Reduced Number of Parameters for Predicting Post-Stroke Activities of Daily Living Using Machine Learning Algorithms on Initiating Rehabilitation

Ali Mohammad Alqudah and Munder Al-Hashem
Department of Biomedical Systems and Informatics Engineering, Yarmouk University, Irbid, Jordan
E-mail: ali_qudah@hotmail.com, munderalhashem@gmail.com

Amin Alqudah
Department of Computer Engineering, Yarmouk University, Irbid, Jordan
E-mail: amin.alqudah@yu.edu.jo

The estimation of the Barthel Index scale (BI) is a significant method for measuring the performance of Activities Daily Living (ADL), where the prediction of ADL is crucial for providing rehabilitation care management and recovery for patients after stroke, therefore in this paper, nine various Machine Learning (ML) algorithms were implemented in a medical dataset contains 776 records from 313 patients 208 of them are men: 208 and 150 are women with multiple features collected from them for predicting and classifying the BI status as clinical decision support for determining the ADL of post-stroke patients. Meanwhile, we have applied feature selection using the chi-squared test to reduce the number of features in the dataset. The results showed that the Decision Tree (DT), XGBoost (XGB), and AdaBoost (ADB) classifiers performed the highest performance achieved with 100% correctness in terms of accuracy, sensitivity, specificity, error, and Area Under Curve (AUC) on both the full and reduced features datasets.

Povzetek: V prispevku je predlagana metodologija za zmanjšanje števila parametrov za napovedovanje primernih aktivnosti po možganski kapi.

1 Introduction

A stroke is a medical condition that occurs when the blood flow to the brain`s parts is reduced or interrupted and frustrating the brain tissues from receiving nutrients and oxygen which leads to dying the brain cells in few minutes [1]. Stroke is a dilemma that needs urgent care and attention from around. Many people in the whole world are suffering from stroke and almost two-thirds of those individuals survive but need rehabilitation and recovery. Patients with long-standing disabilities after stroke could face many load difficulties in their life living whether physical, society, family and mental, therefore, the main object of rehabilitation and recovery is by observing the patients functions after a stroke and monitoring the level of independence to achieve the greatest potential Activities of Daily Living (ADL) [2].

Predicting the ADL is pivotal and crucial for effective use and careful deal with patients after stroke especially in the first months. Furthermore, ADL provides an overview of how the person`s independence in its life and the sufficient support and the health care provided by both the government and patient's family [2]. For example, in 2016 only every 40 seconds in the USA only there is a new onset of stroke events appeared [3], meanwhile, in 2012 in Taiwan there was around 230 patients were admitting hospital every day due to the occurrence of acute stroke

[3]. Due to this, if the disabilities and impairments remained for a long time in patients whose affected by a stroke this may lead to a massive mental, physical, and also a heavy financial load for the patients and their families and this all can be override by just make a rehabilitation for stroke patients as fast as possible [3]. The main target of rehabilitation the stroke patient is to ensure that they returned to their life after sophisticated training of them to make them able to make activities of daily living (ADL) independently. Moreover, the first stage of rehabilitation is to accurately predict ADL because the accurate prediction of ADL is vital to make optimization for stroke management in the first two to three months following a stroke attack [3].

During the rehabilitation process of the patient with strokes, different assessments and parameters should be checked for patients of Post-acute Care-Cerebrovascular Diseases (PAC-CVD) program to provide important and detailed information about the patient's health from various aspects [3]. The collected parameters include Barthel index (BI), modified Rankin scale (MRS), functional oral intake scale (FOIS), mini nutrition assessment (MNA), instrumental activities of daily living scale (IADL), Berg balance test (BBT), gait speed, 6-min walk test (6MWT), Fugl–Meyer upper extremity

assessment (FuglUE), modified Fugl–Meyer sensory assessment (FuglSEN), mini-mental state examination (MMSE), motor activity log (MAL), and concise Chinese aphasia test (CCAT). Using these multiple assessment parameters (features) we can generate and predict more comprehensive data and a conclusion about the current statuses of patients than single assessment features. However, providing an accurate tool for the prediction of ADL is not available due to the contribution of other features on the prognosis of ADL. To solve this problem, machine learning (ML) systems can be used 1.

ML is considered as an application of Artificial Intelligence (AI) that provides the ability to automatically learn and improve from the experienced learned. ML directly focuses on building a mathematical model to be used in the prediction process based on patterns extracted from the experienced learned data whether to be a large or complex dataset. Latterly, ML algorithms are extensively used in real-life applications and the healthcare field. The Healthcare field is directly related to human life that`s because using the ML will intensify, enhance, improve, and reduce the error level of rates in the medical diagnosis [1, 2, 3]. This research presented several ML algorithms for predicting the ADL after the stroke of patients to give information about the person`s independence based on multiple features by predicting BI score. Where the BI score is a widely used and very well-known parameter and tool to assess the independence of functions of ADL especially with post-stroke patients. To predict apply machine learning we use a dataset of 18 parameters 17 of them as features and the last one represents the BI score where the dataset contains 776 records from 313 patients 208 of them are men: 208 and 150 are women. We used nine different ML algorithms which are as follows; Logistic Regression (LR), K Nearest Neighbor (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), XGBoost (XGB), AdaBoost (ADB), and Artificial Neural Network (DNN). The algorithms are estimated by using training and testing methodology and the performance evaluation is calculated utilizing the accuracy, sensitivity, specificity, error, Receiver Operating Characteristic (AUC), and confusion matrix.

