APEM
journal

Advances in Production Engineering & Management

University of Maribor

Advances in Production Engineering & Management

Identification Statement

APEM Editorial

OPEN ACCESS

Statements and opinions expressed in the articles and communications are those of the individual contributors and not necessarily those of the editors or the publisher. No responsibility is accepted for the accuracy of information contained in the text, illustrations or advertisements. Chair of Production Engineering assumes no responsibility or liability for any damage or injury to persons or property arising from the use of any materials, instructions, methods or ideas contained herein.

Advances in Production Engineering & Management

Contents

Scope and topics

Advances in Production Engineering & Management (*APEM journal*) is an interdisciplinary refereed international academic journal published quarterly by the *Chair of Production Engineering* at the *University of Maribor*. The main goal of the *APEM journal* is to present original, high quality, theoretical and application-oriented research developments in all areas of production engineering and production management to a broad audience of academics and practitioners. In order to bridge the gap between theory and practice, applications based on advanced theory and case studies are particularly welcome. For theoretical papers, their originality and research contributions are the main factors in the evaluation process. General approaches, formalisms, algorithms or techniques should be illustrated with significant applications that demonstrate their applicability to real-world problems. Although the *APEM journal* main goal is to publish original research papers, review articles and professional papers are occasionally published.

Fields of interest include, but are not limited to:

Additive Manufacturing Processes
Advanced Production Technologies
Artificial Intelligence in Production
Assembly Systems
Automation
Big Data in Production
Block Chain in Manufacturing
Computer-Integrated Manufacturing
Cutting and Forming Processes
Decision Support Systems
Deep Learning in Manufacturing
Discrete Systems and Methodology
e-Manufacturing
Evolutionary Computation in Production
Fuzzy Systems
Human Factor Engineering, Ergonomics
Industrial Engineering
Industrial Processes
Industrial Robotics
Intelligent Manufacturing Systems
Joining Processes
Knowledge Management
Logistics in Production

Machine Learning in Production
Machine-to-Machine Economy
Machine Tools
Machining Systems
Manufacturing Systems
Materials Science, Multidisciplinary
Mechanical Engineering
Mechatronics
Metrology in Production
Modelling and Simulation
Numerical Techniques
Operations Research
Operations Planning, Scheduling and Control
Optimisation Techniques
Project Management
Quality Management
Risk and Uncertainty
Self-Organizing Systems
Smart Manufacturing
Statistical Methods
Supply Chain Management
Virtual Reality in Production

APEM
journal

Flexible Job-shop Scheduling Problem with parallel operations using Reinforcement Learning: An approach based on Heterogeneous Graph Attention Networks

Lv, Q.H.[a], **Chen, J.**[b], **Chen, P.**[b], **Xun, Q.F.**[b], **Gao, L.**[a,*]

[a]School of Mechanical Science and Engineering, HuaZhong University of Science and Technology, P.R. China
[b]Academy of Mathematics and Systems Science, Chinese Academy of Science, P.R. China

A B S T R A C T

The Flexible Job-shop Scheduling Problem (FJSP) has received considerable scholarly attention as a classic problem. However, in practical industrial manufacturing scenarios, it is common for an operation to have multiple preceding parallel operations. This not only necessitates adhering to the sequential relationships inherent in FJSP but also requires ensuring that preceding operations are completed simultaneously whenever feasible. We term this scenario as the Flexible Job-shop Scheduling Problem with Parallel Operations (FJSP-PO), a pervasive challenge encountered across nearly every production line in real-world discrete manufacturing applications. Despite its prevalence, there is a noticeable scarcity of research on FJSP-PO in existing literature. Given the objective of synchronizing multiple preceding operations, FJSP-PO presents a broader solution space and more intricate optimization challenges compared to traditional FJSP. To address this, we propose an Attention Restart method based on Heterogeneous Graph Attention Networks (AR-HGAT). Leveraging a heterogeneous graph network structure and reinforcement learning, AR-HGAT learns the implicit features of operations and machines through node-level and semantic-level attention mechanisms. The AR mechanism is utilized to determine the optimal scheduling of operations at specific time slots. Compared to existing FJSP methods, our AR-HGAT approach demonstrates superior performance in terms of inference time and solution effectiveness. Furthermore, we conducted a comparative analysis using authentic operational data from companies and contrasted it with results obtained from an online tree search algorithm, thereby providing empirical validation of the effectiveness of the proposed AR-HGAT method.

1. Introduction

Flexible scheduling can adopt production plans in response to changing demands and resources. It enhances production efficiency, optimizes resource utilization, and reduces costs and waste. In the context of Industry 4.0, increasing consumer demands for customization and timeliness align with the advancements in big data and artificial intelligence, providing essential support for flexible production. Addressing flexible production challenges in larger and more complex environments has become imperative for intelligent factories, necessitating the autonomous ability to perceive, analyze, reason, make decisions, and control [1]. The Flexible Job-shop Scheduling Problem (FJSP) plays a crucial role in ensuring low variation in diversified manufacturing and is widely applied in semiconductor manufacturing, automobile assembly, and mechanical manufacturing

systems [2]. Particularly suitable when task-resource relationships are flexible and variable, the FJSP remains a focus of research efforts. As an extension of the FJSP, the FJSP-PO is characterized by the inherent complexity of products, involving multiple components with a generally fixed product structure and component matching relationships. In real scenarios, certain machine resources may have the shortest waiting time for processing operations, aiming to minimize the cost of production line work, such as final assembly line. The presence of parallel operations in this manufacturing context necessitates the delay of some operations. The research on FJSP-PO not only effectively promotes the research on FJSP, but also expands the application scope of FJSP. But to the best of our knowledge, FJSP-PO has rarely been studied.

The FJSP, as a well-known NP-hard problem, with various methods employed for its solution. The priority dispatching rule (PDR), a classical method, iteratively dispatches jobs based on specified priority rules. Heuristics, known for their effectiveness in large-scale problem-solving, represent another popular research direction. While PDRs are commonly used, their effectiveness depends on the dispatching rules employed. Simple rules like first in first out (FIFO) are inadequate for today's production environment, prompting research into new and effective priority rules [3, 4]. Artificial intelligence (AI) emerges as a valuable tool, facilitating the identification of innovative and potentially more effective rules [5]. The classical solutions for the Flexible Job-shop Scheduling Problem (FJSP) can be categorized into three main types: exact algorithms, heuristics, and meta-heuristics [6].

In exact algorithms, most of which are formulated by integer linear programming (ILP) or mixed ILP (MILP) models. Birgin *et al*. proposed a MILP model for an extended version of the Flexible Job-shop Scheduling problem [7]. Some presented mixed integer linear programming and constraint programming models for the minimization of the makespan [8]. The branch-and-bound approach has also been employed to address FJSP, as seen in the work by Kim and Yano [9], who developed an efficient branch-and-bound algorithm for Flexible Manufacturing Systems (FMS), which allocates operations to machine groups to maximise throughput while satisfying tool or component storage constraints. In heuristics, [10] introduced a multi-region division sampling strategy-based multi-objective optimization algorithm that integrates with a genetic algorithm (GA) for addressing the Flexible Job-shop Scheduling Problem (FJSP). This approach aims to optimize multiple objectives concurrently. Furthermore, [11] proposed a Greedy Randomized Adaptive Search Procedure (GRASP) designed for the integrated scheduling of dynamic flexible Job-shops. Notably, this method incorporates a novel preventive maintenance policy to enhance the adaptability of the scheduling process. In response to unexpected scenarios, such as changes in task functions, a heuristic model for solving dynamic flexible workshop scheduling problems (DFJSP) was proposed, inspired by the artificial bee colony algorithm [12]. This approach is tailored to handle unforeseen circumstances that may arise during scheduling. Ortíz *et al*. [13] In response to unexpected scenarios, such as changes in task functions, a heuristic model for solving dynamic flexible workshop scheduling problems (DFJSP) was proposed, inspired by the artificial bee colony algorithm [12]. This approach is tailored to handle unforeseen circumstances that may arise during scheduling. Some researchers have directed their attention towards Priority Dispatching Rules (PDRs). Teymourifar *et al*. [14] presented a practical method aimed at extracting effective rules for a more general type of dynamic workshop scheduling problem. This problem involves jobs arriving at the workshop at different times, and machine failures occurring randomly. Notably, they also considered limited buffer conditions, thereby introducing additional complexity to the problem. In meta-heuristics, Simulated annealing [15], local search [16, 17] and genetic algorithms [18] are widely applied to solve the problem. Cinar *et al*. [19] proposes a genetic algorithm based on priority representation for Flexible Job-shop Scheduling Problems. Kaweegitbundit and Eguchi [20] proposed an effective scheduling method with the goal of minimizing average delay, which uses a genetic algorithm containing heuristic rules. Chen *et al*. [21] proposed a self-learning genetic algorithm (SLGA) based on reinforcement learning (RL) to intelligently adjust key parameters during the calculation process. Zhang *et al*. [22] proposed an improved genetic algorithm to solve multi-objective FJSP problems, considering processing time, machine setup time, and transportation time.

Overall, when studying FJSP, the inherent complexity often makes precise algorithms challenging, particularly when dealing with large-scale problems. Consequently, the research landscape in this domain tends to emphasize heuristic and meta-heuristic algorithms. Additionally, learning-based methods are concurrently employed to effectively manage uncertainties inherent in the scheduling process. Deep Reinforcement Learning (DRL), a combination of deep learning and reinforcement learning, aims to enable machines to learn optimal strategies through continuous trial and error. DRL's ability to process high-dimensional input data and automatically extract features contributes to accurate prediction and decision-making. Its applications span various fields, including autonomous driving, game intelligence, and robot control. Recent works explore the use of DRL methods to automatically generate PDRs for scheduling problems [23-27]. Treating PDR decision-making as Markov Decision Processes (MDPs), these approaches train scheduling strategies to consider the future, overcoming the myopia associated with PDRs. Deep neural networks capture global scheduling states, moving beyond local information. Despite its advantages, applying DRL to FJSP problems poses challenges due to the complex decision-making situations and intricate relationships between machines and operations. A study proposes an architecture based on heterogeneous graph neural networks to address these complexities and capture intricate relationships [28]. The distinctive advantage of the DRL algorithm investigated in this article, compared to previous studies, lies in its applicability to FJSP with parallel operations. This capability allows for a more comprehensive and effective approach to scheduling in scenarios where parallel operations are prevalent.

To summarize, this article makes the following contributions:

- This paper introduces a novel problem, FJSP-PO, where at a specific time, not only can one operation be assigned to different machines, but also multiple operations without contextual dependence can be assigned to different machines simultaneously, introducing parallel operations. This extension enhances the FJSP scenario, making it more suitable for the complexities of contemporary flexible manufacturing.
- Addressing the challenge of scheduling parallel operations, the goal is to minimize the waiting time between operations while ensuring the minimum makespan. To achieve this, a new strategy is proposed to delay the start of some operations, creating a more compact production schedule.
- Recognizing the dynamic nature of operation scheduling in real scenarios, where scale varies constantly, the model's generalization performance is crucial. Leveraging the topology of heterogeneous graphs, this paper incorporates both node-level and semantic-level information into the learning process. Experimental results validate the model's effectiveness across different scales.
- Given the diverse constraints in practical operation scheduling scenarios, this paper separates business constraints from the scheduling model. It establishes a feasible framework in actual production, transforming constraints into mask tensors. This approach simplifies the implementation of the scheduling model and allows for the easy expansion of constraints within the business scenario.

The organizational structure of this paper is as follows: Section I introduces the predecessor work on operations scheduling, Section 2 presents the latest reinforcement learning-based methods for solving operation scheduling, Section 3 describes real business scenarios and challenges encountered in discrete scenes with parallel operations, Section 4 provides a detailed explanation of the overall framework and methodology for solving the problem, Section 5 verifies the effectiveness and generalization of our proposed model through multiple sets of experiments, and Section 6 summarizes the paper comprehensively.

2. Related work

As of our knowledge cutoff date, the predominant focus in operation scheduling pertains to flexible scheduling within discrete scenarios. However, there is a noticeable dearth of studies specifically addressing operation scheduling in discrete scenarios involving parallel operations. In this section, we will provide a concise review of recent scheduling methods based on DRL for FJSP. With the advancement of artificial intelligence, researchers are increasingly leveraging Deep Reinforcement Learning (DRL) to tackle intricate scheduling problems. In DRL-based scheduling, the representation of states is a crucial aspect. Some studies have utilized vectors or matrices for state representation, but a drawback is the limitation imposed by fixed matrix dimensions, constraining the algorithm's scalability for larger problems [23], [29], [30]. Kwon *et al.* [31] employed a matrix encoding network and self-attention [32] to handle projects, addressing scale limitations. However, this method still encounters challenges related to the number of problem stages.

To overcome the scale limitations inherent in matrix representation, scholars have turned their attention to Graph Neural Networks (GNN) [33]. GNNs can handle graphs of varying sizes, offering a solution to the limitations associated with matrix representation. Zhang *et al.* [34] proposed an approach to automatically learn Priority Dispatching Rules (PDR) through end-to-end deep reinforcement learning agents, utilizing JSSP's disjunction graph representation. Hameed and Schwung [35] used features extracted using Graph Neural Network (GNN) to train RL agents, providing a rich encoding of the current production state in the Euclidean space. Su *et al.* [36] introduced a framework based on GNNs and DRL to solve dynamic JSSP (DJSP) with machine failures and random processing time, utilizing evolutionary strategy (ES) to find an optimal strategy that is more stable and robust than traditional DRL algorithms.

While DRL exhibits advantages in solving production scheduling problems, much of the current research has focused on Job-shop Scheduling (JSP), with limited methods available for Flexible Job-shop Scheduling Problem (FJSP). Luo *et al.* [37] proposed an online scheduling framework called a two-level deep Q network (THDQN) for the dynamic multi-objective flexible Job-shop Scheduling Problem with new job inserts (DMOFJSP). However, it relies on manual dependency for the selection of PDRs. Han and Yang [38] introduced an end-to-end DRL method based on 3D discrete graphs to solve FJSP. Nevertheless, the attention-based strategy network they designed processes raw features without considering the graph structure, limiting its ability to extract useful information for high-quality decision-making.

Recently, a number of interesting studies on Flexible Job-shop Scheduling Problem (FJSP) have emerged. For instance, Peng and Zheng [39] have focused on optimizing the common dual-resource-constraint problem using the Wild Horse algorithm. Sun *et al.* [40] have applied reinforcement learning to handle job insertion issues, achieving real-time processing capabilities. In the context of more complex distributed permutation scenarios, Ren *et al.* [41] have creatively employed the Q-Learning algorithm to solve the problem. Additionally, Koblasa *et al.* [42] have delved deeper into FJSP solutions using classic Evolutionary Algorithms (EA). Moreover, Yildiz *et al.* [43] have proposed more suitable neural network structures to predict the makespan in FJSP. Of course, Karacan *et al.* [44] have adopted ensemble learning approaches, combining multiple methods to achieve notable results. However, recent papers have not specifically optimized the partial parallel processes encountered in current industrial applications. Therefore, this paper aims to build upon existing work and further innovate on FJSP-PO, which is more closely aligned with practical business scenarios.

3. Problem description

In this section, we will provide a brief overview of the operation scheduling scenario, which encompasses both the Flexible Job-shop Scheduling Problem (FJSP) and the proposed extension, FJSP-PO. Then we will delve into a detailed exploration of the specific problems that require resolution. Finally, we express the FJSP-PO in mathematical language.

3.1 FJSP and FJSP-PO

In discrete manufacturing, the scheduling of operations across various jobs is orchestrated to meet optimization objectives. In this scenario, a single operation can be executed on different machines, each with potentially distinct processing times. This framework is commonly applied in flexible manufacturing, with the overarching optimization goal typically focused on minimizing makespan.

In a standard discrete scenario, a solitary production line prevails, denoted as FJSP. However, practical situations often involve multiple production lines, presenting a more intricate challenge referred to as FJSP-PO. The cycle times for different operations may vary in both FJSP and FJSP-PO. In FJSP, a singular operation can be processed across multiple machines. In contrast, FJSP-PO introduces greater complexity, allowing not only a single operation to be processed across multiple machines but also enabling multiple operations without contextual dependence to be concurrently processed on different machines.
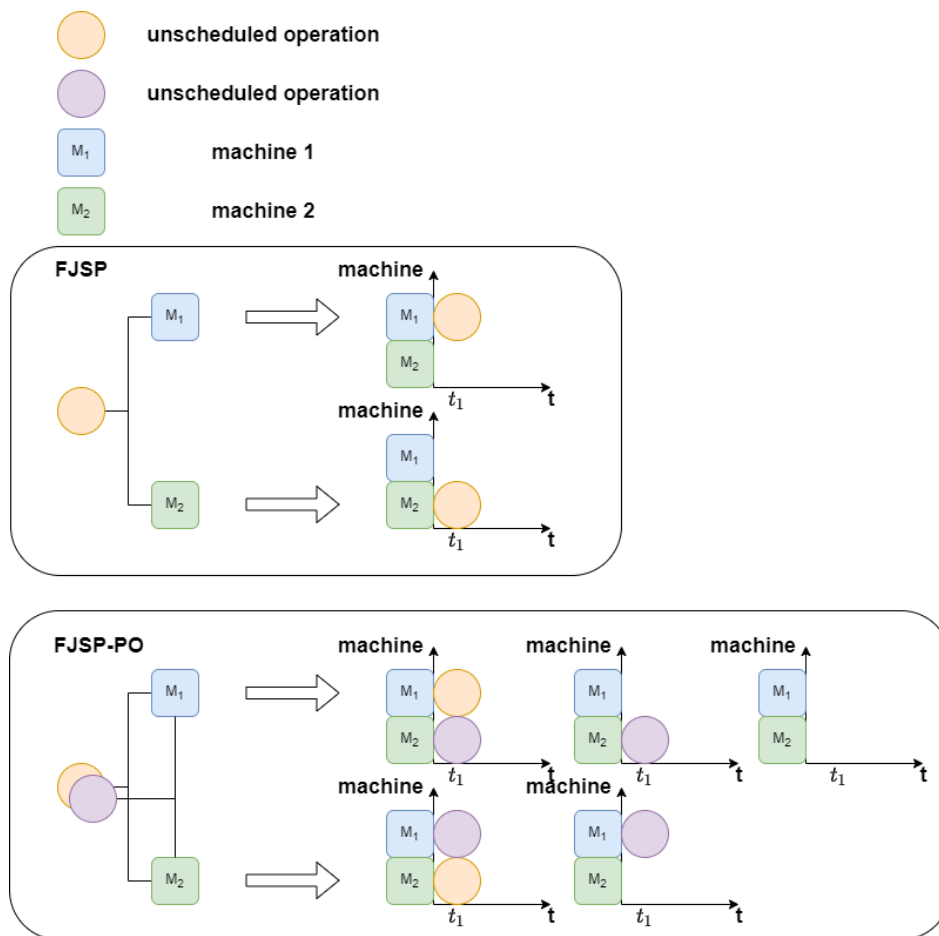
Fig. 1 FJSP and FJSP-PO

As depicted in Fig. 1, in FJSP, during time t, a specific operation can opt to be processed either on the 1st or 2nd machine. However, at time t, only one operation is permitted to commence processing. In contrast, in FJSP-PO, two operations can be processed simultaneously at time t. For instance, if one operation is being processed on the 1st machine, the other operation can simultaneously be processed on the 2nd machine, and vice versa. It's evident that FJSP-PO encompasses FJSP, with the former being more intricate than the latter. And the solution space of the FJSP-PO problem is much larger than that of the FJSP problem.
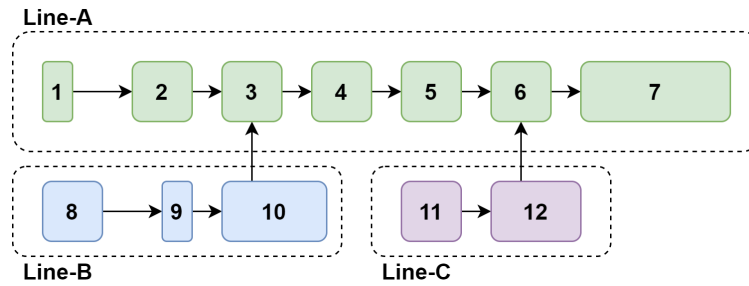
Fig. 2 Multiple lines in FJSP-PO

In an FJSP-PO, the following scenarios may exist.

In FJSP-PO, as shown in Fig. 2, this job has a total of 12 operations, comprises three lines, the Line-A comprises 7 operations, the Line-B comprises 3 operations, the Line-C comprises 2 operations, in order to facilitate elaboration, assume that there are 12 machines altogether, each machine can process all the operations. The arrow indicate the link relationship between operations. For example, the 2nd operation depends on the 1st operation, so the 2nd operation must wait for the 1st operation to complete before starting processing, and the 3rd operation must wait for both the 2nd operation and the 10th operation to be completed before starting processing. Similarly, the 6th operation needs to wait for the 5th and 12th operations to be completed before starting processing.

In practical production scenarios, to optimize costs, a production line is typically initiated as late as possible while still adhering to the product delivery deadline. Consequently, minimizing the waiting time between operations within a line becomes crucial.

Compared with the FJSP, the main features of the FJSP-PO include the following three points:

- There may be multiple lines, so some operations may have more than one precursor operation node,
- At some point, there may be multiple operations that can be processed at the same time,
- As shown in Fig. 1, the solution space for FJSP-PO is larger.

3.2 Problem details

In prior studies, reinforcement learning was employed to address operation scheduling in discrete scenarios featuring only a single line, without the inclusion of parallel operations. However, this approach encountered two distinct challenges:

- How to support the scheduling of parallel operations?
- How to make the waiting time between operations as short as possible?
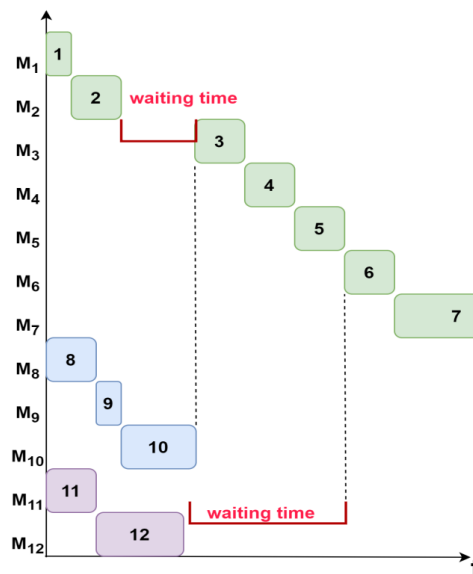
Fig. 3 Operation scheduling in first step

Our objective is to schedule the 12 operations, as illustrated in Fig. 3. In the second step, we aim to minimize the waiting time between these operations, as depicted in Fig. 4.

In Fig. 3, due to the presence of operations in Line-B, the 3rd operation must wait for the completion of all dependent operations. Consequently, the 3rd operation cannot immediately follow the execution of the 2nd operation. Similarly, the 6th operation in Line-A requires the completion of the 5th operation in Line-A and the 12th operation in Line-C before commencing. To minimize waiting times for operations in Line-A, the start time of Line-A operations must be postponed, as shown in Fig. 4.

To address these challenges, we propose a novel architecture and algorithm designed to tackle the complexities of the FJSP-PO problem.
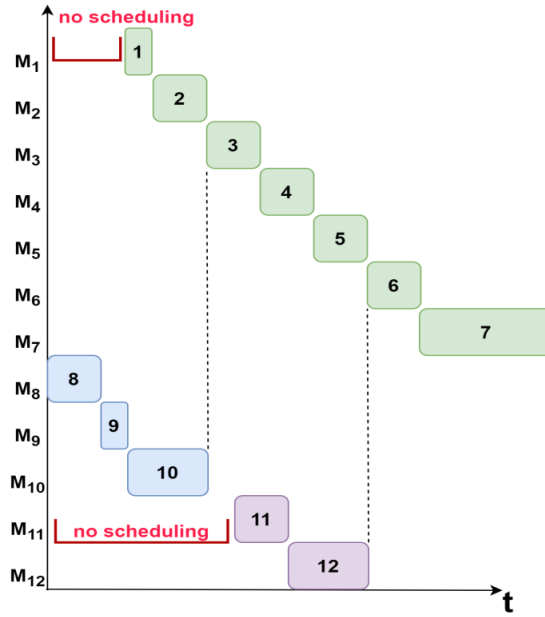
Fig. 4 Operation scheduling in second step

3.3 Operation scheduling scenarios

There are I jobs, K machines, L constraints and Z optimize goals. J_i the i-th job, $J_i \in \{J_1, J_2, \cdots, J_I\}$, O_{ij} the j-th operation of the i-th job, $i \in [1, I], j \in [1, B_i]$. The purpose of operation scheduling is to assign all operations O_{ij} in the job J_i to the appropriate machine M_k at a determined time, and it is processed for a certain amount of time λ_{ijk}, and to optimize all objectives while all constraints are met. The symbol definition is represented as follows:

$$minimize \sum_z \alpha_z F_z, z \in [1, Z], \alpha_z \in [0,1)$$
$$s.t. \sum_l C_l \tag{1}$$

M_k: the k-th machine, $M_k \in \{M_1, M_2, \cdots, M_K\}$
λ_{ijk}: the process time for the operation O_{ij} processed in machine M_k
C_l: the l-th constraint, $C_l \in \{C_1, C_2, \cdots, C_L\}$
F_z: the z-th goal function, $F_z \in \{F_1, F_2, \cdots, F_Z\}$

Reinforcement learning tasks are usually described using the Markov Decision Process (MDP) as shown Fig. 5 and typically contain five important elements: agent, environment, state, action, and reward. Through interactions with the environment, the agent acquires a strategy grounded in the reward feedback received from the environment. Consequently, the agent takes actions within the current environmental state with the aim of maximizing the expected reward value over time. Therefore, reinforcement learning revolves around attaining the optimal strategy as follows:

$$\pi^* = argmax_\pi \sum_t \zeta (s_t, a_t) \tag{2}$$

t: time t

π: the current strategy agent adopted

π^*: the optimal strategy agent learned

s_t: the state of the environment at the time t

a_t: the action taken by the agent at the time t, and $a_t = \pi(s_t; \theta)$, and $a_t \in A_t$, A_t is action space, θ is a parameter in the strategy function π

$\zeta(s_t, a_t)$: the reward function to produce reward value based on the current state and the action taken
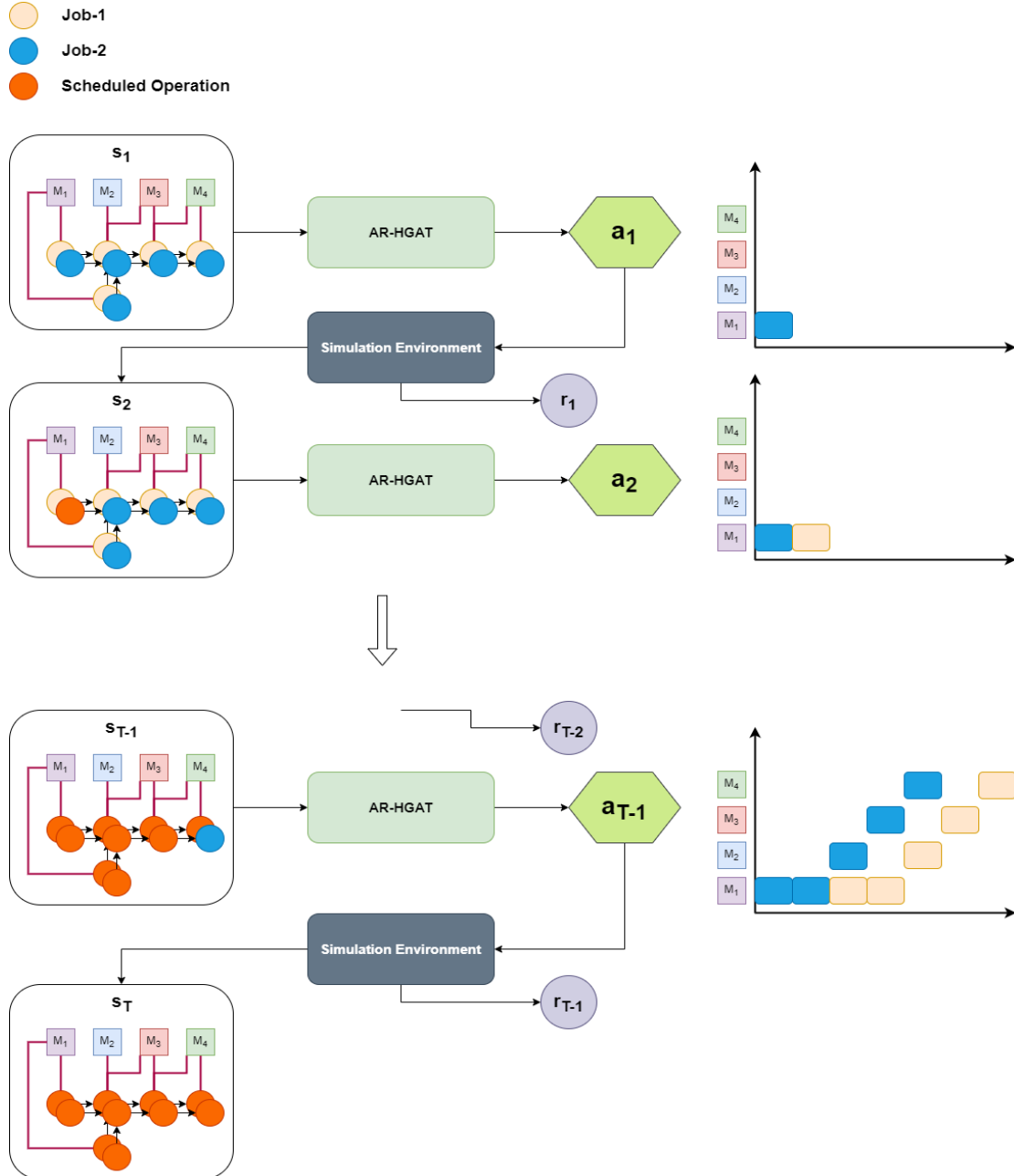
Fig. 5 Markov process

The action at time t consists of the selected operation and the assigned machine, represented by $a_t = \left(O_{ij}, M_k\right)_t$.

The simplest way is for the agent to take an action a_t based on a policy π at time t and receive reward feedback r_t from the environment. It can be simply considered that $\zeta(s_t, a_t)$ is equal to r_t. A more complex $\zeta(s_t, a_t)$ can be conceptualized as a function represented by a neural network.

To address the scheduling problem using reinforcement learning, it is necessary to define the agent, environment, state, action, and reward in the scheduling problem. Similar to the approach in paper [11], the agent is a strategy function π to learn, the environment is a scheduling simulation environment, and the state is the current status of the simulation environment.

However, the paper [11] does not cover the operation scheduling scenario with parallel operations. In scenes with parallel operations, some operations may need to wait, so the empty action ϕ (No action is taken) can be executed at time t. We introduce a job at the end, and there is only one special operation represented by O_{I1}, so we define $\phi = (O_{I1}, M_k)_t$.

Since the action ϕ is now included, we need to determine when to take this action and for how long. The strategy for determining when to take this action is called attention reboot (AR).

Simultaneously, we aim to determine the duration for which this action will persist. We have improved the graph neural network proposed by [11], referred to as AR-HGAT.

As observed, our objective is to learn the parameters of the proposed AR-HGAT and obtain the actions required in the current state at each time t, aiming to maximize the desired reward value.

