FocusOn(e, s) \\ &\cap (DesireOf(o, e) \cup DesireOf_s DeservedOf_o e) \end{aligned} \quad (3)$$

$$\begin{aligned} Emtion(pity) &= EventConsequece(txt_i, e) \cap FocusOn(e, s) \\ &\cap (DesireOf(o_i, \sim e) \cup DesireOf_s \sim DeservedOf_o e) \end{aligned} \quad (4)$$

Equation (3) is the recognition rule of "happy for others". When the evaluator s thinks that the result of event e meets the objective expectation of evaluation object o, the emotion of the text can be recognized as "happy for others". Equation (4) is the recognition rule of regret emotion. When the evaluator s thinks that the result of event e does not meet the objective expectation of evaluation object o, the emotion of the text can be recognized as "regret".

"Attribution" emotion is a multi-type emotion. When the behavior of the evaluation object conforms to the behavior standard of the evaluation subject, the evaluation subject will produce positive emotions such as appreciation to the evaluation object. In this study, the IdealOf function is used to judge the emotional orientation of object behavior.

$$\begin{aligned} Emtion(admiration) &= EventAction(txt_i, a) \cap FocusOf(a, o) \\ &\cap MotivationOf(a, m) \cap IdealOf(s, m, positive) \end{aligned} \quad (5)$$

$$\begin{aligned} Emtion(reproach) &= EventAction(txt_i, a) \cap ActionOf(a, o) \\ &\cap MotivationOf(a, m) \cap IdealOf(s, m, negative) \end{aligned} \quad (6)$$

Equations (5) and (6) are the recognition rules of "appreciation" and "regret", respectively. If IdeaOf is positive, the emotion of the text is "appreciation"; if IdeaOf is negative, the result of text recognition is "regret".

"Compounds" emotion means compound emotion. When the event result and object behavior interact, the

information receiver may have multiple emotions for the information. When the evaluator associates his own experience, or the event touches his own interests through the event, the information receiver may have the emotion of "gratitude" or "anger", and the recognition rules are:

$$Emtion(gratitude) = Event(txt_i, e, a) \cap ActionOf(a, o) \cap MotivationOf(a, m) \cap IdealOf(s, m, positive) \cap DesireOf(s, e) \quad (7)$$

$$Emtion(anger) = Event(txt_i, e, a) \cap ActionOf(a, o) \cap MotivationOf(a, m) \cap IdealOf(s, m, negative) \cap DesireOf(s, \sim e) \quad (8)$$

The source of "attraction" emotion is the preference and dislike of the information receiver to the object image. The information receiver will ignore the event results and the object behavior and express their own opinions on the object image. This kind of emotion is very common in netizens' comments on entertainment events. This type of emotion is mainly "like" and "hate". The recognition rules are:

$$Emtion(love) = EventObject(txt_i, o) \cap AttractionOf(s, o, positive), \quad (9)$$

$$Emtion(hate) = EventObject(txt_i, o) \cap AttractionOf(s, o, negative). \quad (10)$$

The emotional rules of network public opinion based on the OCC model constructed in this study are as follows. When the evaluation criterion of emotion classification is event result and the evaluation subject is satisfied with the results of others, there will be feelings of "happy for others"; when the evaluation subject is not satisfied with the results of others, there will be "regret" emotion; when the evaluation subject is not satisfied with its results, there will be "sad" emotion. When the evaluation subject is satisfied with its results, there will be "happy" emotion. When the evaluation object is taken as the evaluation standard, whether the evaluation subject likes the evaluation object or not determines the emotion as "like" or "hate". When the object's behavior is taken as the evaluation standard, the approval of the evaluation subject determines the emotion as "appreciation" or "condemnation". When the object's behavior and event consequences are taken as the comprehensive evaluation criteria and they are satisfied with their results and agree with the object's behavior, they will have "gratitude" emotion; when they are not satisfied with their results and do not agree with the object's behavior, they will have "anger" emotion.

2.2 Identification of emotional tendency of Internet public opinion

The popularity of the Internet enables the public to express their opinions freely on the public platform. At present, the market is full of a large number of social news software, which provides a lot of channels for the public to express opinions. Generally speaking, the online public opinion text for a social event or social phenomenon is more than 10000 or 100000, covering multiple applications. The traditional classification method of Internet public opinion sentiment is manual annotation. Professionals summarize and analyze the comments made by Internet users with different software to judge the public's emotional tendency towards the event [11, 12]. However, the recognition efficiency of this method is low, the recognition results are easily affected by the subjective emotions of the staff, and the accuracy of the recognition results is difficult to guarantee. In this study, the Word2Vec text representation method and convolutional neural network are used to construct the network public opinion sentiment rule recognition model; with the help of computers, the objective and rapid identification of network public opinion emotion can be realized [13, 14].

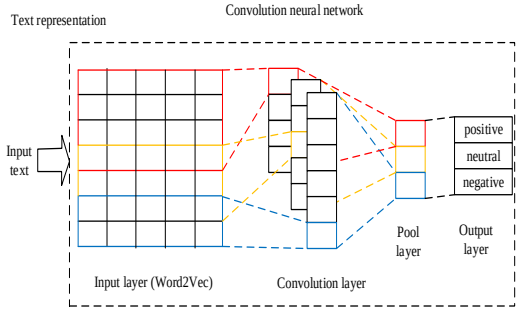
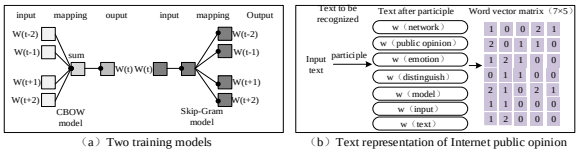


Figure 2: Recognition model of Internet public opinion sentiment rules.

The basic model of network public opinion sentiment rule recognition is shown in Figure 2. The core idea of the model is to predict the sentiment tendency of the text by calculating the vector matrix formed by the text. The key to extracting high-level abstract features hidden in text data using a convolutional neural network is to select the appropriate text representation method [15]. This study uses Word2Vec as the training model of word vector feature extraction. The model runs fast and can train hundreds of billions of words in 24 hours.



（a）Two training models （b）Text representation of Internet public opinion

Figure 3: Representation method and training model of Internet public opinion text.

Figure 3 (a) is the logical block diagram of the Word2Vec training model. The training and prediction of the Word2Vec model is based on the CBOW (Continuous Bag of Words) model and skip-gram model [16, 17]. $w(t)$ in the figure indicates the position t of the current word in the text statement, and other words except the word form the context. The principle of word recognition in the Word2Vec model is as follows: the Word2Vec model uses word frequency as a leaf node, uses words with similar word frequency in the hidden layer to activate similar content, and word frequency is inversely proportional to the number of active hidden layers. Figure 3 (b) is an example of the output result of emotion recognition for a text. Assuming that the text has seven words and the word vector dimension generated by the Word2Vec training model is 5, the text represents a two-dimensional matrix of word vector with the result of 7 × 5. The matrix can be used as the input vector of the next classifier.

After the text is transformed into a vector matrix by the Word2Vec model, to predict the emotion of the text with the vector matrix, this study uses the convolutional neural networks (CNN) to classify the emotional tendency

of the subjective evaluation of the evaluation subject. The basic structure of convolution neural network includes convolution layer, pooling layer, full connection layer, and output layer [18, 19]. The unique structure of a convolutional neural network has the characteristics of local connection, weight sharing, and time-space subsampling, which can effectively reduce the number of calculated weights and is suitable for solving the classification problem of multivariable. The operation process and output results of text data in the convolution layer and pooling layer of a convolution neural network are shown in Figure 4.

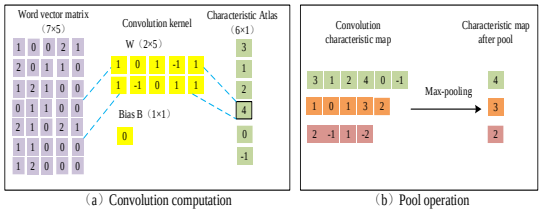


Figure 4: Operation of network public opinion information text in a convolution neural network.

Figure 4 (a) shows the operation of the convolution layer on information text. The convolution layer is a 2 × 5 convolution kernel, and the weight parameter matrix is generated randomly. The input text is a 7 × 5 feature matrix; after convolution processing, a 6 × 1 size feature map can be obtained.

The two-dimensional matrix generated by the Word2Vec model is the input of the convolution layer, the dimension of 7 × 5 vector matrix is reduced to 6 × 1 matrix, and the weight parameters are reduced, which reduces the operation difficulty and operation time of the convolution neural network. The two-dimensional matrix generated by the Word2Vec model is the input of the convolution layer, the dimension of the 7 × 5 vector matrix is reduced to the 6 × 1 matrix. This operation reduces the weight parameters, reduces the difficulty of the convolution neural network operation, and greatly reduces the operation time. In the actual training of network public opinion information text, sometimes the value of the input word vector matrix does not change obviously, and the linear model lacks enough expression ability, which leads to the output deviation from the correct value. In this study, the nonlinear factors are introduced to distinguish the nonlinear data, and the relu function of the unsaturated nonlinear method is used as the activation function, so that the convolution neural network can effectively process the data [20, 21].

Figure 4 (b) is a schematic diagram of the operation of the pooling layer on information text. The pooling layer can maintain the invariance of the map to a certain extent after operations such as translation and scaling while retaining the text information, reducing the feature dimension, and improving the operation speed [22]. After pooling the 6 × 1 matrix output from the convolution layer, the number of neurons in the feature map remains unchanged, and the feature map of 7 words has six row elements. The pool forming process of pool layer pair characteristic map is:

$$Z_i = f(W \square pool(C_i) + B) \quad (11)$$

where $pool(\square)$ is the characteristic map after pooling treatment. In this study, the maximum value of neurons in each area is selected, and the maximum value is defined as $pool_{\max}(C_i) = \max\{R_x\}$.

As shown in Figure 4 (b), the maximum characteristic elements of 6 × 1, 5 × 1, and 4 × 1 feature maps are 4, 3, and 2, respectively. The expression of the full connection function of the convolution neural network is:

$$Z = f(Z_1, Z_2, ..Z_i) \quad (12)$$

After the full connection operation of the feature map is completed, the classification function is used to convert the output of the defined linear function into the sentiment category of network public opinion, and the proportion of positive, neutral, and negative emotional tendency of the text can be speculated. As this study focuses on the negative emotions of online public opinion texts, if the proportion of negative texts exceeds the sum of positive and neutral texts, the event will be regarded as having a greater negative impact on the society; if the proportion of positive texts exceeds the sum of negative texts and neutral texts, the event will have a positive and healthy impact on the society. The logic of updating parameters of the recognition module is as follows: some hidden nodes of the model do not participate in the recognition process, but their weight values will be retained and only participate in the recognition operation at the next sample input [23].

3 An empirical analysis of the emotional tendency analysis model of Internet public opinion

3.1 Experimental design

This study took Dolce & Gabbana humiliation incident in November 2018, Xiangshui explosion in March 2019, and baixiangguo girl's murder in November 2020 as cases. The three incident numbers were DG_2018, XS_2019, and BXG_2020. This study grabbed 24808 text data for the 2018 event, 8856 data texts for the XS_2019 event, and 65537 data texts for the BXG_2020 event.

First of all, the obtained data were cleaned to remove the unclear and confusing texts. The effective data of the three events were 23043, 4573, and 16773, respectively. After cleaning, the effective data were labeled by sentiment classification to form the case database data set of network public opinion events. Data preprocessing mainly included word segmentation and data length processing of Chinese texts. The sample size ratio of training set and data set was 9:1. There were 20738, 4116, and 15096 data texts in the training sets for the three events, and 2305, 457, and 1677 data texts in the test sets, respectively. After extracting multi-dimensional features from data text, it was necessary to construct an SVM classifier and adjust its parameters to the best state. The trained SVM model was used for predicting the network

public opinion sentiment tendency of three kinds of events.

3.2 Hyperparametric analysis

In this study, ten-fold cross-validation was used to test the setting range of the hyper-parameters of the constructed network public opinion sentiment tendency classification model. The initial hyper-parameter ranges were as follows: the dimension of word vector was 300; the shape of convolution kernel was 2, 3, and 4; the number of corresponding convolution kernels was 100, 100, and 100; dropout was 0.5; L2 norm was 0; mini_ batch was 32; the training time was 200, 500, and 200. Using this set of hyperparameters, DG_ 2018, XS_ 2019, and BXG_2020. The accuracy rate of emotion orientation recognition was 83.67%, 67.83%, and 83.92%, respectively.

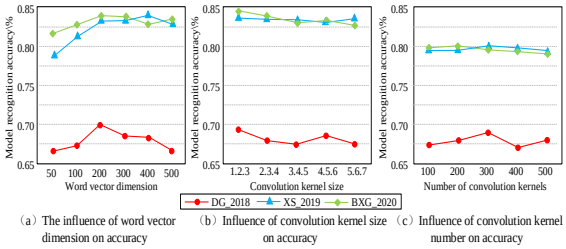


Figure 5: The influence of word vector dimension, convolution kernel size, and number on recognition accuracy.