This paper is organized as follows: Section II describes the related work; Section III explains the materials and method of the post-stroke dataset used in our study and various ML algorithms applied to the multiple features. In section IV we present the results of the suggested methodology; Section V represents a discussion of the results of the proposed research. Finally, we conclude the paper in Section VI.

2   Literature review

In this section, the most recent works related to the topic of stroke rehabilitation prediction using machine learning are discussed. Janne et al [2] presented a study on the early post-stroke stage for final ADL to recognize the variables and the outcome of Activities of Daily Living (ADLs) after stroke. The method was by distinguishing the high and low quality for the risk of bias and qualitative synthesis. The results showed that the median risk of bias was 17 out of 27.

Yin Lin et al [3] proposed a method for predicting the activities of daily living (ADL) of post-stroke patients. Many Machine Learning (ML) algorithms are used such as Logistic Regression (LR), Support Vector Machine (SVM), Linear Regression, and Random Forest (RF) to estimate the Barthel index (BI). The results showed the RF and LR were higher for the Area Under the Curve (AUC) with 0.79 rather than the SVM algorithm with AUC of 0.77. Moreover, the Mean Absolute Error (MAE) of both linear regression and SVM was 9.95 and 9.86, respectively.

Chu Pai et al [4] proposed a study of testing and predicting the time change of Activities Daily Living (ADLs) of post-stroke patients within 24 hours and the whole months of the year. The ADLs were estimated by the Barthel Index (BI). A latent growth curve model was utilized to analyze the dynamic variations in ADLs. The results showed that the time following a stroke increases, survivors attend to continuously upgrade with interest to ADLs. Moreover, the below levels of the primary ADLs were related to the larger growth of ADLs over time.

Douiri et al [5] developed a prediction monitoring model for the post-stroke patient's recovery and estimated its clinical use. Activities of Daily Living (ADLs) were evaluated utilizing the Barthel Index (BI) for many weeks after the stroke. Penalized linear blended models were generated to predict the functional recovery of patients. The results of the prediction recovery curve showed good accuracy for Root Mean Squared Deviation (RMSD) of 3 BI points to one year. The passive predictive values of risk poor recovery with BI less than 8 at three and 12 months.

Bertolin et al [6] presented a study of predicting acute stroke patients after mild to moderate. Many predictor variables were measured for predicting the ADLs. The results showed physical variables described more variety in ALDs than the communication, memory, and thinking results. The Short-Blessed Test (SBT) was the unique meaningful independent foreteller of communication, memory, and thinking, while the National Institute of Health Stroke Scale (NIHSS) was the only one that measures safely predicted ALDs.

Glalanella et al [7] presented a study on investigating the Functional Independence Measure (FIM) of ADLs as potential predictors of the outcomes after stroke through inpatient rehabilitation. Regression analysis was used through Two backward stepwise to estimate the most independent variables. The results showed that social interaction, grooming, upper body dressing, and bowel control are the most independent and important variables for reviewing the FIM of ADLs.

Lee et al [8] proposed a study for developing a computational method to distinguish the potential of predicting the Quality of Life (QOL) after post-stroke rehabilitation. Five classifiers were used by using personal factors and nine functional issues to estimate the QOL. The results showed that the Particle Swarm-Optimized Support Vector Machine (PSO-SVM) gives the most extraordinary accuracy (58.33%), the highest cross-

validated accuracy (74.29%) which results as the best classifier in predicting the QOL of stroke patients.

3   Materials and methods

In this section, we will explain the dataset used and our recommended methodology, then making a comparison with the various classifier's performances.

3.1   Materials

In this paper, the data proposed in paper [3] has been used. The dataset contains 776 patient's data and 15 features. The features contain the following gender, age, acute ward stay, LOS, BI admission, MRS, FOIS, MNA, IADL, BBT, 6MWT, FugIUE, MMSE, MAL Quality, BI discharge. These values are collected from patients and based on them the output (Class) is determined. Table 1 shows the features and class values for the used dataset. collected from the patient based on the following criteria:

1   Stroke onset time must be within the last month.
2   Hemodynamic parameters should be stable within the last 72 hours.
3   No neurological deterioration within the last 72 hours.
4   Sufficient cognition function and ability to learn rehabilitation exercise (Especially MRS between 2 and 4).

While they mentioned that the exclusion criteria for the patient include:
The dataset authors mentioned that the features are
1.   Stroke onset time> 1 month.
2.   Patients with end-stage renal disease or those receiving dialysis therapy.

After applying these requirements above, every three weeks before discharge all qualifying patients receive a detailed post-stroke admission status test. The ratings contain the following characteristics (MRS, BI, FOIS, MNA, QoL, IADL, BBT, gait speed, 6MWT, FuglUE, FuglSEN, MMSE, MAL, and CCAT). During the recovery ward, both tests were used to illustrate the gravity of the BI status or the BI score forecasts for the rehabilitation unit discharge. Besides, an explanatory declaration is made of patients' age and longevity in the acute stroke ward before referral to PAC-CVD features [3]. The summary of the dataset is shown in Table 2.

3.2   Methodology

The suggested study as shown in Figure 1 below is by firstly solving the problem of multiclass of BI score, then applying nine classification algorithms and make a comparison between them to predict the BI score which represents the ADL of stroke patients.