4. Method

In this section, as shown in Fig. 6, we present a unified framework designed to address the Flexible Job-shop Scheduling Problem with Parallel Operations (FJSP-PO). This framework is versatile, accommodating multiple production lines, and facilitates the seamless integration of complex constraints commonly encountered in real-world business scenarios.
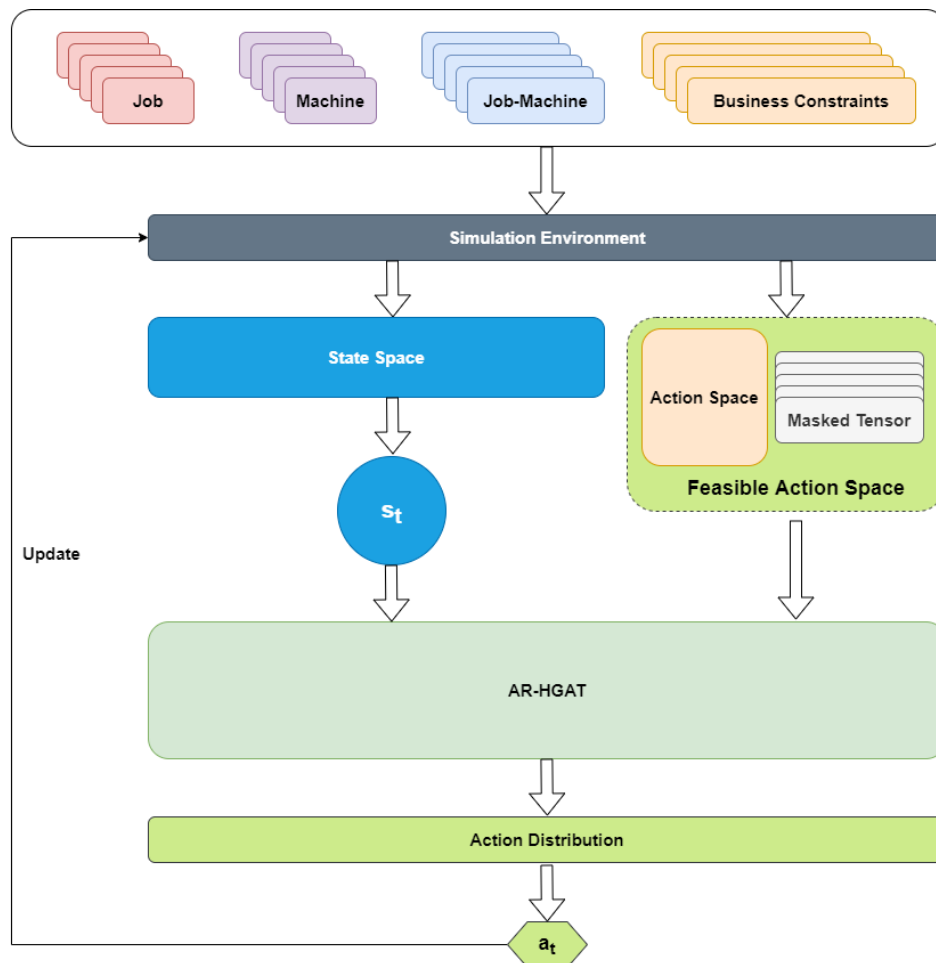
Fig. 6 Unified scheduling framework

To address the Flexible Job-shop Scheduling Problem with Parallel Operations (FJSP-PO), we initiate a simulation environment. This environment plays a pivotal role in initializing the state space, action space, reward space and constraint mask tensor. It facilitates the transfer of states,

allowing the retrieval of the state at the current time t. Subsequently, the AR-HGAT model proposed in this paper is employed to obtain the legal actions at the current time t. The environment is then updated, transitioning to the next state to initiate the subsequent time step.

A key aspect of our approach involves segregating complex constraints from scheduling algorithm models. Each complex constraint can be dynamically updated to the feasible action selection tensor through masking. This design is particularly significant in real-world business scenarios, as it allows non-algorithm professional engineers to articulate corresponding business constraints.

In the following sections, we will provide a detailed introduction to the state space, action space, reward space and constraint mask tensor, model structure, training process, and forecasting process employed in our unified framework for solving FJSP-PO.

4.1 State space

The state space comprises states recorded at different time steps, denoted as t. Each state encapsulates four main categories of information: operation features, machine features, operation link information, and operation-machine pair information, which is as same as the paper [25]. This is illustrated as follows:
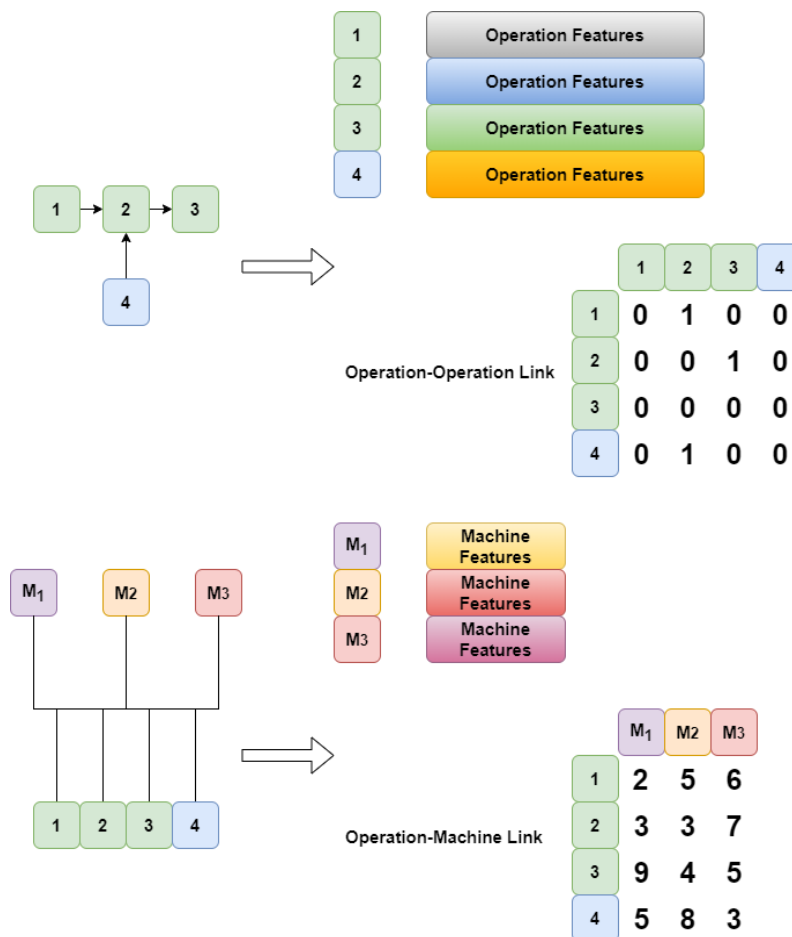
Fig. 7 State representation

As illustrated in Fig. 7, each operation encompasses several features, including:

- The operation status, where 1 indicates the operation is scheduled, and 0 indicates that the operation is not scheduled.
- The number of machines in proximity to the operation, representing the count of machines that can be assigned to the operation.
- The average processing time of the operation on the assignable machines, or the processing time of the operation on the assigned machine if specified.

- The number of operations that are not scheduled in the job to which the operation belongs.
- The time at which the operation is completed in the job.
- The start time of the operation

Similarly, each machine includes various features:

- The number of operations that can be processed by the machine.
- The time when the machine can start.
- Machine utilization as of the current time step.

In addition, the operation link relationship is represented by an adjacency matrix, and the processing time of the operation on different machines is represented by the operation-machine connection matrix, where the numerical value in the matrix represents the processing time of the operation on the machine, where 0 indicates that the operation cannot be assigned to the machine.

4.2 Action space

In the action space, each action corresponds to a pair of operation-machine. It comprises two key components: the operation and the machine, signifying that the operation can be assigned to the specified machine. For instance, the action space depicted in Fig. 7 is represented as in Fig. 8.

Given that each operation can be assigned to any machine for processing, the combination of each operation with each machine results in 3 distinct actions. Consequently, the final action space encompasses a total of 12 actions.
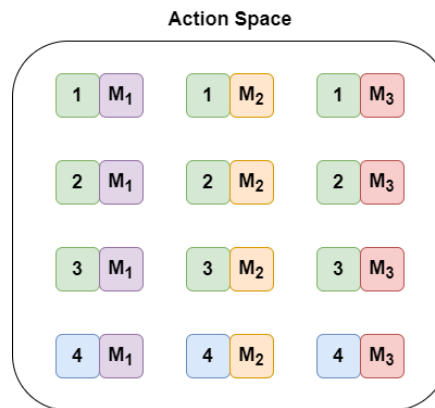
Fig. 8 Action representation

4.3 Reward space

In traditional FJSP, the maximum completion time is typically used as the reward metric. However, in FJSP-PO, not only the maximum completion time needs to be considered, but also the gap between operations. Therefore, the interval between operations needs to be incorporated into the reward function. For further details, please refer to the training process section.

4.4 Constraint mask tensor

To enhance scheduling efficiency, we pre-filter actions within the action space before selection. Additionally, to ensure scalability with respect to business constraints, we consolidate each business constraint into a mask tensor format. Taking the operation depicted in Fig. 7 as an illustration, let's consider a business constraint: prioritizing the operation along the green line. This entails two constraints – one pertaining to operation sequencing and the other to the aforementioned business constraint. These are expressed as follows at time $t = 0$:

Each operation is capable of being processed on all machines, hence every value within the Action Tensor is assigned as 1. Mask Tensor 1 identifies the operations available for scheduling at time $t = 0$, whereas Mask Tensor 2 encapsulates the business constraint of prioritizing operations along the green line. Finally, we conduct element-wise multiplication among these three tensors to derive the ultimate constraint mask tensor as shown in Fig. 9.
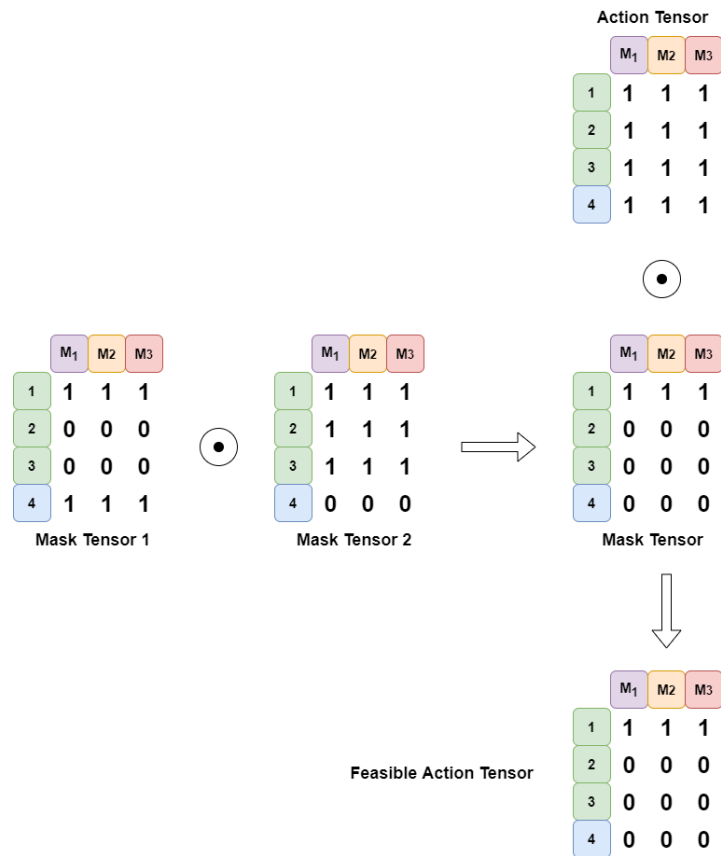
Fig. 9 Mask tensor

4.5 Model structure

We propose an AR-HGAT model structure, as shown in the Fig. 10.
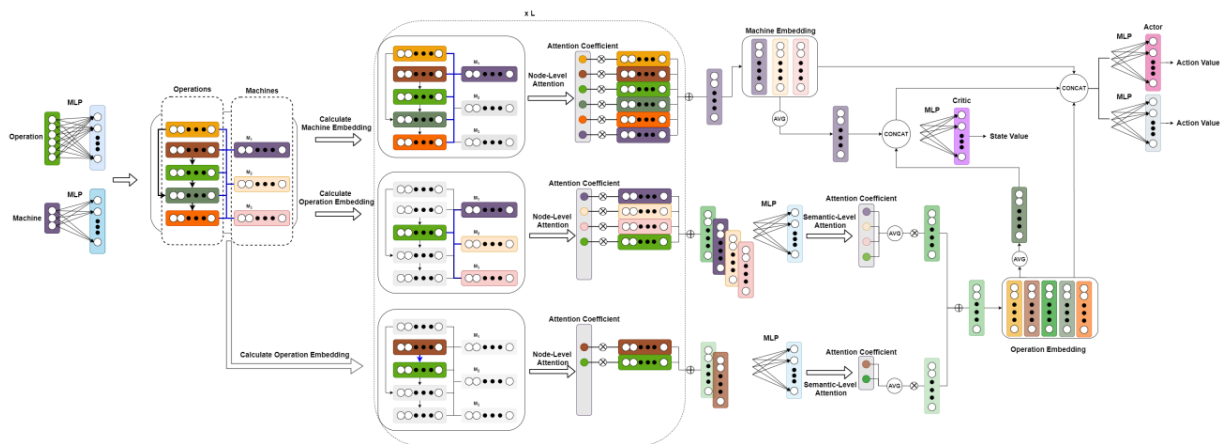
Fig. 10 AR-HGAT model structure

The proposed network architecture primarily consists of a multi-layer perceptron (MLP) and a graph neural network. The MLP is responsible for transforming operation features, while another MLP transforms machine features. These transformed operation and machine features are then processed through a graph attention network, initially employing node-level attention. The node-level attention for operations is bifurcated into two segments: one segment gathers machine information connected to the operation, while the other segment collects information about dependent operations. Subsequently, semantic-level attention is applied to these two segments of node-level information for operations to obtain the final operation embedding. After obtaining embeddings for all operations and machines, average pooling operations are separately conducted on machine and operation embeddings. The resulting outputs are then fed into another

MLP, referred to as the Critic network, to derive the state value of the current moment. Simultaneously, the pooling results are input into an MLP, termed the Actor network, alongside machine and operation embeddings to determine the value of each action at the current moment. Additionally, another MLP is utilized to determine the duration for which an operation needs to wait at the current moment. AR mechanisms and mask tensors will be integrated into the Actor and Critic networks, as elaborated in the subsequent section.

4.6 Training process

We adopt the Proximal Policy Optimization (PPO) [45] algorithm to update the parameters of the AR-HGAT. And the training process as shown in Fig. 11 can be divided into several steps: data collection, data sampling, model inference, calculation of advantage function, calculation of loss, and updating AR-HGAT model parameters. In the data collection stage, the agent interacts with the environment under the guidance of strategy π, and obtains multiple rounds of interaction data, in which s_t, $\tilde{P}_{a_t}$, and r_t represent the state, action probability distribution and reward value at time t, respectively. In the data sampling stage, a batch of data is randomly sampled from multiple rounds of interactive data, and then we obtain a batch of training data, and then the next process is to train the model, firstly, based on the input of the training data s_t, the predicted output of the model: action probability distribution $\tilde{P}_{a_t}$ and state value q_t, then we calculate the advantage function value of the training sample Adv_t, and finally we calculate the loss value through the PPO algorithm and update the AR-HGAT model parameters.

In the AR-HGAT model, it is mainly divided into two stages, the first stage is the embedding calculation of the machine and the operation, wherein the embedding calculation of the machine is as follows:
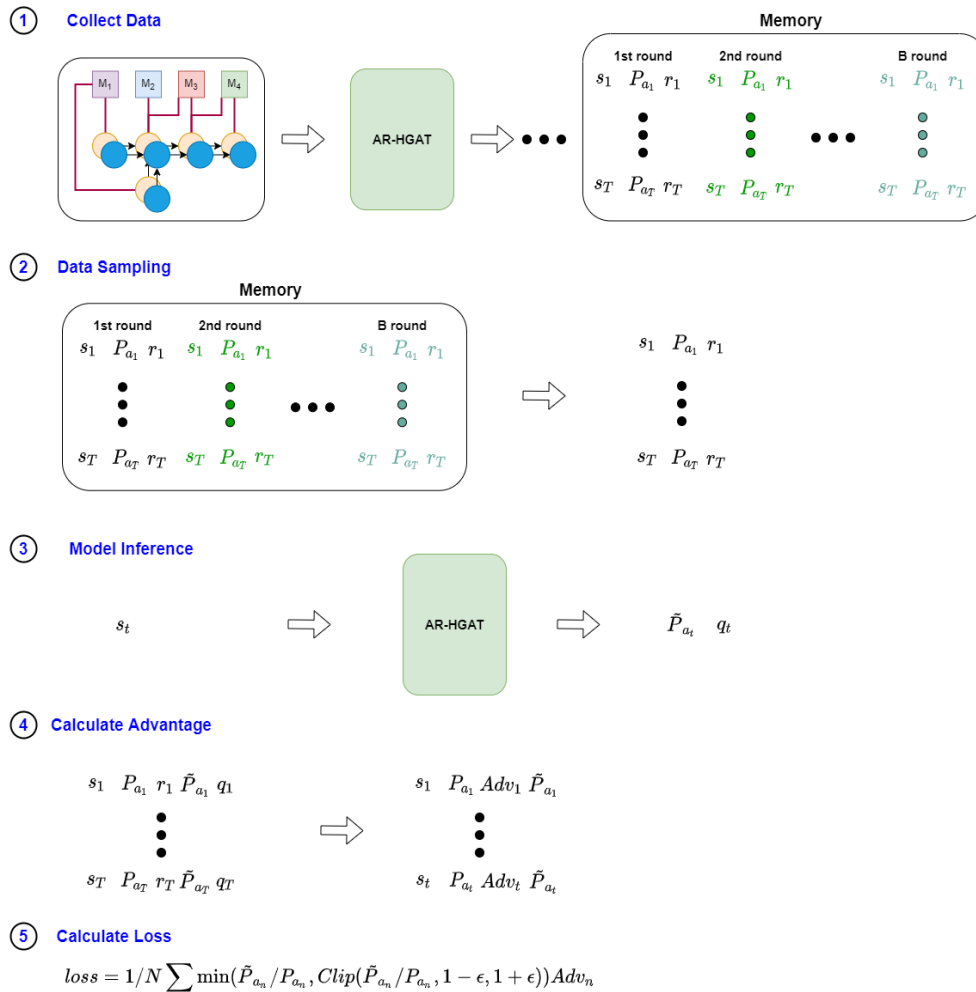
Fig. 11 Training process

1. Calculate attention coefficients

Operation O_{ij} that can be handled by machine M_k, the attention coefficients between operation $O_{ij} \in N_{(M_k)}$ and M_k is represented as:

$$e_{ijk} = LeakyReLU\big(a^T[W^M v_k || W^O \mu_{ij} || W^E \lambda_{ijk}]\big)$$
$$i \in [1, I], j \in [1, B_i], k \in [1, K] \tag{3}$$

The attention coefficients of machine M_k about itself is represented as:

$$e_{kk} = LeakyReLU(a^T[W^M v_k || W^M v_k])$$
$$k \in [1, K] \tag{4}$$

$a \in \mathbb{R}^{2d}$
$W^M \in \mathbb{R}^{d \times 3}$
$W^O \in \mathbb{R}^{d \times 6}$
$W^E \in \mathbb{R}$
$\lambda_{ijk} \in \mathbb{R}$

LeakyReLU is a common activation function in neural network and

v_k: the original features of the k-th machine
μ_{ij}: the original features of the j-th operation of the i-th job
λ_{ijk}: the processing time of operation O_{ij} on machine M_k

2. Normalize attention coefficients

By normalizing the attention coefficients:

$$a_{ijk} = exp(e_{ijk}) / \Big(\sum_{O_{pq} \in N_{(M_k)}} exp(e_{pqk}) + e_{kk}\Big)$$
$$i \in [1, I], j \in [1, B_i], k \in [1, K] \tag{5}$$

$$a_{kk} = exp(e_{kk}) / \Big(\sum_{O_{pq} \in N_{(M_k)}} exp(e_{pqk}) + e_{kk}\Big)$$
$$k \in [1, K] \tag{6}$$

3. Learning machine embedding

Eventually we can get the machine embedding as follows:

$$v'_k = \sigma\Big(\alpha_{kk} W^M v_k + \sum_{O_{ij} \in N_{(M_k)}} \alpha_{ijk}[W^O \mu_{ij} || W^E \lambda_{ijk}]\Big)$$
$$k \in [1, K] \tag{7}$$

and $v'_k \in \mathbb{R}^d$.

Considering that the embedding calculation of the operation not only depends on the assignable machine, but also depends on its pre-operation, the embedding calculation of the operation is mainly divided into node-level attention and semantic-level attention. Node-level attention aggregates information from machines and dependent operations, and semantic-level attention combines information from different attributes as follows:

1. Node-level attention from machine to operation

Operation O_{ij} that can be handled by machine M_k, the attention coefficients between operation O_{ij} and the assignable machine $M_k \in N_{(O_{ij})}$ is the same as Equation 3.

The attention coefficients of operation O_{ij} about itself is represented as:

$$e_{ij,ij} = LeakyReLU\big(a1^T[W^O \mu_{ij} || W^O \mu_{ij}]\big)$$
$$i \in [1, I], j \in [1, B_i] \tag{8}$$

Where $a1 \in \mathbb{R}^{2d}$.

By normalizing the attention coefficients:

$$a1_{ijk} = exp(e_{ijk})/\left(\sum_{M_p \in N_{(O_{ij})}} exp(e_{ijp}) + e_{ij,ij}\right)$$

$$i \in [1, I], j \in [1, B_i], k \in [1, K]$$

(9)

$$a1_{ij,ij} = exp(e_{ij,ij})/\left(\sum_{M_p \in N_{(O_{ij})}} exp(e_{ijp}) + e_{ij,ij}\right)$$

$$i \in [1, I], j \in [1, B_i]$$

(10)

Eventually we can get the node-level embedding from machine to operation as follows:

$$\mu1'_{ij} = \sigma(\alpha1_{ij,ij}W^O\mu_{ij} +$$

$$\sum_{M_k \in N_{(O_{ij})}} \alpha 1_{ijk}[W^O\mu_{ij}||W^E\lambda_{ijk}])$$

$$i \in [1, I], j \in [1, B_i]$$

(11)

2. Node-level attention from operation to operation

Operation O_{ij} needs to wait some operations to complete before it can start, and we use $D_{(O_{ij})}$ to represent all dependent operations of operation O_{ij}, and the attention coefficients between operation O_{ij} and the dependent machines $O_{ir} \in D_{(O_{ij})}$ are represented as:

$$e_{ir,ij} = LeakyReLU\left(a2^T[W^O\mu_{ij}||W^O\mu_{ir}]\right)$$

$$i \in [1, I], j \in [1, B_i], r \in [1, B_i]$$

(12)

Where $a2 \in \mathbb{R}^{2d}$ and the attention coefficients of operation O_{ij} about itself is the same as Equation [node_level_oo].

By normalizing the attention coefficients:

$$a2_{ir,ij} = \frac{exp(e_{ir,ij})}{\left(\sum_{O_{ip} \in D_{(O_{ij})}} exp(e_{ip,ij}) + e_{ij,ij}\right)}$$

$$i \in [1, I], j \in [1, B_i], r \in [1, B_i]$$

(13)

$$a2_{ij,ij} = exp(e_{ij,ij})/\left(\sum_{O_{ip} \in D_{(O_{ij})}} exp(e_{ip,ij}) + e_{ij,ij}\right)$$

$$i \in [1, I], j \in [1, B_i]$$

(14)

Eventually we can get the node-level embedding from dependent operations to operation as follows:

$$\mu2'_{ij} = \sigma\left(\alpha2_{ij,ij}W^O\mu_{ij} + \sum_{O_{ir} \in D_{(O_{ij})}} \alpha 2_{ir,ij}[W^O\mu_{ir}]\right)$$

$$i \in [1, I], j \in [1, B_i]$$

(15)

3. Semantic-level attention

Drawing on paper [46], we fuse the information gathered by the machine and the information gathered by the dependent operations. First, we get the importance of gathering information from machine to operation as follows:

$$\rho^1 = 1/|\Psi^1| \sum_{u \in \Psi^1} a 3^T Tanh(W1^T u)$$

(16)

Where Φ^1 represents a set, which contains the machine connected to operation O_{ij} and the operation O_{ij} itself. And $\Psi^1 \in \left[\mu1'_{ij}||\cup_{M_k \in N_{(O_{ij})}} v'_k\right]$

And Tanh is a common activation function in neural network and $a3 \in \mathbb{R}^d, W^1 \in \mathbb{R}^{d \times d}$.

Then we get the importance of gathering information from dependent operations to operation as follows:

$$\rho^2 = 1/|\Psi^2| \sum_{u \in \Psi^2} a 3^T Tanh(W1^T u)$$

(17)

Where Φ^2 represents a set that contains operation O_{ij} and its dependent operations. And $\Psi^2 \in \left[\mu2'_{ij} || \cup_{O_{ir} \in D_{(O_{ij})}} \right]$.

The second stage is to calculate the probability distribution of the action at the current moment t through the Actor network, and at the same time, taking into account the expansion of business constraints, before the softmax layer of the Actor network, a mask operation is added to limit the output of the illegal action.

In addition, due to the existence of parallel operations, some operations can wait, so a special action ϕ is taken, for which an AR mechanism is also added to the Actor network, so as to solve the problem of when a special action ϕ should be taken, and when the Actor network selects a special action, the MLP network is enabled to obtain the processing time of the special action ϕ. It is shown below:

1. Setting attention reboot factor

We expect the action ϕ to be rebooted at the right time, and the degree of attention for the action ϕ is defined as follows:

$$\gamma = 1\{!\,\phi\} \,\&\, \left(1\{M^f == M^\phi\}\right) * \epsilon \tag{18}$$

and

$$1\{!\,\phi\} = \begin{cases} 0 & \text{executing the action } \phi \\ 1 & \text{other} \end{cases} \tag{19}$$

Where M^f represents the machine that processing the first operation of each line, and M^ϕ represents the machine that process action ϕ, and ϵ is a very small number. And when γ equal ϵ, we need reboot the action ϕ, otherwise close the action ϕ.

2. Obtain action distribution

Firstly, machine embedding and operation embedding of the last network layer of the HGAT (Heterogeneous Graph Attention Network) is obtained at time t:

$$h_t = \left[\frac{1}{\mathcal{O}} \sum_{O_{ij} \in \mathcal{O}} \mu'^{(L)}_{ij} \,||\, \frac{1}{\mathcal{M}} \sum_{M_k \in \mathcal{M}} v'^{(L)}_k \right] \tag{20}$$

To keep things simple, we build an Multi-Layer Perceptron (MLP) neural network to get the step size to take the action ϕ as follows:

$$\Gamma_t = \left| MLP_{\theta_\phi}(\bar{h}_t) \right| \tag{21}$$

and $\bar{h}_t$ is the average of h_t.

And the probability of each action a_t under state s_t is represented as:

$$P(a_t, s_t) = MLP_\omega \left[\mu'^{(L)}_{ij} || v'^{(L)}_k || h_t \right] \tag{22}$$

Because we want to know when to perform this action ϕ, so we deal with this action separately:

$$\begin{aligned} sig &= binary\big(P(a_t = \phi, s_t)\big) \\ coeff &= (1 + sig) * \gamma - sig * \epsilon \\ P(a_t = \phi, s_t) &= P(a_t = \phi, s_t)/coeff \end{aligned} \tag{23}$$

Where γ is attention reboot factor and

$$binary(x) = \begin{cases} 1 & x >= 0 \\ 0 & x < 0 \end{cases} \tag{24}$$

In order to make the loss fall more smoothly, we treat the score as follows:

$$P(a_t, s_t) = P(a_t, s_t)/\sqrt{4 * d} \tag{25}$$

Finally we can get the strategy as follows:

$$\pi_\omega(a_t|s_t) = \frac{exp(P(a_t, s_t))}{\sum_{a'_t \in A_t} exp\big(P(a'_t, s_t)\big)}, \forall a_t \in A_t \tag{26}$$

3. Calculate state-value

In order to calculate the value of the advantage function, in the third stage, the Critic network is added to obtain the value in different states, which is expressed as q_t, as shown below:

$$q_t = MLP_{\theta_c}(\bar{h}_t) \tag{27}$$

The training process is mainly guided by the reward value, and because we added the action ϕ, we hope to make the operations as continuous as possible under the premise of meeting the shortest of the latest completion time. Therefore, we set up two sub reward functions as follows:

1. Makespan

Let's say that at T time we complete the scheduling of all operations, and the makespan at each time t is expressed as C_t, we define C_t as follows:

$$C_t = maximize\left(ET_{O_{ij}}\right)$$
$$t \in [1,T], i \in [1,I], j \in [1,B_i] \tag{28}$$

Where $ET_{O_{ij}}$ represents the finish time of operation O_{ij}, and we need minimize C_T. Specially, $C_0 = 0$ at time $t = 0$, so we have:

$$C_T = \sum_t (C_t - C_{t-1}), t \in [1,T] \tag{29}$$

Therefore, we need minimize each item $C_t - C_{t-1}$, or maximize the item $C_{t-1} - C_t$, and the reward for makespan at time t is represented as:

$$r_t^1 = C_{t-1} - C_t, t \in [1,T] \tag{30}$$

2. ϕ-Reward

Because we add the action ϕ, because the step size Γ_t at time t will affect the makespan C_t, and the C_T is represented as:

$$C_t' = C_t + \xi_t$$
$$C_T = \sum_t (C_t' - C_{t-1}' - \kappa * \Gamma_t)$$
$$t \in [1,T] \tag{31}$$

Where ξ_t indicates the amount of progress change since taking the action ϕ, and hyper-parameter κ balance the impact of the action ϕ on makespan, so we can define the second reward function as follows:

$$r_t^2 = \kappa * \Gamma_t, t \in [1,T] \tag{32}$$

Therefore, the total reward at each time t is shown as follows:

$$r_t = r_t^1 + r_t^2 \tag{33}$$

And the design of advantage function and loss is shown as follows:

$$Adv_t = \sum_{r=0} (\beta^r r_{t+r})$$
$$Adv_t = Adv_t - q_t$$
$$t \in [1,T] \tag{34}$$

$$loss = \frac{1}{N}\sum_{n=1} (\min[ratio_n,$$
$$Clip(ratio_n, 1-\epsilon, 1+\epsilon)]Adv_n)$$
$$n \in [1,N], ratio_n = \tilde{P}_{a_n}/P_{a_n} \tag{35}$$

Where Adv_t represents the value of the advantage function at time t, P_{a_n} represents the action probability distribution obtained in the data collection stage, $\tilde{P}_{a_n}$ represents the action probability distribution predicted by the AR-HGAT model, N represents the number of training samples, q_t is state value, and $Clip$ is the clipping function, which is defined as follows:

$$Clip(x, min, max) = \begin{cases} min & x < min \\ max & x > max \\ x & other \end{cases} \tag{36}$$

Finally, we describe the training process with pseudo-code as Algorithm 1.

Algorithm 1 Training process pseudocode

1: Initialize the model, the simulation environment, and set the hyper-parameters.

2: **for** epoch from 1 to max_epochs: **do**

3:   **if** epoch % update_env equals to 0: **then**

4:     Regenerate random inputs s_1

5:   **end if**

6:   Interact with the environment and collect the dataset to get(s_t, P_{a_t}, r_t)

7:   **if** epoch % update_model equals to 0: **then**

8:     Input s_t to the model to get $\widetilde{P_{a_t}}$

9:     Calculate loss function value by Equation 35

10：   Update the parameters of the AR-HGAT model Clear training data

11： **end if**

12:   **if** epoch % validate_model equals to 0: **then**

13:     Regenerate random inputs s_1

14:     Input s_1 to the model and interact with the environment to get the final reward value

15:     **if** the final reward value is the best: **then**

16:       Save the best model

17:     **end if**

18:   **end if**

19: **end for**

4.7 Forecasting process

The whole prediction process is the same as the markov process, as shown in Fig. 5, and the pseudo-code of the whole inference process is as Algorithm 2.

Algorithm 2 Forecasting process pseudocode

1: Load the best model and sample s_1

2: Initialize Actions $= []$, $t = t + 1$

3: **while** s_t is not terminated: **do**

4:   Obtain action by $a_t = \text{AR} - \text{HGAT}(s_t)$

5:   Actions.add(a_t)

6:   Input at to simulation environment obtain next state

7: $t = t + 1$

8: **end while**

9: OUTPUT Actions

---

5. Experiment and discussion

To validate the effectiveness of our proposed model and algorithm, we will conduct three experiments. Because FJSP-PO is an extension of FJSP, it is also necessary to conduct experimental comparisons on FJSP, for which we selected two sets of experiments. The first experiment aims to evaluate the effectiveness of AR-HGAT on FJSP using a simulated datasets, as listed in Table 1 six problem size (600 instances) are considered. For each instance is sampled by drawing from the corresponding uniform distribution [28]. The second experiment focuses on comparing the effectiveness of AR-HGAT on FJSP using benchmark datasets, including 10 mk instances [47] and three groups of la instances (120 instances) [48], as indicated in Table 2. Additional details about these instances of simulated datasets and benchmark datasets can be found in the respective references. Lastly, the third experiment aims to assess the effectiveness of AR-HGAT on FJSP-PO using the company's proprietary real business datasets as Table 3. Due to the absence of benchmark data and a designated methodology within the framework of FJSP-PO, we propose the application of a tree search algorithm as a comparative approach to address this issue. And the coding part is publicly available.