Following the principle of a single variable, the dimension of the word vector was set as 50, 100, 200, 300, 400, and 500, respectively, convolution kernel size was 1, 2, 3, 2, 3, 4, 3, 4, 4, 5, 4, 5, 6, 6, and the number of convolution kernels was 100, 200, 300, 400, and 500. Under this set of parameters, the accuracy of identifying the emotional orientation of online public opinion is shown in Figures 5 (a), (b), and (c). When the dimension of word vector changed from 50 to 500, the accuracy of the emotion recognition model first increased and then decreased. When the word vector dimension was 200, the recognition accuracy of the recognition model for three events reached the highest, and the value was higher than that under the initial hyper-parameter setting. When the size of the convolution kernel was 1, 2, 3, the recognition model had the highest recognition accuracy for three events. The classification effect of XS_2019 and BXG_2020 events are the best when the number of convolution kernels was 300. The classification effect of DG_2018 events was the best when the number of convolution kernels was 200. There was no significant difference between the classification effect of XS_ 2019 and BXG_2020 events when the convolution kernel was 200 and 300. The reason for this phenomenon is as follows: when the number of convolution kernels was small, some important features could not be included in the network learning range; when the number of convolution kernels was large, the speed of network training reduced. Under the comprehensive consideration, when the number of convolution kernels was 300, the recognition effect of the emotion recognition model was the best.

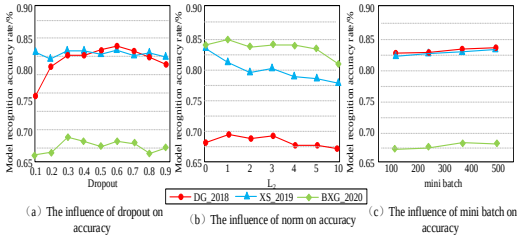


Figure 6: Influence of L2, Dropout value, and mini batch on recognition accuracy.

According to the principle of a single variable, dropout was set as 0.1, 0.2, 0.3,..., 0.9; L2 norm was 0, 1, 2, 3, 5, 10, and mini_batch was 32, 64, 96, 128. Under this set of parameters, the accuracy of identifying the emotional tendency of online public opinion is shown in Figures 6 (a), (b), and (c). When dropout was 0.1, 0.2, 0.8, 0.9, the effect of the emotion recognition model in recognizing BXG-2020 event was poor. The reason for this phenomenon is as follows: when the dropout value was too high, the number of random output neurons was too large, and the probability of fitting was large; when the dropout value was too low, the number of random output neurons was too small to obtain enough neuron characteristics. Therefore, when the dropout value was between 0.3 and 0.7, the recognition model classification effect was better. The dropout value of 0.6 was selected for the experiment. L2 norm can restrict the model space by constraining the parameters to prevent over-fitting. When the L2 norm was 0 and 1, the classification effect of the recognition model was the best. When the L2 norm value was greater than 3, the recognition accuracy showed a downward trend. In a certain range, the larger the value of mini_batch was, the higher the utilization rate of model memory was, and the fewer the iterations were. However, under the same accuracy requirements, the time was longer. With the increase of mini_batch value, the corresponding recognition accuracy reaches 100%.

According to the experimental results of the emotion recognition model under different hyper-parameter settings, the optimal hyper parameter settings are as follows: vector dimension was 200; dropout was 0.6; convolution kernel size was 1, 2, 3; convolution kernel number was 300; L2 norm was 1; mini_ batch value was 128. Under this parameter, the recognition accuracy of the emotion recognition model was 85.87%, 73.65%, and 85.87%, respectively. Compared with the initial hyper-parameter setting, the recognition accuracy was improved by 2.18%, 5.82%, and 1.95%, respectively, showing that the classification effect effectively improved.

3.3 Comparison test results

In order to verify the feasibility of the network public opinion sentiment orientation recognition model constructed in this study, this study compared the two annotation methods, OCC sentiment annotation and word

vector representation. The comparison of the recognition effect between the two methods is shown in Figure 7.

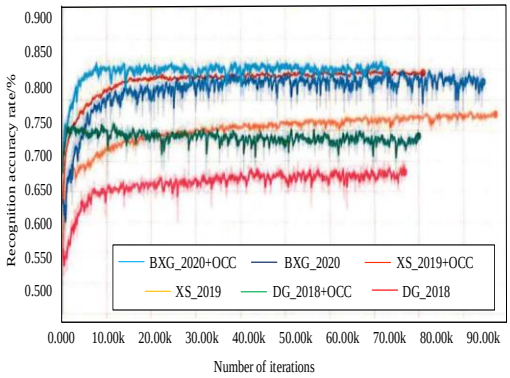


Figure 7: Effect of emotion tagging in OCC on recognition accuracy.

Figure 7 shows the accuracy results of the two annotation methods for identifying three events after 90000 iterations. It was concluded from Figure 7 that when the network public opinion emotion recognition model based on the OCC sentiment annotation method was used for identifying DG_2018, XS_2019, and BXG_2020, the accuracy rates were 84.32%, 73.65%, and 85.87%, respectively. The accuracy rates of the artificial tagging method in identifying these three events were 76.38%, 65.64%, and 82.35%, respectively. The accuracy of emotion recognition was improved by 3.00% ~ 8.00% by using the OCC emotion tagging method. The above result was because some event reviews in the research database might not express emotion with too many emotive words, the results of manual annotation were lack of standardization, and the emotion classification of some comments by staff was not accurate, which affected the recognition accuracy of subsequent recognition models. The OCC emotion labeling method was based on the concept of machine learning, and complete and correct emotion recognition rules made its annotation accuracy higher than the manual method.

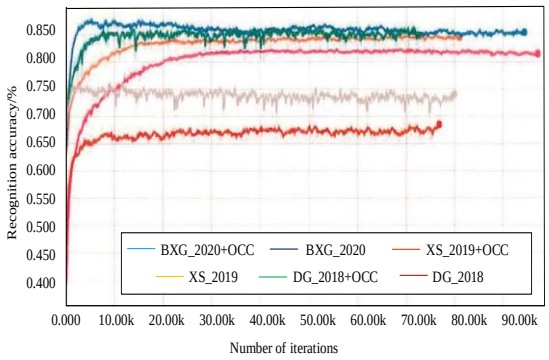


Figure 8: Influence of text representation on recognition accuracy.

In order to compare the effect of word vector generated by Word2Vec on emotion classification, this study used the vocabulary index and the "tf.nn.embedding_lookup" method in TensorFlow to train the word vector from the original data to complete the embedding operation. The two methods were named "cnn_rand" and "cnn_wod2vec". The comparison results are shown in Figure 8. The accuracy of the model which adopted the cnn_rand method to identify DG_2018, XS_2019, and BXG_2020 events was 84.32%, 73.65%, and 85.87%, respectively. The accuracy of the model which adopted the cnn_wod2vec method to identify DG_ 2018, XS_ 2019 and BXG_2020 events was 81.87%%, 67.68%, and 86.17%, respectively. The network public opinion emotion recognition model which adopted the cnn_wod2vec method had a better recognition performance in identifying DG_2018 and XS_2019 events; the model which adopted the cnn_rand method was better in identifying BXG_2020 event, and the recognition accuracy of the model was 0.30% higher than that of the model which adopted the cnn_word2vec method. Overall, the recognition model constructed by generating the word vector with Word2Vec was better, and the Word2Vec method was more suitable for the emotion recognition model constructed in this study.

Classifier/ event	DG_2018	XS_2019	BXG_2020
CNN	84.32%	73.65%	85.87%
SVM	66.60%	58.20%	61.38%

Table 1: Influence of sentiment classifier on the effect of network public opinion sentiment recognition.

The public's emotional tendency to the three events was identified using the CNN classifier and SVM classifier, and the comparison results are shown in Table 1. The data in Table 1 showed that the recognition accuracy of CNN classifier was significantly higher than that of the SVM classifier [24]. It was because the CNN classifier adopted the mode of forward-propagation prediction and back-propagation error, and its unique chain rule improved the output accuracy of the deep neural network model. It was seen from Table 1 that the accuracy rate of the CNN classifier was 17.72%, 15.45%, and 24.49% higher than the SVM classifier. It is inferred that the feature mode of word embedding generated by Word2Vec may not be suitable for text classification using the SVM classifier.

4 Conclusion

Unlike traditional machine learning, deep learning has higher accuracy in mining data, and its unique parameter sharing mechanism can significantly reduce the training time of models. In order to improve the efficiency and accuracy of network public opinion sentiment orientation recognition, this study used a convolutional neural network in deep learning to automatically mine network public opinion sentiment. This study was based on the characteristics of a small amount of text and a high frequency of emotional words. The network public opinion sentiment rule base was constructed, and the network public opinion sentiment tendency recognition model was obtained by training with a convolution neural network. The experimental results showed that the OCC emotion tagging method improved the accuracy of emotion recognition by 3.00% ~ 8.00% compared with the

manual annotation method; compared with the vocabulary index method, Word2Vec was more suitable as the input of the network public opinion emotion recognition model; the accuracy rates of the CNN classifier in identifying DG_2018, XS_2019 and BXG_2020 events were 17.72, 15.45, and 24.49 percentage points higher than the SVM classifier respectively. The identification model constructed in this study has a high accuracy and can provide information decision support in network public opinion. However, the model has a low accuracy in identifying the emotional orientation of netizens with too little text information data; thus, it is necessary to further optimize the recognition model.

5 References

[1] Wu S, Cui X, Hu Y. (2015). Study on the internet public opinion evolution based on SNA. *Journal of Sichuan University (Engineering Science Edition)*, 47, pp. 138-142. https://doi.org/10.15961/j.jsuese.2015.01.019.

[2] Huang B. (2020). Analyze the Influence of Internet Public Opinion on Public Policy. *Open Access Library Journal*, 07, pp. 1-9. https://doi.org/10.4236/oalib.1106674.

[3] Song S, Guo Z, Wang X. (2020). The correlation between social transformation economic risk and internet public opinion. *Behaviour and Information Technology*, pp. 1-11. https://doi.org/10.1080/0144929X.2020.1722750.

[4] Zhang W, Du Y, Li Z, Chen JD. (2018). DeepOpinion: a system for deep analysis and guidance of internet public opinion. *Journal of Beijing University of Chemical Technology (Natural Science Edition)*, 45, pp. 94-98.

[5] Li XZ, Qian YZ, Deng Y, Song WG, Liao JF, Yang MS, Lian YL. (2017). Monitoring and analysis on internet public opinion of agro-products quality and safety in China, 2016. *Chinese Science Bulletin*, 62, pp. 1095-1102. https://doi.org/10.1360/N972017-00015.

[6] Hong W, Li Q, Wu L. (2017). Food safety internet public opinion transmission simulation and management countermeasures considering information authenticity. *System Engineering Theory and Practice*, 37, pp. 3253-3269. https://doi.org/10.12011/1000-6788(2017)12-3253-17.

[7] Gesang D, Qiao SJ, Han N, Zhang XS, Yang Y, Yuan CA, Kang J. (2015). An Internet Public Opinion Hotspot Detection Algorithm Based on Single-Pass. *Journal of the University of Electronic Science and Technology of China*, 44, pp. 599-604. https://doi.org/10.3969/j.issn.1001-0548.2015.04.021.

[8] Zhang Q. (2019). A Collaborative Group Decision Model for Internet Public Opinion Emergency with Interval Value. *American Journal of Modeling and Optimization*, 7, pp. 14-19. https://doi.org/10.12691/ajmo-7-1-3.

[9] He H. (2018). Research on prediction of internet public opinion based on grey system theory and fuzzy neural network. *Journal of Intelligent and Fuzzy Systems*, 35, pp. 1-8. https://doi.org/10.3233/JIFS-169591.

[10] Gao X, Fu L. (2017). Methods of uncertain partial differential equation with application to internet public opinion problem. *Journal of Intelligent & Fuzzy Systems*, 33, pp. 1-11. https://doi.org/10.3233/JIFS-17409.

[11] Wang A, Liu X, Sun X, Wang J. (2017). Research of internet public opinion based on hybrid algorithm of LDA and VSM. *C e Ca*, 42, pp. 1508-1513.

[12] Peng ZH. (2016). The Approaches of Internet Public Opinion Research. *Libraly Journal*, 35, pp. 63-68.

[13] Sang DV, Bao C. (2018). Effective Deep Multi-source Multi-task Learning Frameworks for Smile Detection, Emotion Recognition and Gender Classification. *Informatica*, 42, pp. 345-356. https://doi.org/10.31449/inf.v42i3.2301.

[14] Gjoreski M, Gjoreski H, Kulakov A. (2014). Machine Learning Approach for Emotion Recognition in Speech. *Informatica*, 38, pp. 377-384. https://www.researchgate.net/publication/270893946.

[15] Zhang F, Li SC, Guan Y. (2016). Hot issues about drug price in china: from the view of internet public opinion monitoring. *Value in Health*, 19, pp. A273-A274. https://doi.org/10.1016/j.jval.2016.03.1953.

[16] Zhu H, Liu P, Shan X. (2015). Analysis of internet-based public opinion in China, 2012. *Journal of Molecular Neuroscience*, 49, pp. 614-617. https://doi.org/10.1163/9789004276536_010.

[17] Li L. (2020). Research on the Transfer Rules of Internet Users' Negative Emotional State in Financial Public Opinion. *Open Journal of Business and Management*, 08, pp. 282-301. https://doi.org/10.4236/ojbm.2020.81017.