3.3   Data preparation

Based on the BI at discharge from the PAC-CVD ward, it shows that the BI score is categorized with different values. The BI score with a value of 20 represents as independent ADL, the BI score with a value range of 15 - 19 represent mild dependent ADL, and the BI score with a value range of 10 - 15 represent as moderate dependent
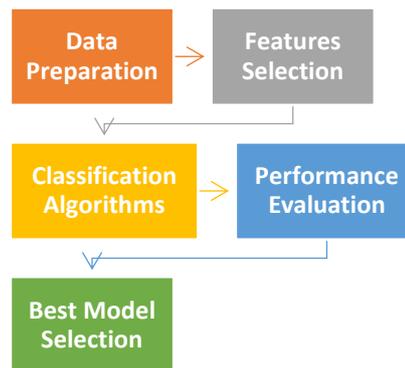

Figure 1: The block diagram of the proposed study.

ADL. Therefore, the multiclass classification problem of BI score can be solved by discretizing the BI into 3 ordinal classes.

3.3.1   Features selections

In this paper for results enhancement and time reduction in future patient BI score prediction, we will perform a Chi-squared test for features selections. The Chi-squared test is well known statistical test of independence that is usually used to test and determine whether two categorical variables are independent or not. Features selection concept using this test is very simple where we find if the feature is independent of the target or not. If the feature is dependent, then the feature will likely not be useful in predicting the target and not selected otherwise if they are not independent the feature will most likely have predictive power on the target and selected [9].

3.3.2   Classification algorithms

Classification in general is a plan that consists of orderly grouping something according to some features or standards. it is a type of supervised machine learning in which an algorithm "learns" to classify new observations from examples of labeled data and is used to develop models that predict what group or class that something belongs to [10-13]. Classification is a two-step process. During the first step, the model is created by applying a classification algorithm on a training data set then in the second step, the extracted model is tested against a predefined test data set to measure the model's trained accuracy. So, classification is the process to assign a class label from a dataset whose class label is unknown [13-16]. The following subsections will discuss the classifiers used in our paper.

Support Vector Machine (SVM)
Support vector machine (SVM) is a supervised learning algorithm under machine learning, and it is used for both classification and regression tasks. But most often the supporting vector machine is used in the classification. The Support Vector Machine works by looking at a set of training examples, and then marking each of them as belonging to one or another of two classes, the SVM training algorithm builds a model that maps new examples to one class or another, making it a non-probability binary linear classifier [10].

Table 1: Dataset Features, Range, and Description.

Feature	Range	Description
Gender	Male: 1, Female: 2	Gender
Age	35 -85 Years	Age
Acute Ward Stay	7-20 Days	length of stay in acute stroke ward
LOS	8-25days	length of stay in rehabilitation ward
BI Admission	20 mean independent ADL, 15-19 mean mild dependent, 0-14 mean moderate dependent, and 5–9 mean severe dependent.	Barthel index on admission to PAC-CVD ward
MRS	0- 2 mean mild disability, and 3 - 5 mean moderate to severe disability.	Modified Rankin Scale
FOIS	Range from 1 to 7 where Level 1 mean nothing by mouth to Level 7 which mean total oral diet with no restrictions.	functional oral intake scale
MNA	<8 mean malnutrition, 8–11 mean risk of malnutrition, and >11 mean no malnutrition.	mini nutrition assessment
IADL	0 mean low function (dependent), and 8 mean high function (independent).	instrumental activities of daily living scale
BBT	Range from 0 to 4 determining sitting and standing balance)	Berg balance test
Gait	Range from 0.3 m/s to 1 m/s.	gait speed
6MWT	Range from 210 m to 300 m.	6-min walk test
FuglUE	0 mean cDNNot perform, 1 mean partially, and 2 mean fully.	Fugl–Meyer upper extremity assessment
FuglSEN	0 mean cDNNot perform, 1 mean partially, and 2 mean fully.	Motor recovery assess the sensorimotor impairment in individuals who have had stroke
MMSE	MMSE >=24 mean normal, MMSE =19-23 mean mild, MMSE = 10-18 mean moderate, and MMSE <=9 mean severe.	mini-mental state examination
MAL Quality	0 mean never, 1 mean poor, MAL =2 mean fair, and 3 mean normal.	Motor activity log, quality of use arm function
BI Discharge	20 mean independent ADL, 5-19 mean mild dependent, 10-14 mean moderate dependent, and BI =5–9 mean severe dependent.	Barthel index at discharge from PAC-CVD ward
Class		Patient Classification Value

Naïve Bayes (NB)

Naive Bayes Classifiers are based on Bayesian classification techniques. This is based on Bayles's theorem, which is an equation defining the relationship between the conditional probability of statistical quantities. In the case of the Bayesian classification, we are interested in finding the likelihood of a category considering some of the characteristics that have been observed. In general, Naive Bayes is a simple, but powerful and widely used machine learning classifier. It is a probabilistic classifier that allows classifications in a Bayesian setting using the Maximum A Posteriori decision law. It can also be interpreted by a very simple Bayesian network. Naive Bayes Classifiers are highly common for text classification and are a standard approach to problems such as spam detection. The naïve Bayes classifier does not need any parameters setup [11].

Decision Tree (DT)

The Decision Tree algorithm is part of the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can also be used to solve regression and classification problems [12]. The purpose of the Decision Tree is to create a training model that can be used to predict the class or value

Table 2: Dataset Features Statistical Description.