5.1 Experiments settings

Configuration

All the experiments in this paper are conducted on a container using the Linux x86_64 system. The system is equipped with no GPUs, 20GB of memory, and 32 cores. The training configuration is outlined in Table 4 and Table 5. For the sake of simplicity of the network structure, we keep the MLP network structure of the calculation step size consistent with the Critic network. And the distribution of the number of operations for each job is $U(0.8 * machines, 1 * machines)$, the *machines* represents the total number of machines. In order to verify the generalization of our proposed method, we only kept one line for each job in the training phase, and considering that when there was only one line, adding special action ϕ had no effect, so we did not introduce special action ϕ in the training process. As described in the FJSP-PO problem in Section II, in order to more easily verify the effectiveness of the proposed method, we use a sample of real data, and each operation can only be assigned to one machine for processing. During our experimental comparisons, we employ a sampling strategy wherein, akin to the training process, sampling is conducted in each state based on probabilities to obtain actions. While the results obtained exhibit some level of randomness, this approach may lead to the discovery of superior solutions.

Baselines

In this study, a comparison will be conducted among the traditional rule-based method, classical operations research algorithms, and the reinforcement learning-based operation scheduling algorithm mentioned in the study [28]. The traditional rule-based methods are well-known PDRs, including FIFO, Most Operations Remaining (MOR), Shortest Processing Time (SPT) and Most WorK Remaining (MWKR). In addition, OR-Tools is a well-known operations research algorithm toolkit, and we have compared our proposed AR-HGAT method with the relevant genetic algorithms in OR-Tools, and the algorithm proposed in the study [28], To our knowledge, there are currently no publicly available datasets or methods tailored specifically for addressing FJSP-PO. Consequently, we have developed and implemented a tree search algorithm for this problem to serve as a comparative benchmark against our proposed AR-HGAT method. We opted for the tree search algorithm due to its ability to effectively balance scheduling performance and computational time through straightforward control over search breadth and depth.

5.2 Data detail

Part of the comparison data is derived from randomly generated simulation data, as shown in Table 1. Another portion of the data is obtained from benchmark datasets, which include the 10 mk instances (mk01-mk10) and three groups of la instances (rdata, edata, and vdata, each with 40 instances) proposed in paper [28]. These instances vary in size, ranging from 10×6 to 30×10.

Table 1 Simulated datasets

Size ($n \times m \times k$)	n_i	$|M_{ij}|$	$\bar{p}_{ij}$
$10 \times 5 \times 100$	U(4,6)	U(1,5)	U(1,20)
$20 \times 5 \times 100$	U(4,6)	U(1,5)	U(1,20)
$15 \times 10 \times 100$	U(8,12)	U(1,10)	U(1,20)
$20 \times 10 \times 100$	U(8,12)	U(1,10)	U(1,20)
$30 \times 10 \times 100$	U(8,12)	U(1,10)	U(1,20)
$40 \times 10 \times 100$	U(8,12)	U(1,10)	U(1,20)

Table 2 Benchmark data

Name	Instances
MK	10
rdata	40
edata	40
vdata	40

The dataset comprises instances with n jobs and m machines. There are a total of k instances in this configuration. Number of operations for job j_i. Number of compatible machines for operation O_{ij}. Average processing time of operation O_{ij}.

To assess the effectiveness of AR-HGAT on FJSP-PO, a dataset was compiled for testing. The data originates from our factory production line, where each operation corresponds one-to-one with a machine. This implies that each operation can only be assigned to a single machine, and each machine exclusively handles one type of operation. Additionally, the processing time of each operation on the machine can vary. To explore diverse scenarios, extra lines and constraints were introduced based on this foundation. The details of the instance are shown as Table 3.

The following two points need to be noted:

- The constraint is the maximum constraint per day, which is set to 3, that is, only 3 jobs can be produced per day.
- For simplicity, the machine calendar does not have a rest period, and only the processing time of the process at the machine is retained (regardless of changeover time, waiting time, disassembly time, etc.)

Table 3 FJSP-PO Data

Size ($n \times m$)	n_i	$|M_{ij}|$	$\bar{p}_{ij}$	l_i	Constraints
10×19	190	1	U(1,20)	1	N
10×19	190	1	U(1,20)	1	Y
12×55	660	1	U(1,20)	11	N
12×55	660	1	U(1,20)	11	Y

n_i: Number of operations in Job J_i.
$|M_{ij}|$: Number of compatible machines for operation O_{ij}.
$\bar{p}_{ij}$: Average processing time of operation O_{ij}.
l_i: Number of lines in Job J_i.
Constraints: Whether there are constraints.

5.3 Training configuration

In order to verify the effectiveness of the AR-HGAT method we proposed, we will train the model on the simulation data generated by ourselves, and the details of the training data and the important model hyperparameters are shown in Table 4 and Table 5.

The training loss curves based on Table 4 and Table 5 are shown in Fig. 12. The training process was iterated 400 times, and the model was verified every 20 times, and it can be seen from the figure that the loss value has a good convergence, and the makespan of the validation set also shows a downward trend.

Table 4 Training data detail

Jobs	10
Machines	5
Processing time	U(1,20)

Table 5 Training Hyper-parameters

Batch size	20
Validation batch size	100
d	8
κ	0.01
Hidden size of Actor	64
Hidden size of Critic	64
Number of hidden layer in Actor	3
Number of hidden layer in Critic	3
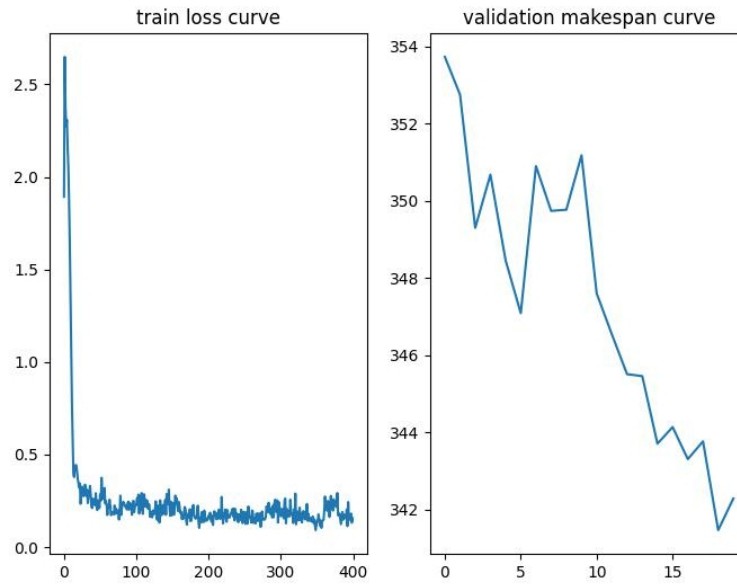
Fig. 12 Training loss curve for 10×5

5.4 Algorithm comparison

Simulated data comparison for FJSP

In order to verify the effectiveness of the AR-HGAT algorithm model on FJSP, we compare the three indicators of makespan (C_T), gap and operation scheduling time with the related algorithms mentioned in paper [28], and the gap is calculated as follows:

$$Gap = \left(C_T / C_{opt} - 1\right) \cdot 100 \%$$ (37)

Where C_opt is the local optimal solution, or even the optimal solution.

We compared AR-HGAT with the best results based on this dataset in the paper [28], and the results are as follows:

Table 6 Simulated data comparison

Size		OR-Tools	MOR	SPT	FIFO	MWKR	DRL	AR-HGAT
10×5	C_T	96.59 (15 %)	116.69	129.06	119.62	115.29	105.61
	Gap	-	20.89 %	33.57 %	23.94 %	19.44 %	9.38 %
20×5	C_T	188.45 (0 %)	217.17	229.89	216.13	216.98	207.5
	Gap	-	15.29 %	22.08 %	14.73 %	15.17 %	10.13 %
15×10	C_T	145.42 (5 %)	173.4	198.2	185.08	169.18	160.36
	Gao	-	19.29 %	36.30 %	27.27 %	16.34 %	10.31 %
20×10	C_T	197.24 (0 %)	221.86	254.59	234.21	220.85	214.87
	Gap	-	12.53 %	29.09 %	18.82 %	11.99 %	8.97 %
30×10	C_T	294.1 (0 %)	320.18	347.4	328.5	319.89	312.2
	Gap	-	8.90 %	18.12 %	11.74 %	8.79 %	6.18 %
40×10	C_T	397.36 (0 %)	425.19	443.3	427.22	425.7	415.14
	Gap	-	7.02 %	11.57 %	7.53 %	7.15 %	4.49 %

Table 7 Running time (in seconds) comparison

Size	OR-Tools	MOR	SPT	FIFO	MWKR	DRL	AR-HGAT
10×5	1597.88	0.15	0.15	0.15	0.16	4.52	5.79
20×5	1800.00	0.29	0.31	0.29	0.31	6.93	7.66
15×10	1724.00	0.42	0.44	0.42	0.46	7.99	8.24
20×10	1800.00	0.58	0.60	0.58	0.61	8.01	8.92
30×10	1800.00	0.87	0.90	0.87	0.93	8.28	9.52
40×10	1800.00	1.17	1.20	1.16	1.25	8.37	9.85

The percentage in column OR-Tools represents the percentage of the optimal solution that can be obtained in 1800 seconds of computational time. In the study [25], two models are proposed, one is DRL-G and the other is DRL-S, and we only take the model with the best performance in the dataset and represent it as DRL. And the last column represents the AR-HGAT algorithm model we proposed.

From the comparison results, it can be seen that for all the simulation data, the algorithm in OR-Tools has the best results, after considering the time, two-thirds of the dataset cannot be solved in 1800 seconds, and the rule-based algorithm, MOR, SPT, FIFO and MWKR have the advantage in solving time, but there is a large gap between the solution and the better solution. And it can be seen that the proposed method is better than the MOR, SPT, FIFO, MWKR and DRL methods in solving the problem, and is better than the method in OR-Tools in terms of solution time, and the gap between the proposed method and the best solution is not much different. As can be seen from Table 6 and Table 7, for 6 datasets of different scales, for rule-based methods such as MOR, SPT, FIFO, MWKR, its solution time is better than that of deep learning-based methods, such as the DRL method and our proposed AR-HGAT method, and compared with the DRL method, because our method takes into account semantic-level information, the inference time is relatively longer, but compared with the solution obtained by the rule-based and DRL methods, the solution obtained by our proposed method is better than that of the 10×5 scale, 20×5 scale, 15×10 scale, 20×10 scale, 30×10 scale and 40×10 scale should be at least 1.75 %, 0.59 %, 1.74 %, 1.95 %, 0.88 % and 0.78 % higher, respectively. Therefore, the proposed method can obtain a better solution in a controllable time.

Benchmark data comparison for FJSP

In order to verify the generalization performance of the proposed algorithm, that is, the prediction ability of the out-of-distribution dataset, we compare multiple algorithms based on the well-known benchmark dataset in FJSP, and the comparison results are shown as [. And the last row represents the AR-HGAT algorithm model we proposed.

From the experimental results, it can be seen that because the proposed AR-HGAT method can process multiple instances in parallel, the overall inference time is better than that of the traditional operations research algorithm provided in OR-Tools, and the solution obtained is also better than that of the classical PDR method. As can be seen from Table 8, for 4 datasets of different scales, and compared with the solution obtained by the rule-based and DRL methods, the solution obtained by our proposed AR-HGAT method is better than that of the mk, rdata, edata, vdata, should be at least 1.61 %, 0.78 %, 1.51 %, 0.47 % higher, respectively. Therefore, the proposed method can obtain a better solution in a controllable time.

Real data comparison for FJSP-PO

Considering our real business scenario, we introduce a new problem termed FJSP-PO, an extension of FJSP tailored to our practical operation scheduling needs. While our proposed AR-HGAT method's effectiveness is validated in basic FJSP through simulation and benchmark data comparison, there lacks a public benchmark dataset and corresponding adaptation algorithm for the new FJSP-PO. To address this gap, we employ a tree search algorithm to solve FJSP-PO and train the AR-HGAT model using extensive simulation data. Therefore, We use the heuristic tree search scheduling algorithm implemented in Java as a benchmark (iterative 10 times), Comparing with our proposed AR-HGAT method based on actual business data. Compared to the tree search algorithm, AR-HGAT offers greater convenience in ease of use and research and development simplicity. Considering potential differences in dataset distributions, training on a more consistent dataset may yield improved results. Notably, subsequent iterations of our proposed AR-HGAT method necessitate minimal customized R&D effort, requiring only dataset collection, preprocessing, and iterative model optimization based on new data.

Table 8 Benchmark comparison

Method	mk			la(rdata)			la(edata)			la(vdata)		
	C_T	Gap (%)	Time (s)	C_T	Gap (%)	Time (s)	C_T	Gap (%)	Time (s)	C_T	Gap (%)	Time (s)
OR-Tools	174.9	1.42	905.45	938.38	0.28	1203.89	1026.7	-0.19	226.53	924.4	0.38	1367.77
MOR	202.31	29.59	0.40	1064.04	14.71	0.45	1211.23	17.91	0.45	969.25	6.09	0.45
SPT	238.8	45.73	0.41	1184.8	27.70	0.46	1305.35	26.19	0.46	1085.46	18.20	0.46
FIFO	206.09	30.72	0.39	1082.08	16.63	0.44	1255.46	22.07	0.45	980.69	7.50	0.44
MWKR	200.17	28.19	0.45	1046.2	12.53	0.51	1179.9	14.92	0.50	962.01	5.11	0.50
DRL	190.4	19.02	3.51	986.1	5.75	4.16	1116.33	8.18	4.78	932	1.44	4.27
AR-HGAT	**187.39**	**17.14**	**4.87**	**978.42**	**4.93**	**5.97**	**1099.72**	**6.57**	**4.98**	**927.68**	**0.97**	**4.76**

Table 9 Real data comparison

No.	Tree Search		AR-HGAT	
	makespan	time	makespan	time
1	443	8	401	9
2	259542	5	259426	10
3	580	12	563	33
4	259443	12	259501	35

From the experimental results as shown in Table 9, we found that in 75 % of the instances, the method we proposed was better than the results based on the tree search algorithm within the time frame acceptable to the user. For the first instance, the results of our proposed AR-HGAT algorithm model are 10.5 % better than the traditional algorithm. Because the implementation of the tree search algorithm is in Java language, while our method is in Python language, considering the difference in language, it may also lead to the difference in the overall inference time. Based on real-world operation scheduling data, our experimental results demonstrate that the AR-HGAT method we introduce not only yields favorable outcomes in conventional FJSP scenarios but also exhibits generalizability to the newly proposed FJSP-PO. This presents a novel technical approach for real process scheduling scenarios, breaking away from the constraints of customized research and development reliant on traditional methods across diverse scenarios. Leveraging the AR-HGAT method, models are iteratively refined to meet business objectives based on data, facilitating direct transferability of the AR-HGAT model from one business scenario to another for assessing its impact and guiding subsequent iterative optimization processes.

6. Conclusion and future work

Our objective is to develop a comprehensive scheduling model capable of managing discrete scenarios with parallel operations, encompassing the processing of multiple lines, and accommodating real-world business constraints. The methodology involves segregating each constraint from the algorithm module, substituting them with mask tensors, and deriving an optimal solution that maximizes the reward function within the feasible solution space. Experimental findings showcase favorable scheduling outcomes compared to conventional algorithmic approaches. Concurrently, our research introduces a novel technical pathway for real operation scheduling scenarios, grounded in data and extended to novel operation scheduling contexts through iterative model optimization. While our results are promising, there are areas for further exploration and experimentation, including testing more intricate business constraints and minimizing waiting times between different lines. Furthermore, addressing operational disparities and optimizing the model for expedited handling of larger operations represent avenues for future enhancement.

References

[1] Maderna, R., Pozzi, M., Zanchettin, A.M., Rocco, P., Prattichizzo, D. (2022). Flexible scheduling and tactile communication for human–robot collaboration, *Robotics and Computer-Integrated Manufacturing,* Vol. 73, Article No. 102233, doi: 10.1016/j.rcim.2021.102233.
[2] Gao, K.Z., Suganthan, P.N., Chua, T.J., Chong, C.S., Cai, T.X., Pan, Q.K. (2015). A two-stage artificial bee colony algorithm scheduling flexible job-shop scheduling problem with new job insertion, *Expert Systems with Applications,* Vol. 42, No. 21, 7652-7663, doi: 10.1016/j.eswa.2015.06.004.

[3] Bandara, D., Mayorga, M.E., McLay, L.A. (2014). Priority dispatching strategies for EMS systems, *Journal of the Operational Research Society,* Vol. 65, No. 4, 572-587, doi: 10.1057/jors.2013.95.

[4] Kanet, J.J., Hayya, J.C. (1982). Priority dispatching with operation due dates in a job shop, *Journal of Operations Management,* Vol. 2, No. 3, 167-175, doi: 10.1016/0272-6963(82)90004-3.

[5] Geiger, C.D., Uzsoy, R., Aytuğ, H. (2006). Rapid modeling and discovery of priority dispatching rules: An autonomous learning approach, *Journal of Scheduling,* Vol. 9, 7-34, doi: 10.1007/s10951-006-5591-8.

[6] Xie, J., Gao, L., Peng, K., Li, X., Li, H. (2019). Review on flexible job shop scheduling, *IET Collaborative Intelligent Manufacturing,* Vol. 1, No. 3, 67-77, doi: 10.1049/iet-cim.2018.0009.

[7] Birgin, E.G., Feofiloff, P., Fernandes, C.G., De Melo, E.L., Oshiro, M.T.I., Ronconi, D.P. (2014). A MILP model for an extended version of the flexible job shop problem, *Optimization Letters,* Vol. 8, 1417-1431, doi: 10.1007/s11590-013-0669-7.

[8] Lunardi, W.T., Birgin, E.G., Laborie, P., Ronconi, D.P., Voos, H. (2020). Mixed Integer linear programming and constraint programming models for the online printing shop scheduling problem, *Computers & Operations Research,* Vol. 123, Article No. 105020, doi: 10.1016/j.cor.2020.105020.

[9] Kim, Y.-D., Yano, C.A. (1994). A new branch and bound algorithm for loading problems in flexible manufacturing systems, *International Journal of Flexible Manufacturing Systems,* Vol. 6, 361-381, doi: 10.1007/BF01324801.

[10] Wang, H., Sheng, B., Lu, Q., Yin, X., Zhao, F., Lu, X., Luo, R., Fu, G. (2021). A novel multi-objective optimization algorithm for the integrated scheduling of flexible job shops considering preventive maintenance activities and transportation processes, *Methodologies and Application,* Vol. 25, 2863-2889, doi: 10.1007/s00500-020-05347-z.

[11] Baykasoğlu, A, Madenoğlu, F.S. (2021). Greedy randomized adaptive search procedure for simultaneous scheduling of production and preventive maintenance activities in dynamic flexible job shops, *Soft Computing*, Vol. 25 14893-14932, doi: 10.1007/s00500-021-06053-0.

[12] Shahgholi Zadeh, M., Katebi, Y., Doniavi, A. (2018). A heuristic model for dynamic flexible job shop scheduling problem considering variable processing times, *International Journal of Production Research,* Vol. 57, No. 10, 3020-3035, doi: 10.1080/00207543.2018.1524165.

[13] Ortíz, M.A., Betancourt, L.E., Negrete, K.P., De Felice, F., Petrillo, A. (2018). Dispatching algorithm for production programming of flexible job-shop systems in the smart factory industry, *Annals of Operations Research,* Vol. 264, 409-433, doi: 10.1007/s10479-017-2678-x.

[14] Teymourifar, A., Ozturk, G., Ozturk, Z.K., Bahadir, O. (2020). Extracting new dispatching rules for multi-objective dynamic flexible job shop scheduling with limited buffer spaces, *Cognitive Computation,* Vol. 12, 195-205, doi: 10.1007/s12559-018-9595-4.

[15] Lim, K.C.W., Wong, L.P., Chin, J.F. (2022). Simulated-annealing-based hyper-heuristic for flexible job-shop scheduling, *Engineering Optimization,* Vol. 55, No. 10, 1635-1651, doi: 10.1080/0305215X.2022.2106477.

[16] Lv, L., Shen, W. (2023). An improved NSGA-II with local search for multi-objective integrated production and inventory scheduling problem, *Journal of Manufacturing Systems*, Vol. 68, 99-116, doi: 10.1016/j.jmsy.2023.03.002.

[17] Hajibabaei, M., Behnamian, J. (2021). Flexible job-shop scheduling problem with unrelated parallel machines and resources-dependent processing times: A tabu search algorithm, *International Journal of Management Science and Engineering Management*, Vol. 16, No. 4, 242-253, doi: 10.1080/17509653.2021.1941368.

[18] Campo, E.A., Cano, J.A., Gómez-Montoya, R., Rodríguez-Velásquez, E., Cortés, P. (2022). Flexible job shop scheduling problem with fuzzy times and due-windows: Minimizing weighted tardiness and earliness using genetic algorithms, *Algorithms,* Vol. 15, No. 10, Article No. 334, doi. 10.3390/a15100334.

[19] Cinar, D., Oliveira, J.A., Topcu, Y.I., Pardalos, P.M. (2016). A priority-based genetic algorithm for a flexible job shop scheduling problem, *Journal of Industrial and Management Optimization,* Vol. 12, No. 4, 1391-1415, doi: 10.3934/jimo.2016.12.1391.

[20] Kaweegitbundit, P., Eguchi, T. (2016). Flexible job shop scheduling using genetic algorithm and heuristic rules, *Journal of Advanced Mechanical Design, Systems, and Manufacturing,* Vol. 10, No. 1, Article JAMDSM0010, doi: 10.1299/jamdsm.2016jamdsm0010.

[21] Chen, R., Yang, B., Li, S., Wang, S. (2020). A self-learning genetic algorithm based on reinforcement learning for flexible job-shop scheduling problem, *Computers & Industrial Engineering,* Vol. 149, Article No. 106778, doi: 10.1016/j.cie.2020.106778.

[22] Zhang, G., Hu, Y., Sun, J., Zhang, W. (2020). An improved genetic algorithm for the flexible job shop scheduling problem with multiple time constraints, *Swarm and Evolutionary Computation*, Vol. 54, Article No. 100664, doi: 10.1016/j.swevo.2020.100664.

[23] Wang, L., Hu, X., Wang, Y., Xu, S., Ma, S., Yang, K., Liu, Z., Wang, W. (2021). Dynamic job-shop scheduling in smart manufacturing using deep reinforcement learning, *Computer Networks,* Vol. 190, Article No. 107969, doi: 10.1016/j.comnet.2021.107969.

[24] Wang, L., Pan, Z., Wang, J. (2021). A review of reinforcement learning based intelligent optimization for manufacturing scheduling, *Complex System Modeling and Simulation,* Vol. 1, No. 4, 257-270, doi: 10.23919/CSMS.2021.0027.

[25] Zhang, Y., Zhu, H., Tang, D., Zhou, T., Gui, Y. (2022). Dynamic job shop scheduling based on deep reinforcement learning for multi-agent manufacturing systems, *Robotics and Computer-Integrated Manufacturing,* Vol. 78, Article No. 102412, doi: 10.1016/j.rcim.2022.102412.

[26] Chien, C.-F., Lan, Y.-B. (2021). Agent-based approach integrating deep reinforcement learning and hybrid genetic algorithm for dynamic scheduling for Industry 3.5 smart production, *Computers & Industrial Engineering,* Vol. 162, Article No. 107782, doi: 10.1016/j.cie.2021.107782.

[27] Pan, Z., Wang, L., Wang, J., Lu, J. (2021). Deep reinforcement learning based optimization algorithm for permutation flow-shop scheduling, *IEEE Transactions on Emerging Topics in Computational Intelligence,* Vol. 7, No. 4, 983-994, doi: 10.1109/TETCI.2021.3098354.

[28] Song, W., Chen, X., Li, Q., Cao, Z. (2022). Flexible job-shop scheduling via graph neural network and deep reinforcement learning, *IEEE Transactions on Industrial Informatics,* Vol. 19, No. 2, 1600-1610, doi: 10.1109/TII.2022.3189725.

[29] Lin, C.-C., Deng, D.-J., Chih, Y.-L., Chiu, H.-T. (2019). Smart manufacturing scheduling with edge computing using multiclass deep Q network, *IEEE Transactions on Industrial Informatics,* Vol. 15, No. 7, 4276-4284, doi: 10.1109/TII.2019.2908210.

[30] Liu, C.-L., Chang, C.-C., Tseng, C.-J. (2020). Actor-critic deep reinforcement learning for solving job shop scheduling problems, *IEEE Access,* Vol. 8, 71752-71762, doi: 10.1109/ACCESS.2020.2987820.

[31] Kwon, Y.-D., Choo, J., Yoon, I., Park, M., Park, D., Gwon, Y. (2021). Matrix encoding networks for neural combinatorial optimization, (Computer Science, Machine Learning), doi: 10.48550/arXiv.2106.11113.

[32] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need, In: *Proceedings of 31st Conference on Neural Information Processing Systems (NIPS 2017),* Long Beach, California, USA, 1-11.

[33] Wu, Z., Pan, F., Chen, F., Long, G., Zhang, C., Yu, P.S. (2021). A comprehensive survey on graph neural networks, *IEEE Transactions on Neural Networks and Learning Systems,* Vol. 32, No. 1, 4-24, doi: 10.1109/TNNLS.2020.2978386.

[34] Zhang, C., Song, W., Cao, Z., Zhang, J., Tan, P.S., Xu, C. (2020). Learning to dispatch for job shop scheduling via deep reinforcement learning, In: *Proceedings of the 34th Conference on Neural Information Processing Systems,* Vancouver, Canada, 1621-1632.

[35] Hameed, M.S.A., Schwung, A. (2023). Graph neural networks-based scheduler for production planning problems using reinforcement learning, (Computer Science, Machine Learning), doi: 10.48550/arXiv.2009.03836.

[36] Su, C., Zhang, C., Xia, D., Han, B., Wang, C., Chen, G., Xie, L. (2023). Evolution strategies-based optimized graph reinforcement learning for solving dynamic job shop scheduling problem, *Applied Soft Computing,* Vol. 145, Article No. 110596, doi: 10.1016/j.asoc.2023.110596.

[37] Luo, S., Zhang, L., Fan, Y. (2021). Dynamic multi-objective scheduling for flexible job shop by deep reinforcement learning, *Computers & Industrial Engineering,* Vol. 159, Article No. 107489, doi: 10.1016/j.cie.2021.107489.

[38] Han, B.A., Yang, J.J. (2021). A deep reinforcement learning based solution for flexible job shop scheduling problem, *International Journal of Simulation Modelling,* Vol. 20, No. 2, 375-386, doi: 10.2507/IJSIMM20-2-CO7.

[39] Peng, F., Zheng, L. (2023). An improved multi-objective Wild Horse optimization for the dual-resource-constrained flexible job shop scheduling problem: A comparative analysis with NSGA-II and a real case study, *Advances in Production Engineering & Management,* Vol. 18, No. 3, 271-287, doi: 10.14743/apem2023.3.472.

[40] Sun, Z.Y., Han, W.M., Gao, L.L. (2023). Real-time scheduling for dynamic workshops with random new job insertions by using deep reinforcement learning, *Advances in Production Engineering & Management,* Vol. 18, No. 2, 137-151, doi: 10.14743/apem2023.2.462.

[41] Ren, J.F., Ye, C.M., Li, Y. (2021). A new solution to distributed permutation flow shop scheduling problem based on NASH Q-Learning, *Advances in Production Engineering & Management,* Vol. 16, No. 3, 269-284, doi: 10.14743/apem2021.3.399.

[42] Koblasa, F., Kralikova, R., Votrubec, R. (2020). Influence of EA control parameters to optimization process of FJSSP problem, *International Journal of Simulation Modelling,* Vol. 19, No. 3, 387-398, doi: 10.2507/IJSIMM19-3-519.

[43] Yildiz, İ., Saygin, A., Çolak, S., Abut, F. (2023). Development of a neural network algorithm for estimating the makespan in jobshop production scheduling, *Tehnički Vjesnik – Technical Gazette,* Vol. 30, No. 4, 1257-1264, doi: 10.17559/TV-20220818161430.

[44] Karacan, I., Karacan, I., Senvar, O., Bulkan, S. (2021). An integrated solution approach for Flow Shop Scheduling, *Tehnički Vjesnik – Technical Gazette,* Vol. 28, No. 3, 786-795, doi: 10.17559/TV-20200208192653.

[45] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. (2017). Proximal policy optimization algorithms, (Computer Science, Machine Learning), doi: 10.48550/arXiv.1707.06347.

[46] Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., Yu, P.S. (2019). Heterogeneous graph attention network, In: *Proceedings of WWW '19: The World Wide Web Conference,* San Francisco, California, USA, 215-233, doi: 10.1145/3308558.3313562.

[47] Brandimarte, P. (1993). Routing and scheduling in a flexible job shop by tabu search, *Annals of Operations Research,* Vol. 41, 157-183, doi: 10.1007/BF02023073.

[48] Hurink, J., Jurisch, B., Thole, M. (1994). Tabu search for the job-shop scheduling problem with multi-purpose machines, *Operations-Research-Spektrum,* Vol. 15, 205-215, doi: 10.1007/BF01719451.

Optimization of reliability and speed of the end-of-line quality inspection of electric motors using machine learning

Mlinarič, J.[a,b,*], **Pregelj, B.**[a], **Boškoski, P.**[a], **Dolanc, G.**[a], **Petrovčič, J.**[a]

[a]Jozef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[b]Jozef Stefan International Postgraduate School, Ljubljana, Slovenia

ABSTRACT

Consistently maintaining high-end product quality in the production process is challenging. End-quality inspection must be highly sensitive to detect even minimal deviations, while being fast and accurate. However, quality inspection systems often face calibration intricacies, are time-consuming, and rely heavily on expert knowledge. They handle substantial data flows and inspect numerous features, some of which contribute minimally to the final grade. To address these challenges, the paper proposes employing statistically supervised machine learning methods for classification. Decision trees, Random forests, Bagging, and Gradient boosting classifiers are recommended for feature selection and accurate diagnosis, particularly for electric motor classification. By utilizing the feature importance attribute for feature selection, the proposed approach compares model accuracies, reducing ramp-up and commission times significantly. The study found that all suggested classifiers achieved high accuracy in classifying electric motors in end-of-line quality inspection system. Moreover, they effectively reduced the number of features and optimize database operations. Utilizing a reduced feature set streamlined diagnostic algorithms, accelerated learning, and improved model interpretability, enhancing overall efficiency and comprehension. Furthermore, analysing the feature importance attribute could simplify diagnostic hardware and expedite quality inspection by eliminating unnecessary steps. Newly generated models can also verify expert decisions on feature selection and limit adjustments, enhancing efficiency in production processes.

1. Introduction

Electric motors are one of the most mass-produced devices and they are produced at highly automated manufacturing lines equipped with 100 % end-of-line (EoL) quality inspection, [1, 2]. Maintaining constant product quality, detecting faulty products and preventing them to be delivered to the customers or further built into devices and systems is highly important. With each production/integration step of faulty part, replacement costs increase substantially [1]. For this reason, a great attention is put to the design and implementation of fully automated EoL quality inspection systems. These must be reliable and at the same time fast enough not to hinder the production line pace. In the presented case, quality inspection of electric motors is performed in a non-invasive way, where several variables are measured during short test run of the motor. The measured variables comprise electric parameters (voltage, current and power), speed, torque, vibrations at different points of motor body and sound at different rotational speeds [1]. The

mentioned signals are sampled by high frequency (e.g. 50 kHz) and further processed by signal processing methods (e.g. Digital Filtering, Fast Fourier Transformation, etc.). This reduces the amount of data considerably while preserving relevant information, but still results in a high number of calculated parameters – features, representing basis for motor fault detection and isolation. High number of features can be impractical from several reasons: a) not all features carry useful information, b) some feature may carry the same information, c) comparing a high number of features against their thresholds values may be time consuming and finally, d) determining (learning) feature threshold values is demanding and time-consuming process.

With an ultimate goal of developing reliable and fast quality inspection methods, this paper deals with the problem of reduction of feature space to a limited subspace of relevant features, carrying enough information for motor quality inspection. The space of features can be reduced by implementing machine learning methods which select only the relevant features. Since not all relevant features contribute the same amount of information to final classification, additionally evaluation of feature selection describes the influence of each observed feature. Therefore, the features with the minimal influence can be eliminated from learning procedure. This can significantly decrease computational demand during learning (ramp-up and commission time) and operation phases. Moreover, in certain cases it can even lead to elimination of particular measurements (sensors), thus simplifying the inspection system hardware and software as well as speeding-up the EoL testing procedure. Study presented in this paper is based on the real industrial data derived from real EoL quality inspection systems installed at the production site of one of renowned European mass producer of electric motors. EoL quality inspection line, which is subject of this paper, was designed and implemented by the authors of the paper.