[18] Li L, Zhu X, Hao Y, Wang S, Gao X, Huang Q. (2019). A Hierarchical CNN-RNN Approach for Visual Emotion Classification. *ACM Transactions on Multimedia Computing Communications and Applications*, 15, pp. 1-17. https://doi.org/10.1145/3359753.

[19] Ni HB. (2020). Face Recognition Based on Deep Learning Under the Background of Big Data. *Informatica* 44, pp. 491-495.

[20] Gupta A, Srinivasan SM. (2020). Constructing a Heterogeneous Training Dataset for Emotion Classification. *Procedia Computer Science*, 168, pp. 73-79. https://doi.org/10.1016/j.procs.2020.02.259.

[21] Zhang W, He X, Lu W. (2020). Exploring Discriminative Representations for Image Emotion Recognition With CNNs. *IEEE Transactions on Multimedia*, 22, pp. 515-523. https://doi.org/10.1109/TMM.2019.2928998.

[22] Rao Y, Xie H, Li J, Jin F, Wang F, Li Q. (2016). Social emotion classification of short text via topic-level maximum entropy model. *Information & Management*, 53, pp. 978-986. https://doi.org/10.1016/j.im.2016.04.005.

[23] Lin YP. (2020). Constructing a Personalized Cross-Day EEG-Based Emotion-Classification Model Using Transfer Learning. *IEEE Journal of Biomedical and Health Informatics*, 24, pp. 1255-1264. https://doi.org/10.1109/JBHI.2019.2934172.

[24] Wang D, Xu G. (2020). Research on the Detection of Network Intrusion Prevention With Svm Based Optimization Algorithm. *Informatica*, 44, 269-273. https://doi.org/10.31449/inf.v44i2.3195.

https://doi.org/10.31449/inf.v45i1.3426 Informatica **45** (2021) 157–162 **157**

Information Visualization Analysis of Public Opinion Data on Social Media

Feng Chen and Shi Zhang
Luxun Academy of Fine Arts, Liaoning 116650, China
E-mail: chengfang09583@163.com

Keywords: social media, public opinion, visualization, Weibo, emotional analysis

Received: January 29, 2021

Public opinion data on social media contains much useful information, which can be visually displayed through visualization. This study mainly focused on Weibo and analyzed the keyword extraction of text and the analysis of emotional tendency. Keywords were extracted using the term frequency-inverse document frequency (TF-IDF) method, and the emotional tendency of the text was calculated based on the HowNet emotion dictionary and BosonNLP emotion dictionary. Finally, relevant data were collected by taking "Jiang Ziya" as the keyword for visualization analysis. It was found that the discussion on "Jiang Ziya" gradually reduced in the research period, and the extracted keywords were relatively positive. The visualization results of word cloud showed that there were many positive comments on "Jiang Ziya", but there were also negative comments. Finally, the calculation of emotional tendency showed that 69% of the texts showed a positive emotional tendency, and 31% of the texts were negative, indicating that netizens' emotional tendency towards "Jiang Ziya" was mainly positive. The study results make some contributions to the visualization of public opinion data and can be further applied in practice.

Povzetek: Razvita je metoda za vizualizacijo mnenj v socialnih omrežjih, tj. na Weibo.

1 Introduction

With the development of network technology, the popularity of networks has improved, and the number of network users is also growing. It is not only a tool for people to learn and work. Because of its anonymous, timely, and interactive characteristics, the network has become a new platform for people to obtain, share, and exchange opinions. Compared with traditional media, network-based social media plays a great role in information spreading and exchange. Everyone can express their opinions and spread the news through social media, promoting the rapid development of online public opinion. Moreover, the particularity of social media makes network public opinions have concealment and abruptness [1]. It is an important issue to guide and monitor public opinions correctly. Social media includes Weibo, WeChat, forums, etc., which contains many public opinion data. These data contain netizens' emotion and attitude towards events, which has a strong influence and will change with the development of events. However, if the public opinion of some events develops to some scale or the trend is not conducive to social stability, it may cause chaos of public opinion and lead to the occurrence of adverse public opinion events. Therefore, to create a stable public opinion environment, it is necessary to monitor, manage, timely warning, and guide public opinion, thereby establishing a harmonious network environment. At present, the research on these data includes hot spot discovery [2], crisis early warning, public opinion prediction [3], etc. Tan et al. [4] analyzed the campus network public opinion, explored technologies, such as Chinese word segmentation and topic recognition, and combined analytical hierarchy process (AHP) with wavelet neural network to monitor the network public opinion. Taking the Sade event as an example, they found that the method had a good estimation accuracy. Tang et al. [5] studied the role of fuzzy sets in network public opinion analysis, compared different functions and advantages of different fuzzy sets, and discussed the future trend of this field. Chai and Cheng [6] proposed an improved AHP-entropy method to evaluate the risk of network public opinion, which comprehensively considered the subjective weight and objective weight, and verified the effectiveness of the method by experiments. Zhang et al. [7] predicted network public opinion with the gray model, corrected the results of the gray model with the back-propagation neural network, and carried out a simulation experiment on the model by taking a hot topic as an example. The results showed that the method could effectively and accurately predict public opinion. The public opinion data on social media involves a lot of content, but the current research is mostly static and one-sided, which can not show the information dynamically and stereoscopically. As an effective way of data expression, visualization technology has good application in information processing, but it has less application in processing public opinion on social media. Thus, this study took Weibo as the research subject, analyzed the keyword extraction and emotional tendency, displayed the information through visualization, with the intention of understanding the reliability of the

visualization method in the public opinion data processing.

2 Public opinion and visualization

The public opinion data on social media is the information collection of a specific event, and it will change with the development of time. It is ① massive: in the network, an event can cause a great deal of discussion in a very short time, thus generating a huge amount of information; ② diverse: public opinion data has a variety of forms, and the concept of Internet users can also be expressed by videos and pictures in addition to text; ③ dynamic: public opinion is in rapid dynamic evolution.

Given the characteristics of public opinion data, traditional information processing methods can not effectively deal with them [8]. As a relatively mature information processing method, visual analysis combines computer graphics, data mining, etc., and it can show massive and abstract information in a visual way [9] to help people find the hidden information in the data. For public opinion data, visualization can display the distribution, development, and change of public opinion through images so that people can analyze the public opinion data dynamically and globally. A chart is a basic form of visualization. At present, the commonly used visual chart include ① bar chart, which displays the difference of data through rectangles with different lengths; ② histogram, which is used for understanding the distribution of things; ③ pie chart, which is used for showing the proportion of items; ④ trend chart, which can display the development trend of time; ⑤ theme river chart, which is used for displaying the change of events in a period; ⑥ word cloud, which is the visualization of keywords and can directly display the main idea of the text.

3 Processing of Weibo public opinion data

3.1 Keyword extraction of Weibo text

Weibo is a very widely used social media. Users can participate in the discussion of events as long as they register the account. Everyone can be both the publisher of information and the receiver of information. With the development of the network, the influence of Weibo has become increasingly larger. More and more events arouse the widespread concern of people across the country through the form of hot search on Weibo. Due to the characteristics of Weibo, the Weibo text is limited to 140 words, which is concise and comprehensive and is a fragmented description of an event. These texts include users' emotions and attitudes towards the event. Keywords in the text is a summary of the theme, which contains the key information of the text. The extraction of keywords is conducive to understand the opinions and emotions of users.

Before keyword extraction, first of all, text segmentation is needed. Text segmentation refers to dividing a complete sentence into separate words. Weibo text is mainly in Chinese, with few words and prominent colloquialism. In this study, natural language processing and information retrieval (NLPIR) (Institute of Computing Technology, Chinese Lexical Analysis System) [10] was used for word segmentation. It is a Chinese word segmentation system developed by the Chinese Academy of Sciences. In addition to word segmentation, it will also mark properties of words, which is more conducive to subsequent text processing.

After word segmentation, due to the existence of punctuation, space, and function words in the text, in order to improve the efficiency of text processing, it is necessary to remove stop words. The specific steps are as follows: ① function words, such as adverbs, prepositions, and conjunctions, and notional words, such as numerals and quantifiers, have no practical significance; therefore, these words with little public opinion information should be filtered; ② in the process of word segmentation, some symbols, numbers, and separate words will be segmented; therefore, separate words, numbers, and symbols with a length of 1 should be filtered; ③ Weibo texts usually carry some special words, such as "comment", "link", "forward", etc., which contains few public opinion information and cannot reflect the emotion of users, and these words should also be filtered.

In this study, keywords were extracted by the term frequency-inverse document frequency (TF-IDF) model [11]. In a text, if a word has high TF (high frequency in the same text) and high IDF (rarely appears in other documents), it can be used as a keyword. For word t, the calculation method of TF is: $TF = \frac{count(t)}{count(d_i)}$, where $count(t)$ refers to the number of t in document d_i and $count(d_i)$ refers to the total number of words in document d_i, and the calculation method of IDF is: $IDF = \frac{num(corpus)}{num(t)+1}$, where $num(corpus)$ refers to the number of documents in corpus and $num(t)$ refers to the number of documents containing word t in the corpus.

3.2 Analysis of emotional tendency

In analyzing public opinion data, emotional analysis is a key part, reflecting the users' emotional tendency to an event. The extraction of keywords can help understand users' key emotions to the event, but it is difficult to judge the specific emotional tendency. Based on HowNet emotional dictionary [12] and BosonNLP emotional dictionary, this study analyzed the emotional tendency of Weibo text. HowNet emotional dictionary includes 836 positive emotion words and 1254 negative emotion words. BosonNLP emotional dictionary takes the emotional value as the expression of emotional tendency, positive number as positive emotion word, and negative number as negative emotion word. BosonNLP emotional dictionary includes 114767 words.

A Weibo text is decomposed into several sentences, and then the emotional polarity of every sentence is calculated. In a known document named D, there are n sentences. The document can be written as $D = \{s_1, s_2, \cdots, s_n\}$. Firstly, the emotional value of every

sentence is calculated: $F(s_i) = \sum s_{w_i}$, where s_{w_i} stands for the emotional value of word w_i in the sentence. Then, the emotional value of the whole text is: $F(s) = \sum F(s_i)$. If $F(s) > 0$, it indicates that the text has a positive emotional tendency; if $F(s) < 0$, it indicates that the text has a negative tendency; $F(s) = 0$, it indicates that the text is neural.

The verbs and adjectives are separated from the text as emotional words for calculation using the word segmentation system: $s_{w_i} = \frac{fp_{w_i}}{(fp_{w_i}+fn_{w_i})} \times \frac{N_p}{(N_p+N_n)} - \frac{fn_{w_i}}{(fp_{w_i}+fn_{w_i})} \times \frac{N_p}{(N_p+N_n)}$, where fp_{w_i} stands for the ratio of w_i to positive emotional words, fn_{w_i} stands for the ratio of w_i to negative emotional words, N_p stands for the number of positive emotional words in the emotional dictionary, and N_n stands for the number of negative emotional words. For the calculated result, $s_{w_i} > 0$ is determined as the positive emotional word, $s_{w_i} < 0$ as the negative emotional word, and $s_{w_i} = 0$ as the neural word.

4 Analysis of Weibo information visualization

The film "Jiang Ziya" was released in the mainland of China and North America on October 1, 2002. The film was originally scheduled to be released at the Spring Festival in 2020 but was canceled due to the influence of the epidemic. After its release, the film has caused extensive discussion on Weibo. The reading quantity of "Jiang Ziya" on Weibo has reached 430 million, and there are about 227000 discussions. In this study, "Jiang Ziya" was taken as the keyword. Through the octopus data collector, Weibo texts were collected as public opinion data from September 30, 2020, to October 15, 2020. First of all, the top ten heat searches related to "Jiang Ziya" in this period are shown in Table 1.

It was seen from Table 1 that the release of "Jiang Ziya" had caused extensive discussion on Weibo. The highest ranking of "Jiang Ziya" on the hot search list is 5, and the lowest is 35. In this period, the spread trend of "Jiang Ziya" is shown in Figure 1.

It was seen from Figure 1 that the popularity of "Jiang Ziya" was the highest on the day of its release, and the number of Weibo texts related to it reached 2261. Then, discussions on "Jiang Ziya" began to decline. On October 8, according to the box office of the National Day, the box office reached nearly 3.7 billion yuan; the box office of "Jiang Ziya" was 1.324 billion yuan, which aroused a new round of discussion. Subsequently, with the extension of the showing time, the number of Weibo texts related to "Jiang Ziya" gradually decreased.

Keywords were extracted from the collected Weibo texts using the TD-IDF method. The top 10 keywords are shown in Table 2.

It was seen from Table 2 that most netizens' comments on "Jiang Ziya" were positive and they thought that "Jiang Ziya" was a breakthrough in animated films and had high expectations for it. In order to display

Ranking	Topic of conversation	Maximum heat	The highest ranking
1	"Jiang Ziya" extended showing	985577	11
2	The box office of "Jiang Ziya" exceeds 1.5 billion	227522	35
3	The box office of "Jiang Ziya" exceeds 1 billion	424056	20
4	Details in "Jiang Ziya"	1089140	6
5	The box office of "Jiang Ziya"	655630	16
6	Imitated makeup of Daji Su in "Jiang Ziya"	1429278	5
7	"Jiang Ziya" renews the 1st-week box office record of animation film	1124068	9
8	Post-credit scenes of "Jiang Ziya"	913289	22
9	The box office of "Jiang Ziya" exceeds 0.6 billion	309196	30
10	Poster of Dujie City for "Jiang Ziya"	415238	20

Table 1: Hot searches related to "Jiang Ziya" on Weibo.