	Statistical Value	Standard Deviation	Mean	Minimum	Maximum	25% Percentile	50% Percentile	75% Percentile
Feature	Age	4.104472	13.84021	7	20	10	14	18
	Acute Ward Stay	5.297994	16.55928	8	25	12	17	21
	LOS	3.502869	14.59278	9	20	11	15	18
	BI Admission	1.703339	2.597938	0	5	1	3	4
	MRS	2.016674	4.063144	1	7	2	4	6
	FOIS	4.036867	8.271907	2	15	5	8	12
	MNA	2.616618	3.840206	0	8	1	4	6
	IADL	1.409733	1.967784	0	4	1	2	3
	BBT	26.66156	254.357	210	300	231	253	277
	6MWT	0.809418	1.018041	0	2	0	1	2
	FugIUE	7.027365	18.75258	7	30	13	19	25
	MMSE	1.104851	1.449742	0	3	0	1	2
	MAL Quality	3.092184	16.43041	11	20	14	17	20
	BI Discharge	14.79479	60.33376	35	85	48	61	73

of the target variable by learning basic decision rules derived from previous results (training data). In Decision Trees, to predict the class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. Based on a comparison, we follow the branch corresponding to the value and leap to the next node. Decision trees classify the examples by sorting them down from the root tree to some leaf/terminal node, with the leaf/terminal node giving the example classification. Each node in the tree serves as a test case for some attribute, and each edge coming down from the node corresponds to the possible answers to the test case. This process is recursive and is repeated for every subtree rooted in a new node [12].

Random Forests (RF)
The Random Forest (RF) Classifier, first proposed by Breiman [13], is one of the most common classification tools and an excellent set of machine learning techniques. The key principle of the RF classifier is to construct a classification tree based on a few randomly selected features of randomly selected samples with a bagging technique. Designed trees are used to vote for an input vector to get a class name. RF classifiers are built by several simple learners, where each basic learner is an individual binary tree following recursive partitioning. RF has many advantages; it has better precision than other classifiers, allows large-scale data efficiency, does not bypass, and can be conveniently extended to multi-class inputs. The RF classifier has demonstrated superior classification efficiency over the other suggested methods since it was proposed [10, 11].

eXtreme Gradient Boosting (XGBoost)
XGBoost is a Machine Learning algorithm based on the decision-tree, using a gradient enhancement method. Artificial neural networks tend to outdo all other algorithms or systems in prediction problems involving unstructured data (images, messages, etc.). In small to medium structured/tabular data, however, decision-tab based algorithms are now regarded as best-in-class. For advances in tree-based algorithms over the years please see the map below. At its inception, this algorithm has not only been regarded as the guiding force behind the hood but many leading-edge applications in the industry [14].

Adaptive Boosting (AdaBoost)
AdaBoost is an ensemble learning system that was originally designed to improve the performance of binary classifiers (also known as "Meta-Learning"). AdaBoost uses an iterative approach to learn from and improve the

errors of weak classifiers. AdaBoost is a common stimulation strategy, attempting to combine several weak categorizers to create a powerful category. A single classifier cannot reliably predict an object's class, but we can build up such a strong model by grouping several weak classifications with each of them eventually learning from the incorrectly categorized objects of the others. The above classifier may be any of your basic classifiers, from Decision Trees (often the default) to Logistic Regression, etc [15].

Deep Neural Network (DNN)

A deep neural network (DNN) is an artificial neural network (DNN) with several layers between the input and output layers. There are various types of neural networks, but they are all made up of the same components: neurons, synapses, weights, beliefs, and functions. These elements are identical to the human brain and can be learned like any other ML algorithm [16, 17]. DNNs can model complex non-linear relationships. DNN architectures create compositional models in which the object is expressed as a layered composition of primitives. Extra layers permit the composition of features from lower layers, theoretically modeling complex data with fewer units than an equally powerful shallow network. The deep architecture contains a variety of versions of a few simple approaches. Different architectures have achieved success in particular fields. It is not always possible to assess the performance of different architectures unless they are tested on the same data sets [18]. DNNs are usually feed-forward networks in which data flows from the input layer to the output layer without looping back. Next, the DNN maps simulated neurons and assigns arbitrary numerical values, or "weights," to the relations between them. The weights and inputs are multiplied, and the product is returned between 0 and 1 [19]. If a certain pattern was not correctly understood by the network, an algorithm would change the weights. This way, the algorithm will make those parameters more influential before it decides the right mathematical manipulation of the data to be completely processed [20].

Performance evaluation

Evaluating the performance of machine learning algorithms is an essential part of any project. The model leads to giving you pleasing results when estimated using metric indices such as accuracy, specificity, sensitivity, error, and AUC [10-20]. Calculating these metrics by using True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) can the outputs of the proposed system compared to reference data. and consequently, the accuracy, sensitivity, precision, specificity, error, and AUC were evaluated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Error = 100 - Accuracy$$

Best model selection

Table 3: Training and testing dataset division.