The paper is organized as follows: In Section 2, the subject of inspection and existing quality inspection procedure/system are briefly described. The structure of the measured data record and resulting feature set generated by the inspection of one motor is described. This is then followed by the Section 3, where machine learning algorithms (Decision tree, Random forest, Bagging and Gradient boosting) for feature selection are presented. In Section 4 the presented algorithms are evaluated and compared.

2. Problem description: Subject of inspection and quality inspection system

The subjects of inspection are brushless DC (BLDC) motors for domestic and automotive applications. An example of such motor used for vacuum cleaning applications, is shown on the Fig. 1. The addressed motors are manufactured by the renowned mass producer (Domel Slovenia, [3]). The production takes place at fully automated assembly line equipped with modular EoL quality inspection system, presented on the Fig. 2. More details can be found in [2], which describes similar system.
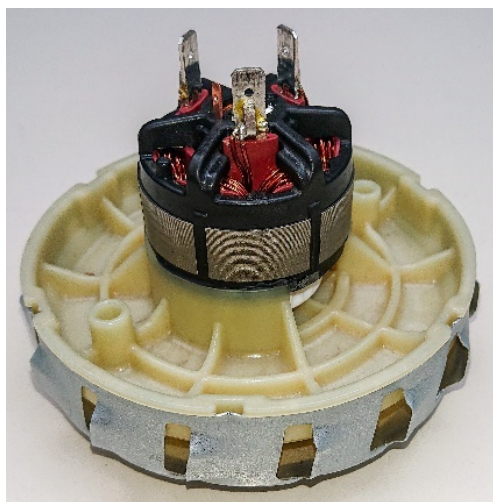
Fig. 1 Example of the BLDC motor (subject of EoL quality inspection)

Fig. 2 Modular EoL quality inspection system

The EoL quality inspection system assures "100 % quality inspection" meaning that each produced motor undergoes the test procedure. In general, the faults fall in two categories: electrical and mechanical. The latter can be further divided to rotor, bearing and turbine faults. Rotor and bearing faults are comprehensively elaborated in [4] whereas the explanation of turbine faults can be found in [1, 2]. The main steps of quality inspection procedure are described in following sections 2.1, 2.2 and 2.3 and in the Fig. 3.

2.1 Measurement and data acquisition

In the EoL quality inspection system, each motor is started several times and during short test runs various motor parameter are measured by the automatic measuring and data acquisition system (Fig. 3, square 1). The following parameters are measured by sensors: electric parameters (winding voltage and current, power, power supply current and voltage), vacuum pressure, rotation speed, vibrations at several points of motor body, sound at low and high rotational speed and also environmental conditions (ambient temperature and pressure) to compensate their effect on motor performance. The results of measurements are time series (waveforms) of particular parameter, and they represent "raw signals". Depending on observed parameter and derived features, signals are acquired at specific sampling frequency (typically 10-60 kHz) and measurement duration (typically from 0.1 up to 1 s), resulting in timeseries of various lengths (from 1000 samples to 30000 samples).

2.2 Feature extraction by signal processing

To reduce the amount of data and to extract the relevant information, raw signals are processed by signal processing methods, such as filtering (low-pass, high-pass, band-pass filters), downsampling, averaging, frequency analysis, etc. The outcome of signal processing is a set of "features", which are detailed in [4] and shown on the Fig. 3 (square 2). They are in general:

- Root-Mean-Square (RMS) values of band-pass-filtered waveforms;
- Power of signals at particular frequencies;
- Aggregated/actual values obtained from specific measurement equipment.

Details of feature extraction and signal processing algorithm are not described in this paper as they are subject of past research and development, elaborated in detail in [1, 5]. In this particular case the signal processing algorithm generates 80 features, where each feature is represented by floating-point numeric value.

2.3 Diagnostic result generation

Based on the on the values of the features, final diagnostic result of the inspected motor is generated by simple rules, as follows from the Table 1. For each feature it is checked if it is inside specified range.

Table 1 Diagnostic result generation

Measurements	Features	Diagnostic result
Completed	All features are within specified ranges	Motor GOOD
	One or more features are outside specified range	Motor BAD
Not completed, due to: • measurement faults • sensor faults • motor manipulation fault • motor transport faults • etc.		UNDEFINED

At the start of the production of new motor type, the motors from test-production set (series 0) are assessed as good or bad by the skilled experts. Based on experiences, the experts select features that are going to be used in diagnostic result generation. For this case-study system and motor type, 36 of total 80 features were chosen, and for each of chosen features two limit values (low and high) are set. In practice, all this is done manually by skilled experts. This is time-consuming and highly depends on expert skills. In addition, this method requires regular updates and fine-tuning of limit values when the mass production starts and production volumes increases [4].

Fig. 3 illustrates the entire procedure of measurement and data acquisition, feature extraction and diagnostic result generation. The whole process for one motor can be executed under 30 s, but due to parallel execution of diagnosis steps, the motor inspection rate is 10 s, which means that every 10 s one motor exists the EoL inspection system.

The described quality inspection algorithm successfully detects motors with insufficient quality, but it has some drawbacks:

- High number of original features (80) leads to high number of feature range limits that must be defined;
- Some features carry similar information (redundancy);
- Some features carry no useful information;
- Skilled expert is required to remove redundant features and features that carry no relevant information and to adjust limit values of feature in use. This is difficult and time consuming and becomes an issue during start of production and commissioning of new motor types.

Based on that there are four main goals of the study presented in this paper:

- Automatic selection relevant of features (removing redundant features and features that do not carry relevant information);
- Decrease the dependence on human expert skills;
- Automatic determination of feature limit values;
- Generating classification models with set of features that hold 95 % of useful information;
- Reducing ramp-up and commission time of the quality inspection system.
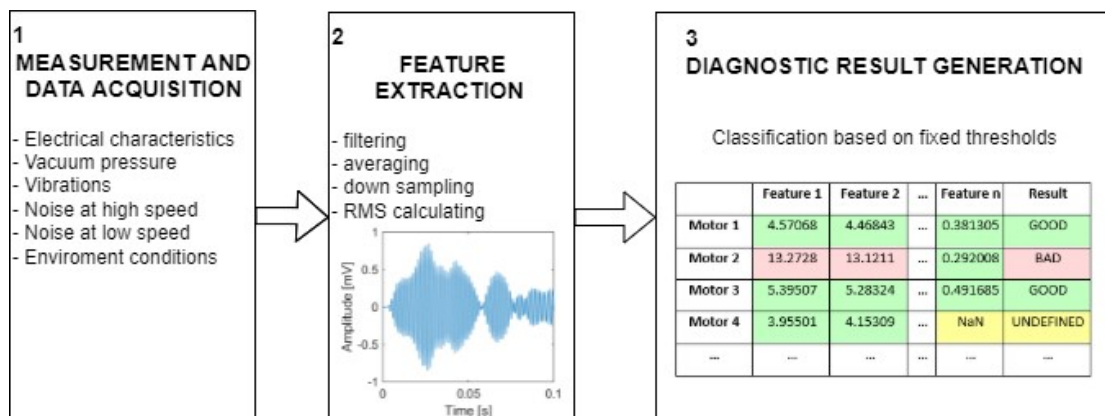
Fig. 3 Data transformation during the procedure

3. Feature selection using machine learning methods

In this section, dedicated machine learning methods will be used to automatically select the relevant features and set their threshold values. In general, feature selection is an effective way to deal with dimensionality [6] and it is often used in areas where a huge amount data is being obtained, such as identifying genes [7-9], image classification and analysing [10-13] text classification [14-16], in recent times also for monitoring manufacturing processes and quality control [17-19]. Feature selection aims to identify and retain the most relevant features while discarding redundant or uninformative ones by determining the "degree of usefulness" of a specific feature. By reducing the number of features, also the number feature limit values to be adjusted is reduced. The risk of the associated quality inspection errors [6] is also decreased. However, within the set of informative features, some may be significantly more informative than others.

The methods presented not only eliminate non-informative features but also sort the remaining features according to their informativeness. The goal of this paper is therefore to assess whether the quality inspection can be successfully performed by using only the limited number of the most informative features. Feature selection reduces the dimensionality of the data; therefore, data mining algorithms can be operated faster and more efficient [6]. Reduced amount of input data simplifies the interpretability of tree-like machine learning methods [20]. Additionally, such simplified classification methods and reduced input datasets decrease the ramp-up and commission time of quality inspection system and whole production line. Since some redundant and non-informative features are removed, sensors associated with removed features can potentially be eliminated. Optimization and reordering of the diagnostic steps based on feature importance can speed-up the quality inspection procedure.

3.1 Supervised machine learning classification methods

Supervised machine learning methods were selected since the labelled data for learning is available. These methods offer several advantages, including high reliability based on statistics, robustness, and reduction of the need for expert knowledge (e.g. knowledge about physical background of the system). However, in order to establish a supervised learning method, a sufficient amount of data from the production process is required. Therefore, these methods are suitable for manufacturing lines for mass production, like the one presented in this paper, where a lot of data is generated. In this paper 4 different methods were tested and compared:

- Decision tree classifier (DT);
- Random forest classifier (RF);
- Bagging classifier (BG);
- Gradient boosting classifier (GB).

The decision tree classifier partitions the instance space through a recursive process, forming a tree model where top nodes (roots) lack incoming edges, while internal nodes (test nodes) split the space based on attribute values. Internal nodes symbolize decision points, and bottom nodes (leaves) indicate decision outcomes [21, 22].

Random forests is a powerful ensemble learning method that combines multiple tree predictors. It belongs to family of averaging methods, meaning, the driving principle is to build several estimators independently and then to average their predictions [23-25].

Bagging, short for bootstrap aggregating, is a powerful and straightforward method for constructing an ensemble of classifiers. It also belongs to family of averaging methods. It combines multiple classifiers' outputs for improved accuracy by training each on a subset of instances randomly drawn from the training set [22, 26, 27].

Gradient boosting is a technique for improving the performance of weak learners [28]. It belongs to the family of boosting methods, meaning, base estimators are built sequentially, and one tries to reduce the bias of the combined estimator. It enhances weak learners sequentially, aiming to reduce the combined estimator's bias. The technique combines weak models for a powerful ensemble, particularly effective in decision trees [22, 25].

While the decision tree classifier (DT) stands alone, the remaining three classifiers (RF, BA,

GB) belong to the ensemble-based methods category. All methods here are "tree-like" classifiers and can be represented as decision trees using IF-THEN rules. The methods automatically generate IF-THEN rules for classification, including threshold values for each observed feature.

3.2 Data for learning and evaluation

To implement and test the methods, the data is needed. Data was generated by existing EoL quality inspection system (Fig. 2) during inspection of a total of 37440 motors. Generated data contains raw time series of measured signals and extracted features (mentioned 80 features), as follows from the Section 2. For machine learning algorithms, all features are used. The whole data set of features can be presented as 37440 × 80 matrix. The quality inspection results (1=Motor GOOD, 2 = Motor BAD, 0 = UNDEFINED) represents 37440 × 1 vector. The entire data set was divided into two parts: training data (75 % of all data) and testing data (25 % of all data). The situation is represented by the Table 2.

The data records of all 37440 motors were randomly distributed between training data and testing data to compensate for possible time drift of product quality. The same training and testing datasets were used during the training and testing of all 4 machine learning methods. Dataset remained unchanged throughout the entire training process. All involved features are named by symbols and anonymized to prevent the disclosure of sensitive technical data.

Table 2 Training and testing data arrangement

	F1	F2	...	F80	R	
M1	X	X	X	X	X	
M2	X	X	X	X	X	Training data
...	X	X	X	X	X	(75 %)
M28080	X	X	X	X	X	
M28081	X	X	X	X	X	
...	X	X	X	X	X	Testing data
M37440	X	X	X	X	X	(25 %)

M – motors, F – features, R – quality inspection results

3.3 Implementation

Fig. 4 illustrates the proposed process of feature selection.
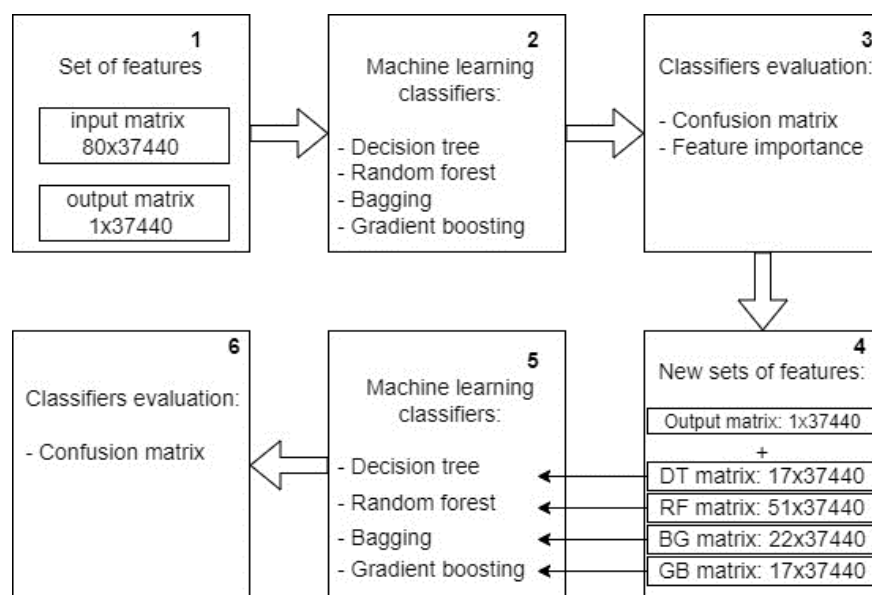
Fig. 4 Flow chart of the procedure

The process is carried-out in six steps:

1. Generation of original feature set (an input matrix of size 80 features × 37440 motors and output matrix of size one feature – grade × 37440 motors) obtained by data acquisition and signal processing, described in Section 2.
2. First machine learning to train the classifiers. Feature selection is utilized to eliminate redundant features. Depending on the type of classifier, the range of selected features is reduced to 38 (DT classifier), 78 (RF classifier), 42 (BG classifier), and 35 (GB classifier).
3. Evaluation of trained classifiers to present their performance and capability to evaluate features' classification impact and importance. From obtained results it followed:
    a. Set of features with 95 % informativeness additionally reduce space of features. Depending on the type of classifier, the range of selected features is additionally reduced to 17 (DT classifier), 51 (RF classifier), 22 (BG classifier), and 17 (GB classifier).
    b. Certain features, despite their low importance, still persist.
       Therefore, it was decided to check the performance of reduced classifiers with features that contains 95 % of all useful information.
4. Generation of reduced sets of training and testing data with features with 95 % of useful information for each classifier, output matrix remains the same from step 1.
5. Second machine learning to train the classifiers with new reduced dataset.
6. Evaluation of new classifiers and comparison to the results of classifiers from step 2.

All classification methods were implemented and tested in Python using the *scikit-learn* (sklearn) library. Cross-validation was employed during training (step 2 and step 5) to ensure robust model evaluation and to prevent model overfitting. For visualization and data manipulation, *matplotlib*, *numpy*, and *pandas* libraries were also utilized.

3.4 Presentation of the results and comparison of the methods

Following the training phases, the trained algorithms were evaluated using the testing data, and the predicted output classes (quality inspection result) were compared to the actual output classes. For each method, outcomes are presented in the form of well-known Confusion Matrix (CM). The CM provides numerical and visual representation of the classification algorithm's accuracy. It consists of columns representing the predicted output classes and rows representing the actual output classes. In the presented case, since there are three classes, the size of the CM is 3 × 3 as shown in the Table 3. The diagonal elements represent correctly classified instances, while the off-diagonal elements represent miss-classified instances.

Table 3 Confusion matrix structure

		PREDICTED		
		UNDEFINED	Motor BAD	Motor GOOD
ACTUAL	UNDEFINED	▩		
	Motor BAD		▩	
	Motor GOOD			▩

The CM provides valuable information about the miss-classification, but it does not directly capture the cost associated with each type of miss-classification. To address this, the Miss-classification cost matrix can be introduced. This matrix assigns specific costs to different types of miss-classifications, considering the relative importance or impact of miss-classifying different classes. The Miss-classification cost matrix has the same dimension as the Confusion matrix and consists

of numerical values, as shown in the Table 4. The diagonal elements of the Miss-classification cost matrix are set to zero (correct classification represent no cost). The off-diagonal elements represent the costs associated with miss-classified instances, with larger values indicating higher risks or costs associated with miss-classification. In our case the highest cost is associated with situation when actual faulty or undefined motor is recognized as good one and delivered to the customer (actual class BAD or UNDEFINED, predicted class GOOD). The cost of this miss-classification is set to 10. Bad motor that is predicted as undefined (and the opposite) presents low cost of miss-classification, therefore these cases are graded with 0.5. Good motor, ranked as bad or undefined, does not present any risk for the customer, but it represents an unnecessary waste of motors, therefore it is marked with 1.

To calculate the overall miss-classification cost, the CM the Miss-classification cost matrix are multiplied (element-by-element multiplication) resulting in new 3 × 3 matrix. Total cost is calculated by adding up all 9 elements of resulting matrix. This value provides a measure of the total cost incurred due to miss-classification. Ideally, a well-performing classifier would have a miss-classification cost close to zero, indicating minimal miss-classification and associated risks.

The accuracy of a classifier is a measure of its performance and is calculated as the ratio between the number of correctly classified elements and the total number of elements. The desired accuracy is close to 1 (100 %), indicating that almost all elements were classified correctly.

Table 4 Cost matrix

		PREDICTED		
		UNDEFINED	Motor BAD	Motor GOOD
ACTUAL	UNDEFINED	0	0.5	10
	Motor BAD	0.5	0	10
	Motor GOOD	1	1	0

4. Analysis of the results

During machine learning, all methods generated own IF-THEN rules for classification. In 3.1 it is explained that chosen methods are "tree-like" and can be explained with IF-THEN rules. In the Fig. 5, an example of the decision tree of the decision tree classifier is presented. The figure shows a diagnosis procedure with a tree-like set of rules and leaves. Since the generated decision tree is very extensive, one rule for the first branch is explained. At the enlarged part of the figure, the auto generated threshold value for one particular feature (BE_H2) is illustrated. At enlarged part the gini index value [6, 21, 29] is also illustrated and used as splitting criteria [6]. Further, at this branch 27064 samples are involved in classification process, where 600 of them are marked as UNDEFINED, 1355 as Motor BAD and 25109 as Motor GOOD.

Each of the methods generates similar decision tree scheme where rules are defined by observed features and their threshold values. Threshold values are set automatically during machine learning and since they are presented as real values. They can be easily checked and readjusted. Such examination of the decision tree structure of each classifier enhances comprehension of the decision-making process employed by each method. This understanding is particularly valuable in industrial applications where interpretability is paramount, as it allows users to gain insights and interpret the decision-making process with clarity.
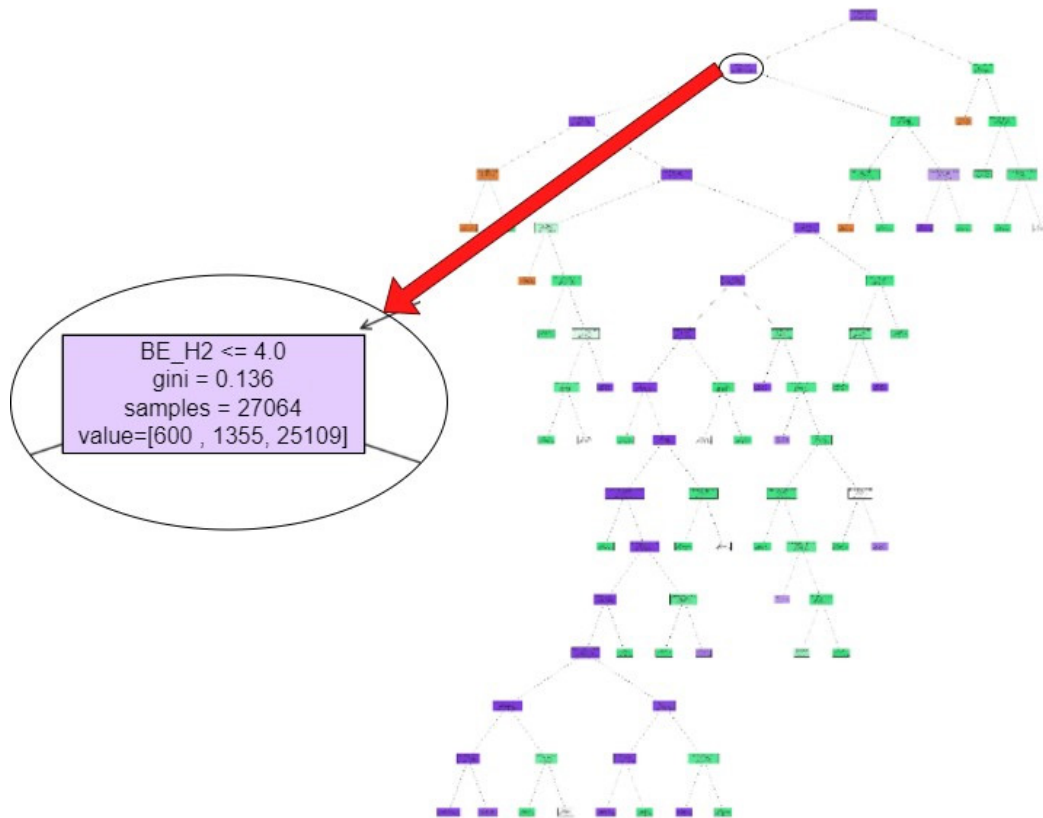
Fig. 5 Decision tree of the decision tree classifier

Results of feature selection are collected in Table 5, where the comparison between classifiers is presented. The results show that all methods successfully executed feature selection in the step 2 (Fig. 4). The *gradient boosting* classifier selected the lowest number of important features, meaning, this method requires the smallest amount of data for successful classification. The least successful method here is *random forest*.

However, the best performance shows the classifier with the lowest number of wrong classified motors and with the lowest miss-classification cost. Therefore, *bagging* overperformed all other classifiers and *gradient boosting* yield the worse performance. Fig. 6, Fig. 8, Fig. 10 and Fig. 12 present confusion matrices for each classifier, generated at step 2.

Table 5 Comparison of observed classifiers

	Decision tree	Random forest	Bagging	Gradient boosting
No. of selected features with original dataset after step 2	38	78	42	35
Number of features for 95 % informativeness	17	51	22	17
Influence of 10 most important features	92 %	69 %	88.6 %	91.4 %
Accuracy of classifiers with original dataset (step 3)	99.2 %	99.33 %	99.46 %	99.19 %
Accuracy of classifiers with reduced dataset (step 6)	99.16 %	99.33 %	99.46 %	99.16 %
No. of wrong classified motors of classifiers with original dataset (step 3)	75	53	51	76
No. of wrong classified motors of classifiers with reduced dataset (step 6)	79	53	51	79
Miss-classification cost of classifiers with original dataset (step 3)	426	477	312	679
Miss-classification cost of classifiers with reduced dataset (step 6)	430	468	348	637

Further insights into feature selection across different classifiers show that the most influential features hold the majority of information, useful for classification. Third row in the Table 5 shows a part of information contained in the 10 most important features of each method. In all methods (except RF) top ten features contain the majority of information and to achieve 95 % of data informativeness, 17-51 features are required. This comprehensive evaluation provides insights into the strength and weakness of each classifier, considering both accuracy and miss-classification cost. At the Table 6, where the 10 most influential features for each classifier in the order of importance are listed, it is shown that all classifiers recognize the majority of features as important (Figs. 14-17 visually illustrate the informativeness of each observed classifier for 10 most important features). At the Table 7 all features involved in machine learning are listed and in columns each of them is marked whenever it appears to be important for each classifier. With gravy are coloured rows of features that are important for all observed classifiers.

These findings can be highly beneficial for tuning classification parameters. Instead of adjusting the thresholds for all observed features, only the most influential features need to be addressed. Additionally, since some features do not contribute to the final classification decision, they do not need to be stored in the company database. This results in the reduced computational burden by focusing only on the most informative features, a smaller data flow between the local computer and the company database, reduces the potential for communication errors, and ultimately requires less storage space.

Table 6 List of 10 most important features for each classifier

	Decision tree	Random forest	Bagging	Gradient boost
1. important feature	BE_H1	BE_H2	BE_H1	BE_H1
2. important feature	BE_H2	BE_H1	BE_H2	BE_H2
3. important feature	BE_H3	BE_H3	BE_H5	VRC
4. important feature	VA	BE_H5	VRC	VA
5. important feature	VRC	BE_H4	VA	VA_H1
6. important feature	HW_H1S	BE_H6	HW_H1S	HW_H1S
7. important feature	VRL_H2	VRC	VA_H1	BE_H3
8. important feature	HW_H6	VA	BE_H3	BE_H6
9. important feature	AVR_V	HW_H1S	BE_H6	VRL_H2
10. important feature	VW_V	VA_H1	HW_H6	FR_H2

Table 7 All features involved in machine learning and their recognition as important for each classifier

Feature name	DT	RF	BG	GB
VA	X	X	X	X
VA_H1		X	X	X
VA_H2		X		
VA_H3		X		
VA_H4		X		
VA_H5		X		
VA_H6	X	X		
VA_H7		X		
VA_H8		X		
VA_H9		X		X
VA_H10		X		
VA_H11		X		
VA_H12		X		X
VA_H13	X	X		
VA_H14		X		
VA_H15	X	X		
AVR_U		X	X	X
AVR_V	X	X	X	X
AVR_W		X		
BM_U			X	
BM_V	X	X	X	X
BM_W	X	X		
FR_H1	X	X	X	X
FR_H2	X	X	X	X
BE_H0				
BE_H1	X	X	X	X
BE_H2	X	X	X	X
BE_H3	X	X	X	X
BE_H4	X	X	X	X
BE_H5	X	X	X	X
BE_H6	X	X	X	X
CP	X	X	X	
VW_U		X	X	X
VW_V	X	X	X	
VW_W	X	X	X	X
VRC	X	X	X	X
VRC_H1		X		
VRC_H2	X	X		
VRC_H3		X		
VRC_H4	X	X	X	

VRC_H5		X		X
VRC_H6		X		
VRC_H7		X		
VRC_H8		X		
VRC_H9	X	X	X	
VRC_H10		X		
VRC_H11		X		X
VRC_H12	X	X		
VRC_H13		X	X	X
VRC_H14		X		X
VRC_H15		X	X	X
VRL		X		
VRL_H1		X	X	
VRL_H2	X	X		X
VRL_H3		X		
VRL_H4		X		
VRL_H5		X	X	X
VRL_H6	X	X	X	X
VRL_H7		X	X	
VRL_H8		X	X	

VRL_H9	X	X	X	X
VRL_H10		X	X	X
VRL_H11		X		
VRL_H12		X	X	
VRL_H13	X	X		
VRL_H14		X		
VRL_H15		X		
MV	X	X	X	X
CW_U	X	X		
CW_V	X	X	X	
CW_W	X	X	X	
HW_H1S	X	X	X	X
HW_H18	X	X	X	X
HW_H2	X	X	X	X
HW_H2S	X	X	X	X
HW_H3	X	X	X	X
HW_H4	X	X	X	X
HW_H6	X	X	X	X
HW_H9		X	X	
REV				

At last, the performance of classifiers, trained with features of 95 % information (step 5), is evaluated (step 6). The result in Table 5 shows that the accuracies and performances of classifiers do not change a lot. However, despite the RF and GB classifier, the cost of miss-classification increased, meaning, those classifiers miss-classified performed slightly worse. (as shown at Fig 6-13). Despite minor fluctuations in accuracy, the overall performance remains relatively stable what indicate the robustness of chosen methods. However, the increase in miss-classification costs for certain classifiers indicates potential areas for optimization in future iterations.
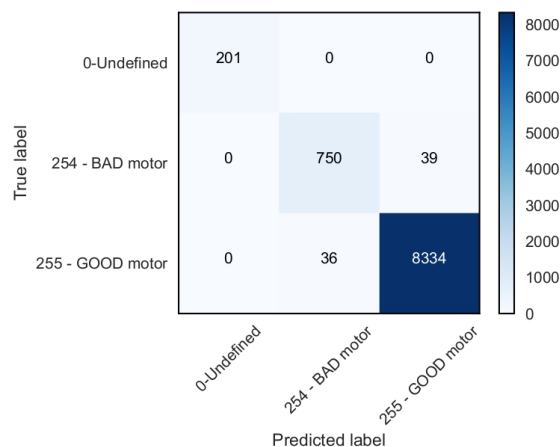
Fig. 6 Confusion matrix for decision tree classifier (38 features)
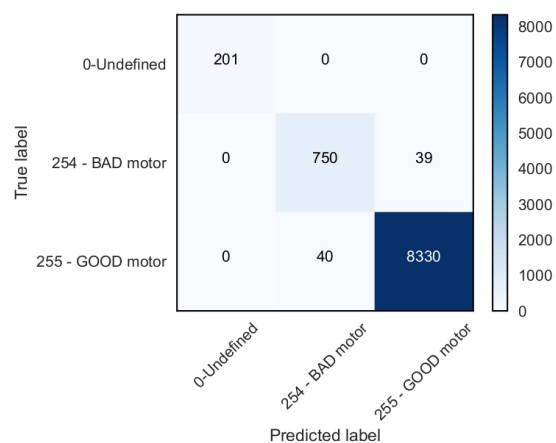
Fig. 7 Confusion matrix for decision tree classifier (17 features)

Fig. 8 Confusion matrix for random forest classifier (78 features)

Fig. 9 Confusion matrix random forest classifier (51 features)

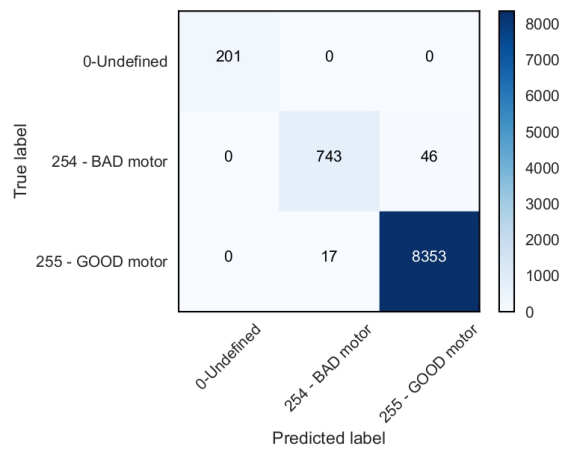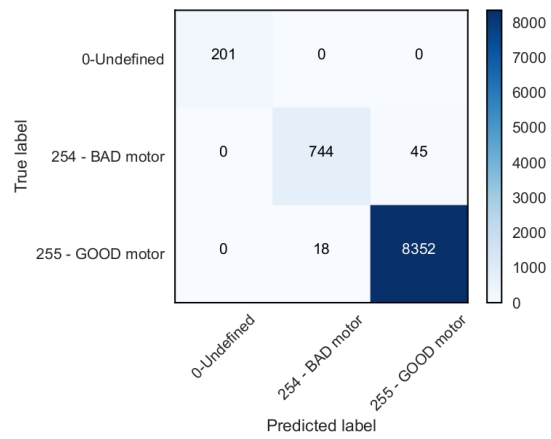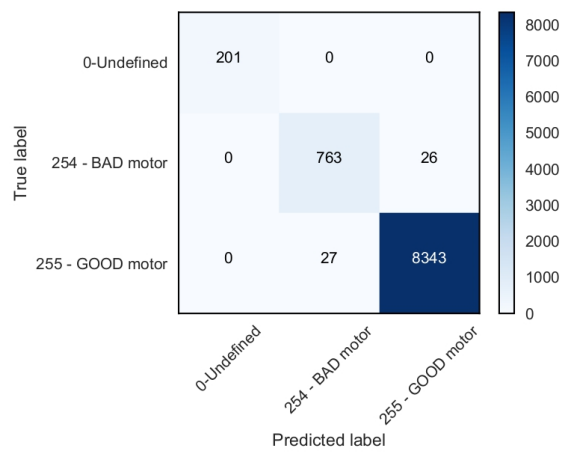**Fig. 10** Confusion matrix for bagging classifier (42 features)
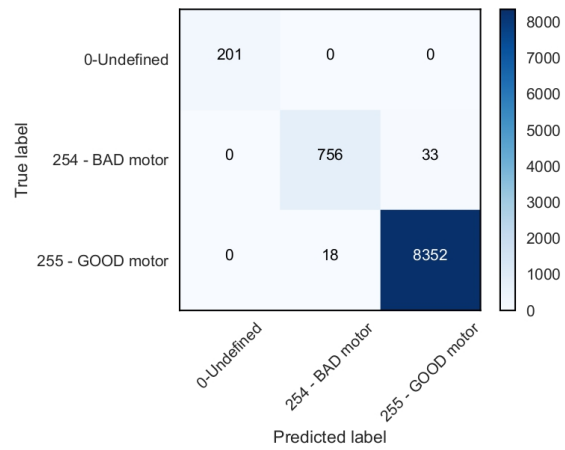
Fig. 11 Confusion matrix for bagging classifier (22 features)
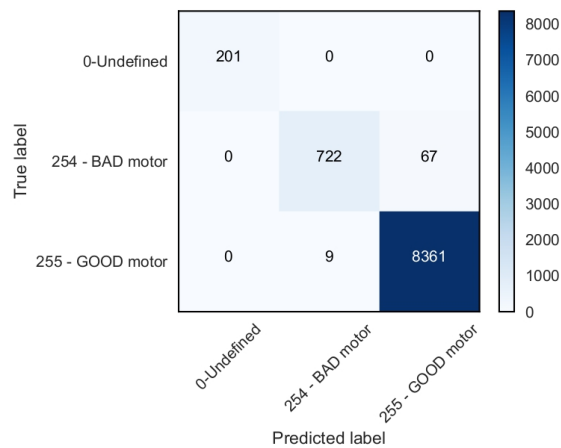
Fig. 12 Confusion matrix for gradient boost classifier (35 features)
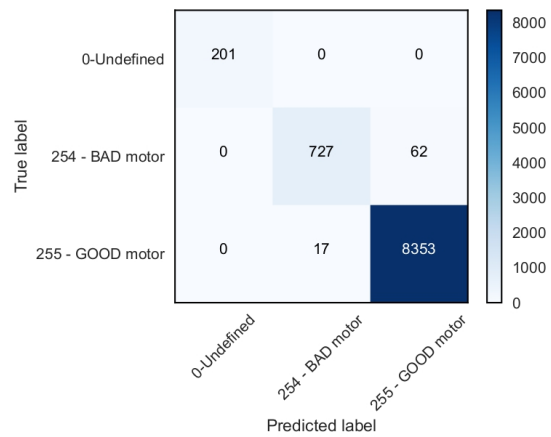
Fig. 13 Confusion matrix for gradient boost classifier (17 features)