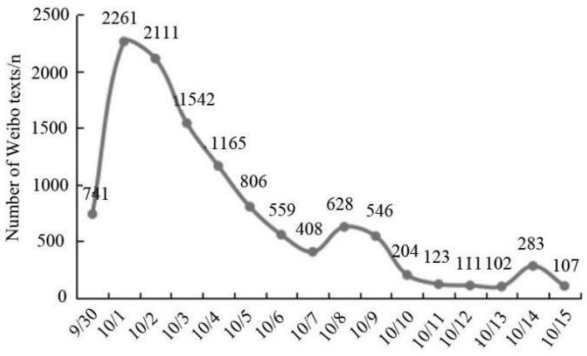


Figure 1: The spread trend of "Jiang Ziya".

keywords more intuitively, they were displayed in the form of the word cloud, and the results are shown in Figure 2.

1	Jiang Ziya
2	Film
3	Animation
4	Plot
5	Good
6	China
7	Expect
8	Hot blood
9	Special effects
10	Post-credit scene

Table 2: Keyword extraction results.

Figure 2: Keyword visualization.

It was seen from Figure 2 that the keywords in the texts could be seen directly through the word cloud. The larger the font was, the more times it appeared. The smaller the font was, the fewer times it appeared. The most prominent words in Figure 2 are "Jiang Ziya", "film", "plot", "animation", "good", "expect", etc., which showed that netizens generally had great expectations on the film. The plot of "Jiang Ziya" has been widely discussed by netizens. However, it was also seen from Figure 2 that there were words such as "polarization", "not good", "low", and "poor", which indicated that some netizens had negative comments on the film and they thought that the film was not good. As shown in Table 3, netizens mainly had two emotional tendencies.

By analyzing the emotional tendency of Weibo text, the number of positive and negative emotional tendencies is obtained by calculating the emotional value, as shown in Figure 3.

It was seen from Figure 3 that texts with a positive emotional tendency ($F(s) > 0$) accounted for 69% (8071) among the collected Weibo texts and texts with a negative emotional tendency ($F(s) < 0$) accounted for 31% (3626), which showed that the emotional tendency of netizens towards "Jiang Ziya" was mainly positive. Generally speaking, netizens love and praise the film.

Negative	We have waited for its release so long, but it is not good.
	"Jiang Ziya" is not good. I slept twice in the process of watching the film. I fell asleep in the cinema for the first time.
	Our expectations of "Jiang Ziya" are very high, and we have other feelings, even more than our expectations of "Nezha: Birth of the Demon Child". Under this expectation, even if the two films are equally good, we will still feel that "Jiang Ziya" is not good because it does not meet our psychological expectations.
	"Jiang Ziya" is not good. It may be because we have waited for its release so long. It is very general.
	How bad is "Jiang Ziya"? I played with my mobile phone when watching the film and left before the post-credit scene.
Positive	"Jiang Ziya" that was released recently is very good.
	I watched the film! Jiang Ziya!!!!!!!!!!!!!!!! So gooooooood! Nezha in the post-credit scene, hahaha! Chinese animation is getting better and better! "Jiang Ziya" expected # Jiang Ziya's post-credit scene, deep sea# it has a feeling of Chinese style, and the hotel makes me remember the film of Miyazaki Hayao.
	The picture is super beautiful. Every frame is super beautiful and fluent, and the plot is wonderful too. I read the comments before watching the film, and someone said that the character of Jiang Ziya is unlikable and the film will not be good with Jiang Ziya as the main character. But now, I think such a comment is wrong. He is likable, the main characters are likable, the supporting roles are likable, and the plot is likable. It doesn't matter what you think, and it matters what I think.
	After watching "Jiang Ziya", I want to say that it is great. I hope it will get better and better. The animation effect is excellent. The adoption of plots is also good, but it can be better. Such excellent works should be encouraged as the revitalization of Chinese animation just start. I am looking forward to the rest of the story in Dujie City.
	"Jiang Ziya" is good, and the elk is so cute!

Table 3: Netizens' emotional tendency towards the film.

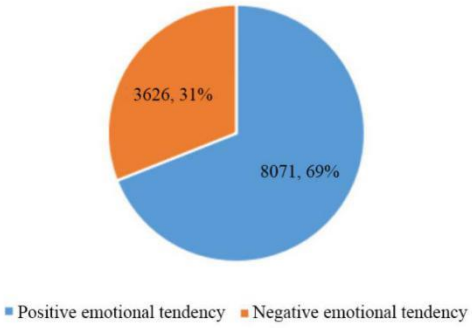


Figure 3: Analysis results of emotional tendencies.

5 Conclusion

In this study, the visualization of public opinion data was analyzed. Taking Weibo texts as the subject, the text data were processed through keyword extraction and emotional tendency analysis and visualized. Then, Weibo texts were collected by taking "Jiang Ziya" as the keyword. The results showed that this method proposed in this study could extract the keyword of the text, visualize texts in the form of the word cloud, and calculate the emotional tendency of texts, which plays a very good role in correctly grasping public opinion data and understanding the emotional attitude of netizens.

References

[1] Fei Y, Qian Z, Xiao G (2017). Evolution mechanism and countermeasures of network public opinion of group emergencies based on data mining method. Boletin Tecnico/Technical Bulletin, 55, pp. 196-202.

[2] Chen Y C, Hui L, Wu C I, Liu H Y, Chen S C (2017). Opinion leaders discovery in dynamic social network. 2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media).

[3] Dai J, Li Y (2017). Modeling and Simulating of Network Public Opinion Evolution Based on Dynamic Reference Point of Prospect Theory. 2017 6th International Conference on Measurement, Instrumentation and Automation (ICMIA 2017).

[4] Tan Y, Lin Q, Luan Y, Chen T, Qiao Y, Luan Y (2019). Campus Network Public Opinion Monitoring System Based on Reptile Technology. IOP Conference Series: Earth and Environmental Science, 252, pp. 052136 (8pp). https://doi.org/10.1088/1755-1315/252/5/052136.

[5] Tang J, Wang J, Li F (2020). Research Progress of Network Public Opinion Based on Fuzzy Set from the Perspective of Big Data. Journal of Physics: Conference Series, 1631, pp. 012108 (6pp). https://doi.org/10.1088/1742-6596/1631/1/012108.

[6] Chai W L, Cheng M (2016). The Research on the Network Public Opinion Risk Assessment based on the CWAHP-Entropy Method. International Journal of Security & Its Applications, 10, pp. 197-208. https://doi.org/10.14257/ijsia.2016.10.4.19.

[7] Zhang X (2016). Network Public Opinion Data Mining Model of Hierarchical Multi Level. Journal of Computational and Theoretical Nanoscience, 13, pp. 9498-9501. https://doi.org/10.1166/jctn.2016.5872.

[8] Yuan F, Yang J, Zheng Q (2019). Research on Network Public opinion Analysis Platform Architecture Based on Big Data. IOP Conference Series: Earth and Environmental Science, 252, pp. 032014 (6pp). https://doi.org/10.1088/1755-1315/252/3/032014.

[9] Wang L (2015). Big Data and Visualization: Methods, Challenges and Technology Progress. Canadian Journal of Electrical & Computer Engineering, 34, pp. 3-6. https://doi.org/10.1109/CJECE.2009.5443861.

[10] Kay S, Zhao B, Sui D (2015). Can Social Media Clear the Air? A Case Study of the Air Pollution Problem in Chinese Cities. Professional Geographer, 67, pp. 351-363. https://doi.org/10.1080/00330124.2014.970838.

[11] Chen K, Zhang Z, Long J, Zhang H (2016). Turning from TF-IDF to TF-IGM for term weighting in text classification. Expert Systems with Applications An International Journal, 66, pp. 245-260.

[12] Jiang X, Qiu L (2013). A Tibetan Ontology Concept Acquisition Method Based on HowNet and Chinese-Tibetan Dictionary. 2013 International Conference on Asian Language Processing.

https://doi.org/10.31449/inf.v45i1.3431 Informatica **45** (2021) 163–168 **163**

Prediction and Estimation of Book Borrowing in the Library: Machine Learning

Jinbao Sun
Graphic Center, Henan Mechanical and Electrical Vocation College, Zhengzhou, Henan 451191, China
E-mail: baojian134597@yeah.net

Keywords: data mining, library, book borrowing, radial basis function neural network, prediction

Received: February 3, 2021

In the library, the prediction and estimation of book borrowing plays an important role in library work. Based on the data mining method, this paper analyzed the prediction and estimation of book borrowing. Firstly, the radial basis function neural network (RBFNN) was analyzed. Then, the improved ant colony algorithm (IACO) was used to obtain the optimal parameters of RBFNN, and then the IACO-RBFNN model was established to realize the prediction and estimation of book borrowing. The results showed that the improved model had advantages in training time, iteration times, and error compared with BPNN and RBFNN. The results of book prediction and estimation showed that the results obtained by the IACO-RBFNN model were closer to the actual book borrowing situation, with smaller error and higher precision (97.09%), and its precision was 11.18% and 4.74% higher than BPNN and RBFNN respectively. The training time and testing time of the IACO-RBFNN model were 5.12 s and 1.03 s, respectively, which were significantly shorter than the other two methods. The results show that the IACO-RBFNN model has a good performance in the prediction and estimation of book borrowing and can be further promoted and applied in practice.

Povzetek: Opisana je metoda strojnega učenja za napovedovanje izposoje knjig v knjižnici.

1 Introduction

The library is an important facility in a school. It can meet the needs of teachers and students in teaching and scientific research by collecting and sorting books. With the expansion of the school scale, the amount of books borrowed in the library is also growing. In the management of the library, the borrowing amount can reflect the work quality of the library to a certain extent and has a reference value for the purchase of new book resources. Therefore, it is of great significance to predict, estimate, and analyze the borrowing amount [1]. Data mining refers to the process of finding hidden and useful information from massive data, which has been widely used in data prediction and estimation [2]. Shan et al. [3] studied the on-line prediction of tool wear, designed a method based on least squares support vector machine regression, and found through experiments that the method had better accuracy than a neural network. Qazi et al. [4] analyzed the role of artificial neural networks in predicting solar radiation. Through the analysis of 24 literature, they found that the prediction error of the artificial neural network was smaller than 20% and it could process a variety of input meteorological parameters. Zhang et al. [5] combined the long short-term memory method with the recurrent neural network to predict the remaining life of lithium-ion batteries. Through experiments and comparison, they found that the method could predict the remaining life of the lithium-ion batteries effectively. Manek et al. [6] used the back propagation neural network (BPNN), generalized regression neural network (GRNN), and radial basis function neural network (RBFNN) to predict the rainfall in Thanjavur district of southern province Tamil Nadu, India, and found that the RBFNN could get the best prediction results. Ramos et al. [7] predicted delayed cerebral ischemia (DCI) in patients with aneurysmal subarachnoid hemorrhage, combined logistic regression model, machine learning model, and automatic encoder, trained and tested the model with 317 cases of data, and found that the method could effectively improve the prediction of DCI in patients. Souri et al. [8] studied the fault prediction of the Internet of things and proposed a model combining multi-layer perceptron and particle swarm optimization algorithm. The experiment showed that the method had short operation time and small memory consumption. Aiming at the problem of urban traffic flow prediction, Hu et al. [9] established GSTAR-SVM model with wavelet transform and predicted the short-term traffic flow. Through experiments, they found that the model had high prediction accuracy. Iqbal et al. [10] evaluated the performance of seven machine learning methods in predicting dengue outbreak and found through an experiment that the LogitBoost integration model had the highest classification accuracy (92%), a sensitivity of 90%, and a specificity of 94%. At present, the application of methods such as data mining and machine learning in library management is seldom, and artificial method is highly dependent, which is not conducive to the scientific management of a large number of books. Therefore, based on RBFNN, this study applied RBFNN to the prediction and estimation of book borrowing and optimized it with

the ant colony optimization (ACO) algorithm to improve the accuracy of prediction and estimation. This work aims to guide book purchase and management of libraries.

2 Book borrowing prediction and estimation model

2.1 RBF neural network

The problem of book borrowing prediction and estimation is affected by many factors and has nonlinear characteristics. However, the models used in this problem, such as the regression analysis model [11] and grey model [12], are all linear estimation models, which has poor estimation accuracy. A neural network is a kind of data mining, which is a simulation of a biological neural network. A neural network is a kind of nonlinear estimation model with excellent nonlinear approximation ability [13]. BPNN [14] and RBFNN [15] have been widely used in prediction and estimation. Compared with BPNN, RBFNN has more advantages in operation speed and structure and has been successfully applied in fields such as human face recognition [16] and defect detection [17]. Therefore, this study used RBFNN to establish the prediction and estimation model of book borrowing.