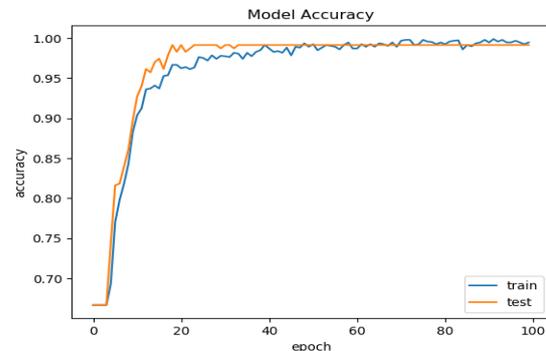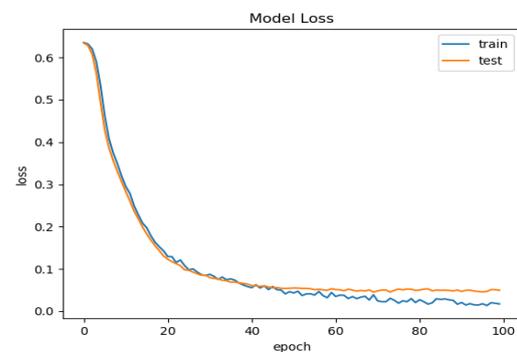| Post Stroke Dataset	Training set	Testing set
	620	156

Figure 2: Training and Validation Accuracy.

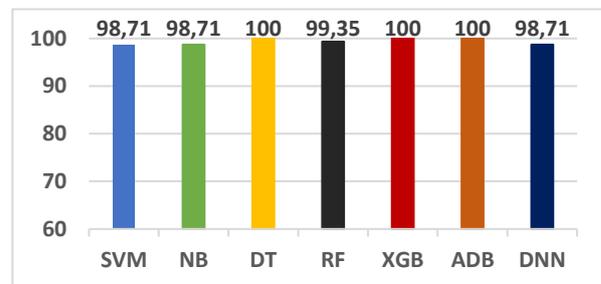Figure 3: Training and Validation Loss.

Figure 4: Accuracy performance evaluation.

For each ML algorithm that trains to the training dataset, the advantage of the learning algorithm here comes through the observed patterns of the training data that can do such as mapping to the input data attaches to the target, which is the required answer to be predicted, then creating ML model which makes capturing for these patterns. Moreover, the testing dataset is next implemented to the ML models to check and test their performances, but after comparing the results from the nine different algorithms that used, the most proper and appropriate model with the best accuracy, sensitivity, specificity, error, and AUC result is chosen to estimate and predict the BI as was discretized into 3 ordinal classes [10].

4   Results

The proposed methodology was employed with 80% - 20% training and testing division on the post-stroke dataset as shown in Table 3. Nine different ML classifiers are used for classifying the digitizing BI score that represents the ADL. The multi-classification process was successful in predicting the BI score with a high level of performance.

4.1   Features selection results

Before starting training and testing the used classifiers with and without features selection method first we applied the features selection methodology and select the top five features and the worst five features. The top five features selected by Chi-squared test methodology were ordered as follows: BI Admission, BI Discharge, 6MWT, Age, and FOIS. While the worst five were ordered as follows: Acute Ward Stay, FugIUE, MRS, Gender, and MAL Quality. The top five features are only used when we used the features selected classifiers.

4.2   Classifiers results without features selection

The nine classifiers were implemented for predicting as follows; the SVM classifier was performed with a linear kernel function. Moreover, the RF classifier was implemented too with 5 trees in the forest. Finally, the DNN classifier was trained by utilizing the Adaptive Moment Learning Rate (ADAM) solver with an initial learning rate value of 0.001. Figure 2 shows the training and validation process while Figure 3 shows the training and validation loss.

The hyperparameters of SVM, RF, DNN classifiers were chosen based on the principle of GridSearchCV which is a library function that helps to loop through predefined hyperparameters and fit the model on the training set. So, in the end, it can select the best parameters from the listed hyperparameters. The other classifiers that were also used were NB, DT, XGB, and ADB. The results of all algorithms employed show an extremely high level of results based on all the dataset features used. Figure 4 shows the results of accuracy, while Figure 5 shows the results for the error, where Figure 6 shows the results for the AUC, and finally, Figure 7 shows the results for the sensitivity and specificity of the three discretized classes that all obtained from the nine different ML algorithms.

4.3   Classifiers results with features selection

The same classifiers settings are performed on the selected features using the Chi-squared test where only five features are fed to the classifiers on the following results. Figure 8 shows the training and validation process while Figure 9 shows the training and validation loss.
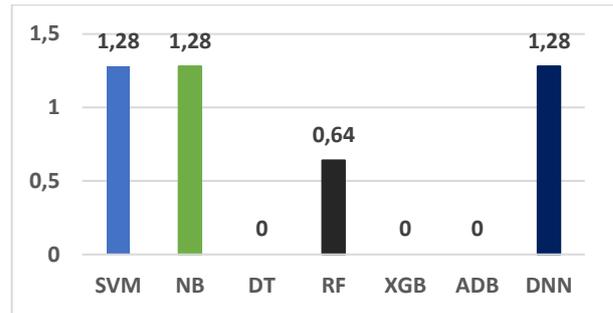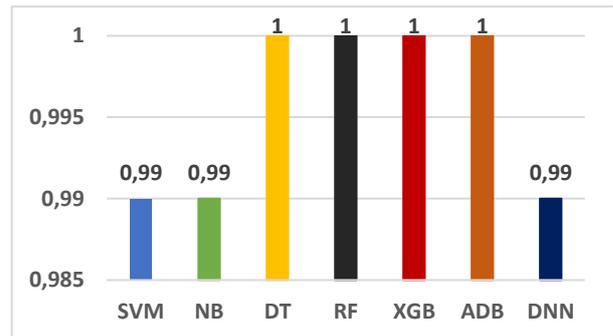

Figure 5: Error performance evaluation.