Feature importance for Decision tree classifier
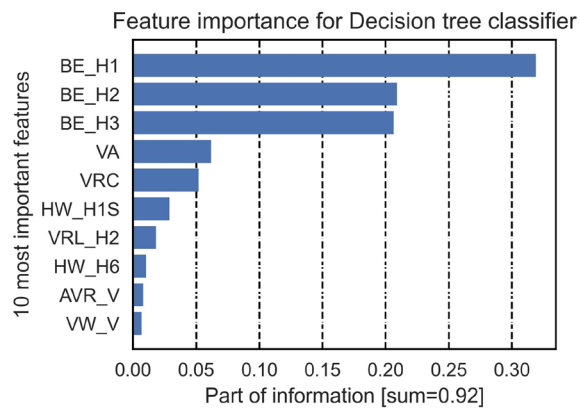
Fig. 14 Feature importance for decision tree classifier

Feature importance for Random forest classifier
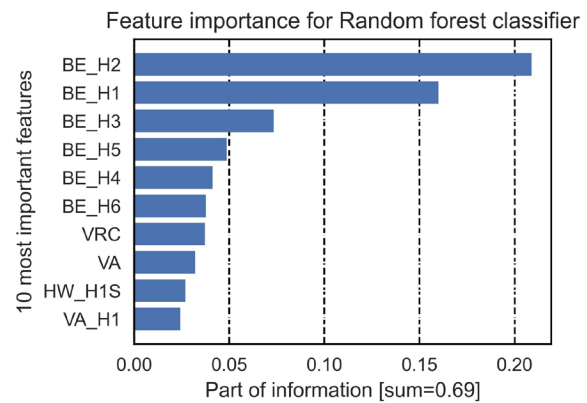
Fig. 15 Feature importance for random forest classifier

Feature importance for Bagging classifier

Fig. 16 Feature importance for bagging classifier
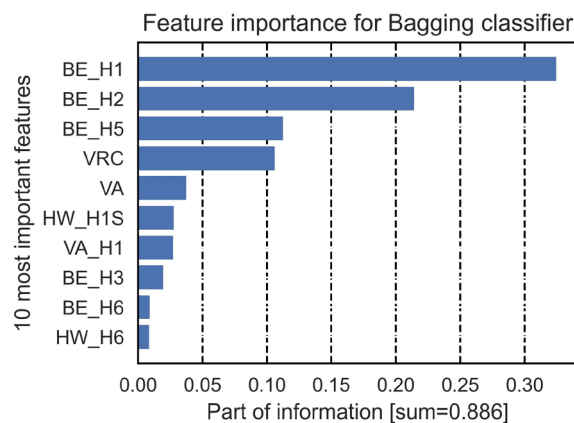
Feature importance for Gradient boost classifier
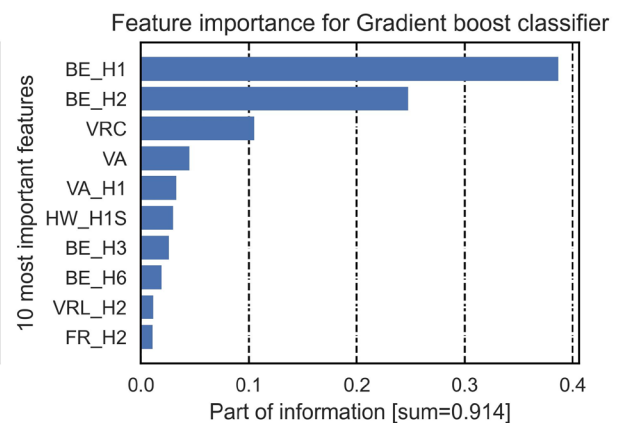
Fig. 17 Feature importance for gradient boost classifier

5. Conclusion

This study introduced and compared various classifiers for feature selection purposes used for automated end-of-line quality inspection of electric motors within the real manufacturing line. Decision tree, Random forest, Bagging, and Gradient boosting classifiers were implemented and assessed based on their complexity (number of selected features), accuracy, and the impact of the important features. Initial goal of the study was achieved successfully. All four tested classifiers demonstrated high accuracy, proving their suitability for electric motor classification in an end-of-line quality inspection system. All investigated classifiers successfully reduced the number of features and thus optimized the database operation. Further, the second step of feature selection, with a reduced dataset featuring features that hold 95 % of useful information, yielded high accuracy of trained classifiers. This reduced feature set simplifies the diagnostic algorithm, speeds-up its' learning, improves the interpretability of the observed models and makes them more understandable and explainable. In addition, new classification models, learned with reduced dataset, simplify the end-of-line quality inspection, decrease the ramp-up and commission time, eliminate unnecessary steps of diagnosis, reduce equipment complexity (in some cases eliminate the need for particular sensors), reduce costs, and minimize data flow. Consequently, company databases are optimized. Due to fully automated learning procedures, reliance on specialized experts is reduced. Developed classification models can also be used as a verification of experts' decision regarding feature selection and threshold values adjustment. In summary, this study encompasses insights into feature selection, practical implications for industrial applications considering methods robustness and comprehensive evaluation of different classifier, considering accuracy and miss-classification cost, aiding in decision making when selecting the most suitable classifier for specific application.

During research and implementation, two interesting and useful future topics of research were identified: 1. Transferability of classification models, and 2. Condition monitoring of production lines.

Within the first topic it could be investigated if classification models derived for one motor type could be used for quality inspection of similar motor types, or if they could increase learning procedure of new motor types. Based on the manufacturer's experiences, the diagnosis procedure across various manufacturing lines follows a similar approach, leading to the detection of similar faults and malfunctions across different products. Furthermore, common features are identified, suggesting that the feature selection process from observed classifiers of a particular motor type could be applied to other product. This transferability would be beneficial especially for small-series products. Some motor types are manufactured in small-series (e.g. up to thousand pieces per year), therefore it is challenging to establish accurate classification model with limited amount of learning data. Exploring the applicability of these findings across various motor types can speed-up the creation of quality inspection algorithms for new motor types in the future. As new motor types are developed frequently and produced in varying quantities, the transferability of the methods can establish a standardized approach to implement quality inspection algorithms for new motor types, reducing costs and thus enhancing the whole manufacturing process.

Second topics regards possibilities of condition monitoring of production line. In normal conditions (when there is no degradation of the manufacturing process) features importance attributes do not significantly change with time. On the other hand, an increase of importance of particular feature may indicate the issue of particular manufacturing operation or an issue of input material or components. Periodic evaluation of feature importance attributes can therefore help to detect faults or degradations of various steps of manufacturing process or issues with input materials and components.

Acknowledgments

References

[1] Juričić, Đ., Petrovčič, J., Benko, U., Musizza, B., Dolanc, G., Boškoski, P., Petelin, D. (2013). End-quality control in the manufacturing of electrical motors, In: Strmčnik, S., Juričić, Đ. (eds.), *Case studies in control, Advances in industrial control,* Springer, London, United Kingdom, 221-256, doi: 10.1007/978-1-4471-5176-0_8.

[2] Benko, U., Petrovčič, J., Mussiza, B., Juričić, Đ. (2008). A system for automated final quality assessment in the manufacturing of vacuum cleaner motors, *IFAC Proceedings Volumes,* Vol. 41, No. 2, 7399-7404, doi: 10.3182/20080706-5-KR-1001.01251.

[3] Domel, Domel d.o.o., from *https://www.domel.com/sl,* accessed September 26, 2023.

[4] Boškoski, P., Petrovčič, J., Musizza, B., Juričić, Đ. (2011). An end-quality assessment system for electronically commutated motors based on evidential reasoning, *Expert Systems with Applications,* Vol. 38, No. 11, 13816-13826, doi: 10.1016/j.eswa.2011.04.185.

[5] Benko, U., Petrovčič, J., Juričić, Đ. (2005). In-depth fault diagnosis of small universal motors based on acoustic analysis, *IFAC Proceedings Volumes,* Vol. 38, No. 1, 323-328, doi: 10.3182/20050703-6-cz-1902.01856.

[6] Rokach, L., Maimon, O. (2014). *Data mining with decision trees, Theory and applications,* 2nd Edition, World Scientific, New Jersey, USA, doi: 10.1142/9097.

[7] Kim, S., Xing, E.P. (2009). Statistical estimation of correlated genome associations to a quantitative trait network, *PLOS Genetics,* Vol. 5, No. 8, Article No. e1000587, doi: 10.1371/journal.pgen.1000587.

[8] Beisvag, V., Jünge, F.K.F., Bergum, H., Jølsum, L., Lydersen, S., Günther, C.-C., Ramampiaro, H., Langaas, M., Sandvik, A.K., Lægreid, A. (2006). GeneTools - application for functional annotation and statistical hypothesis testing, *BMC Bioinformatics,* Vol. 7, No. 1, Article No. 470, doi: 10.1186/1471-2105-7-470.

[9] Kuehl, P.M., Weisemann, J.M., Touchman, J.W., Green, E.D., Boguski, M.S. (1999). An effective approach for analyzing "prefinished" genomic sequence data, *Genome Research,* Vol. 9, No. 2, 189-194, doi: 10.1101/gr.9.2.189.

[10] Núñez, J., Llacer, J. (2003). Astronomical image segmentation by self-organizing neural networks and wavelets, *Neural Networks,* Vol. 16, No. 3-4, 411-417, doi: 10.1016/s0893-6080(03)00011-x.

[11] Chen, E.-L., Chung, P.-C., Chen, C.-L., Tsai, H.-M., Chang, C.-I. (1998). An automatic diagnostic system for CT liver image classification, *IEEE Transactions on Biomedical Engineering,* Vol. 45, No. 6, 783-794, doi: 10.1109/10.678613.

[12] Makadia, A., Pavlovic, V., Kumar, S. (2008). A new baseline for image annotation, In: Forsyth, D., Torr, P., Zisserman, A. (eds.), *Computer Vision – ECCV 2008. ECCV 2008. Lecture notes in computer science,* Vol. 5304, Springer, Berlin, Heidelberg, Germany, doi: 10.1007/978-3-540-88690-7_24.

[13] Chen, Z.-Y., Lin, W.-C., Ke, S.-W., Tsai, C.-F. (2015). Evolutionary feature and instance selection for traffic sign recognition, *Computers in Industry,* Vol. 74, 201-211, doi: 10.1016/j.compind.2015.08.007.

[14] Deng, X., Li, Y., Weng, J., Zhang, J. (2019). Feature selection for text classification: A review, *Multimedia Tools and Applications,* Vol. 78, No. 3, 3797-3816, doi: 10.1007/s11042-018-6083-5.

[15] Baccianella, S., Esuli, A., Sebastiani, F. (2014). Feature selection for ordinal text classification, *Neural Computation,* Vol. 26, No. 3, 557-591, doi: 10.1162/NECO_a_00558.

[16] Baecchi, C., Uricchio, T., Bertini, M., Del Bimbo, A. (2016). A multimodal feature learning approach for sentiment analysis of social network multimedia, *Multimedia Tools and Applications,* Vol. 75, No. 5, 2507-2525, doi: 10.1007/s11042-015-2646-x.

[17] Chen, J., Wang, T., Gao, X., Wei, L. (2018). Real-time monitoring of high-power disk laser welding based on support vector machine, *Computers in Industry,* Vol. 94, 75-81, doi: 10.1016/j.compind.2017.10.003.

[18] Guinea, D., Ruiz, A., Barrios, L.J. (1991). Multi-sensor integration—An automatic feature selection and state identification methodology for tool wear estimation, *Computers in Industry,* Vol. 17, No. 2-3, 121-130, doi: 10.1016/0166-3615(91)90025-5.

[19] Hidalgo-Mompeán, F., Gómez Fernández, J.F., Cerruela-García, G., Márquez, A.C. (2021). Dimensionality analysis in machine learning failure detection models. A case study with LNG compressors, *Computers in Industry*, Vol. 128, Article No. 103434, doi: 10.1016/j.compind.2021.103434.

[20] EdwardI, G., Foster, D.P. (2000). Calibration and empirical Bayes variable selection, *Biometrika,* Vol. 87, No. 4, 731-747, doi: 10.1093/biomet/87.4.731.

[21] Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J. (1984). *Classification and regression trees,* 1st Edition, Chapman and Hall/CRC, New York, USA, doi: 10.1201/9781315139470.

[22] Rokach, L. (2010). *Pattern classification using ensemble methods,* World Scientific, Singapore, doi: 10.1142/7238.

[23] Breiman, L. (2001). Random forest, *Machine learning,* Vol. 45, No. 1, 5-32, doi: 10.1023/A:1010933404324.

[24] Breiman, L. (1998). Arcing classifier (with discussion and a rejoinder by the author), *Annals of Statistics,* Vol. 26, No. 3, 801-849, doi: 10.1214/aos/1024691079.

[25] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The elements of statistical learning, Data mining, inference, and prediction,* Second Edition, Springer, New York, USA, doi: 10.1007/978-0-387-84858-7.

[26] Breiman, L. (1996). Bagging predictors, *Machine Learning,* Vol. 24, No. 2, 123-140, doi: 10.1007/bf00058655.

[27] Breiman, L. (1999). Pasting small votes for classification in large databases and on-line, *Machine Learning,* Vol. 36, No. 1, 85-103, doi: 10.1023/A:1007563306331.

[28] Okun, O. (2011). *Feature selection and ensemble methods for bioinformatics: Algoritmic classification and implementations,* IGI Global, Hershey, Pennsylvania, USA, doi: 10.5555/2050025.

[29] Gelfand, S.B., Ravishankar, C.S., Delp, E.J. (1991). An iterative growing and pruning algorithm for classification tree design, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 2, 163-174, doi: 10.1109/34.67645.

[30] Igual, L., Segui, S. (2017). *Introduction to data science; A Python approach to concepts, Techniques and applications,* Springer, Cham, Switzerland.

[31] Carletti, M., Masiero, C., Beghi, A., Susto, G.A. (2019). Explainable machine learning in Industry 4.0: Evaluating feature importance in anomaly detection to enable root cause analysis, In: *Proceedings of 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC),* Bari, Italy, 21-26, doi: 10.1109/SMC.2019.8913901.

A modified bi-objective NSGA-II approach to sustainability in reconfiguration planning of dynamic cellular manufacturing systems

Sibanda, M.M.[a,*], Padayachee, J.[a]

[a]School of Mechanical Engineering, Howard college, University of Kwazulu-Natal, Durban, South Africa

A B S T R A C T

Manufacturing plant layouts are developed to facilitate optimal process flow. Modern manufacturing systems must meet present production demands and be adaptable to changes in process flow in the future. Dynamic Cellular Manufacturing Systems (DCMS) increase the flexibility of layouts by reconfiguring cell structure and equipment distribution, to effectively adjust part routings for optimal process flow. Frequent reconfiguring of plant layout may not always be feasible or economical, however, when new product releases are planned, reconfiguring the plant layout to optimise the workflow may be extremely beneficial. This paper presents a Non-dominated Sorting Genetic Algorithm (NSGA-II) approach to solving a DCMS problem in a sustainable, and responsible manner. A bi-objective integer programming model was developed over multiple planning horizons with fluctuating product demands. This model aims to achieve sustainability by reducing the cost of production, mitigating the environmental impact of production, and minimise negative social impacts on labourers that work in such environments. A penalty function approach was used to enforce the model constraints during optimisation. This study details trade-offs between the economic factors of a DCMS, the environmental implications of reconfiguring such a system, and the social impacts of reconfigurations on the workforce.

1. Introduction

Cellular Manufacturing Systems (CMS) are a widely adopted method of manufacturing. The first step in developing a CMS is the process of Cell Formation (CF), followed by layout design [1]. CF is composed of two primary tasks: machine-cell allocation and part-family formation. Machine-cell allocation involves optimally grouping machines within cells. Part-family formation groups parts to be produced according to their production requirements, size, shape, or other geometric characteristics. Layout design also has two main tasks, namely intercell and intracell layout design. Batch manufacturing is used for several reasons including, but not limited to, efficient production, product tracking, short production cycles, short delivery time and volume constraints. It accounts for a significant fraction of all production as it helps manufactures cope with a wide variety of small manufacturing lot sizes. As the demand for these small lot sizes vary over time, CMS must be reconfigured to ensure that the current layout stays relevant and optimal for each

demand period. Although a layout may be optimal for a particular demand period, when the demand changes that system may not be optimal for the new demand period. This would result in bottlenecks and inefficient production.

According to Lokesh and Jain [1] about 75 % of the revenue of Hewlett Packard was from products that were manufactured less than 3 years prior. This highlights a trend of short product lifecycles for many manufacturers today. A study from Defersha and Chen [2] revealed that considerable financial resources are spent annually on reconfiguring the layouts of production plants to improve the optimality of CMS. These great efforts are regularly made because part and material handling costs account for 20-50 % of production costs, suggesting that CMS need to be operated with optimum configurations to reduce production costs. They further suggested that these production costs could be reduced by 10-30 % by dynamically changing the layout of CMS.

In keeping with the world's Sustainable Development Goals (SDG), particularly goal 8 and 12, manufacturers may need ways to comply with new or stricter sustainability regulations [3]. Sustainability is an increasingly important consideration for manufacturers, as it involves the ability to produce goods and services in ways that meet present needs without compromising the ability of future generations to meet their own needs. The three pillars of sustainable manufacturing are economic limitations, environmental constraints, and social influences [3]. We acknowledge that sustainability can be measured in many ways, for our application we will consider carbon emissions as a general, all-inclusive measure of environmental sustainability, and the workload on machine operators as our limiting metric for social responsibility.

The motivation for this work stems from the need to address sustainability and social responsibility in manufacturing operations. In today's competitive market, cost optimisation is a key business concern, hence, this aspect of manufacturing is, and always has been, the primary focus in most research. However, we recognise that single cost objective models could have adverse environmental impacts, hence, incorporating environmental sustainability as a second objective in the proposed model provides a suitable trade-off to these related aspects of manufacturing. Introducing environmental considerations could lead to solutions with a lower carbon footprint, reduced energy consumption and waste generation. Furthermore, we understand that the implementation of DCMS can have significant social implications on the workforce. By incorporating social constraints into a DCMS model, researchers and organizations can identify solutions that minimise negative impacts on the well-being of employees; promote worker satisfaction, improve safety, and facilitate a smooth transition between DCMS periods. Ultimately, this research contributes to a more sustainable future for the manufacturing sector as a whole.

This paper is organised into seven sections and presents a bi-objective mathematical model that utilizes a penalty approach to minimise the operating costs and negative environmental impacts of a DCMS. A unique NSGA-II heuristic is used for the optimisation. The permissible social impact on the work force is introduced in the model through a set of inequality constraints. The remainder of the paper is organised as follows: section 2 presents a literature review while section 3 details the problem description. The mathematical model is described in section 4, and section 5 presents the solution approach. In section 6 we present and discuss the results from an optimisation study. This paper concludes with future work and recommendations in section 7.

2. Literature review

CMS can be formulated under static or dynamic conditions. Static conditions imply that product demand is known and does not change over the planning horizon. Dynamic conditions imply that the demand not only changes over the planning horizon, but can be modelled as either deterministic or probabilistic. When dynamic conditions are used for the modelling of CMS, it is then termed as a DCMS. The concept of DCMS was introduced by Rheault *et al.* [4] to overcome the limitations of CMS that only focused on static demand periods. Both cell formation and layout designs are reconfigured in each period to optimise the factory layout and reduce production costs. The reconfigurations often include adding and/or removing machines to or from the system. Although the dynamic modelling of CMS is more realistic, constraints on time, finances and the physical arrangement of the manufacturing system can make frequent reconfiguration infeasible or uneconomical.

Various approaches have been used to solve DCMS models across literature. Each method having its own advantages and limitations depending on the problem being addressed, and the number of factors modelled. Table 1 details the factors considered in the formulation of DCMS models by other researchers. Models are typically formulated as nonlinear optimisation problems. The speed in which a solution is generated, and the quality of the solution produced have been the primary attributes benchmarked across the different techniques. It was noted by Defersha and Chen [2] and is commonly acknowledged among researchers that some techniques are only suitable for certain sized problems. This has been observed in cases where some solvers did not produce a solution for a large problem regardless of an extended computational time allocation. It is noted that the computational capacity of the machines used for the study could have influenced the outcome, however, the results suggest that this could not be the main reason. Some researchers [5] focused their efforts on solving their models with a different solver than that used by previous authors, while others used a few different solvers common in literature to investigate the performance of each solver for the same problem. Commercial software constructed from mathematical programming algorithms have been used by most of the early authors, however, the use of meta-heuristic techniques has gradually increased in present day.

Table 1 Review of model parameters

Authors	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
This paper	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Saxena & Jain [1]		√	√	√	√			√	√	√		√	√	√	√	√	√	√	√	√	√	√
Defersha & Chen [6]		√	√	√	√			√	√			√	√	√	√	√				√	√	√
Niakan *et al.* [7]		√	√	√			√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Safaei & Tavakkoli-Moghaddam [8]		√	√	√	√			√	√	√		√		√		√	√	√	√	√	√	√
Safaei *et al.* [9]	√		√	√	√	√		√	√	√		√	√		√	√	√	√	√	√	√	√
Ossama *et al.* [10]		√	√	√		√		√	√			√	√	√	√	√				√	√	√
Bayram & Sahin [11]		√	√	√		√		√	√	√		√	√	√	√	√				√	√	
Kia et al [12]		√	√	√				√	√	√		√	√	√	√	√				√	√	
Shahram *et al.* [13]		√	√	√		√	√	√				√	√	√	√	√				√	√	√
Aramoon *et al.* [14]		√	√	√		√		√	√			√	√	√	√	√	√			√	√	√

a: Selecting the best route; **b**: Allowing alternative routing; **c**: Deterministic demand fluctuation; **d**: Dynamic cell reconfiguration; **e**: Intercell workload balancing; **f**: Sequence of operations; **g**: Setup time; **h**: Cell size constraint; **i**: Intercell material movement; **j**: Intracell material movement; **k**: Operator allocation; **l**: Machine capacity; **m**: Identical machines within a cell; **n**: Identical machines in the entire system; **o**: Machine investment cost; **p**: Unit operation time; **q**: Machine operation cost; **r**: Intercell batch size movement; **s**: Intracell batch size; **t**: Multiperiod planning; **u**: Machine relocation; **v**: Process batch size.

In recent publications, DCMS, Genetic Algorithm (GA), and varied multi-objective models have been utilised to identify potential improvements and obstacles within diverse organisations. The work of [15] presents a unique layered GA with a reinitialisation approach to preserving population diversity. They found their model to improve avoiding premature stall in local minima. The work of [16] incorporates a focus on environmental implications of a supply chain and how goods sourced with environmental consideration showed robust resource allocation. Furthermore, [17] used a variable neighbourhood search GA to enhance workload balancing in an extended travelling salesman problem. Their work showed how the workload balancing could be intelligently introduced for better computational efficiency. A NSGA-III was used by [18] to improve downtime in a production environment. They compared its performance to that of a NSGA-II and found that the additional objective function improved the quality of solutions generated. To help present-day manufacturers deal with diverse and volatile demand fluctuations, the work of [19] presents a mathematical model with a GA to show how work from academics can be applied in industry. With the aid of a numerical case study, they successfully showed how literature ideas are effective in industry scenarios. This paper seeks to merge the key elements from these reviewed works. They each promote different ideas for improved solutions; hence, we propose a mathematical model with an environmental objective function, and social considerations modelled in constraint form, including a workload balancing constraint. The proposed NSGA-II is modified to cater for unique solutions and is applied over three industry problems to show how it is generally applicable, and not biased to a single data set.

3. Problem description

We propose a DCMS model where part operations are performed sequentially on any capable machine. The operation capabilities of each machine are known and are deterministic. Each part operation can be performed by more than one machine, creating the possibility of alternate process routings. The operation time of all part operations, on all capable machines, are known and deterministic. Furthermore, we assume that the planning horizon is for a predetermined number of periods, with each machine having a predetermined production capacity, expressed in hours, for that period. These production periods are of equal duration. Duplicating machines within and across cells to meet capacity requirements is permissible without limit to the number of machines procured, so long as the cell size constraints are satisfied.

We assume that parts are transported between cells in batches and the transportation cost per batch is known. The per batch Green House Gas (GHG) emissions are also known. The model assumes that intra cell material handling is manually managed by human porters, or is managed by equipment with a negligible power consumption compared to the inter cell material handling equipment. The demand in each period must be satisfied within that period, no backorders, or inventory is kept between periods. All machine types can be relocated between cells. The cost and emissions generated for moving a machine from one cell to another are known. Relocation impacts include uninstalling, moving and reinstallation efforts altogether.

To simplify the model, it is assumed that machine relocation time is negligible and does not affect the available time on the machine for that period. Meaning that plant shutdown time between periods is not included in the total time over the planning horizon. Machines are independently grouped, meaning that there are no machine colocation or separation constraints. We further assume that one worker/operator is assigned to each machine. When a machine is purchased, an experienced worker is hired to operate the machine. When a machine is retired, its operator is retrenched. When a machine is relocated, the assigned operator is relocated with the machine. Machines are added at the beginning of each period and retired/relocated at the end of each period. All machines will be retired at the end of the last demand period to prepare the plant for the next project. Lot splitting is not permissible and job setup time is included in the part processing time.

We present the model notation as:

Sets

h – time index, $h = 1,2,3, \ldots, H$ (H is total number of periods)
p – part index, $p = 1,2,3, \ldots, P_h$ (P_h is total number of parts in period h)
j – index of operations on parts, $j = 1,2,3, \ldots, O_p$ (O_p is total number of operations needed for part p)
m – machine index, $m = 1,2,3, \ldots, M$ (M is total number of machines)
c – cell index, $c = 1,2,3, \ldots, C$ (C is total number of cells)

Decision variables

x_{jpmch} – 1 if operation j of part p by machine m is done in cell c during time h, otherwise 0
N_{mch} – number of type m machines placed in cell c at time h
${K^+}_{mch}$ – number of type m machines added to cell c at the beginning of time h
${K^-}_{mch}$ – number of type m machines removed from cell c at the end of time h
a_{jpm} – 1 if operation j of part p can be done on machine m; otherwise 0

Parameters

LB – Cell size lower bound
UB – Cell size upper bound
D_{ph} – Demand of part p during period h
B_p^{inter} – Intercell batch size for part p
B_p^{intra} – Intracell batch size for part p
T_m – Available time on machine m (h)

t_{jpm} – time for operation j of part p on machine m (h)

α_m – Overhead cost of machine type m

β_m – Variable operating cost for each unit time on machine m (R/h)

γ^{inter} – Batch intercell material handling cost

γ^{intra} – Batch intracell material handling cost

δ_m – Relocation cost for machine type m

φ^{inter} – Batch intercell carbon emissions (kgCO2)

τ_m – Carbon emissions from adding and removing machine type m (kgCO2)

μ_m – Carbon emissions from idle time of machine type m (kgCO2/h)

ε_m – Variable carbon emissions for each operating unit time, on machine type m (kgCO2/h)

σ_m – Carbon emissions from relocating machine type m (kgCO2)

q – Workload balancing factor (taken as ± 0.75 for 75 %)

η – maximum number of different operations an operator can be assigned (taken as 3)

We applied the model to three different problems from literature and solved each problem with the addition of supplementing data to incorporate the second objective. The environmental data for each problem is contained in Table 2, while the reader is referred to the literature [20, 9] for the rest of the data. A linear weighting was observed, such that low-cost machines had a higher environmental impact, while costly machines had a lower environmental impact.