RBFNN is a three-layer forward network. It is assumed that the input layer of RBFNN has n nodes, $X = (x_1, x_2, \ldots, x_n)$, its output layer has m nodes, $Y = (y_1, y_2, \ldots, y_m)$, its hidden layer has h nodes, then the output of RBFNN can be written as:

$$y_i = f_i(x_i) = \sum_{k=1}^{N} w_{ik}\phi_k(\|x - c_k\|_2), i = 1,2,\cdots,m,$$

where w_{ik} refers to the weight between the hidden layer and output layer, φ_k is the activation function, and c_k is to the center vector of the basis function. In RBFNN, the most commonly used activation function is Gaussian function. Compared with other functions, the Gaussian function is simpler and radial symmetric and has better smoothness. The formula of the Gaussian function is:

$$\phi(x) = exp\left(-\frac{x^2}{2\sigma^2}\right),$$

where σ^2 is a variance. In this case, the output of RBFNN can be written as:

$$y_i = f_i(x_i) = \sum_{k=1}^{N} w_{ik}\, exp\left[-\left(\frac{\|x-c_k\|_2^2}{2\sigma^2}\right)\right], i = 1,2,\cdots,m.$$

To sum up, it can be found that parameters w_{ik}, c_k, and σ have a great influence on the performance of RBFNN, which is also the key and difficult point to establish the RBFNN model. In order to find the optimal parameters, this study selected the ACO algorithm.

2.2 Ant colony algorithm

ACO algorithm is a heuristic algorithm based on simulated ant colony behaviors [18]. In the process of ants' foraging, pheromones will be released. In the process of searching for paths, ants will find the path with high pheromone concentration and release pheromone at the same time, which will make the pheromone concentration on the path higher and higher, and all ants will gather on one path finally. For ant k, the probability of ant k from city i to city j at time t can be written as:

$$\rho_{ij}^k(t) = \begin{cases} \frac{\delta_{ij}^{\alpha}(t)\eta_{ij}^{\beta}}{\sum_{j\in N_j^k}\delta_{ij}^{\alpha}(t)\eta_{ij}^{\beta}}, j \in allowed_k \\ 0, otherwise \end{cases}.$$

The update process of pheromone can be written as:

$$\delta_{ij}(t+n) = p\delta_{ij}(t) + \sum_{k=1}^{m}\Delta\delta_{ij}^k,$$

$$\Delta\delta_{ij}^k = \begin{cases} L_k^{-1}, j \in allowed \\ 0, otherwise \end{cases}.$$

In the above formula, the parameters involved and their meanings are shown in Table 1.

Parameter	Meaning
δ_{ij}	Pheromone concentration
η_{ij}	Heuristic factor
α	The importance of pheromone
β	The importance of heuristic factor
N_j^k	A city without passing by
$allowed_k$	Feasible solution set
p	Pheromone volatilization factor
L_k	The length of a path that ant k passes

Table 1: Parameter table

In the ACO algorithm, the value of volatilization factor p is between 0 and 1. When the value of p is too large, it may affect the global search ability of the algorithm. Therefore, this study used an adaptive method to improve the ACO algorithm. The initial value of p is set as 0.9, and then it changes followed the following formula:

$$p_t = \begin{cases} 0.9p(t-1), 0.9p(t-1) \geq p_{low} \\ p_{low}, p_{low} \end{cases},$$

where p_{low} is the minimum value of p.

2.3 Improved ACO-RBFNN model

The parameters of RBF were optimized using the improved ACO (IACO) algorithm. It is assumed that there are m parameters including w_{ik}, c_k, and σ, and they are randomly sorted, which is set as R_i, $i \in [1, \mathrm{m}]$. R_i

is taken as the food source, and the optimal parameters are searched according to the IACO algorithm. When all the ants concentrated on the same route, the parameters obtained at that moment were optimal for RBF. The flow

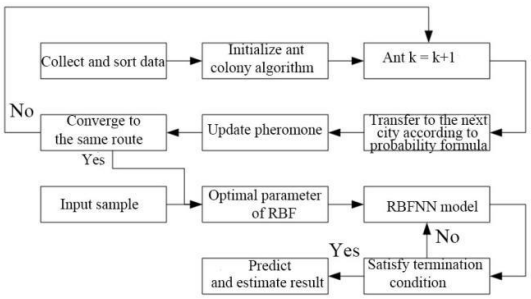


Figure 1: The IACO-RBFNN model.

chart of the IACO-RBFNN model designed for book borrowing prediction and estimation is shown in Figure 1.

As shown in Figure 1, the optimal parameters of RBFNN are obtained using the IACO algorithm after the collected data are processed, those parameters are used for establishing the RBFNN model, and the model is trained by inputting training samples until the model iterates out the most accurate result. The obtained result is the prediction and estimation result of book borrowing.

3 Experimental analysis

3.1 Experimental data

Taking the Graphic Center of Henan Mechanical and Electrical Vocation College as an example, the book borrowing data from January 2018 to June 2020 (24 months) were collected through the library information system. During the training of the IACO-RBFNN model, the data from January 2018 to December 2019 was used as training samples, the number of books borrowed in one month as one sample. The data of the fourth month were predicted based on the data of the first three months, for example, estimating the data of April 2018 based on the data of January 2018 ~ March 2018, i.e., the number of books borrowed in January, February, and March 2018 were taken as the input of the RBFNN model, and the number of books borrowed April 2018 was taken as the

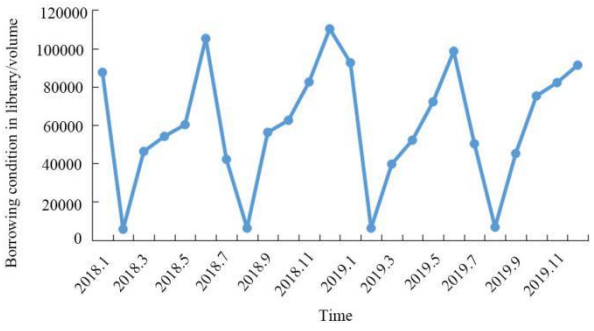


Figure 2: Data of the training sample.

output of the RBFNN model. The training sample is shown in Figure 2.

It was seen from Figure 2 that there is a law in the borrowing of books. In January, June, and December of each year, the number of books borrowed is relatively large, while the number in February and August is small. The above phenomenon may be related to the particularity of the school. In the months of the final examination, the borrowing demand for books is great, while the borrowing demand significantly decreases during the winter and summer vacation.

In order to speed up the operation of the formula, it is necessary to standardize the collected data using the following formula:

$$x' = \frac{x - x_{min}}{xmin_{max}},$$

where x' refers to the normalized data, x is the original data, and x_{max} and x_{min} are the maximum and minimum values of the original data.

3.2 Experimental results

Firstly, a nonlinear function, $y = f(x)$, $x \in [-1,1]$, was used to verify the performance of the IACO-RBFNN model, and it was compared with the traditional BPNN and traditional RBFNN models. The target accuracy was 0.01, and the maximum number of iterations was 500. The

	BPNN	RBFNN	IACO-RBFNN
Training time/s	36.78	10.16	2.34
Number of iterations	245	148	34
Error	1.26	0.53	0.26

Table 2: Comparison of model performance.

performance of the three methods is shown in Table 2.

It was seen from Table 2 that the IACO-RBFNN model had obvious advantages in performance. Firstly, in terms of training time, the BPNN model took 36.78 s in training, the RBFNN model took 10.16 s, and the IACO-RBFNN model only took 2.34 s, which was significantly shorter than the other two models; secondly, in terms of the number of iterations, the BPNN model needed 245 times of iterations to get the optimal value, the RBFNN model needed 148 times, and the IACO-RBFNN model only needed 34 times; finally, from the perspective of error, BPNN model > RBFNN model > IACO-RBFNN model. In a comprehensive view, the model designed in this paper had the best performance.

In the book borrowing prediction and estimation from January 2020 to June 2020, the results of the three models are shown in Figure 3.

It was seen from Figure 3 that there was a gap between the estimated results of the BPNN and RBFNN models and the actual borrowing situation. The error of the

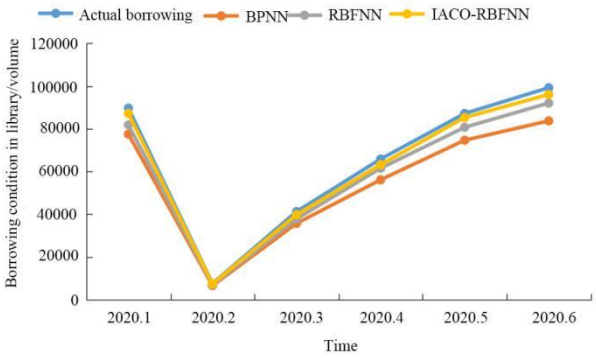


Figure 3: Book borrowing prediction results of three models.

RBFNN model was larger than the BPNN model, which indicated that the RBFNN model had a better performance. Compared with the RBFNN model, the estimated result of the IACO-RBFNN model was closer to the actual borrowing result, which showed that the RBFNN model had a significantly improved performance after improvement by the IACO algorithm, and it had better accuracy in the prediction and estimation of book borrowing. In order to further verify the effectiveness of the model designed in this study, the error and precision of the three models were calculated, and the results are shown in Table 3.

It was seen from Table 3 that the error of the BPNN model was the largest, while that of the IACO-RBFNN model was the smallest. In the prediction and estimation, the average error of the BPNN, RBFNN, and IACO-RBFNN models was 9404 books, 4967 books, and 1955 books, respectively, and the average error of the IACO-RBFNN model was 79.21% less than that of the BPNN model and 60.64% less than that of the RBFNN model. The average precision of the BPNN and RBFNN models was 85.91% and 92.35%, respectively, while the average precision of the IACO-RBFNN model was 97.09%, which

	BPNN model		RBFNN model		IACO-RBFNN model	
	Error	Precision	Error	Precision	Error	Precision
January 2020	12175	86.42 %	7836	91.26 %	2457	97.26 %
February 2020	962	87.26 %	577	92.35 %	139	98.16 %
March 2020	5553	86.54 %	3437	91.67 %	1452	96.48 %
April 2020	9702	85.26 %	4305	93.46 %	2646	95.98 %
May 2020	12516	85.64 %	6467	92.58 %	1874	97.85 %
June 2020	15518	84.36 %	7183	92.76 %	3165	96.81 %
Average value	9404	85.91 %	4967	92.35 %	1955	97.09 %

Table 3: Comparison results of error and precision.

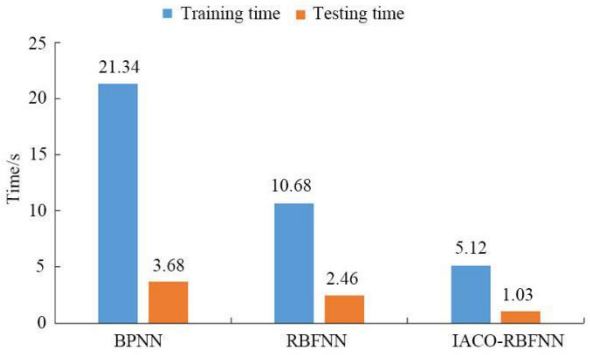


Figure 4: Comparison of operation time.

was 11.18 higher than the BPNN model and 4.74% higher than the RBFNN model. Thus, it was concluded that the IACO-RBFNN model was the most effective in the prediction and estimation of book borrowing.

Finally, the operation time of the model was compared, as shown in Figure 4.

It was seen from Figure 4 that the operation time of the BPNN model was the longest, followed by the RBFNN and IACO-RBFNN models. The training time and testing time of the BPNN model were 21.34 s and 3.68 s, respectively; the operation time of the RBFNN model significantly reduced. The training time of the IACO-RBFNN model was 5.12 s, which improved 52.06% compared to the RBFNN model, and the testing time of the model was 1.03 s, which improved 58.13% compared to the RBFNN model. It was found that the algorithm improved by the IACO algorithm had improved precision and significantly shortened operation time and showed a better performance in the prediction and estimation of book borrowing.

4 Discussion

The development of information technology has brought new changes to many fields. Many industries have established information systems to realize information management, so does the library [19]. In the process of library informatization, a large amount of information has been accumulated, but most digital libraries cannot effectively develop and utilize these data and make the information accumulate, which brings great difficulties to the resource management and data processing of the library. In order to develop the library better, a method is urgently needed to realize the analysis and utilization of these data, and the emergence of data mining solves this problem [20].

This study mainly analyzed RBFNN. For the parameter selection of RBFNN, many algorithms have been applied, such as the gravity search algorithm [21], genetic algorithm [22], grey wolf optimization algorithm [23], etc. This paper selected the ACO algorithm, improved the ACO algorithm to optimize the parameters of RBFNN, carried out experiments with the actual book lending data, and compared the IACO-RBFNN model with the BPNN and RBFNN models. The results suggested that the IACO-RBFNN model needed fewer times of iterations, shorter training time, and smaller error compared to the BPNN and RBFNN models, indicating

that the IACO-RBFNN model had more obvious advantages in performance. In the prediction and estimation of book borrowing, the prediction result of the RBFNN model was closer to the actual situation than that of the BPNN model, suggesting that the performance of the RBFNN model was better than that of RBFNN. Then, the comparison between the IACO-RBFNN model and the RBFNN model found that the prediction result of the IACO-RBFNN model was closer to the actual situation. It was seen from Table 3 that the IACO-RBFNN model had smaller prediction error and higher precision, and the average precision of the method was 97.09%, which was 11.18% higher than that of the BPNN model and 4.74% higher than that of the RBFNN model. In the comparison of the operation time, the IACO-RBFNN model was significantly shorter, i.e., it could obtain results with high precision in a short time, which showed that the IACO-RBFNN model had better usability in the prediction and estimation of book borrowing.