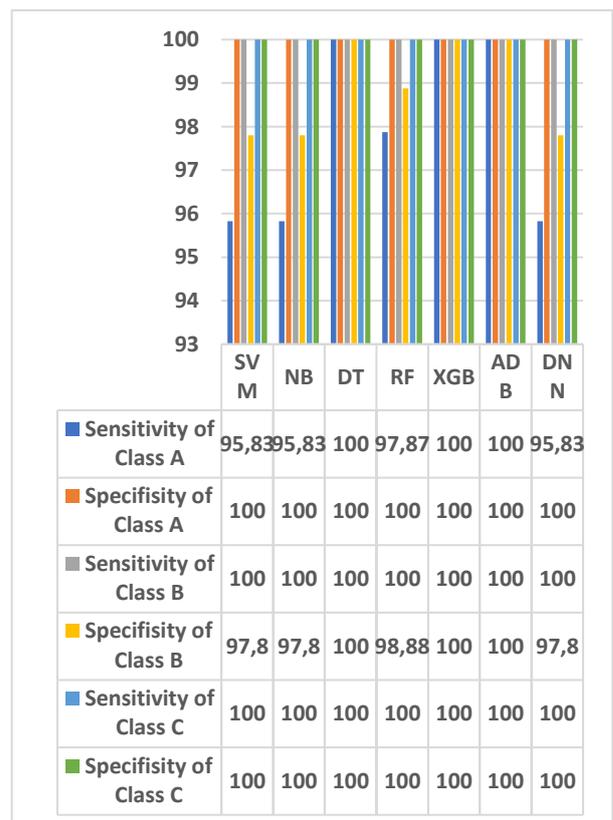Figure 6: AUC performance evaluation.

Figure 7: Sensitivity and specificity performance evaluation of the three classes.
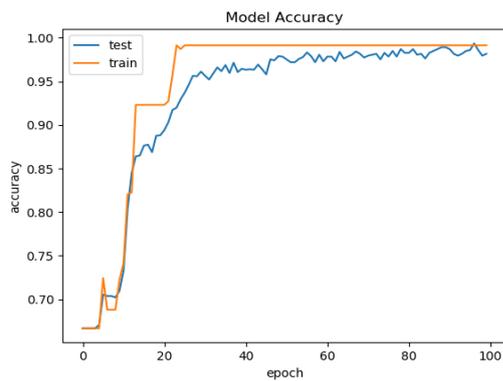
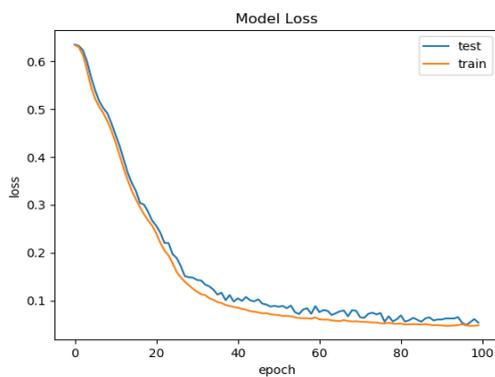Figure 8: Training and Validation Accuracy.

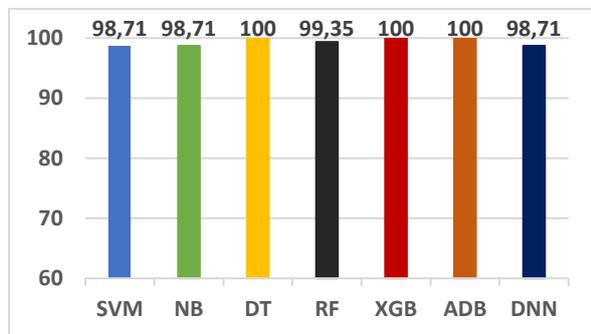Figure 9: Training and Validation Loss.

Figure 10: Accuracy performance evaluation.

The hyperparameters of SVM, RF, DNN classifiers were also chosen based on the principle of GridSearchCV which is a library function that helps to loop through predefined hyperparameters and fit the model on the training set. So, in the end, it can select the best parameters from the listed hyperparameters. The other classifiers that were also used were NB, DT, XGB, and ADB. The results of all algorithms employed show an extremely high level of results based on all the dataset features used. Figure 10 shows the results of accuracy, while Figure 11 shows the results for the error, where Figure 12 shows the results for the AUC, and finally, Figure 13 shows the results for the sensitivity and specificity of the three discretized classes that all obtained from the nine different ML algorithms.
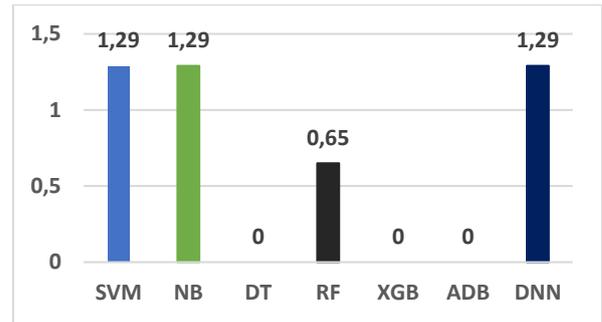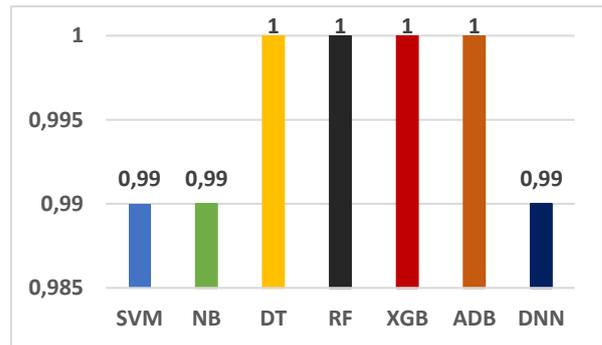

Figure 11: Error performance evaluation.