Table 2 Environmental objective input data

Problem	Machines	M1	M2	M3	M4	M5	M6	M7	M8	M9
1	Sourcing emissions	5087	3149	4313	4408	6200	4509	2480	3815	n/a
2		5425	6200	5710	4822	5425	4931	4340	4520	4822
3		6200	5320	4960	5320	6200	4650	n/a	n/a	n/a
1	Relocation emissions	656	406	556	568	800	581	320	492	n/a
2		500	600	550	350	500	400	250	300	350
3		310	266	150	266	310	233	n/a	n/a	n/a
1	Idle time carbon emissions	11	7	9	9.5	13.6	10	5.4	8.4	n/a
2		6.5	8	7	5	6.5	5.5	3	4	5
3		8	6	4	6	8	2	n/a	n/a	n/a

Problem	Parts	P1	P2	P3	P4	P5	P6	P7	P8
1	Unit intercell emissions	2	2	2	2	n/a	n/a	n/a	n/a
2		1.5	2.1	1.8	1.1	0.95	2.2	2	n/a
3		1.8	1.8	2.25	1.2	2.25	1.13	1.13	1.5

4. Mathematical model

Minimise:

$$
\begin{aligned}
Z_1 = & \sum_{h=1}^{H} \sum_{m=1}^{M} \sum_{c=1}^{C} N_{mch}\alpha_m \\
& + \sum_{h=1}^{H} \sum_{c=1}^{C} \sum_{p=1}^{P} \sum_{j=1}^{Op} \sum_{m=1}^{M} \beta_m D_{ph} t_{jpm} x_{jpmch} \\
& + \frac{1}{2} \sum_{h=1}^{H} \sum_{p=1}^{P} \sum_{j=1}^{OP-1} \sum_{c=1}^{C} \left[\frac{D_{ph}}{B_p^{inter}}\right] \gamma^{inter} \left| \sum_{m=1}^{M} x_{(j+1)pmch} - \sum_{m=1}^{M} x_{jpmch} \right| \\
& + \frac{1}{2} \sum_{h=1}^{H} \sum_{p=1}^{P} \sum_{j=1}^{OP-1} \sum_{c=1}^{C} \left[\frac{D_{ph}}{B_p^{intra}}\right] \gamma^{intra} \left(\sum_{m=1}^{M} \left| x_{(j+1)pmch} - x_{jpmch} \right| \right. \\
& \left. - \left| \sum_{m=1}^{M} x_{(j+1)pch} - \sum_{m=1}^{M} x_{jpch} \right| \right) + \frac{1}{2} \sum_{h=1}^{H} \sum_{m=1}^{M} \sum_{c=1}^{C} \delta_m (K^+{}_{mch} + K^-{}_{mch})
\end{aligned}
\tag{1}
$$

$$Z_2 = \sum_{h=1}^{H} \sum_{c=1}^{C} \sum_{m=1}^{M} \tau_m \left| N_{mc(h+1)} - N_{mch} \right|$$

$$+ \sum_{h=1}^{H} \sum_{m=1}^{M} \sum_{c=1}^{C} \sigma_m (K^+_{mch} + K^-_{mch})$$

$$+ \sum_{h=1}^{H} \sum_{c=1}^{C} \sum_{m=1}^{M} \mu_m \left\{ \sum_{p=1}^{P} \sum_{j=1}^{OP} T_m N_{mch} - D_{ph} t_{jpm} x_{jpmch} \right\} \tag{2}$$

$$+ \frac{1}{2} \sum_{h=1}^{H} \sum_{p=1}^{P} \sum_{j=1}^{OP-1} \sum_{c=1}^{C} \left[\frac{D_{ph}}{B_p^{inter}} \right] \varphi^{inter} \left| \sum_{m=1}^{M} x_{(j+1)pmch} - \sum_{m=1}^{M} x_{jpmch} \right|$$

Subject to:

$$\sum_{c=1}^{C} \sum_{m=1}^{M} a_{jpm} x_{jpmch} = 1 \qquad \forall j, p, h \tag{3}$$

$$\sum_{p=1}^{P} \sum_{j=1}^{OP} D_{ph} t_{jpm} x_{jpmch} \leq T_m N_{mch} \qquad \forall m, c, h \tag{4}$$

$$LB \leq \sum_{m=1}^{M} N_{mch} \leq UB \qquad \forall c, h \tag{5}$$

$$\sum_{m=1}^{M} \sum_{p=1}^{P} \sum_{j=1}^{OP} x_{jpmch} \geq \frac{q}{C} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{p=1}^{P} \sum_{j=1}^{OP} x_{jpmch} \qquad \forall c, h \tag{6}$$

$$\sum_{p=1}^{P} \sum_{j=1}^{OP} x_{jpmch} \leq \eta N_{mch} \qquad \forall m, c, h \tag{7}$$

$$N_{mc(h-1)} + K^+_{mch} - K^-_{mch} = N_{mch} \quad \forall m, c, h \tag{8}$$

$$x_{jpmch} \in \{0,1\}, N_{mch}, K^+_{mch}, K^-_{mch} \geq 0 \in \mathbb{Z} \tag{9}$$

Eq. 1 is a cost minimisation objective, consisting of five terms. The first term is the machine overhead/rental cost. The second term represents the machine operational cost per unit time. The third and fourth terms calculate the intercell and intracell material handling costs respectfully. The machine relocation cost is calculated in the fifth term. Eq. 2 represents an environmental impact minimisation objective, consisting of four terms measured in $kgCO_2$. The first term represents the carbon emissions from the machine manufacturing, transportation to and from the plant on purchase/retirement, and from the recycling/disposal of its components. The second term represents emissions from machine relocation within the system. The third term measures the emissions from the machine while idling. The fourth term measures the emissions from intercell material handling. Eq. 3 ensures that each part operation is assigned to a single machine and is performed only once. Eq. 4 bounds the model from violating the capacity of the machines. Eq. 5 introduces the lower and upper bounds for the number of machines permissible in a cell. Eq. 6 is the social constraint for intercell workload balancing. Eq. 7 is the social constraint on the number of different part operations a single operator can be assigned to perform. Eq. 8 ensures that the machine placement and relocation is calculated accurately across demand periods.

5. Solution approach

GAs are an evolutionary metaheuristic technique that progressively iterates from a randomly generated initial population, towards an optimal solution set through crossover and mutation operations [18]. The unique GA developed to solve the mathematical model will be detailed hereafter in eight sections. Fig. 1 depicts the holistic program flow of the optimisation process.
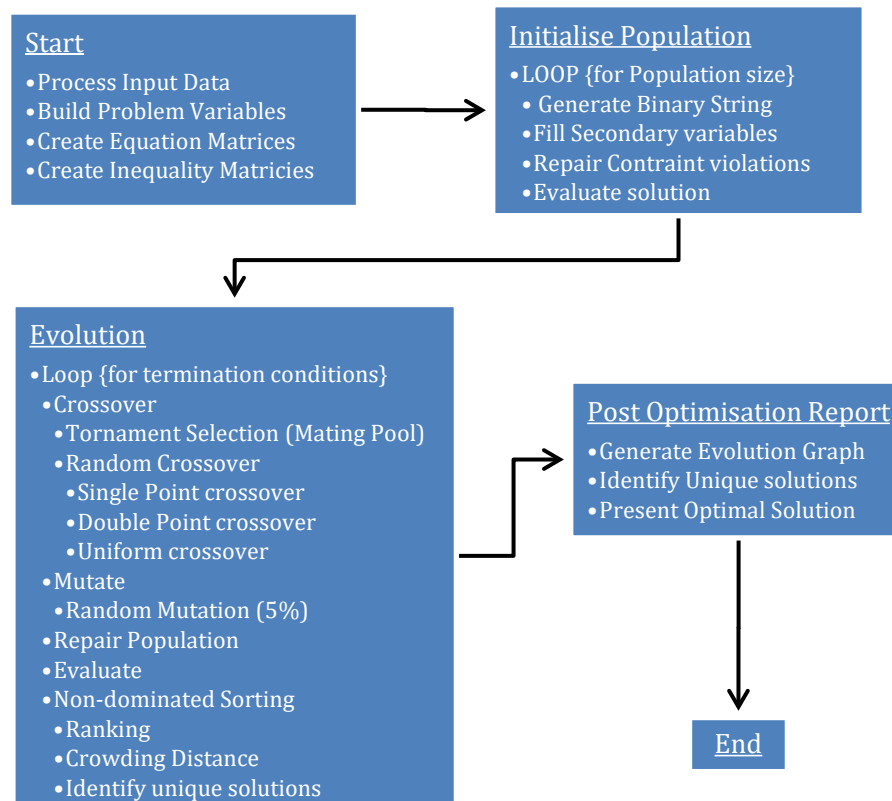
Fig. 1 GA optimisation algorithm flow

5.1 Chromosome encoding

Within the population, each solution's chromosome structure was encoded as a single row array of integers representing all variables. Each array was split into two sections, separating the binary part production decision variables x_{jpmch} from the rest. The part production decision variables were used as the primary variables while the other variables were changed accordingly, to satisfy the problem constraints. Hence, during initialisation, crossover and mutation, only primary variables were changed and/or modified, while the rest were adapted by a custom repair function to suit the new arrangements. An overview of the repair function is given hereafter.

5.2 Initial population

To generate the initial population, a random binary $M \times N$ matrix was constructed, where M was the input population size and N was the number of primary decision variables. Each row, which represented a solution in the population, then had the rest of the variables populated accordingly. The solutions were individually scrutinised against the constraints using the penalty approach, and a penalty value was assigned to each solution based on the number of constraints violated. An attempt to repair solutions that violated constraints was initialised after the chromosomes were evaluated. The initial population was then filled with the solutions that had a penalty value of zero from the lot. This initialisation strategy is similar to that used by Kia [20].

5.3 Tournament selection

Instead of a roulette wheel approach [19], a tournament selection [21] process was held for each generation to select the prime solutions for the mating pool. A random bi-contestant structure was observed for parent selection. Better ranking solutions won the contest and were added to the mating pool, however, in cases where contestants were of equal rank, solutions with a higher crowding distance were selected. In cases where contestants had the same rank and an equal crowding distance, a random solution was selected between the two. Each solution in the population participated in two tournaments, such that there would be two copies of the best solution in the mating pool, and the worst solution of the population would never enter the mating pool. Tournaments were conducted until the mating pool was the same size as the original population.

5.4 Crossover

Crossover was performed with 80 % probability. To generate offspring, three crossover operations were utilised, namely, Single Point Crossover, Double Point Crossover and Uniform Crossover [15]. Two solutions were randomly selected from the mating pool and a crossover operator was randomly selected from the three operators. The crossover operations were performed only over the binary primary variables as discussed. After crossover, each offspring was evaluated and if it violated any constraints, it was repaired and then re-evaluated.

5.5 Mutation

A 5 % mutation rate over the primary genes of crossover offspring was observed. A uniform mutation strategy was employed [20], which randomly mutated genes across all the binary decision variables. Since the primary variables were binary, the mutation changed the genes randomly selected for mutation from either 1 to 0 or vice versa. After the mutation, mutant solutions that violated the constraints were repaired and all solutions were evaluated and grouped with the rest of the solutions for ranking.

5.6 Repair function

A repair function was developed to fix solutions that violated any of the constraints during initialisation, crossover, and mutation. When a solution was under repair, initially the primary variables were repaired such that constraint Eq. 3 was observed. Thereafter, machines were placed in the appropriate cells in keeping with constraint Eq. 4. All machines that did not have parts allocated to them were removed from the solution, and lastly, the demand periods were then joined together to determine machine relocation variables. Thereafter, repaired solutions were checked against cell size constraints Eq. 5. The penalty approach was effective at maintaining most of the constraints, however, not all of them. The cell upper bound constraints Eq. 5 were particularly difficult to repair without redefining the primary variables and increasing the computational effort of the algorithm needlessly. Hence, when a solution did not satisfy all the constraints after the first repair attempt, it was discarded. If the cell lower bound was violated, a minimal number of random machines were placed in the cell to satisfy the constraint.

5.7 Penalty function and evaluation

In keeping with the non-dominated sorting method of evaluation, and in conjunction with the penalty approach, each solution was graded in three major categories namely, penalty value, rank, and crowding distance. Other attributes of each solution which were also evaluated were solution uniqueness, number of solutions that dominate that solution, and the number of solutions that were dominated by that solution. For a given solution Z, where A, A_{eq}, and b_{eq} are matrix representations of the problem constraints, and M is a large positive value, the penalty expression used for the inequality constraints was of the form:

$$P_1(Z) = M \cdot \max [0, [A - b]]^2 \tag{10}$$

while the penalty expression for the equality constraints was modelled:

$$P_2(Z) = M \cdot sum\left\{\left[sum\left[A_{eq} - b_{eq} \right]\right]^2\right\} \tag{11}$$

5.8 Unique non-dominated sorting

The weighting of the objective functions was enforced using principles from the NSGA-II approach to avoid localised stalling and protracted computational time. We extended on this method, by sorting the final population for each generation with the best unique solutions from the combined populations, i.e., the previous generation population, the offspring population, and the mutant population. During experimentation, this modified approach not only improved the range of solution exploration, but also improved population diversity across generations. The method initially groups solutions according to pareto front rankings; thereafter, the solutions are sorted according to their individual crowding distances within each rank. Solutions are then compared to each other and the top unique 200 (population size) solutions were passed on to the next generation. When there were less than 200 unique solutions the algorithm added duplicate solutions from the top-ranking solution going down, to fill the required number.

For a single minimisation objective function, solution $Z(x_1)$ dominates solution $Z(x_2)$ if the value of x_1 is less than that of x_2. When a second minimisation objective is introduced, $Z(x_1, y_1)$ dominates $Z(x_2, y_2)$ if and only if $(x_1 \leq x_2$ and $y_1 \leq y_2)$ and $(x_1 < x_2$ or $y_1 < y_2)$. The Boolean result from this expression was used to identify non-dominated solutions across populations and assign rank accordingly. Where N is the total number of fronts/ranks for each generation, F is the pareto front, and M is the total number of objectives, the crowding distance (d_i) for each solution was calculated using the equation:

$$d_i = \sum_{n=1}^{N} \sum_{m=1}^{M} \frac{F_n(Z_m)(i+1) - F_n(Z_m)(i-1)}{F_n{}^{max} - F_n{}^{min}} \quad \forall i \tag{12}$$

6. Results and discussion

The problems were solved on MATLAB R2021a running on a personal computer with a 2.70 GHz, 7th generation Intel i7 processor and 8 GB RAM. The GA was run for 1000 generations to ensure complete evolution stall for each problem. The results are detailed in Table 3.

Table 3 Post optimisation data

No. of parts	No. of machines	LB / UB	No. of constraints	No. of non-zero variables	No. of unique solutions	Min. cycle time (h)	Average Machine utilisation (%)	Lowest obj. 1 val. R(ZAR)	Lowest obj. 2 val. kgCO$_2$
4	8	1/5	1251	1104	31	7059	56-74	970,910	161,682.5
7	9	1/7	2223	1782	73	66	52-67	5,236,400	445,827.8
8	6	1/5	1773	1908	128	689	61-70	342,518	105,210.1

Valuable insights can be extracted from the consistency of the concave shape of the pareto fronts resulting from the optimisation, depicted in Fig. 2 and Fig. 3. It suggests that the model does not favour a particular data set, and that the results are not problem biased. Furthermore, it suggests a conflicting relationship between the objective functions, implying a non-linear trade-off between the two. This is expected since the objective functions have a complex non-linear relation to each other.

The pareto front's concavity is owed to the different aspects each term in each objective function seeks to optimise. In Eq. 1, the first term seeks to minimise the number of machines held in each period, while the first term in Eq. 2 seeks to minimise the changes in the machines held. These terms compete for dominance over machine allocations. Although the fifth term in Eq. 1 and the second term in Eq. 2 both seek to reduce machine relocation, they too will conflict over the particular machines to be moved as one objective favours moving cost effective machines and the other environmentally friendly machines. This conflict stems from the difference in weighting for the machines for each objective. Another dynamic relationship observed is be-

tween term 2 of Eq. 1 and term 3 of Eq. 2, as one seeks to minimize the utilization of machines while the other seeks to maximise machine utilisation. Some linearity is shared by term 3 of Eq. 1 and term 4 of Eq. 2 as they both seek to minimise intercell material handling. However, they are limited by the social objective imposed by constraint Eq. 7 which prohibits the overloading of machine operators with multiple different operations. This inclines the model to create larger cells with a higher number of machines and less intercell movement, or fewer machines with higher intercell material handling. The cell upper bound constraint Eq. 5 also influences the model on the machine placements by adding a ceiling to the permissible cell sizes.

(a)   (b)
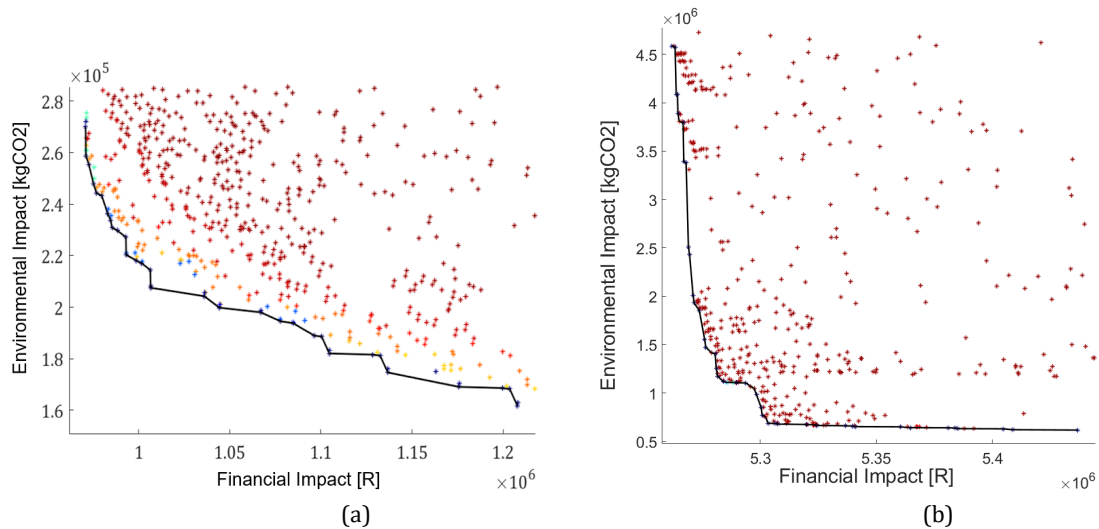
Fig. 2 Pareto fronts of: (a) problem 1, (b) problem 2
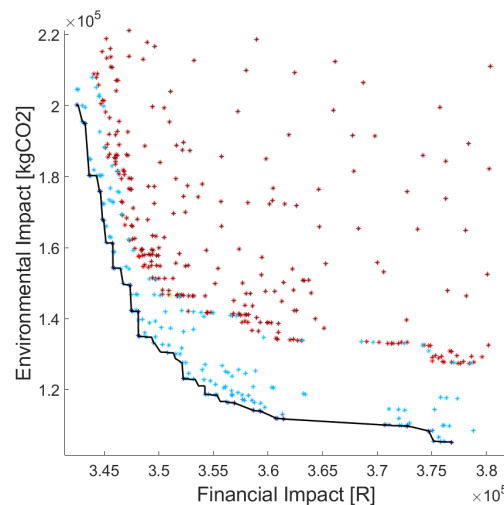
Fig. 3 Pareto front of problem 3

It is worth noting that when these problems were solved by previous researchers, their work did not consider environmental and social objectives, hence, their work is not directly relatable to ours on a one-to-one comparison. However, from a holistic perspective, the pareto front shows that the most cost-effective solutions are not the most environmentally friendly ones. Rather, solutions are a compromise between the two objectives. Since the social objective is embedded in the model constraints, it is observed in each solution. Although we have suggested selecting each problem's solution based on minimum cycle time, or highest machine utilisation, the existence of multiple unique solutions presents opportunity for additional post optimisation processing that could reveal more insights and improved practicality. Some post optimisation processing could include a sensitivity analysis on the different parameters of each problem, or a clustering analysis to identify solution groups with similar trade-offs. However, the most practical post optimisation assessment would be developing quality metrics to narrow down pre-

ferred solutions. These metrics could analyse the volume and/or direction of material flow per solution, and/or the traffic of machine relocations. These metrics are not considered in the model but could greatly affect the safety in a production environment, and the human efforts needed to calibrate equipment once it has been moved. Ultimately, the pareto front of unique solutions to choose from provides decision makers options for informed decision making. Once a system has been optimised, a project budget together with an environmental assessment of acceptable emissions can be used to select the best layout according to project specifications on a case-by-case basis.

7. Conclusion

It is known that DCMS are NP hard problems. We proposed a multi-objective mathematical model to solve a novel multiperiod DCMS model which was NP-complete. The problems were formulated as binary integer programming models and solved heuristically by a NSGA-II approach. The model has two objective functions that aim to minimise economic cost, and environmental impact of production. A social aspect of the model was introduced through a set of constraints. The penalty method was used to enforce the model constraints. Three numerical example problems from literature were solved to demonstrate the effectiveness of the genetic algorithm, and a pareto plot was generated for each problem. From post experiment analysis, it was observed that all the pareto fronts were concaved and that the generated plots suggested solutions which were a weighted compromise between the two objectives. The results indicate that incorporating an environmental objective produces alternatives that are not only economically viable but also socially and environmentally responsible. Future enhancements to the model could include factors such as intracell workload balancing, lot splitting, inventory holding, and outsourcing.

Acknowledgements

References

[1] Saxena, L.K., Jain, P.K. (2011). Dynamic cellular manufacturing systems- a comprehensive model, *The International Journal of Advanced Manufacturing Technology,* Vol. 53, 11-34, doi: 10.1007/s00170-010-2842-9.

[2] Defersha, F.M., Chen, M. (2008). A parallel multiple Markov chain simulated annealing for multi-period manufacturing cell formation problems, *The International Journal of Advanced Manufacturing Technology,* Vol. 37, 140-156, doi: 10.1007/s00170-007-0947-6.

[3] Rasshif, M.N., Wicaksono, P.A., Hartini, S. (2023). Desain sustainable value stream mapping untuk evaluasi kinerja keberlanjutan perusahaan obat, *Industrial Engineering Online Journal,* Vol.12, No. 4, 1-9.

[4] Rheault, M., Drolet, J.R., Abdulnour, G. (1996). Dynamic cellular manufacturing system (DCMS), *Computers & Industrial Engineering,* Vol. 31, No. 1-2, 143-146, doi: 10.1016/0360-8352(96)00098-8.

[5] Khettabi, I., Benyoucef, L., Boutiche, M.A. (2022). Sustainable multi-objective process planning in reconfigurable manufacturing environment: Adapted new dynamic NSGA-II vs New NSGA-III, *International Journal of Production Research*, Vol. 20, No. 60, 6329-6349, doi: 10.1080/00207543.2022.2044537.

[6] Defersha, F.M., Chen, M. (2006). Machine cell formation using a mathematical model and a genetic-algorithm-based heuristic, *International Journal of Production Research,* Vol. 12, No. 44, 2421-2444, doi: 10.1080/00207540500337833.

[7] Niakan, F., Baboli, A., Moyaux, T., Botta-Genoulaz, V. (2015). A bi-objective model in sustainable dynamic cell formation problem with skill-based worker assignment, *Journal of Manufacturing Systems,* Vol. 38, 46-62, doi: 10.1016/j.jmsy.2015.11.001.

[8] Safaei, N., Tavakkoli-Moghaddam, R. (2009). Integrated multi-period cell formation and subcontracting production planning in dynamic cellular manufacturing systems, *International Journal of Production Economics,* Vol. 120, No. 2, 301-314, doi: 10.1016/j.ijpe.2008.12.013.

[9] Safaei, N., Saidi-Mehrabad, M., Tavakkoli-Moghaddam, R., Sassani, F. (2008). A fuzzy programming approach for a cell formation problem with dynamic and uncertain conditions, *Fuzzy Sets and Systems,* Vol. 159, No. 2, 215-236, doi: 10.1016/j.fss.2007.06.014.

[10] Ossama, M., Youssef, A.M.A., Shalaby, M.A. (2014). A multi-period cell formation model for reconfigurable manufacturing systems, *Procedia CIRP,* Vol. 17, 130-135, doi: 10.1016/j.procir.2014.01.120.

[11] Bayram, H., Sahin, R. (2016). A comprehensive mathematical model for dynamic cellular manufacturing system design and linear programming embedded hybrid solution techniques, *Computers & Industrial Engineering,* Vol. 91, 10-29, doi: 10.1016/j.cie.2015.10.014.

[12] Kia, R., Baboli, A., Javadian, N., Tavakkoli-Moghaddam, R., Kazemi, M., Khorrami, J. (2012). Solving a group layout design model of a dynamic cellular manufacturing system with alternative process routings, lot splitting and flexible reconfiguration by simulated annealing, Vol. 39, No. 11, 2642-2658, doi: 10.1016/j.cor.2012.01.012.

[13] Sharifi, S., Chauhan, S.S., Bhuiyan, N. (2014). A dynamic programming approach to ga-based heuristic for multi-period CF problems, *Journal of Manufacturing Systems,* Vol. 33, No. 3, 366-375, doi: 10.1016/j.jmsy.2014.02.004.

[14] Aramon Bajestani, M., Rabbani, M., Rahimi-Vahed, A.R., Baharian Khoshkhou, G. (2009). A multi-objective scatter search for a dynamic cell formation problem, *Computers & Operations Research,* Vol. 36, No. 3, 777-794, doi: 10.1016/j.cor.2007.10.026.

[15] Amjad, M.K., Butt, S.I., Anjum, N., Chaudhry, I.A., Faping, Z., Khan, M. (2020). A layered genetic algorithm with iterative diversification for optimization of flexible job shop scheduling problems, *Advances in Production Engineering & Management,* Vol. 15, No. 4, 377-389, doi: 10.14743/apem2020.4.372.

[16] Fang, I.W., Lin, W.T. (2021). A multi-objective optimal decision model for a green closed-loop supply chain under uncertainty: A real industrial case study, *Advances in Production Engineering & Management,* Vol. 16, No. 2, 161-172, doi: 10.14743/apem2021.2.391.

[17] Wang, Y.D., Lu, X.C., Shen, J.R. (2021). Improved genetic algorithm (VNS-GA) using polar coordinate classification for workload balanced multiple Traveling Salesman Problem (mTSP), *Advances in Production Engineering & Management,* Vol. 16, No. 2, 173-184, doi: 10.14743/apem2021.2.392.

[18] Xu, E.B., Yang, M.S., Li, Y., Gao, X.Q., Wang, Z.Y., Ren, L.J. (2021). A multi-objective selective maintenance optimization method for series-parallel systems using NSGA-III and NSGA-II evolutionary algorithms, *Advances in Production Engineering & Management,* Vol. 16, No. 3, 372-384, doi: 10.14743/apem2021.3.407.

[19] Guzman, E., Poler, R., Andres, B., (2023). A metaheuristic approach combining genetic algorithm and mixed integer linear programming model for production and distribution planning in the supply chain, *Advances in Production Engineering & Management,* Vol. 18, No. 1, 19-31, doi: 10.14743/apem2023.1.454.

[20] Kia, R., Khaksar-Haghani, F., Javadian, N., Tavakkoli-Moghaddam, R. (2013). Solving a multi-floor layout design model of a dynamic cellular manufacturing system by an efficient genetic algorithm, *Journal of Manufacturing Systems,* Vol. 33, No. 1, 218-232, doi: 10.1016/j.jmsy.2013.12.005.

[21] Balakrishnan, J., Cheng, C.H., Conway, D.G., Lau, C.M. (2003). A hybrid genetic algorithm for the dynamic plant layout problem, *International Journal of Production Economics,* Vol. 86, No. 2, 107-120, doi: 10.1016/S0925-5273(03)00027-6.

Unsupervised machine learning application in the selection of measurement strategy on Coordinate Measuring Machine

Štrbac, B.[a,*], **Ranisavljev, M.**[a], **Orošnjak, M.**[b], **Havrlišan, S.**[c], **Dudić, B.**[d,e], **Savković, B.**[a]

[a]University of Novi Sad, Faculty of Technical Sciences, Department of Production Engineering, Novi Sad, Serbia
[b]University of Novi Sad, Faculty of Technical Sciences, Department of Industrial Engineering and Management, Novi Sad, Serbia
[c]Josip Juraj Strossmayer University of Osijek, Mechanical Engineering Faculty, Slavonski Brod, Croatia
[d]Comenius University Bratislava, Faculty of Management, Bratislava, Slovakia
[e]Faculty of Economics and Engineering Management, University Business Academy, Novi Sad, Serbia

ABSTRACT

It is indisputable that some type of coordinate measurement system (CMS) is generally used to assess the quality of dimensional and geometric characteristics. Considering the required accuracy, flexibility, and speed of measurement, a CMM with a scanning sensor may offer the best performance. These measurement systems are very complex, and many factors affect the reliability of the measurement results. A Metrologist's choice represents the greatest variability in the measurement strategy. Previous research has shown that the measurement results can be changed up to 100 % by choosing a different measurement strategy when evaluating the form error. This paper conducts a detailed study of the impact of the measurement strategy on the cylindricity error when measuring eleven workpieces with the same nominal characteristics, but different real characteristics described by roughness and the reference value of cylindricity. To examine the influence and importance of certain factors and their levels, design of experiment (DoE) and unsupervised machine learning techniques of PCA (Principal Component Analysis) and Multiple Correspondence Analysis (MCA), were used. The results suggest that depending on the real geometry of the workpiece, different factors with different percentages influence the output characteristic. The objective was to choose a uniform measurement strategy when measuring cylindricity on the CMM, while the prior information about the actual geometry of the workpiece is lacking.

1. Introduction

Quality assessment of dimensional and geometric product specifications in the field of precision engineering is generally performed on some type of coordinate measuring system (CMS). In the last two decades, various versions of CMS have appeared, such as coordinate measuring machine with contact and non-contact sensors, photogrammetry, measuring arms, non-contact (optical and laser) scanners, industrial CT, etc. [1]. Each type has its advantages and disadvantages, which are primarily reflected in the accuracy of measurement, flexibility, the amount of information collected from the workpiece (points), inspection time, price, etc. [2, 3]. Considering all the above requirements, perhaps a coordinate measuring machine (CMM) equipped with a contact-scanning sensor is the most optimal solution for the verification of the narrow specification requirements imposed by the functional requirements of the workpieces. However, CMMs are very complex

measurement systems and there are numerous influencing factors and their interactions that contribute to a large range of dispersion of results, which can be correlated with measurement uncertainty [4]. It seems that the assessment of measurement uncertainty remains within the framework of laboratory measurements for calibration and even theory for practical industrial measurements. So far developed methods for measurement uncertainty assessment such as GUM or series of ISO 15530 standards either have many limitations and/or require calibrated parts, repeated measurements, development of simulation methods, etc. [5-11]. For this reason, measurement uncertainty is generally not given with the results of practical measurements, although it is the only indicator of the quality of the measurement results, and the maximum permissible error (MPE) is often taken as an indicator of the quality of the measurement. Also, repeatability and reproducibility studies are not effective with CMMs because they are automated measuring systems.

In the total uncertainty budget of the CMM, e.g. according to ISO 15530-3, which is considered a reference standard, it figures the uncertainty of calibration of the workpiece, repeated measurements, temperature effects, while the measurement strategy is not considered because it is the same for all 20 repeated measurements required by the standard [6]. However, a certain issue is the selection of the measurement strategy because the specification operators are not complete, and the selection of the measurement strategy is left to the operator [12]. The results of some measurement strategies may show compliance with the specification while others may not. Also, the actual dimensions of the parts from the production that are measured always remain unknown, so no judgment can be made about the "good" or "bad" chosen measurement strategy. Under the measurement strategy of CMM can be considered the number and position of measurement points on the observed feature, the criteria for creating a fitted geometry, the choice of the tip of the measuring probe and, if they are scanning contact sensors, the scanning speed and filtering. Yes, current research has shown that these parameters in interaction with the form deviation of the measured primitive really affect the measurement uncertainty [13, 14]. Uncertainty factors originating from geometrical errors of the CMM, and measurement sensor cannot be controlled because they are inherent to the measuring instrument and will always be present in the overall measurement result, but the uncertainties originating from the choice of measurement strategy can be greatly influenced by choosing the optimal measurement strategy for ordered measurement task. In papers [15, 16], the selection of the optimal number of points was made for obtaining measurement values that are close to "actual". Certainly, the number of points tending to infinity would be the best for describing actual geometry, but it is not economically justified, and in these publications, it was shown that depending on the processing procedure, the number of optimal points also depends on the example of flatness measurement. These studies have shown that the choice of strategy when evaluating the form error depends primarily on the processing procedure used to obtain the workpiece, and the value of the mean arithmetic roughness was used as input information.

The aim of this work is not to evaluate the measurement uncertainty but to demonstrate how much the choice of measurement strategy with contact scanning sensors can affect the scattering of results when measuring cylindricity on workpieces that are of the same nominal dimensions but were manufactured with different processes, different settings and are characterized by deviations in the quality of the surface. To examine the variability of measurement results due to a change in measurement strategy, the design of the experiment was used. In addition, using dimensionality reduction tools, i.e. Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA), the optimal solution of measurement strategies has been determined by considering all objects with same nominal dimensions but with deviations in geometry.

2. Methodology

2.1 Experimental setup

To choose the optimal strategy for measuring cylindricity on a CMM in scanning mode, eleven workpieces were manufactured. All workpieces had the same nominal dimensions, an outer diameter of 12 mm and a height of 14.9 mm and are made by different technological procedures,

grinding and turning, with different cutting parameters to achieve variability in terms of form error and roughness of the processed surface, Fig. 1. The samples selected in this way cover a wide range of imperfections of real workpieces that can be found in practice when it comes to cylindrical workpieces. The measure of the diversity of the workpieces is the difference in the mean arithmetic roughness Ra and the reference (actual) cylindricity $CYLt_{ref}$, which was obtained by measuring the workpieces on the Roundtest, Table 1. The results obtained on the Roundtest were used as target (desired) values of cylindricity and were used to optimize the measurements on CMM. Roundtest is a measuring instrument specialized for measuring form error, primarily circularity, and cylindricity, and is used to verify very narrow specification limits and generally has a much higher accuracy than coordinate measuring machines. For the purposes of this experiment, Roundtest RA-2200AS Mitutoyo, radial rotational accuracy $(0.04+6H/10000)$ μm was used; H is measuring height in mm and axial rotational accuracy $(0.04+6X/10000)$ μm; X is the distance from the rotation center in mm. The workpieces were measured with a measuring probe of 0.75 mm in radius, with twelve concentric circles (each circle contained 7000 points), and a Gaussian-50%Low [50 UPR] filter was used. Fig. 2 shows the measurement report from the roundtest for cylinder 7.

Following the roundtest measurements, workpieces were measured on a coordinate measuring machine Carl Zeiss Contura G2 (Oberkochen, Germany), with the RDS rotating head and VAST XXT tactile sensor. The helix measurement strategy was used during the experiment, applying the scanning method. The maximum permissible error of the machine in the scanning mode is $MPE_{THP/\tau} = (2.2 + L/330)$ μm, defined according to DIN EN ISO 10360-4:2001. The measurements were conducted in the temperature-controlled environment of 20 ±0.5 °C. At about 80 % of the total height of the workpieces, points were sampled for the creation of substitute geometry, using the minimal zone method (MZ) and estimation of cylindricity.