Although some useful achievements have been made in this article, there are some deficiencies that need to be solved in future work:

(1) comparing the performance of more data mining methods;
(2) further optimizing the precision of the RBFNN model;
(3) studying more applications of data mining methods in libraries.

5 Conclusion

This study designed the IACO-RBFNN model for the prediction and estimation of book borrowing in the library. Taking the data in the Graphic Center of Henan Mechanical and Electrical Vocation College as an example, the experiment was carried out. The comparison with the BPNN and RBFNN models found that:

(1) the IACO-RBFNN model needed shorter training time and fewer times of iterations and had smaller error;

(2) the predicted result of the IACO-RBFNN model was closer to the actual book borrowing situation;

(3) the average error and average precision of the IACO-RBFNN model was 1955 books and 97.09%, which was better than the other two models;

(4) the training time and testing time of the IACO-RBFNN model were only 5.12 s and 1.03 s, respectively.

It was found from the results that the IACO-RBFNN model had a good performance in the prediction and estimation of book borrowing and could be applied in the actual library work to give guidance for the library work.

References

[1] Xu SX (2020). Association rule model of on-demand lending recommendation for university library. Informatica, 44, pp. 395-399. https://doi.org/10.31449/inf.v44i3.3295.

[2] Javeed S, Mohamed S (2020). Novel Feature Reduction (NFR) Model With Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction. IEEE Access, 8, pp. 184087-184108. https://doi.org/10.1109/ACCESS.2020.3028714.

[3] Guan S, Yan LH, Peng C (2015). Application of regression algorithm of LS-SVM in tool wear prediction. China Mechanical Engineering, 26, pp. :217-222. https://doi.org/10.3969/j.issn.1004-132X.2015.02.016.

[4] Qazi A, Fayaz H, Wadi A, Raj R, Rahim N, Khan W (2015). The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. Journal of Cleaner Production, 104, pp. 1-12. https://doi.org/10.1016/j.jclepro.2015.04.041.

[5] Zhang Y, Xiong R, He H, Pecht MG (2018). Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. IEEE Transactions on Vehicular Technology, 67, pp. 5695-5705. https://doi.org/10.1109/TVT.2018.2805189.

[6] Manek AH, Singh PK (2016). Comparative study of neural network architectures for rainfall prediction. pp. 171-174. https://doi.org/10.1109/TIAR.2016.7801233.

[7] Ramos L A, Steen W E V D, Barros R S, Majoie CBLM, van den Berg R, Verbaan D, Vandertop WP, Zijlstra IJAJ, Zwinderman AH, Strijkers GJ, Olabarriaga SD, Marquering HA (2019). Machine learning improves prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage. Journal of Neurointerventional Surgery, 11, pp. 497-502. https://doi.org/10.1136/neurintsurg-2018-014258.

[8] Souri A, Mohammed A S, Potrus M Y, Malik MH, Safara F, Hosseinzadeh M (2020). Formal Verification of a Hybrid Machine Learning-Based Fault Prediction Model in Internet of Things Applications. IEEE Access, 8, pp. 23863-23874. https://doi.org/10.1109/ACCESS.2020.2967629.

[9] Hu J, Wang S, Mao J (2019). Research on GSTAR-SVM Traffic Prediction Model Based on Wavelet Transform. Journal of Physics: Conference Series, 1345, pp. 032009 (7pp). https://doi.org/10.1088/1742-6596/1345/3/032009

[10] Iqbal N, Islam M (2019). Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers. Informatica, 43, pp. 363-371. https://doi.org/10.31449/inf.v43i3.1548.

[11] Wang CY, Wang WS (2015). Regression Analysis When Covariates Are Regression Parameters of a Random Effects Model for Observed Longitudinal Measurements. Biometrics, 56, pp. 487-495. https://doi.org/10.1111/j.0006-341X.2000.00487.x.

[12] Fu M, Wang W, Le Z, Khorram MS (2015). Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model. Neural Computing & Applications, 26, pp. 1789-1797. https://doi.org/10.1007/s00521-015-1853-8.

[13] Boufadene M, Belkheiri M, Rabhi A, Hajjaji AE (2019). Vehicle longitudinal force estimation using

adaptive neural network nonlinear observer. International Journal of Vehicle Design, 79, 205-. https://doi.org/10.1504/IJVD.2019.103593.

[14] Lyu J, Zhang J (2018). BP Neural Network Prediction Model for Suicide Attempt among Chinese Rural Residents. Journal of Affective Disorders, 246, pp. 465-473.

[15] Xiong T, Bao Y, Hu Z, Chiong R (2015). Forecasting interval time series using a fully complex-valued RBF neural network with DPSO and PSO algorithms. Information Sciences, 305, pp. 77-92. https://doi.org/10.1016/j.ins.2015.01.029.

[16] Dey A, Ghosh M (2019). A Novel Approach to Fuzzy-Based Facial Feature Extraction and Face Recognition. Informatica, 43, pp. 535-543. https://doi.org/10.31449/inf.v43i4.2117.

[17] Jiang HN (2018). Defect features recognition in 3D industrial CT images. Informatica, 42, pp. 477-482. https://doi.org/10.31449/inf.v42i3.2454.

[18] Mandloi M, Bhatia V (2015). Congestion control based ant colony optimization algorithm for large MIMO detection. Expert Systems with Applications, 42, pp. 3662-3669. https://doi.org/10.1016/j.eswa.2014.12.035.

[19] Day A (2018). Research Information Management: How the Library Can Contribute to the Campus Conversation. New Review of Academic Librarianship, 24, pp. 23-34. https://doi.org/10.1080/13614533.2017.1333014.

[20] Wang W, Meng L, Wu L, Zhang J. (2020). Research and Application of Data Mining Technology in Library Office Information Construction. Journal of Physics: Conference Series, 1550, pp. 032001 (6pp). https://doi.org/10.1088/1742-6596/1550/3/032001.

[21] Assareh E, Biglari M (2015). A novel approach to capture the maximum power from variable speed wind turbines using PI controller, RBF neural network and GSA evolutionary algorithm. Renewable & Sustainable Energy Reviews, 51, pp. 1023-1037. https://doi.org/10.1016/j.rser.2015.07.034.

[22] Jia W, Zhao D, Ding L (2016). An optimized RBF neural network algorithm based on partial least squares and genetic algorithm for classification of small sample. Applied Soft Computing, pp. 373-384. https://doi.org/10.1016/j.asoc.2016.07.037.

[23] Shang S, He K N, Wang Z B, Yang T, Liu M, Li X. (2020). Sea Clutter Suppression Method of HFSWR Based on RBF Neural Network Model Optimized by Improved GWO Algorithm. Computational Intelligence and Neuroscience, 2020, pp. 1-10. https://doi.org/10.1155/2020/8842390.

https://doi.org/10.31449/inf.v45i1.3482 Informatica **45** (2021) 169–170 **169**

A Method for Combining Classical and Deep Machine Learning for Mobile Health and Behavior Monitoring

Martin Gjoreski
Department of Intelligent Systems, Jozef Stefan Institute, Jamova 39, Ljubljana, Slovenia
E-mail: martin.gjoreski@ijs.si
https://martingjoreski.github.io

Keywords: machine learning, deep learning, mobile health, behavior monitoring, wearable sensors

Received: March 30, 2021

This paper summarizes the doctoral dissertation of the author, which presents a method for fusing classical and deep machine learning for mobile health and behavior monitoring with wearable sensors.

Povzetek: Prispevek povzame doktorsko disertacijo avtorja, ki temelji na metodi za kombiniranje klasičnega in globokega strojnega učenja za mobilno spremljanje zdravja in obnašanja z nosljivimi senzorji..

1 Introduction

Commercially available smartphones, smart glasses, smartwatches, and smart rings are just a few examples of sensor-packed devices that are enabling the technological revolution currently underway. To further extend the successful applicability of wearable devices in sectors such as mobile health, methods for accurate measurements of psycho-physiological information are required. However, accessing psycho-physiological information using wearable devices remains challenging. One reason is that the relationship between sensor data and human psycho-physiological states is not as unambiguous as the relationship between sensor data and individual physical states is. Thus, we are facing a question: How can we transform wearable sensor data into valuable human health and behavior information? Such information has the potential to improve healthcare, decrease healthcare costs, improve the quality of life and, ultimately, save human lives.

For a decade, deep learning (DL) has dominated the AI world by achieving a breakthrough in several areas such as image processing, natural language processing, and reinforcement learning. Thus, a successful fusion of classical machine learning (ML) and DL methods could lead to beyond state-of-the-art results for mobile health and behavior monitoring.

2 Case studies

The proposed method was applied in seven health and behavior-monitoring domains [1]: stress recognition from physiological sensors, blood pressure estimation from ECG sensors, emotion recognition from physiological sensors and cognitive-load recognition from physiological sensors, chronic heart failure monitoring from heart sounds [2], driver distractions monitoring from physiological and video-based sensors [3], and locomotion recognition from smartphone sensors [4].

3 Method

The proposed method (Figure 1) extracts valuable human health and behavior information from wearable sensor data. The method uses end-to-end learning on sensor data as a standalone approach or in combination with classical ML to produce beyond state-of-the-art performance.

The method uses as input any data collected using wearable sensors from human users. The type of sensors depends on the use-case. Regarding the sensors utilized in the thesis, in the studies on stress and cognitive load monitoring, physiological and acceleration data from a wrist-worn device was used. In the study on emotions, physiological data from wearable sensors was used. In the study on driving distractions, physiological and video-based sensors were used. In the study on locomotion recognition, smartphone sensors were used. In the study on blood pressure estimation, data from chest-worn ECG sensor was used. Finally, in the study on chronic heart failure (CHF) detection, digital stethoscope was used to record heart sounds. Each of these studies had a different hardware setup, while the method is hardware-independent. Regarding the data labels, in the study on CHF, the labels were provided by medical experts. In the rest of the studies, the labels were provided by the users themselves.

The data from the wearable devices is quite often noisy. The usual source of the noise are movement artefacts and sensor misplacement. The filtering strategies include winsorization, detrending, moving average, low-pass, high-pass, band-pass, etc. The sensor data can be transformed into different domains (e.g., time domain and frequency domain), each of them specialized for extracting different types of information from the input data. For example, by combining gyroscope and acceleration data from smartphone data, the acceleration data can be rotated. This produces location-independent acceleration data, which is useful for more robust activity recognition from smartphone sensors. Another simple transformation calculates the acceleration magnitude by combining the sensor from each axis (x-, y- and z-axis).

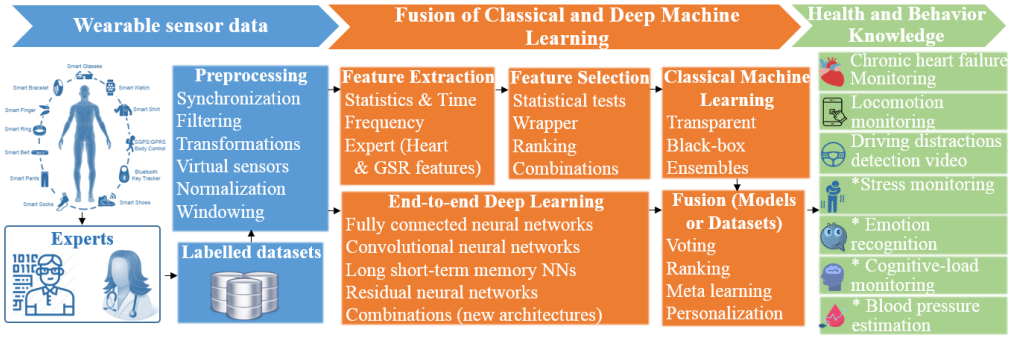


Figure 1: A fusion of classical and deep machine learning for mobile health and behavior monitoring.

Some people have a faster heart rate than other people, some sweat more than others, some walk faster than others, etc. This variability can harm any ML model, especially if there is a small dataset to train the model on. To minimize the variability, different normalization techniques can be employed. The normalization can be done either on the sensor data, or on features. In both cases, the normalization has similar effects, i.e., it scales the values of the variables, and in some cases, it also changes the distribution of the variables.

Informative features should be extracted which are used as input into classical ML algorithms. The feature extraction is an important step as it offers the possibility to encode expert knowledge into the system. In addition to the expert knowledge, for domains where expert knowledge is not well defined, all possible features can be extracted by "borrowing" expert knowledge from similar domains. In the thesis, we experimented with several types of features: statistical features, frequency features, heart-related features and galvanic skin response features.

In the study on monitoring stress, a new feature selection method was proposed by combining ranking and wrapper methods. The method aims to minimize the number of evaluations required by the wrapper method by prioritizing top features ranked by information gain, and by removing low-ranked features which are correlated with the top-ranked features.

The extracted features can be fed into a variety of classical ML algorithms including algorithms that produce comprehensible classifiers (e.g., Decision Trees), black-box classifiers (e.g., SVM) and ensembles (e.g., Extreme Gradient Boosting).