Figure 12: AUC performance evaluation.

	SVM	NB	DT	RF	XGB	ADB	DNN
■ Sensitivity of Class A	95,83	95,83	100	97,87	100	100	95,83
■ Specifisity of Class A	100	100	100	100	100	100	100
■ Sensitivity of Class B	100	100	100	100	100	100	100
■ Specifisity of Class B	97,8	97,8	100	98,88	100	100	97,8
■ Sensitivity of Class C	100	100	100	100	100	100	100
■ Specifisity of Class C	100	100	100	100	100	100	100

Figure 13: Sensitivity and specificity performance evaluation of the three classes.

Table 4: Comparing Proposed Methods with literature.

Reference	Number of Patients	Number of Features	Methods	Results (%)
[2]	48 of 8425 identified citations were included		Median Risk Bias	7 out of 27 (range, 6–22) points
[3]	313 individuals	15	LR SVM RF	The AUC is 0.79 for LR and RF, while SVM is 0.77
[4]	1,021 stroke survivors.	13-items	A latent growth curve model	The time following a stroke increases, survivors attend to continuously upgrade with interest to ADLs.
[5]	495, 1049 Patients	11	Penalized Linear Blended Models	RMSD for 3 BI points
[6]	498	7	6 Methods	
[7]	241	19	Two backward stepwise regression	
[8]	130	14	Back Propagation Artificial Neural Network (BP-ANN)	Accuracy (38.33%) and cross-validated (48.51%)
			Learning Vector Quantization (LVQ)	Accuracy (50.00%) and cross validated (58.96%)
			Self-Organizing Mapping (SOM)	Accuracy (53.33%) and cross validated (66.57%)
			Support Vector Machine (SVM)	Accuracy (53.33%) and cross validated (71.47%)
			Particle Swarm-Optimized SVM (PSO-SVM).	Accuracy (58.33%) and the highest cross-validated 74.29%
Proposed Method No Features Selection	313 with 776 Records	15	SVM NB DT RF XGB ADB DNN	Accuracy 98.71 Accuracy 98.71 Accuracy 100 Accuracy 99.35 Accuracy 100 Accuracy 100 Accuracy 98.71
Proposed Method with Features Selection	313 with 776 Records	15	SVM NB DT RF XGB ADB	Accuracy 98.71 Accuracy 98.71 Accuracy 100 Accuracy 100 Accuracy 100 Accuracy 100

5 Discussions

The present study aimed to investigate the influence of using many several ML algorithms for the classification of the BI score that represents the life living of patients after a stroke based on many features. Throughout the training step, all of the classifiers used were achieved an extraordinary scale of performances. While the testing results show that the used classifiers still show high accuracy. The results of this research based on the figures exhibited in the results showed that DT, XGB, and ADB classifiers achieved the most eminent performance reached 100% correctness in terms of accuracy, sensitivity, specificity, error, and AUC, for multi classify the digitized BI score, while the SVM and DNN classifiers are the worst. Moreover, using the features selection technique decreases the number of collected features in any future data collection from 15 to 5 only with a reduction ratio of 60% with the same results in classifiers except the RF enhanced from 98.71% to 100% in terms of accuracy. The results of feature selection mean that we can reduce the number of collected data and reducing the round time which makes patients more comfortable during data collection. Comparing the results of the proposed methods with other methods in the literature are shown in Table 4 The listed research studies in Table 4 have used the different datasets either collected by authors themself or by others. It is noted that they used a different number of classes, patients, and features set. These factors can affect the performance of the used classification methods significantly. However, most of the listed methods in the literature have achieved accepted recognition methods have high classification rates compared to other methods, the system is tested for time consumption in intel core i5-6700 /3.4 GHz and 12 GB of RAM desktop computer using Python 3.9 on Spyder IDE.

6 Conclusion

In this research, a study on applying nine different machine learning algorithms for the prediction and multi-classification of Barthel index which represents the activities of daily living of post-stroke patients in clinical practice. The research focused on finding the best classifier(s) for diagnosing the dependency of life living of post-stroke patients. Also, we have provided a features reduction methodology using the Chi-squared test to reduce the number of features in the datasets and during the round where they were collected from the patients. Experimental results show that the proposed method achieves very high accuracy when the BI score of three classes is classified even in the full or reduced features dataset. Therefore, the proposed method may be used effectively in hospitals with a lower number of features collected from patients for predicting the status of the life living of patients after a stroke. By comparing the proposed method with other methods in the literature, the present method is proven to be more effective and can provide a powerful tool for automatic stroke patient evaluation using the mentioned features.