Based on previous research and experience data, it was observed that the factors like diameter of the tip of the measuring probe, scanning speed, number of sampled points and filtering have the most influence on the error of cylindricity and are related to the measurement strategy on the CMM. To quantify the influence of individual factors on the variation of the output characteristics, the Methodology of the response surface - Box Behnken experiment design was used. The factors and levels used are shown in Table 2. The choice of factor levels was based on experience from practical measurements and the capabilities offered by CMM hardware and software.

Table 1 Measured values of the roughness parameter Ra and cylindricity, both in μm

Cylinder number	1	2	3	4	5	6	7	8	9	10	11
Ra	0.28	1.27	1.57	1.80	3.69	5.21	4.49	7.01	11.62	11.56	15.37
$CYLt_{ref}$	5.57	9.40	16.86	11.11	25.07	24.12	21.90	29.22	46.40	46.01	116.96

Fig. 1 Sample of workpieces
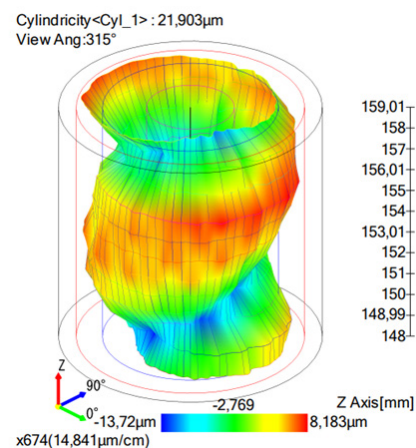
Fig. 2 Cylindricity result on Roundtest

Table 2 Factors used in Box-Behnken DoE in CMM measurement

Factors	Levels		
	-1	0	1
Measuring speed	5	35	100
Number of points	25	75	125
Stylus type diameter	2	3	5
Filter	Without	Cut-off 0.8	Cut-off 8

Measuring speed significantly impacts the measurement accuracy of CMMs. High scanning speeds can introduce dynamic flexibilities that affect measurement accuracy [18]. Ali *et al.* explored the impact of scanning speed and fitting algorithms on measurement results for cylindrical standards. They investigated five different scanning speeds in combination with many available fitting algorithms and concluded that for roundness measurements of the standard, the maximum inscribed fitting technique and 2 mm/s scanning speed give the lowest potential for errors [20]. Urban *et al.* investigated the influence of scanning speeds in the range from 2 up to 40 mm/s and the average deviation from reference size and roundness of gage rings. Evaluation of the experiment revealed that the scanning speed value used in combination with different radii of measured circular path quite fundamentally affects the accuracy and precision of measurement because increasing speed leads to increasing deviations due to the inertia of CMM construction. This increase is more pronounced for smaller diameters. For the roundness measurement, the same trend can be detected. Up to 6 mm/s speed, deviations are approximately the same for different ring gages, but above this speed, significant differences are observed [21]. Stępień [19] also investigated the influence of selected measurement parameters and strategy on the evaluation of form deviation values. Flatness and straightness were measured with five different scanning speeds and a number of sampling points. For the straightness evaluation, he chose a cylindrical workpiece. Like in the previous research, results indicate that form deviation increases along with the increase of scanning speed, without changing the number of sampling points.

The CMM inspection is based on a discrete measurement strategy, which is inherently an approximation process. If the sample size is infinite, the error in approximation approaches zero, and whenever the sample size is finite, the error must be non-zero. Since the number of points must be finite and much smaller than the number of points required by the ideal specification operator, the user of a CMM must plan the right measurement strategy to achieve the highest possible measurement accuracy [22]. The set of sample points is assumed to represent the feature being measured. These sample points are used to create the substitute profile for the feature being measured. The substitute profile is compared with the design profile to determine conformance to the specification. Intuitively, an increased sample size could lead to better characterization of the profile, however, the sample size is often limited by cost and time constraints. Thus, for a given sample size, the measurement strategy used must determine the locations and number of these measurement points such that the actual form may be effectively characterized [23].

Dimensional measurements performed by a CMM typically include all measured data points in an "as collected" manner. In other words, the raw, measured data points were directly analyzed without any pre-processing. However, as data densities have increased there has been an increased awareness that there is a great deal of "noise" in these "high density" data sets. This "noise" can be the result of such things as the surface roughness of the component being measured or due to errors in the measurement system like vibration. In many cases, it is desirable to filter out this "noise" to arrive at a more stable data set that is perhaps more indicative of the attributes that are to be assessed. In this regard, filtering methodologies have been put forth. These methodologies have been very common in the context of roughness and roundness measurement and are now making their way into other dimensional metrology applications [24]. According to Weckenmann [13], one factor in measurement strategy that has been largely ignored and has a significant impact on the results of coordinate measurements is the filtration of measurement points. From the experiments conducted on the geometrically ideal cylinder and the cylinder subject to deviation (mean value and standard deviation) effect of different cut-off values of the

Gaussian filter on the parallelism deviation is evident. For the range from 2000 up to 5000 sampled points, the lowest deviation, on both cylinders is acquired with the use of λc = 8.0 mm cut-off value, and the largest deviation is present where there is no filtration of sampled points.

The Stylus tip measurement diameter introduces mechanical filtration for sampling deviations from the workpiece. This factor is particularly important when measuring roundness because it filters out the effect of roughness and waviness on the form error [9]. Due to the finite size of the stylus, it is unable to access some deep valleys, and thus the measured surface is not the true surface but an approximate one, i.e. the mechanical surface. Normally the rougher the surface is, the more obvious the mechanical filtering effect will be. For the measurement of workpiece surfaces, the distortions caused by the tip mechanical filtration effect noticeably influence measurement accuracy [25].

2.2 Principal component analysis

Principal Component Analysis (PCA) is a multivariate statistical and unsupervised machine learning technique used for dimensionality reduction. Namely, the PCA transforms original set of variables p into a new set of uncorrelated variables by ordination, commonly referred to as principal components (PCs). These PCs are ordered by the amount of variance they capture from the original dataset. Thus, the PC1 accounts for the highest variation present, followed by PC2,…,PCn. It is important to note that ordination of PCs is under the constraint that they orthogonal on each previous PC. Performing the PCA analysis is conducted through a linear algebra and eigenvalue decomposition of a covariance and a correlation matrix. Let us consider that the original dataset X of dimensions $n \times p$, where n represents the number of observations, in our case measurements, while p represents the number of variables, in our case cylinders. Next, the PCA then converts the original set of variables into a set of uncorrelated variables (PCs), which capture the maximum amount of variance [26].

Explaining the procedure analytically, the first step of PCA is to perform standardization of the original dataset by z standardization. The underlying reason is that some variables may have different scales and range, which in turn, influence the PCA results. The common practice is to standardize dataset having the mean of 0 and a standard deviation of 1. The standardized dataset or matrix Z is then expressed as:

$$Z = \frac{X - \mu}{\sigma},$$ (1)

where μ and σ are mean and standard deviation of the original dataset, respectively. After standardization, the covariance matrix C is computed to gain insight into variance and covariance of shared variables. The covariance matrix is then given by

$$C = \frac{1}{n-1} Z^T Z,$$ (2)

where Z^T is just transposed Z data matrix. After obtaining covariance matrix C, the eigenvalue decomposition is performed. In this step, the PCA computes eigenvalues and eigenvectors of the covariance matrix. The eigenvectors represent the direction of the maximum variance (PCs), while eigenvalues depict magnitude of the respected PC variance. The eigenvalue is expressed as:

$$Cv_i = \lambda_i v_i,$$ (3)

where v_i is the eigenvector and the λ_i is the eigenvalue of the covariance matrix. The eigenvectors obtained are ordered by decreasing eigenvalues. The first k principal components account for most of the variance in the data, where k is chosen based on the cumulative percentage of variance explained, in our case using a scree plot. The ordination into a new space is projected by k-dimensional feature space (where $k < p$) based on the selected eigenvectors that form new feature space via PCs. This transformation is obtained by

$$Y = ZV_k,$$ (4)

where V_k is the projected data matrix containing k selected (by order) eigenvectors, while Y represents dataset in the reduced feature space. It is important to note that simplification, i.e. transformation into lower-dimensional space is performed through linear combinations of uncorrelated principal components. Such transformation, although provides a more insightful understanding of the underlying structure of the dataset, is linearly dependent. Lastly, after obtaining lower dimensional space, our idea is to allocate (potential) clusters that share the most variance with projected reference measurement $x_{i,\,ref}$ from which we can obtain the factor set. However, instead of using additional machine learning algorithms, such as k-means, Agglomerative Hierarchical Clustering, or similar distance-based metrics, we turn attention to allocating the similarities between groups by using correspondence analysis techniques. The underlying reason for such an approach is that the method deals with categorical data, unlike clustering techniques that require numerical and continuous data for performing distance-based analysis (e.g., Euclidian distance).

2.3 Multiple correspondence analysis

The Correspondence Analysis (CA) is another dimensionality reduction technique that is, unlike PCA, used for categorical variables. Namely, the CA is performed by transforming original dataset of categorical variables into contingency tables. Thus, it accounts for count data between, for instance, two categorical variables. After obtaining frequencies via contingency table, the CA then projects the association between specific categories using Pearson's χ^2 distance metric [17]. The contingency table(s) consider I categories (rows) of specific variable p_i, where $i = 1, 2,…,I$, and J categories (columns) of variable p_j, where $j = 1, 2,… J$. Thus, the value in the dataset $x_{i,j}$ corresponds to the value of a variable with i rows and j columns, which consits of l instances. This can usually be performed by pivot tables accounting for frequencies between two categorical variables. Unlike proposed simple CA, which is constrained to the use of only two variables, the Multiple Correspondence Analysis (MCA) is able to perform singular value decomposition by using multiple categorical variables.

The process of MCA includes decomposition of matrix of indicator variables formed by transforming original set of categorical variables into a binary set of variables. In this case, however, the MCA computes only column profiles, i.e. J categories. Thus, the MCA performes weighted-PCA of the indicator matrix L, where $L = I \times H$. The levels of categorical variables p_h, such that $h = \{1, 2,…,H\}$, which is a binary representation of column profiles, i.e. categorical levels of variables, while I is number of observations $i = 1, 2,…,I$. If the number of categories are C_h, where $h = 1, 2,…,H$ then there is $C_h - 1$ dimensions. Performing the calculation of the correspondence matrix P, from an indicator matrix L, is derived by dividing each element from the number of observations I to obtain relative frequency, i.e. probabilities as

$$P = \frac{1}{I}L^T L \,, \tag{5}$$

where L^T is the transposed indicator matrix L. As same as in PCA, the singular value decomposition, which are core for dimensionality reduction techniques, is performed on the correspondence matrix P:

$$P = UDV^T, \tag{6}$$

where U and V are the orthogonal matrices representing singular vectors, which are actually coordinates in the reduced data space. The D is the diagonal matrix of containing singular values (square root of eigenvalues), which represent the inertia (a counterpart to variance in continuous variables) of an MCA, of each dimension. The total inertia of the dataset, and each singular value in D, represents the amount of inertia explained corresponding dimensions. Same as PCA, the cumulative inertia (scree plot) is commonly used to estimate the amount PCs retained. Lastly, we used an MCA biplot, by referring to clusters of PCA and binary factor variables, to come to conclusion which factors contribute the most to the reference value.

3. Results and discussion

Minitab 21® software was used for statistical analysis of experimental results. Each workpiece was measured according to the experiment plan and a total of 45 measurements were performed on one workpiece with 9 measurements at the central point. The adequacy of the model (R-sq) for all workpieces exceeds 90 %, which indicates that an adequate selection of factors was made and that the variations are well described by the obtained models. Analysis of variance (ANOVA) was used to determine statistically significant factors on the cylindricity of all workpieces. Fig. 3. proves the fact that the choice of measurement strategy can increase or decrease the measurement result by more than 100 %. The pictures show box-plot diagrams of the dispersion of the cylindricity measurement regarding different workpieces. In Fig. 4. p-values of individual factors are shown. It is possible to observe that the scanning speed and the stylus tip diameter are statistically significant for all workpieces since the p-value is below the significance threshold $\alpha = 0.05$ and a two-sided confidence interval of 95 %. The factor "number of points" is significant in the case of measuring cylindricity with the highest value of surface roughness (cylinder 11), while the filtration of sampled points is significant for cylinders 3 and 11. Also, a significant range of p-values for the number of points and filtration can be observed. Although the observation of the influence of individual factors leads to the conclusion that in most cases factors such as the number of points and the filtering of points should be omitted, the analysis of the interactions of individual factors leads to a different conclusion. For most workpieces, interactions of a number of points and filtration are significant.

Panel variable: Workpiece
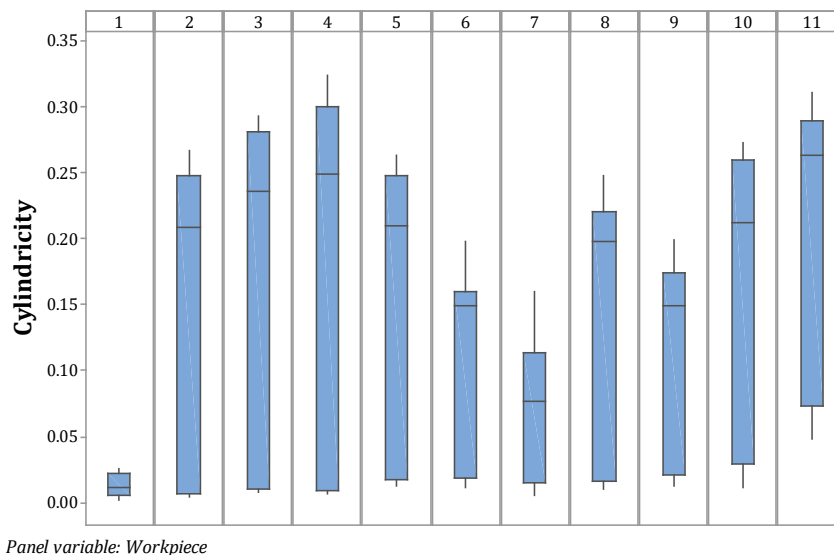
Fig. 3 Variability of measurement results depending on the applied measuring strategy

In Fig. 5 a combined diagram of the main effects for all workpieces is shown. The identification of optimal factor levels using the main effects diagram is usually a simple procedure where those factors that are closest to the optimal value defined by the user are selected. The optimal values, in this case, represent different values of cylindricity that were obtained on the roundtest (Table 1). As mentioned earlier, the workpieces differ significantly due to the different manufacturing processes. For example, in Fig. 6. reference values of cylindricity can be seen for cylinders 1, 5, and 11. For cylinder 1, for example, the optimal levels of the factors are: scanning speed – 5 mm/s; the number of points – 75; stylus tip diameter – 2; filter – cut off 8. For cylinder 5, the optimal factor levels are: scanning speed – 5 mm/s; the number of points – 25; stylus tip diameter – 5; filter – cut off 8. For cylinder 11, the optimal levels of factors are: scanning speed – 5 mm/s; the number of points – 25; stylus tip diameter – 5; filter - cut off 8. Identification of an optimal combination of factor levels for all workpieces is a serious challenge. When choosing a measurement strategy, the CMM programmer, in most cases, will not know the values of the quality characteristics such as the form error and the quality of the machined surface, and this way of finding the optimal values of the factors requires significant resources.

p - values versus individual factors for different cylinders

	Scanning speed	Number of points	Stylus tip diameter	Filter
p-values Cylinder 1	0.000	0.925	0.000	0.899
p-values Cylinder 2	0.000	0.582	0.000	0.648
p-values Cylinder 3	0.000	0.162	0.000	0.041
p-values Cylinder 4	0.000	0.062	0.000	0.089
p-values Cylinder 5	0.000	0.723	0.000	0.552
p-values Cylinder 6	0.000	0.230	0.000	0.548
p-values Cylinder 7	0.000	0.338	0.001	0.071
p-values Cylinder 8	0.000	0.692	0.000	0.710
p-values Cylinder 9	0.000	0.303	0.000	0.667
p-values Cylinder 10	0.000	0.619	0.000	0.248
p-values Cylinder 11	0.000	0.011	0.000	0.012

Fig. 4 *p*-values of individual factors for different cylinders

Fig. 5 Combined main effects plot for different workpieces

This study showed that the choice of factor levels and their percentage in the total variability of the measurement results depends on the real characteristics of the workpiece, even though they have the same nominal characteristics. In practical measurements, the real geometry of the workpiece is unknown, i.e. it is not determined before measuring on the CMM. To estimate optimal solution in the selection of the measurement strategy for measuring the cylindricity, we rely on unsupervised learning and dimensionality reduction tools of PCA (Principal Component Analysis) for quantitative data, and MCA (Multiple Correspondence Analysis) for qualitative data.

3.1 Principal component analysis result

The PCA scree plot (Fig. 6) shows that the first two components explain up to 83.8 % of the variance, while the first three components show 98.9 % of variance, suggesting excellent ordination properties. Interpreting the results in combination with all three components suggest that much of the information is retained. Indeed, observing the distribution of the points on PC' axes, the distributions (Fig. 7) suggest highest frequency points in PC1 and PC2. In fact, there are potential

clusters observed by frequencies all three components, while PC4-PC6 retains only 1.1 % of the variance.
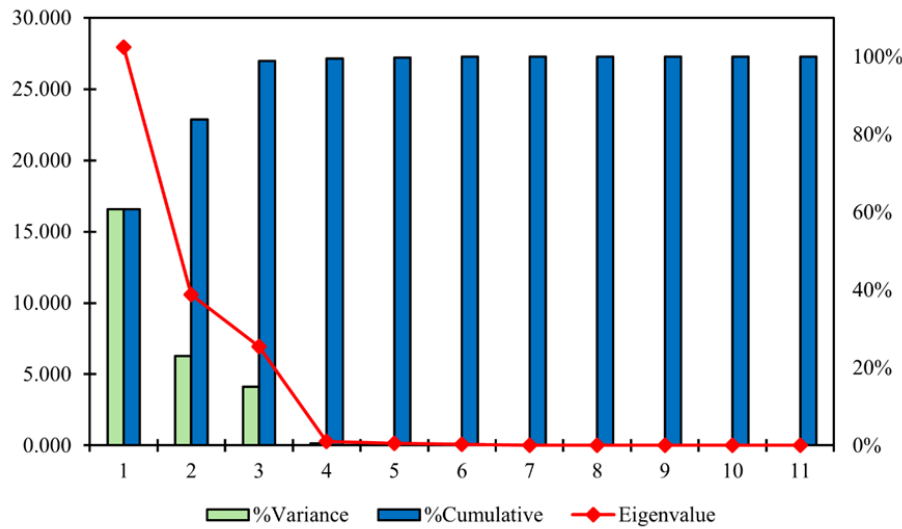
Fig. 6 Combined main effects plot for different workpieces

After projecting the points on the first three PCs, the results suggest three clusters (Fig. 8). In addition, we conducted additional exploratory analysis in the rest of the components, and the spread of the data did not suggest additional clustering, which is suggested by the distribution plots (Fig. 7). Given the reference value X46_Ref, our interest is to allocate clusters that share the most variance with this point. Thus, in all three ordinations these points are retained within the Cluster 1 (points X5, X10, X16, X20, X21, X23, X24, X27, X31, X32, X37, X44, X46).
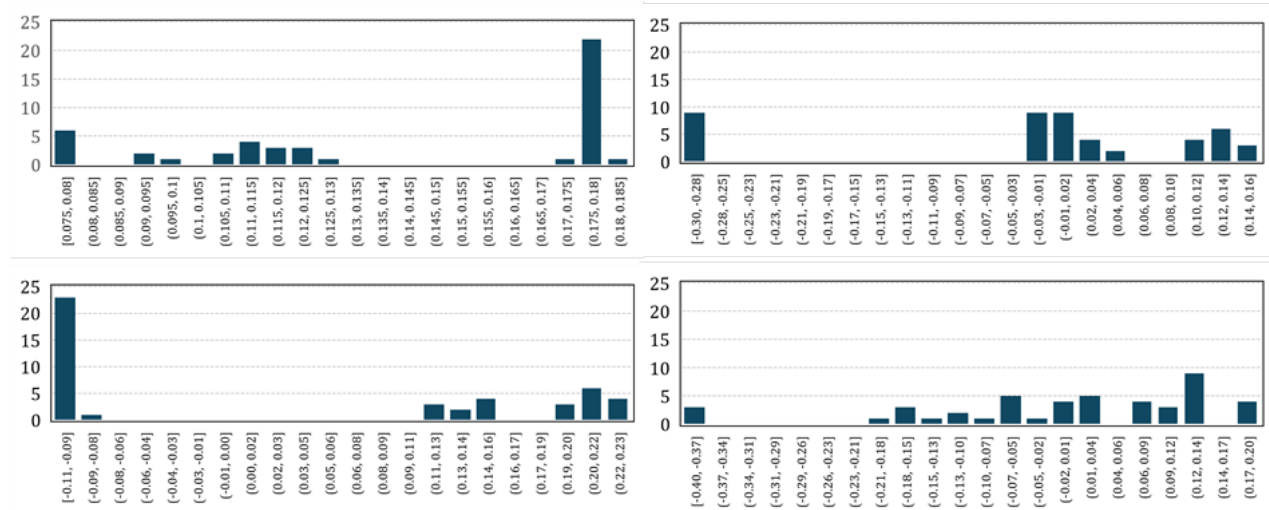
Fig. 7 Distribution plots of PC1 (top-left); PC2 (bottom-left); PC3 (top-right); PC4 (bottom-left)

After performing the PCA, labelling of clusters is performed (Fig. 9). Measurement speed 5 was perfectly associated with Cluster 1 (Fig. 9b). The results of the filter (Fig. 9a), and number of scanning points (Fig. 9d) did not provide good clustering properties, while using different measuring diameters (Fig. 9c), the results show almost perfect classification to Cluster 2. Nevertheless, given that performed clustering using factors from experimental setup did not provide full understanding of factors associating with reference value, we performed an MCA. The idea of performing an MCA analysis is to allocate factors that have the highest contribution to the associated measurements with respect to the reference value of measurements.
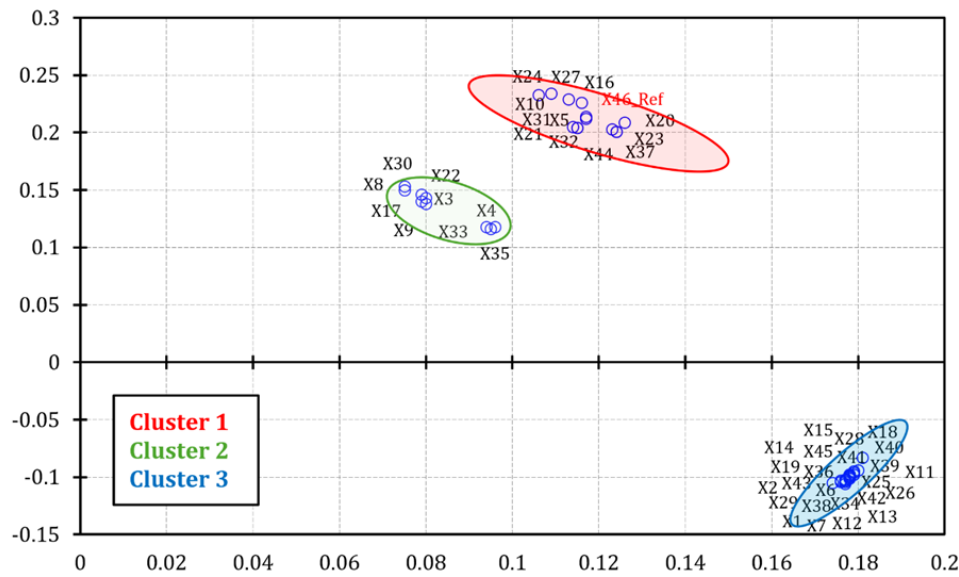
Fig. 8 Ordination using PC1 (x-axis) and PC2 (y-axis)
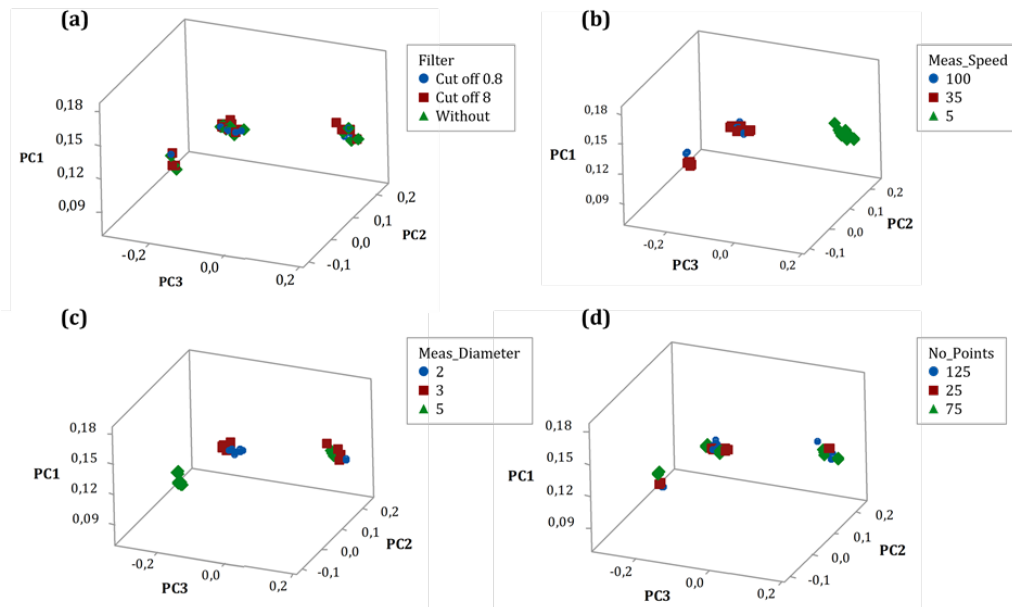
Fig. 9 PCA 3D ordination using classes (a) filter; (b) measurement speed; (c) diameter; (d) number of scanning points

3.2 Multiple correspondence analysis results

After performing a sequential construction of axes with the MCA, the eigenvalue decomposition shows that the first two components capture 38.75 % of the inertia (Fig. 10). Given that our main intent of using an MCA is to allocate factors associated with the Cluster 1, which contains the reference value, we relied on the quality of the interpretation and contribution associated with this cluster (Table 3). The MCA shows > 90 % quality of interpretation of clusters, where measurement speed 5 is perfectly association with Cluster 1, suggesting that measurement speed 5 has the highest impact for aligning the results of measurement to the reference value.

Fig. 10 MCA ordination biplot using PC1 (x-axis) and PC2 (y-axis) including the scree plot (top-left)

Table 3 Association between factors and clusters

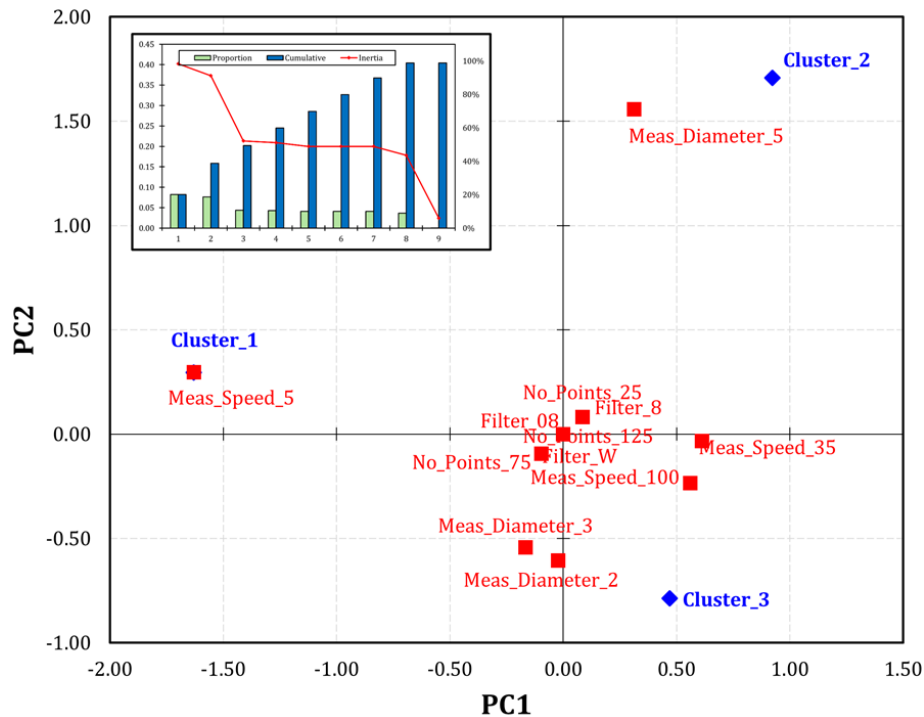| Name	Qual	Mass	Inert	Coord	Corr	Contr	Coord	Corr	Contr
Cluster_1	0.998	0.053	0.073	-1.630	0.966	0.352	0.295	0.032	0.012
Cluster_2	0.942	0.040	0.080	0.923	0.213	0.085	1.708	0.729	0.313
Cluster_3	0.961	0.107	0.047	0.469	0.251	0.058	-0.788	0.710	0.178
Meas_Speed_5	0.998	0.053	0.073	-1.630	0.966	0.352	0.295	0.032	0.012
Meas_Speed_35	0.328	0.093	0.053	0.611	0.327	0.087	-0.034	0.001	0.000
Meas_Speed_100	0.134	0.053	0.073	0.560	0.114	0.042	-0.236	0.020	0.008
No_Points_25	0.005	0.053	0.073	0.085	0.003	0.001	0.082	0.002	0.001
No_Points_75	0.016	0.093	0.053	-0.097	0.008	0.002	-0.093	0.008	0.002
No_Points_125	0.005	0.053	0.073	0.085	0.003	0.001	0.082	0.002	0.001
Meas_Diameter_2	0.134	0.053	0.073	-0.023	0.000	0.000	-0.607	0.134	0.053
Meas_Diameter_3	0.282	0.093	0.053	-0.166	0.024	0.006	-0.542	0.257	0.074
Meas_Diameter_5	0.916	0.053	0.073	0.313	0.036	0.013	1.556	0.881	0.346
Filter_W	0.000	0.067	0.067	0.000	0.000	0.000	0.000	0.000	0.000
Filter_08	0.000	0.067	0.067	0.000	0.000	0.000	0.000	0.000	0.000
Filter_8	0.000	0.067	0.067	0.000	0.000	0.000	0.000	0.000	0.000

Note: Qual = Quality of interpretation; Mass = Proportion of the presence of the variable in complete dataset; Inert = Inertia; Corr = Correlation; Contr = Contribution to the cumulative inertia.