Processed sensor data can be also used to build end-to-end DL models, i.e., models that do not require feature extraction and learn directly from sensor data with the potential to discover new useful patterns in the data, previously unknown to experts. The DL models include existing DL architectures (e.g., Convolutional Neural Networks, Long Short-Term Memory Neural Networks, their combination – ConvLSTMs, Residual Networks – ResNet etc.), and our novel DL architecture Spectro-Temporal Residual Network (STResNet). STResNet is a novel DL architecture for end-to-end learning specialized for multimodal sensor data. More specifically, STResNet can learn from several sensors simultaneously, each of them having a different sampling frequency, and it learns both in the time domain and in the frequency domain.

Regarding the final fusion of the models ML and DL models, depending on the use-case, in some cases, only the highest-ranked model is used to minimize complexity. In other cases, classical meta-learners and voting ensembles are used to maximize accuracy. In third cases, meta-learners that can account for temporal dependencies in the data are used.

4 Conclusion

This paper summarized the dissertation [1] and presented the main idea and findings of the same. The thesis presented: a novel general method that combines expert knowledge, classical machine learning, and deep learning for extracting human physical, physiological, and psychological information from wearable sensor data; a novel deep learning architecture for end-to-end learning (STResNet) specialized for multimodal sensor data; unified application of the method on seven domains; new datasets and publicly available software for mobile health and behavior monitoring with wearable sensors.

References

[1] M. Gjoreski, A fusion of classical and deep machine learning for mobile health and behavior monitoring with wearable sensors, PhD Thesis, IPS Jožef Stefan, Ljubljana, Slovenia, 2020.

[2] M. Gjoreski, A. Gradišek, B. Budna, M. Gams, and G. Poglajen, "Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds," IEEE Access, vol. 8, pp. 20313–20324, 2020. https://doi.org/10.1109/ACCESS.2020.2968900

[3] M. Gjoreski, M. Gams, M. Luštrek, P. Genc, J.-U. Garbas, and T. Hassan, "Machine learning and end-to-end deep learning for monitoring driver distractions from physiological and visual signals," IEEE Access, vol. 8, pp. 70590–70603, 2020. https://doi.org/10.1109/ACCESS.2020.2986810

[4] M. Gjoreski, V. Janko, G. Slapničar, M. Mlakar, N. Reščič, J. Bizjak, V. Drobnič, M. Marinko, N. Mlakar, M. Luštrek, and M. Gams, "Classical, and deep learning methods for recognizing human activities, and modes of transportation with smartphone sensors," Information Fusion, vol. 62, pp. 47–62, 2020. https://doi.org/10.1016/j.inffus.2020.04.004

Extraction and Evaluation of Software Components from Object-Oriented Artifacts

Amit Rathee and Jitender Kumar Chhabra
Computer Engg. Dept., National Institute of Technology, Kurukshetra-136119, Haryana, India
E-mail: amit1983_rathee@rediffmail.com, jitenderchhabra@gmail.com

Thesis summary

Keywords: CBSD, reusability, component extraction, software artifacts, component, JavaBeans, frequent usage patterns

Received: March 7, 2021

A doctoral thesis is summarized in this paper that focuses on strengthening the Component-Based Software Development (CBSD) approach by proposing an efficient approach for extracting and evaluating reusable software components from an Object Oriented (OO) software by utilizing its various artifacts. The carried out research work mainly consists of two main steps: (1) extracting a possible set of components by utilizing optimal software artifacts and clustering techniques; (2) identifying reusable components by evaluating the quality of different components using the proposed reusability metric suite. The carried out research work significantly helps in identifying and extracting the reusable components for the CBSD environment and the proposed metric suite helps in evaluating the quality of all components.

Povzetek: Predstavljen je povzetek doktorata na temo obdelovanja programskih komponent pri objektnem programiranju.

1 Introduction

With a fast-paced changing world, software functionalities demand continuous modification. Software reuse principles significantly help in faster development within allotted budget and CBSD is commonly used for it. The key composing unit (aka reusable unit) in the CBSD environment is called a component and it hides the complexity of its implementation behind its provides and requires interface. Such components possess larger granularity as compared to classes/ objects in object-oriented languages. Hence reusable components should be identified from existing OO legacy software systems and stored in a component library in order to use them for future development. This motivates the researchers to identify the affecting factors and develop some efficient techniques for extracting high quality components and quantifying their overall quality.

2 Methodology

The thesis deals with extracting and evaluating reusable components by utilizing soft computing techniques, efficient selection of software artifacts, and bio-inspired algorithms. It consists of five main steps as depicted in Figure-1.

1. The first step aims at analyzing dependency relations among different software elements (classes and/ or interfaces) based on optimal dependency information extracted by utilizing different software artifacts

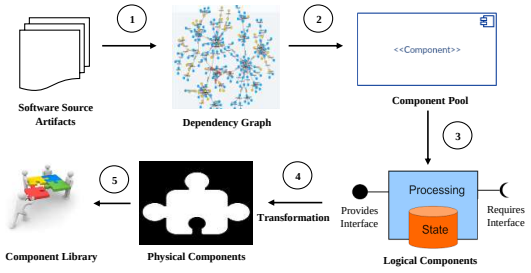


Figure 1: Proposed Research Methodology.

[1, 7, 8]. Based on the study, it was determined that combined use of structural, conceptual, and change-history based (called evolutionary) relations helps in estimating optimal dependency relations with 60% or more weight factor value assigned to evolutionary relations [2, 9]. Further, the authors determined that frequent usage patterns help in measuring more accurate structural dependency relations [3].

2. In the second step, different software elements and their dependencies are modeled as a graph and are clustered by grouping one or more strongly-connected elements into a single cluster such that each cluster is minimally connected with rest of the clusters. Different clustering algorithms are studied and best clustering algorithm is further used [5].

3. In the third step, different obtained clusters are ana-

lyzed to identify interfaces (provides and requires) of the component [4], called as the logical component.

4. The logical components are transformed into the corresponding reusable physical component by following recommendations of a Java Beans component model [6].

5. The authors further propose a set of reusability metric suite for measuring the reusability of a component and use it in the fifth step to identify high quality components for the CBSD environment based on cohesion, coupling, customizability, self-completeness, and interface complexity parameters.

3 Results

Dependency relations of software elements as well as clustering algorithms are analysed using precision, recall, F-measure and modularization metric (TurboMQ). Reusable software components are empirically identified and evaluated using some well-known IR metrics and TurboMQ metric. The results for the proposed metric suite are collected and evaluated over three different categories of software, specifically designed to have different levels of reusability. Moreover, human expertise is also considered for cross verifying the obtained reusability scores.

4 Conclusion

The thesis proposes an efficient novel approach for extracting reusable software components from existing legacy softwares. The thesis proposes a new efficient frequent usage pattern structural dependency measure approach and uses it in combination with other conceptual and evolutionary dependency relations for optimal measurement of the dependency especially useful from component point-of-view. It also proposes a novel metric suite for measuring the reusability of a software component designed as per the specifications of Java Beans. The carried out research work is able to effectively quantify various dependencies and overall quality of software components and can be used by IT companies to develop reusable component repositories.

References

[1] Imad Eddine Araar and Hassina Seridi. Software features extraction from object-oriented source code using an overlapping clustering approach. *Informatica*, 40(2), 2016.

[2] Amit Rathee and Jitender Kumar Chhabra. Clustering for software remodularization by using structural, conceptual and evolutionary features. *Journal of Universal Computer Science*, 24(12):1731–1757, 2018a. doi: 10.1007/978-981-10-5780-9_9.

[3] Amit Rathee and Jitender Kumar Chhabra. Improving cohesion of a software system by performing usage pattern based clustering. *Procedia Computer Science*, 125:740–746, 2018b. doi: https://doi.org/10.1016/j.procs.2017.12.095.

[4] Amit Rathee and Jitender Kumar Chhabra. Reusability in multimedia softwares using structural and lexical dependencies. *Multimedia Tools and Applications*, 78(14):20065–20086, 2019a. doi: https://doi.org/10.1007/s11042-019-7382-1.

[5] Amit Rathee and Jitender Kumar Chhabra. A multi-objective search based approach to identify reusable software components. *Journal of Computer Languages*, 52:26–43, 2019b. doi: https://doi.org/10.1016/j.cola.2019.01.006.

[6] Amit Rathee and Jitender Kumar Chhabra. Mining reusable software components from object-oriented source code using discrete pso and modeling them as java beans. *Information Systems Frontiers*, 22 (6):1519–1537, 2020. doi: https://doi.org/10.1007/s10796-019-09948-4.

[7] Cherait, Hanene, and Nora Bounour. "History-based approach for detecting modularity defects in aspect oriented software." *Informatica* 39.2 (2015).

[8] Amit Rathee, and Jitender Kumar Chhabra. "Software remodularization by estimating structural and conceptual relations among classes and using hierarchical clustering." *International Conference on Advanced Informatics for Computing Research*. Springer, Singapore, 2017. doi: 10.1007/978-981-10-5780-9_9

[9] Amit Rathee and Jitender Kumar Chhabra. "Sensitivity Analysis of Evolutionary Algorithm for Software Reusability." *MENDEL*. Vol. 25. No. 1. 2019. doi: https://doi.org/10.13164/mendel.2019.1.031

Comments About a Paper Titled »A Model and Framework for Online Security Benchmarking«

SrinivasaRao SubramanyaRao
E-mail: srinivasa.subramanya.anu@gmail.com

Keywords: online security benchmarking

Received: March 17, 2021

This paper comments about overlap in contents between a previously published paper in this journal and another paper in another journal.

1 Introduction

We wish to inform the readers of this journal that a paper titled **A model and framework for online security benchmarking** that was published in this journal in 2007 and authored by Graeme Pye and Matthew J Warren [1] has considerable overlap with another paper previously published in another journal and written by the same authors. This paper is titled **E-business security benchmarking: a model and framework** [2].

Neither of the two papers referred to above cites the other and thus it is useful to note this and alert readers to this.

References

[1] Graeme Pye and Matthew J Warren. (2007) , A model and framework for online security benchmarking, *Informatica : Journal of computing and informatics*, Slovensko društvo Informatika (Vol. 31, No. 2, pp. 209-215) http://hdl.handle.net/10536/DRO/DU:30007062.

[2] Graeme Pye and Matthew J Warren. (2007), E-business security benchmarking : a model and framework, *International journal of information and computer security*, Inderscience Publishers (Vol. 1, No. 4, pp. 378-390). http://hdl.handle.net/10536/DRO/DU:30007307. https://doi.org/10.1504/ijics.2007.015499

JOŽEF STEFAN INSTITUTE

Jožef Stefan (1835-1893) was one of the most prominent physicists of the 19th century. Born to Slovene parents, he obtained his Ph.D. at Vienna University, where he was later Director of the Physics Institute, Vice-President of the Vienna Academy of Sciences and a member of several scientific institutions in Europe. Stefan explored many areas in hydrodynamics, optics, acoustics, electricity, magnetism and the kinetic theory of gases. Among other things, he originated the law that the total radiation from a black body is proportional to the 4th power of its absolute temperature, known as the Stefan–Boltzmann law.

The Jožef Stefan Institute (JSI) is the leading independent scientific research institution in Slovenia, covering a broad spectrum of fundamental and applied research in the fields of physics, chemistry and biochemistry, electronics and information science, nuclear science technology, energy research and environmental science.

The Jožef Stefan Institute (JSI) is a research organisation for pure and applied research in the natural sciences and technology. Both are closely interconnected in research departments composed of different task teams. Emphasis in basic research is given to the development and education of young scientists, while applied research and development serve for the transfer of advanced knowledge, contributing to the development of the national economy and society in general.

At present the Institute, with a total of about 900 staff, has 700 researchers, about 250 of whom are postgraduates, around 500 of whom have doctorates (Ph.D.), and around 200 of whom have permanent professorships or temporary teaching assignments at the Universities.

In view of its activities and status, the JSI plays the role of a national institute, complementing the role of the universities and bridging the gap between basic science and applications.

Research at the JSI includes the following major fields: physics; chemistry; electronics, informatics and computer sciences; biochemistry; ecology; reactor technology; applied mathematics. Most of the activities are more or less closely connected to information sciences, in particular computer sciences, artificial intelligence, language and speech technologies, computer-aided design, computer architectures, biocybernetics and robotics, computer automation and control, professional electronics, digital communications and networks, and applied mathematics.

The Institute is located in Ljubljana, the capital of the independent state of S**love**nia (or S♡nia). The capital today is considered a crossroad between East, West and Mediterranean Europe, offering excellent productive capabilities and solid business opportunities, with strong international connections. Ljubljana is connected to important centers such as Prague, Budapest, Vienna, Zagreb, Milan, Rome, Monaco, Nice, Bern and Munich, all within a radius of 600 km.

From the Jožef Stefan Institute, the Technology park "Ljubljana" has been proposed as part of the national strategy for technological development to foster synergies between research and industry, to promote joint ventures between university bodies, research institutes and innovative industry, to act as an incubator for high-tech initiatives and to accelerate the development cycle of innovative products.

Part of the Institute was reorganized into several high-tech units supported by and connected within the Technology park at the Jožef Stefan Institute, established as the beginning of a regional Technology park "Ljubljana". The project was developed at a particularly historical moment, characterized by the process of state reorganisation, privatisation and private initiative. The national Technology Park is a shareholding company hosting an independent venture-capital institution.