Conflict of interest

References

[1] B. R. Wittenauer and L. Smith, "Priority Medicines for Europe and the World " A Public Health Approach to Innovation " Update on 2004 Background Paper Written by Eduardo Sabaté and Sunil Wimalaratna Background Paper 6. 6 Ischaemic and Haemorrhagic Stroke," Who, no. December, 2012.

[2] Veerbeek JM, Kwakkel G, van Wegen EE, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke: a systematic review. Stroke. 42(5):1482-8,2011. http://doi.org10.1161/STROKEAHA.110.604090

[3] Lin WY, Chen CH, Tseng YJ, Tsai YT, Chang CY, Wang HY, Chen CK. Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation. International journal of medical informatics. 111(1):159-64, 2018. http://doi.org/10.1016/j.ijmedinf.2018.01.002.

[4] Pai HC, Lai MY, Chen AC, Lin PS. Change in activities of daily living in the year following a stroke: a latent growth curve analysis. Nursing research. 67(4):286-93, 2018. http://doi.org/ 10.1097/NNR.0000000000000280.

[5] Douiri A, Grace J, Sarker SJ, Tilling K, McKevitt C, Wolfe CD, Rudd AG. Patient-specific prediction of functional recovery after stroke. International Journal of Stroke. 12(5):539-48, 2017. http://doi.org/10.1177/1747493017706241.

[6] Bertolin M, Van Patten R, Greif T, Fucetola R. Predicting cognitive functioning, activities of daily living, and participation 6 months after mild to moderate stroke. Archives of Clinical Neuropsychology. 33(5):562-76, 2018. http://doi.org/10.1093/arclin/acx096.

[7] Gialanella B. Predicting outcome a er stroke: the role of basic activities of daily living. Eur J Phys Rehabil Med. 49:629-37, 2013.

[8] Lee JD, Chang TC, Yang ST, Huang CH, Hsieh FH, Wu CY. Prediction of quality of life after stroke rehabilitation. Neuropsychiatry. 6(6):369-75, 2016. http://doi.org/10.4172/Neuropsychiatry. 1000163.

[9] Jin X, Xu A, Bie R, Guo P. Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. InInternational workshop on data mining for biomedical applications, 9: 106-115, 2006. Springer, Berlin, Heidelberg. http://doi.org/ 10.1007/11691730_11.

[10] Alqudah AM. Ovarian cancer classification using serum proteomic profiling and wavelet features a comparison of machine learning and features

selection algorithms. Journal of Clinical Engineering. 44(4):165-73, 2019. http://doi.org/1 0.1097/JCE.000000000000359.

[11] Ampomah EK, Nyame G, Qin Z, Addo PC, Gyamfi EO, Gyan M. Stock Market Prediction with Gaussian Naïve Bayes Machine Learning Algorithm. Informatica. 15;45(2), 2021. http://doi.org/ 10.31449/inf.v45i2.3407

[12] Tiwari P, Dao H, Nguyen GN. Performance evaluation of lazy, decision tree classifier and multilayer perceptron on traffic accident analysis. Informatica. 13;41(1), 2017.

[13] Alqudah AM. Towards classifying non-segmented heart sound records using instantaneous frequency based features. Journal of medical engineering & technology. 3;43(7):418-30, 2019. http://doi.org/1 0.1080/03091902.2019.1688408.

[14] Babajide Mustapha I, Saeed F. Bioactive molecule prediction using extreme gradient boosting. Molecules. 21(8):983, 2016. http://doi.org/1 0.3390/molecules21080983.

[15] Sun Y, Liu Z, Todorovic S, Li J. Adaptive boosting for SAR automatic target recognition. IEEE Transactions on Aerospace and Electronic Systems. 7;43(1):112-25, 2007. http://doi.org/10 .1109/TAES.2007.357120.

[16] Alqudah AM, Alquran H, Qasmieh IA. Classification of heart sound short records using bispectrum analysis approach images and deep learning. Network Modeling Analysis in Health Informatics and Bioinformatics. 9(1):1-6, 2020. http://doi.org/10.1007/s13721-020-00272-5.

[17] Alqudah AM, Alquraan H, Qasmieh IA. Segmented and non-segmented skin lesions classification using transfer learning and adaptive moment learning rate technique using pretrained convolutional neural network. InJournal of Biomimetics, Biomaterials and Biomedical Engineering, 42 :67-78, 2019. Trans Tech Publications Ltd. http://doi.org/10.4028/www.scientific.net/JBBBE.4 2.67.

[18] Malkawi A, Al-Assi R, Salameh T, Alquran H, Alqudah AM. White blood cells classification using convolutional neural network hybrid system. In2020 IEEE 5th middle east and Africa conference on biomedical engineering (MECBME) 27: 1-5, 2020. IEEE. http://doi.org/10.1109/MECBME47393.2020.92651 54.

[19] Alqudah A, Alqudah AM, Qazan S. Lightweight Deep Learning for Malaria Parasite Detection Using Cell-Image of Blood Smear Images. Rev. d'Intelligence Artif.. 34(5):571-6, 2020. http://doi.org/10.18280/ria.340506

[20] Alquran H, Alqudah AM, Abu-Qasmieh I, Al-Badarneh A, Almashaqbeh S. ECG classification using higher order spectral estimation and deep learning techniques. Neural Network World.1;29(4):207-19, 2019. http://doi.org/10.1 4311/NNW.2019.29.014.