The promoters and operational entities of the project are the Republic of Slovenia, Ministry of Higher Education, Science and Technology and the Jožef Stefan Institute. The framework of the operation also includes the University of Ljubljana, the National Institute of Chemistry, the Institute for Electronics and Vacuum Technology and the Institute for Materials and Construction Research among others. In addition, the project is supported by the Ministry of the Economy, the National Chamber of Economy and the City of Ljubljana.

Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel.:+386 1 4773 900, Fax.:+386 1 251 93 85
WWW: http://www.ijs.si
E-mail: matjaz.gams@ijs.si
Public relations: Polona Strnad

INFORMATICA

AN INTERNATIONAL JOURNAL OF COMPUTING AND INFORMATICS

INVITATION, COOPERATION

Submissions and Refereeing

Please register as an author and submit a manuscript at: http://www.informatica.si. At least two referees outside the author's country will examine it, and they are invited to make as many remarks as possible from typing errors to global philosophical disagreements. The chosen editor will send the author the obtained reviews. If the paper is accepted, the editor will also send an email to the managing editor. The executive board will inform the author that the paper has been accepted, and the author will send the paper to the managing editor. The paper will be published within one year of receipt of email with the text in Informatica MS Word format or Informatica LaTeX format and figures in .eps format. Style and examples of papers can be obtained from http://www.informatica.si. Opinions, news, calls for conferences, calls for papers, etc. should be sent directly to the managing editor.

SUBSCRIPTION

Please, complete the order form and send it to Dr. Drago Torkar, Informatica, Institut Jožef Stefan, Jamova 39, 1000 Ljubljana, Slovenia. E-mail: drago.torkar@ijs.si

Since 1977, Informatica has been a major Slovenian scientific journal of computing and informatics, including telecommunications, automation and other related areas. In its 16th year (more than twentyseven years ago) it became truly international, although it still remains connected to Central Europe. The basic aim of Informatica is to impose intellectual values (science, engineering) in a distributed organisation.

Informatica is a journal primarily covering intelligent systems in the European computer science, informatics and cognitive community; scientific and educational as well as technical, commercial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers accepted by at least two referees outside the author's country. In addition, it contains information about conferences, opinions, critical examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and information industry are presented through commercial publications as well as through independent evaluations.

Editing and refereeing are distributed. Each editor can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the Refereeing Board.

Informatica web edition is free of charge and accessible at http://www.informatica.si.
Informatica print edition is free of charge for major scientific, educational and governmental institutions. Others should subscribe.

Informatica WWW:

http://www.informatica.si/

Referees from 2008 on:

A. Abraham, S. Abraham, R. Accornero, A. Adhikari, R. Ahmad, G. Alvarez, N. Anciaux, R. Arora, I. Awan, J. Azimi, C. Badica, Z. Balogh, S. Banerjee, G. Barbier, A. Baruzzo, B. Batagelj, T. Beaubouef, N. Beaulieu, M. ter Beek, P. Bellavista, K. Bilal, S. Bishop, J. Bodlaj, M. Bohanec, D. Bolme, Z. Bonikowski, B. Bošković, M. Botta, P. Brazdil, J. Brest, J. Brichau, A. Brodnik, D. Brown, I. Bruha, M. Bruynooghe, W. Buntine, D.D. Burdescu, J. Buys, X. Cai, Y. Cai, J.C. Cano, T. Cao, J.-V. Capella-Hernández, N. Carver, M. Cavazza, R. Ceylan, A. Chebotko, I. Chekalov, J. Chen, L.-M. Cheng, G. Chiola, Y.-C. Chiou, I. Chorbev, S.R. Choudhary, S.S.M. Chow, K.R. Chowdhury, V. Christlein, W. Chu, L. Chung, M. Ciglarič, J.-N. Colin, V. Cortellessa, J. Cui, P. Cui, Z. Cui, D. Cutting, A. Cuzzocrea, V. Cvjetkovic, J. Cypryjanski, L. Čehovin, D. Čerepnalkoski, I. Čosić, G. Daniele, G. Danoy, M. Dash, S. Datt, A. Datta, M.-Y. Day, F. Debili, C.J. Debono, J. Dedič, P. Degano, A. Dekdouk, H. Demirel, B. Demoen, S. Dendamrongvit, T. Deng, A. Derezinska, J. Dezert, G. Dias, I. Dimitrovski, S. Dobrišek, Q. Dou, J. Doumen, E. Dovgan, B. Dragovich, D. Drajic, O. Drbohlav, M. Drole, J. Dujmović, O. Ebers, J. Eder, S. Elaluf-Calderwood, E. Engström, U. riza Erturk, A. Farago, C. Fei, L. Feng, Y.X. Feng, B. Filipič, I. Fister, I. Fister Jr., D. Fišer, A. Flores, V.A. Fomichov, S. Forli, A. Freitas, J. Fridrich, S. Friedman, C. Fu, X. Fu, T. Fujimoto, G. Fung, S. Gabrielli, D. Galindo, A. Gambarara, M. Gams, M. Ganzha, J. Garbajosa, R. Gennari, G. Georgeson, N. Gligorić, S. Goel, G.H. Gonnet, D.S. Goodsell, S. Gordillo, J. Gore, M. Grčar, M. Grgurović, D. Grosse, Z.-H. Guan, D. Gubiani, M. Guid, C. Guo, B. Gupta, M. Gusev, M. Hahsler, Z. Haiping, A. Hameed, C. Hamzaçebi, Q.-L. Han, H. Hanping, T. Härder, J.N. Hatzopoulos, S. Hazelhurst, K. Hempstalk, J.M.G. Hidalgo, J. Hodgson, M. Holbl, M.P. Hong, G. Howells, M. Hu, J. Hyvärinen, D. Ienco, B. Ionescu, R. Irfan, N. Jaisankar, D. Jakobović, K. Jassem, I. Jawhar, Y. Jia, T. Jin, I. Jureta, Đ. Juričić, S. K, S. Kalajdziski, Y. Kalantidis, B. Kaluža, D. Kanellopoulos, R. Kapoor, D. Karapetyan, A. Kassler, D.S. Katz, A. Kaveh, S.U. Khan, M. Khattak, V. Khomenko, E.S. Khorasani, I. Kitanovski, D. Kocev, J. Kocijan, J. Kollár, A. Kontostathis, P. Korošec, A. Koschmider, D. Košir, J. Kovač, A. Krajnc, M. Krevs, J. Krogstie, P. Krsek, M. Kubat, M. Kukar, A. Kulis, A.P.S. Kumar, H. Kwaśnicka, W.K. Lai, C.-S. Laih, K.-Y. Lam, N. Landwehr, J. Lanir, A. Lavrov, M. Layouni, G. Leban, A. Lee, Y.-C. Lee, U. Legat, A. Leonardis, G. Li, G.-Z. Li, J. Li, X. Li, X. Li, Y. Li, Y. Li, S. Lian, L. Liao, C. Lim, J.-C. Lin, H. Liu, J. Liu, P. Liu, X. Liu, X. Liu, F. Logist, S. Loskovska, H. Lu, Z. Lu, X. Luo, M. Luštrek, I.V. Lyustig, S.A. Madani, M. Mahoney, S.U.R. Malik, Y. Marinakis, D. Marinčič, J. Marques-Silva, A. Martin, D. Marwede, M. Matijašević, T. Matsui, L. McMillan, A. McPherson, A. McPherson, Z. Meng, M.C. Mihaescu, V. Milea, N. Min-Allah, E. Minisci, V. Mišić, A.-H. Mogos, P. Mohapatra, D.D. Monica, A. Montanari, A. Moroni, J. Mosegaard, M. Moškon, L. de M. Mourelle, H. Moustafa, M. Možina, M. Mrak, Y. Mu, J. Mula, D. Nagamalai, M. Di Natale, A. Navarra, P. Navrat, N. Nedjah, R. Nejabati, W. Ng, Z. Ni, E.S. Nielsen, O. Nouali, F. Novak, B. Novikov, P. Nurmi, D. Obrul, B. Oliboni, X. Pan, M. Pančur, W. Pang, G. Papa, M. Paprzycki, M. Paralič, B.-K. Park, P. Patel, T.B. Pedersen, Z. Peng, R.G. Pensa, J. Perš, D. Petcu, B. Petelin, M. Petkovšek, D. Pevec, M. Pičulin, R. Piltaver, E. Pirogova, V. Podpečan, M. Polo, V. Pomponiu, E. Popescu, D. Poshyvanyk, B. Potočnik, R.J. Povinelli, S.R.M. Prasanna, K. Pripužić, G. Puppis, H. Qian, Y. Qian, L. Qiao, C. Qin, J. Que, J.-J. Quisquater, C. Rafe, S. Rahimi, V. Rajkovič, D. Raković, J. Ramaekers, J. Ramon, R. Ravnik, Y. Reddy, W. Reimche, H. Rezankova, D. Rispoli, B. Ristevski, B. Robič, J.A. Rodriguez-Aguilar, P. Rohatgi, W. Rossak, I. Rožanc, J. Rupnik, S.B. Sadkhan, K. Saeed, M. Saeki, K.S.M. Sahari, C. Sakharwade, E. Sakkopoulos, P. Sala, M.H. Samadzadeh, J.S. Sandhu, P. Scaglioso, V. Schau, W. Schempp, J. Seberry, A. Senanayake, M. Senobari, T.C. Seong, S. Shamala, c. shi, Z. Shi, L. Shiguo, N. Shilov, Z.-E.H. Slimane, F. Smith, H. Sneed, P. Sokolowski, T. Song, A. Soppera, A. Sorniotti, M. Stajdohar, L. Stanescu, D. Strnad, X. Sun, L. Šajn, R. Šenkeřík, M.R. Šikonja, J. Šilc, I. Škrjanc, T. Štajner, B. Šter, V. Štruc, H. Takizawa, C. Talcott, N. Tomasev, D. Torkar, S. Torrente, M. Trampuš, C. Tranoris, K. Trojacanec, M. Tschierschke, F. De Turck, J. Twycross, N. Tziritas, W. Vanhoof, P. Vateekul, L.A. Vese, A. Visconti, B. Vlaovič, V. Vojisavljević, M. Vozalis, P. Vračar, V. Vranić, C.-H. Wang, H. Wang, H. Wang, H. Wang, S. Wang, X.-F. Wang, X. Wang, Y. Wang, A. Wasilewska, S. Wenzel, V. Wickramasinghe, J. Wong, S. Wrobel, K. Wrona, B. Wu, L. Xiang, Y. Xiang, D. Xiao, F. Xie, L. Xie, Z. Xing, H. Yang, X. Yang, N.Y. Yen, C. Yong-Sheng, J.J. You, G. Yu, X. Zabulis, A. Zainal, A. Zamuda, M. Zand, Z. Zhang, Z. Zhao, D. Zheng, J. Zheng, X. Zheng, Z.-H. Zhou, F. Zhuang, A. Zimmermann, M.J. Zuo, B. Zupan, M. Zuqiang, B. Žalik, J. Žižka,

Informatica

An International Journal of Computing and Informatics

Web edition of Informatica may be accessed at: http://www.informatica.si.

Subscription Information Informatica (ISSN 0350-5596) is published four times a year in Spring, Summer, Autumn, and Winter (4 issues per year) by the Slovene Society Informatika, Litostrojska cesta 54, 1000 Ljubljana, Slovenia.
The subscription rate for 2021 (Volume 45) is
– 60 EUR for institutions,
– 30 EUR for individuals, and
– 15 EUR for students
Claims for missing issues will be honored free of charge within six months after the publication date of the issue.

Typesetting: Borut, Peter and Jaša Žnidar; borut.znidar@gmail.com.
Printing: ABO grafika d.o.o., Ob železnici 16, 1000 Ljubljana.

Orders may be placed by email (drago.torkar@ijs.si), telephone (+386 1 477 3900) or fax (+386 1 251 93 85). The payment should be made to our bank account no.: 02083-0013014662 at NLB d.d., 1520 Ljubljana, Trg republike 2, Slovenija, IBAN no.: SI56020830013014662, SWIFT Code: LJBASI2X.

Informatica is published by Slovene Society Informatika (president Niko Schlamberger) in cooperation with the following societies (and contact persons):
Slovene Society for Pattern Recognition (Vitomir Štruc)
Slovenian Artificial Intelligence Society (Sašo Džeroski)
Cognitive Science Society (Olga Markič)
Slovenian Society of Mathematicians, Physicists and Astronomers (Dragan Mihailović)
Automatic Control Society of Slovenia (Giovanni Godena)
Slovenian Association of Technical and Natural Sciences / Engineering Academy of Slovenia (Mark Pleško)
ACM Slovenia (Nikolaj Zimic)

Informatica is financially supported by the Slovenian research agency from the Call for co-financing of scientific periodical publications.

Informatica is surveyed by: ACM Digital Library, Citeseer, COBISS, Compendex, Computer & Information Systems Abstracts, Computer Database, Computer Science Index, Current Mathematical Publications, DBLP Computer Science Bibliography, Directory of Open Access Journals, InfoTrac OneFile, Inspec, Linguistic and Language Behaviour Abstracts, Mathematical Reviews, MatSciNet, MatSci on SilverPlatter, Scopus, Zentralblatt Math

Volume 45 Number 1 March 2021 **ISSN 0350-5596**

Informatica

An International Journal of Computing and Informatics

Informatica **45** (2021) Number 1, pp. 1–175