3.3 Independent Chi-square and ANOVA

To validate our findings we report χ^2 independence test. Namely, we used contingency tables formed between clusters and categorical variables. The obtained findings did not show any statistically significant dependency between clusters and filter. In fact, there was a spread of the same values across all categories ($\chi^2 = 0$; $p = 1.0$). The test statistic considering number of points also did not provide any statistically significant results ($\chi^2 = 0.804$; $p = 0.938$). However, the test considering measuring diameter reports statistically significant association with clusters ($\chi^2 = 33.549$; $p < 0.001$; VS-MPR = 28713.66). Most importantly, as observed in the MCA biplot (Fig. 11), the statistically significant relationship exists between clusters and measurement speed ($\chi^2 = 45.067$; $p < 0.001$; VS-MPR = $4.931 \cdot 10^6$), with reported high strength of the relationship (Cramer's $V = 0.708$). The performed analysis suggests that low measurement speed directly contributes to the improved cylindricity measurement "accuracy".
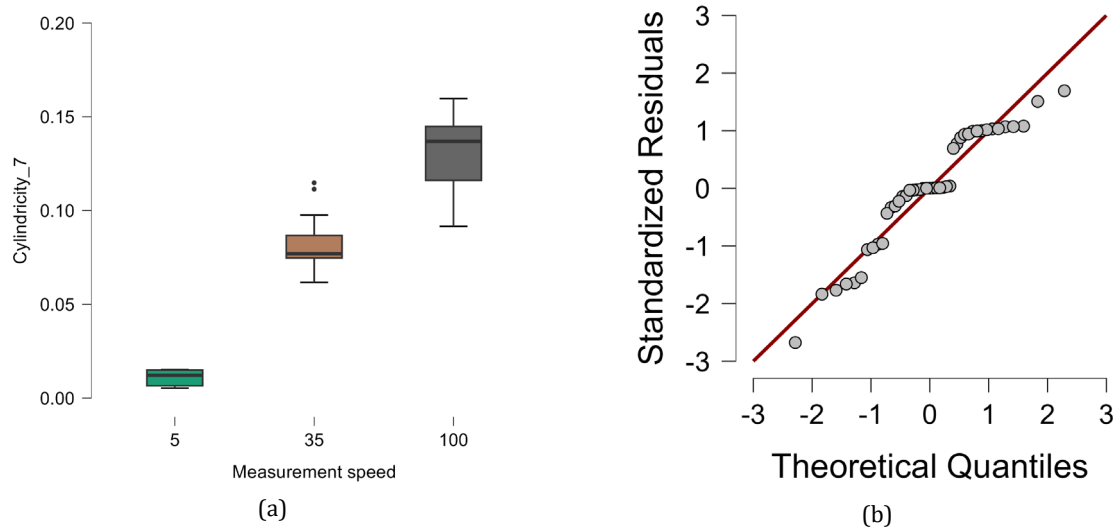
(a)   (b)

Fig. 11 Box plot of measurement speed (a) and QQ plot (b)

Table 4 ANOVA Analysis

Cases	SSQ	df	MSQ	F	p	VS-MPR	η^2	η^2_r	ω^2
Measurement speed	0.589	2	0.294	225.971	< 0.01	$2.141*10^{20}$	0.915	0.915	0.909
Residuals	0.055	42	0.001						

SSQ = Sum of Squares; MSQ = Mean Sum of Squares; VS-MPR = Volk-Sellke Maximum p—Ratio (odds in favor of H_1 over H_0, estimated as 1/-e*p*log(p))

Table 5 Bootstrapped Post Hoc Comparisons – Measurement Speed

Meas. speed	Mean	95%L	95%U	SE	bias*	t	Cohen's d	p_{Tukey}	p_{Bonf}
5 : 35	-0.26	-0.276	-0.235	0.010	> 0.001	-19.665	-7.116	< 0.01	< 0.01
5 : 100	-0.26	-0.276	-0.245	0.008	> 0.001	-17.661	-7.251	< 0.01	< 0.01
35 : 100	-0.06	-0.029	0.023	0.013	> 0.001	-0.135	-0.135	0.927	1.00

***p < 0.001; * Bias corrected accelerated. Bootstrapping is based on 1000 succesful replicates. Mean difference estimate is based on the median of the bootstrap distribution. p value and confidence intervals adjusted for comapring family of 3 estimates (confidence intervals corrected using the Tukey method). The Bonferroni correction for multiple comparison test is also report as p_{Bonf}.

Finally, relying on the results from the obtained cluster 1 (with measurement speed 5), we first subtracted values from the reference point within the cluster and use sum and average values with absolute deviation from the reference point (Fig. 12). What can be concluded from Fig. 12 is that low scanning speeds will give results closer to the reference value, while a uniform solution cannot be obtained for other factors. In general, it can be concluded that there are strong interactions between the number of points, the size of the diameter of the measuring probe, and filtering, and the operator should follow the guidelines from ISO 12180-2:2011 [27] regarding the Nyquist criterion when measuring.
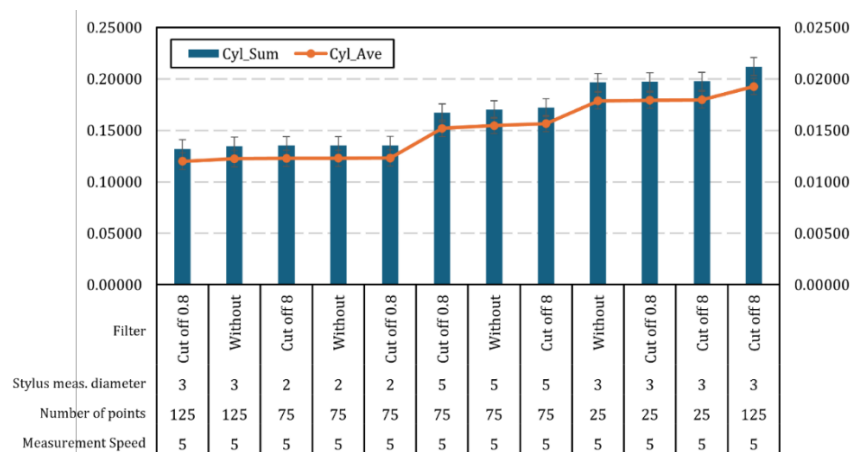

Fig. 12 Box plot of measurement speed (a), and QQ plot (b)

4. Conclusion

This paper deals with a practical industrial problem that CMM operators face every day, i.e. choosing a measurement strategy when measuring form error. A "bad" measurement strategy can lead to fatal decisions about rejecting good workpieces or letting bad workpieces pass the inspection. Since the complete specification operator is generally not indicated on the technical documentation, it is left to the machine operators to choose a strategy based on experience, while keeping in mind the economic imperative in terms of measurement time. Also in this study, the results of the measurements were compared with the reference value, and in this way, it was possible to reach the optimal levels of the considered factors. The general conclusion would be that when measuring cylindricity on a CMM, the scanning speed should be low and the choice of other factors must consider the interaction of the tip diameter of the measuring probe, the number of measuring points and filtering with the real geometry of the workpiece. As directions for future analysis, the research could be extended to other geometric deviations such as straightness, flatness, and circularity. A dependency could also be established when choosing the tip of the measuring probe and the quality of the machined surface of the workpieces, it can be determined how many points need to be measured and which filter to use.

Acknowledgement

References

[1] Jotić, G., Štrbac, B., Toth, T., Blanuša, V., Dovica, M., Hadžistević, M. (2023). The analysis of metrological characteristics of different coordinate measuring systems, *Tehnički Vjesnik – Technical Gazette,* Vol. 30, No. 1, 32-38, doi: 10.17559/TV-20220204091212.

[2] Dube, L., Gupta, K. (2023). Lean manufacturing based space utilization and motion waste reduction for efficiency enhancement in a machining shop: A case study, *Applied Engineering Letters*, Vol. 8, No. 3, 121-130, doi: 10.18485/aeletters.2023.8.3.4.

[3] Grazion, S., Spiryagin, V., Erofeev, M., Kravchenko, I., Kuznetsov, Y., Mukomela, M., Velichko, S., Ašonja, A., Kalashnikova, L. (2022). Diagnostics of defect detection in the initial stages of structural failure using the acoustic emission method of control, *Applied Engineering Letters*, Vol. 7, No. 2, 45-53, doi: 10.18485/aeletters.2022.7.2.1.

[4] Shen, Y., Ren, J., Huang, N., Zhang, Y., Zhang, X., Zhu, L. (2023). Surface form inspection with contact coordinate measurement: A review, *International Journal of Extreme Manufacturing*, Vol. 5, No. 2, Article No. 022006, doi: 10.1088/2631-7990/acc76e.

[5] JCGM 100:2008 (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement,* First edition, International Organization of Legal Metrology, Paris, France.

[6] ISO/TS15530-3:2011. *Geometrical products pecifications (GPS) – Coordinate measuring machines (CMM): technique for determining the uncertainty of measurement,* Part 3: Use of calibrated workpieces or standards, Geneva, Switzerland.

[7] ISO/TS15530-4:2008. *Geometrical product specifications (GPS) – Coordinate measuring machines (CMM): technique for determining the uncertainty of measurement,* Part 4: Evaluating task-specific measurement uncertainty using simulation, Geneva, Switzerland.

[8] Aggogeri, F., Barbato, G., Barini, E.M., Genta, G., Levi, R. (2011). Measurement uncertainty assessment of coordinate measuring machines by simulation and planned experimentation, *CIRP Journal of Manufacturing Science and Technology*, Vol. 4, No. 1, 51-56, doi: 10.1016/j.cirpj.2011.01.007.

[9] Štrbac, B., Ačko, B., Havrlišan, S., Matin, I., Savković, B., Hadžistević, M. (2020). Investigation of the effect of temperature and other significant factors on systematic error and measurement uncertainty in CMM measurements by applying design of experiments, *Measurement,* Vol. 158, Article No. 107692, doi: 10.1016/j.measurement.2020.107692.

[10] Acko, B., Brezovnik, S., Crepinsek Lipus, L., Klobucar, R. (2015). Verification of statistical calculations in interlaboratory comparisons by simulating input datasets, *International Journal of Simulation Modelling*, Vol. 14, No. 2, 227-237, doi: 10.2507/IJSIMM14(2)4.288.

[11] Lipus, L.C., Acko, B., Tompa, J. (2022). Experimental determination of influences on a gauge block's stack length, *Advances in Production Engineering & Management*, Vol. 17, No. 3, 339-349, doi: 10.14743/apem2022.3.440.

[12] Ricci, F., Scott, P.J., Jiang, X. (2013). A categorical model for uncertainty and cost management within the Geometrical Product Specification (GPS) framework, *Precision Engineering*, Vol. 37, No. 2, 265-274, doi: 10.1016/j.precisioneng.2012.09.005.

[13] Weckenmann, A., Knauer, M., Kunzmann, H. (1998). The influence of measurement strategy on the uncertainty of CMM-measurements, *CIRP Annals*, Vol. 47, No. 1, 451-454, doi: 10.1016/S0007-8506(07)62872-8.

[14] Barini, E.M., Tosello, G., De Chiffre, L. (2010). Uncertainty analysis of point-by-point sampling complex surfaces using touch probe CMMs: DOE for complex surfaces verification with CMM, *Precision Engineering*, Vol. 34, No. 1, 16-21, doi: 10.1016/j.precisioneng.2009.06.009.

[15] Raghunandan, R., Venkateswara Rao, P. (2008). Selection of sampling points for accurate evaluation of flatness error using coordinate measuring machine, *Journal of Materials Processing Technology*, Vol. 202, No. 1-3, 240-245, doi: 10.1016/j.jmatprotec.2007.09.066.

[16] Štrbac, B., Rodić, D., Delić, M., Savković, B., Hadžistević, M. (2021). Investigation of functional dependency between the characteristics of the machining process and flatness error measured on a CMM, *Measurement Science Review*, Vol. 21, No. 6, 158-167, doi: 10.2478/msr-2021-0022.

[17] Orošnjak, M., Šević, D. (2023). Benchmarking maintenance practices for allocating features affecting hydraulic system maintenance: A West-Balkan perspective, *Mathematics*, Vol. 11, No. 18, Article No. 3816, doi: 10.3390/math11183816.

[18] Özel, T. (2006). Precision tracking control of a horizontal arm coordinate measuring machine in the presence of dynamic flexibilities, *The International Journal of Advanced Manufacturing Technology*, Vol. 27, 960-968, doi: 10.1007/s00170-004-2292-3.

[19] Stępień, K. (2015). An analysis of influence of sampling strategy and scanning speed on estimation of straightness and flatness deviations with CMMs, *Advanced Technologies in Mechanics*, Vol. 2, No. 2-3, 2-7.

[20] Ali, S.H.R. (2014). Performance investigation of CMM measurement quality using flick standard, *Journal of Quality and Reliability Engineering,* Vol. 2014, No. 1, Article No. 960649, doi: 10.1155/2014/960649.

[21] Urban, J., Beranek, L., Koptiš, M., Šimota, J., Košťák, O. (2020). Influence of CMM scanning speed and inspected feature size on an accuracy of size and form measurement, *Manufacturing Technology*, Vol. 20, No. 4, 538-544, doi: 10.21062/mft.2020.074.

[22] Magdziak, M. (2020). Determining the strategy of contact measurements based on results of non-contact coordinate measurements, *Procedia Manufacturing*, Vol. 51, 337-344, doi: 10.1016/j.promfg.2020.10.048.

[23] Rajamohan, G., Shunmugam, M.S., Samuel, G.L. (2011). Effect of probe size and measurement strategies on assessment of freeform profile deviations using coordinate measuring machine, *Measurement*, Vol. 44, No. 5, 832-841, doi: 10.1016/j.measurement.2011.01.020.

[24] Malburg, M.C. (2002). Fitting, filtering and analysis: Feature extraction in dimensional metrology applications, *International Dimensional Workshop*, *https://digitalmetrology.com/Papers/IDW2002-Slides.pdf*

[25] Lou, S., Brown, S.B., Sun, W., Zeng, W., Jiang, X., Scott, P.J. (2019). An investigation of the mechanical filtering effect of tactile CMM in the measurement of additively manufactured parts, *Measurement*, Vol. 144, 173-182, doi: 10.1016/j.measurement.2019.04.066.

[26] Li, K., Li, D., Ma, H.Q. (2023). An improved discrete particle swarm optimization approach for a multi-objective optimization model of an urban logistics distribution network considering traffic congestion, *Advances in Production Engineering & Management,* Vol. 18, No. 2, 211-224, doi: 10.14743/apem2023.2.468.

[27] ISO 12180-2:2011, *Geometrical product specifications (GPS) – Cylindricity,* Part 2: Specification operators, Geneva, Switzerland.

Impact of fairness concerns on resource-sharing decisions: A comparative analysis using evolutionary game models in manufacturing enterprises

Xu, W.[a], **Xu, S.**[a], **Liu, D.Y.**[b], **Awaga, A.L.**[a], **Rabia, A.**[c], **Zhang, Y.Y.**[d,*]

[a]School of Management, Shenyang University of Technology, Shenyang, P.R. China
[b]Liaoning Road and Bridge Construction Group Co., LTD, Shenyang, P.R. China
[c]Department of Commerce, University of the Punjab, Gujranwala, Pakistan
[d]School of Public Health, Dalian Medical University, Dalian, P.R. China

ABSTRACT

In recent years, the resource-sharing model has developed rapidly, and manufacturing enterprises have paid more attention to profit distribution and resource investment. Studying the role of fairness concern behaviour in enterprise resource-sharing decisions can provide strategic references for enterprises to implement resource-sharing. This article considers the fairness concern behaviour of resource-sharing enterprises and constructs two symmetric evolutionary game models for resource-sharing. Based on the resource-sharing strategy when the enterprise is fair and neutral, a comparative study is conducted on the impact of fairness concern behaviour on the resource-sharing equilibrium strategy. MATLAB is used for numerical simulation. The analysis results indicate that: (1) the system cost coefficient, information level coefficient, information loss coefficient, resource conversion ability coefficient, resource collaboration ability coefficient, and penalty factor all have an impact on the strategic evolution results of both parties in the game enterprise, (2) fairness concerns include heterogeneity, directionality, and reinforcement, which affect the resource-sharing strategies of enterprises. This study provides a reference for shared decision-making in manufacturing enterprises and provides recommendations for their business development.

1. Introduction

In recent years, the development of the sharing economy under the influence of the digital economy has activated the economic vitality of the China's market and received strong support from the government, the digital economy to accelerate the flow of information, technology, capital and other innovation factors to promote enterprise development [1]. Especially in China's rapid development of the manufacturing industry, has occupied a certain position in the world based on the sharing manufacturing model in the era of digital economy came into being. As an important sector of China's economic development, the manufacturing industry is the key to high-quality

economic development and has a significant influence on economic transformation and upgrading [2]. The manufacturing industry itself is increasingly complex in production and manufacturing, making it difficult to maintain an advantage in market competition solely on its own. However, through the shared manufacturing model, resource-sharing among manufacturing enterprises reduces costs, is more efficient, and is more secure. The sharing platform serves as a "matchmaker" to connect the supply and demand sides of manufacturing resources [3], fully realizing the rational utilization of resources for manufacturing enterprises and improving resource utilization efficiency. The resource-sharing behaviour involving multiple, diverse, and multi-manufacturing links gives impetus for the development of enterprises, drives the sustainable development [4] of manufacturing enterprises, and ultimately promotes the transformation and upgrading of the manufacturing economy.

Both parties involved in manufacturing resource-sharing are not completely rational. Different manufacturing enterprises compete fiercely and generate irrational behaviours in the platform market, and such irrational behaviours are particularly evident in the competition among manufacturing enterprises with predominantly homogeneous or similar resources, and the same applies to enterprises that share resources for manufacturing. In this paper, enterprises participating in resource-sharing manufacturing are referred to as symmetric manufacturing enterprises, which manufacture products by sharing their own resources with other resource-sharing enterprises. Manufacturing companies hope to receive corresponding resource returns for the resources invested in shared platforms. Once the expected returns are exceeded or not met, it will affect the company's shared manufacturing strategy. At the same time, when manufacturing enterprises perceive that other enterprises participating in resource-sharing receive higher profits than themselves, it will also have an impact on their sharing behaviour, which is not conducive to the development of shared manufacturing model. This type of enterprise's concern for the difference in profits between itself and other enterprises is a concern for fairness [5]. Fairness concern behaviour is mainly its decision-making process not only seek to maximise their own interests, but also concerned about the benefits of other manufacturing entities and the fairness of the distribution of issues. Fairness concern behaviour has been confirmed by multiple studies, and there is a fairness concern in the current process of manufacturing resource-sharing. This article studies the resource-sharing decisions of manufacturing enterprises under fairness neutrality and fairness concerns. Based on evolutionary game theory, two symmetric evolutionary game models for enterprise resource-sharing are constructed. Taking as a reference the resource-sharing strategy when the enterprises are fairness-neutral, the impact of fairness concern behaviour on the resource-sharing equilibrium strategy is compared and the role of influencing factors is analysed using numerical simulations.

The specific contributions of this article include: (1) investigating the equilibrium strategy of enterprise resource-sharing under fairness concern behaviour, (2) the different resource-sharing decisions of enterprises under fairness neutrality and fairness concerns, (3) investigating the significant impact of fairness concern behaviour on resource-sharing and proposing a method of controlling influencing factors to better achieve enterprise resource-sharing.

2. Literature review

In this section we review existing studies and provide a literature review of shared manufacturing models, fairness concerns behaviour and evolutionary game theory to lay the foundation for the research in this paper.

2.1 Research on shared manufacturing models

In the era of sharing economy development [6], the role of the organisation as an infrastructure provider [7] has been gradually enlarged, shared factory models [8] and manufacturing sharing models [9] have been developed. These promote resource allocation and maximize resource utilization. Billo *et al.* [10] proposed a shared manufacturing model based on shared manufacturing to help large-scale enterprises to consolidate their facilities into shared manufacturing facilities. Khoa *et al.* [11] presented a study of the sharing economy based on perceived value. Based on the

shared manufacturing model, an important challenge facing the sharing economy is how to allocate resources so that both parties of the manufacturing enterprises can obtain the maximum benefits. For the benefit distribution problem of manufacturing enterprises after resource-sharing, Liu *et al.* [12] explored the benefit distribution under the shared manufacturing model by designing a resource optimisation model that includes the benefits of the shared manufacturing platform. Local organisations acted as intermediate 'mediators' in the shared manufacturing process, linking networks within and outside the cluster. It can be seen that the sharing platform in the shared manufacturing model can become an organisation for constructing enterprise alliances and a channel for enterprises to share manufacturing resources.

2.2 Research on fairness concerns behaviour

Early academic researchers mostly assume that people are rational and seek to maximise their interests. However, the distribution of benefits has been found to influence behavioural choices, triggering the fairness concern mechanism, which breaks through the limitations of the existing results, which are mostly fully rational, and makes the results of behavioural decisions more accurate. Therefore, more and more studies have questioned the assumption of complete rationality. For example, the trust game and the gift exchange game show that people care not only about themselves, but about fairness. The fairness concern exists in the resource-sharing strategy of manufacturing enterprises, Ho and Zhang [13] introduced fairness concerns into the supply chain, proving their actual existence and providing a descriptive utility function for fairness concerns. There are also some scholars who incorporate fairness concern into the study of pricing strategy and revenue coordination of supply chain subjects, for example, Wang *et al.* [14] investigated the impact of fairness concern behaviour on retail pricing strategies in supply chains from the perspective of the existence of retailers' fairness concern behaviour, and Zhang *et al.* [15] introduced fairness concern to solve the decision-making problem of green manufacturing in supply chains. In addition, fairness concern behaviour has also been introduced into group decision-making behaviour [16] and cost consensus studies [17].

2.3 Research on evolutionary game modelling of resource-sharing

Scholars have conducted a lot of research on data sharing behaviour, such as Wang *et al.* [18] who constructed an evolutionary game model of data sharing among cooperative enterprises considering the role of government involvement, and Cao *et al.* [19] who considered an evolutionary game model of data sharing among enterprises in the context of smart manufacturing. Among them, Liu *et al.* [20] found that revenue and trust level are key in determining sharing behaviour, while the government plays an important role in regulating data sharing. Zhu *et al.* [21] explored the positive impact of data transformation capability on sharing behaviour. At the same time, creating the right incentives can address the risks of data sharing, such as Zhang *et al.* [22] based on the user characteristics in the context of the smart grid constructed a dynamic incentive mechanism, the effective incentives for data sharing to provide a new way of thinking. The level of information synergy [23] will have an important impact on the development of resource-sharing in manufacturing enterprises. With the development of advanced technology, the role of the platform in the data sharing behaviour is very important, such as Liu *et al.* [24] explored the importance of platform empowerment in the data sharing process. Resource-sharing can effectively promote the sharing of enterprise advantageous resources and reduce the development risk.

In summary, through the existing literature found that data sharing is a hot issue in current research, but most of the research is to consider the data sharing between different subjects, and rarely discuss the symmetric enterprise resource-sharing problem under the shared manufacturing mode from the perspective of fairness concern, and for the act of resource-sharing, the manufacturing enterprise cannot achieve its benefit target maximisation, and will also pay attention to the distribution of benefits at this time, and is not completely rational itself. The "limited rationality" assumption of evolutionary game theory is in line with the research background of this paper.

3. An evolutionary game model of resource-sharing in symmetric enterprises

3.1 Problem description

This article considers a shared supply chain consisting of A-class and B-class manufacturing re-source enterprises, where the enterprises have differing resources. The middle end of a shared supply chain is a shared platform, which belongs to a virtual resource pool, mainly composed of manufacturing resources, production services, digital technology [25], and other components. To help manufacturing enterprises achieve better resource-sharing, it is necessary to connect with enterprises that provide and demand many manufacturing resources. The connected enterprises need to continuously update their core networks and information facilities to achieve real-time sharing of manufacturing resources. Manufacturing resources refer to the collection of elements required by enterprises to complete the entire lifecycle activities of product design, manufactur-ing, assembly, sales, management, etc., roughly divided into hard manufacturing resources and soft manufacturing resources. Hard manufacturing resources mainly include various types of manufacturing equipment, computing equipment, and materials involved in manufacturing activ-ities. Soft manufacturing resources refer to the software tools, production models, knowledge, and data required in the manufacturing process. These manufacturing resources (especially hard manufacturing resources) need to use the Internet of Things and information systems to achieve the acquisition and release of resources, to capture and control manufacturing resources through the Internet. The process of implementing resource-sharing in manufacturing enterprises, as shown in Fig. 1, is a dynamic game process involving a series of technical and non-technical issues such as resource conversion and profit sharing. Transferring production processes that are not proficient to other enterprises, production costs and technical risks can be reduced. For manufac-turing enterprises to implement resource-sharing, these issues must be resolved.
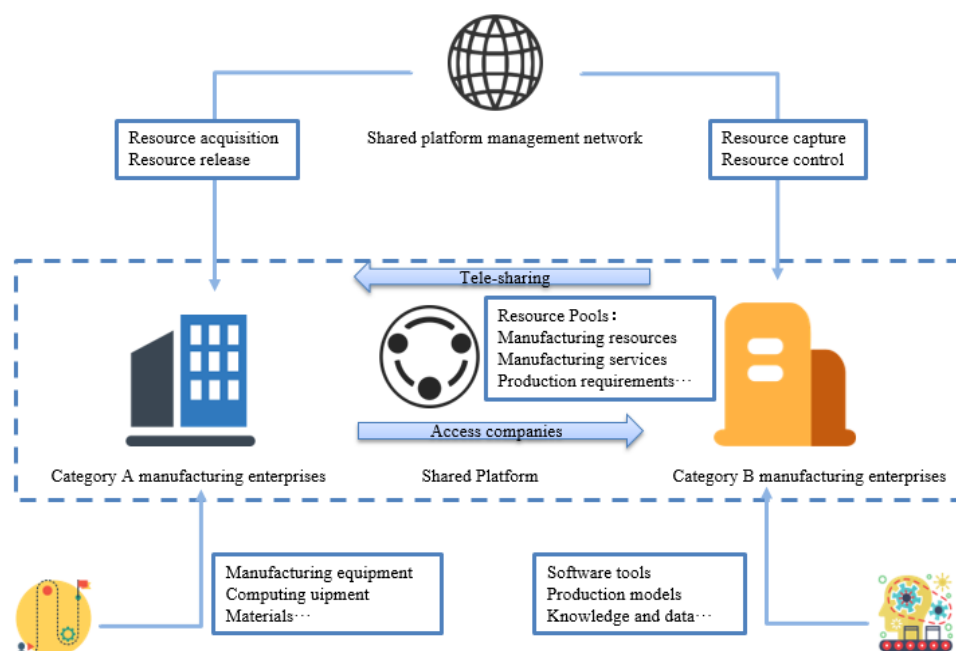
Fig. 1 Manufacturing enterprise flowchart

3.2 Model assumptions and relevant variables

The following hypotheses are proposed to better explain the research questions in this paper and the symbols and meanings of the parameters involved are shown in Table 1.

Assumption 1: A-class manufacturing enterprises mainly provide all kinds of man-ufacturing equipment and material resources used in the manufacturing process, etc. B-class manufacturing enterprises mainly provide all kinds of software tools, product models, professional knowledge and process data necessary for manufacturing activities.

Assumption 2: The sharing platform has unlimited management capacity and can access a large number of manufacturing enterprises to provide them with homogeneous services. At the same

time the population size of the two types of enterprises is homogeneously mixed, and both enterprises are finitely rational and can adopt strategies independently.

Assumption 3: There is no heterogeneity between the products produced by manufacturing enterprises using shared resources or shared manufacturing by multiple enter-prises and the products produced by manufacturing enterprises individually, the market demand is homogeneous, and the resource differences between manufacturing enterprises are reasonable.

Table 1 Parameters symbols and descriptions

Symbol	Descriptions
c	System cost coefficient
τ_A	Information level coefficient
τ_B	Information loss coefficient
t	Resource conversion ability coefficient
s	Resource collaboration ability coefficient
p	Penalty factor
λ	Fairness concern coefficient of A-class enterprises
η	Fairness concern coefficient of B-class enterprises
K	Volume of shared manufacturing resources
ω	Unit value of resources

3.3 Model analysis

This paper examines the sharing strategies of two types of manufacturing enterprises.

(1) When the strategy space is (non-sharing, non-sharing), A-class and B-class enterprises choose not to share resources. The basic income of A-class enterprises is $\pi_A = R_A$. The basic income of B-class enterprises is $\pi_B = R_B$.

(2) When the strategy space is (sharing, non-sharing), A-class enterprises choose resource-sharing, and B-class enterprises choose resource non-sharing. A-class enterprises utilize the Internet of Things and information systems to connect the involved hard manufacturing resources to the shared platform, and the costs incurred in upgrading and maintaining the existing system are recorded as system cost is $SC_A = C_A \times K_A \times \omega_A$. At the same time, as manufacturing enterprises upgrade and maintain their systems, their information management level improves, and the corresponding benefits will also change, recorded as information level benefit is $IR_A = \tau_A \times K_A \times \omega_A$. The basic income of B-class enterprises is $\pi_B = R_B$.

(3) When the strategy space is (non-sharing, sharing), A-class enterprises choose not to share resources, and B-class enterprises choose to share resources. B-class enterprises have a high level of information management in sharing soft manufacturing resources, and the soft manufacturing resources they provide do not require the use of the Internet of Things and information systems. Therefore, the cost invested by B-class enterprises is lower than that of A-class enterprises. The system cost of B-class enterprises is recorded as $SC_B = C_B \times K_B \times \omega_B$ ($C_A > C_B$). Due to the core technological advantages of the resources shared by B-class enterprises, once sharing is open, the enterprise will have to bear the risk of information loss, which recorded as the cost of information loss is $IR_B = \tau_B \times K_B \times \omega_B$. The basic income of A-class enterprises still is $\pi_B = R_B$.

(4) When the strategy space is (sharing, sharing), A-class and B-class enterprises choose resource-sharing. In addition to the above parts, there is also conversion revenue, which refers to manufacturing enterprises converting platform-shared resources into their profits based on their own production, learning, and conversion capabilities. The size of the revenue is influenced by the ability to convert resources. The conversion benefits of A-class and B-class enterprises are $TR_A = t_A \times R_A$ and $TR_B = t_B \times R_B$. Collaborative revenue refers to the revenue generated by manufacturing enterprises sharing resources with other enterprises for production through a shared platform, and the size of the revenue is influenced by the ability of resource collaboration. The collaborative benefits of A-class and B-class enterprises are $SR_A = s_A \times K_A \times \omega_A$ and $SR_B = s_B \times K_B \times \omega_B$. Additional benefits refer to manufacturing enterprises sharing resources through shared platforms, which, due to the improvement of information technology and collaborative manufactur-

ing, enhance their reputation and social recognition. This will improve social relations and increase social capital for manufacturing enterprises, which is also a part of the shared benefits of manufacturing resources between both parties and is recorded as additional benefits: R_0 ($R_0 > 0$). As well as penalty costs, both parties of the enterprise will automatically form a contract or agreement in the process of manufacturing resource-sharing. For enterprises that engage in sharing retention behaviour and do not comply with sharing agreements, the platform will take punitive measures to improve platform management and maintain platform construction. Record this as penalty cost: $PC_i = p_i \times R_i$ ($i = A, B$).

Therefore, for the shared benefits of A-class manufacturing enterprises $\pi_A = R_A + IR_A - SC_A + TR_A + SR_A - PC_A + R_0$. Shared benefits of B-class manufacturing enterprises $\pi_B = R_B - IR_B - SC_B + TR_B + SR_B - PC_B + R_0$. When manufacturing enterprises engage in fairness concern behaviour, the utility function of A-class manufacturing enterprises can be expressed as $Q_A = \pi_A + \lambda(\pi_B - \pi_A)$. The utility function of B-class manufacturing enterprises is expressed as $Q_B = \pi_B + \eta(\pi_A - \pi_B)$.

3.4 Model construction

Based on the above assumptions, the mixed strategy benefits matrices of A-class and B-class manufacturing enterprises under fairness concern behaviour are shown in Table 2. For the convenience of understanding, they are all represented by name symbols in the table below.

The probability of resource-sharing for A-class manufacturing enterprises is x, and the probability of resource-sharing for B-class manufacturing enterprises is y.

For A-class manufacturing enterprises, the expected benefits of choosing sharing resources:

$$
\begin{aligned}
U_1 = y[&(1 - \lambda)(R_A + IR_A - SC_A + TR_A + SR_A - PC_A + R_0) \\
&+ \lambda(R_B - IR_B - SC_B + TR_B + SR_B - PC_B + R_0)] \\
&+ (1 - y)[(1 - \lambda)(R_A + IR_A - SC_A) + \lambda R_B]
\end{aligned}
\tag{1}
$$

The expected benefits of choosing non-sharing resources:

$$
U_2 = y[(1 - \lambda)R_A + \lambda(R_B - IR_B - SC_B)] + (1 - y)[(1 - \lambda)R_A + \lambda R_B]
\tag{2}
$$

Therefore, the expected average return for A-class enterprises:

$$
\bar{U} = yU_1 + (1 - y)U_2
\tag{3}
$$

Similarly, for B-class manufacturing enterprises, the expected benefits of choosing sharing resources:

$$
\begin{aligned}
V_1 = x[&(1 - \eta)(R_B - IR_B - SC_B + TR_B + SR_B - PC_B + R_0) + \eta(R_A + IR_A \\
&- SC_A + TR_A + SR_A - PC_A + R_0)] + (1 - x)[(1 - \eta)(R_B - IR_B - SC_B + \eta R_A]
\end{aligned}
\tag{4}
$$

The expected benefits of choosing non-sharing resources:

$$
V_2 = x[(1 - \eta)R_B + \eta(R_A + IR_A - SC_A)] + (1 - x)[(1 - \eta)R_B + \eta R_A]
\tag{5}
$$

Therefore, the expected average return for A-class enterprises:

$$
\bar{V} = xV_1 + (1 - x)V_2
\tag{6}
$$

The replicated dynamic equations of the evolutionary game for the two types of firms are respectively:

$$
\begin{aligned}
F(x) = \frac{dx}{dt} = x(1\text{-}x)\{&y[R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A \omega_A \\
&- \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B] + (1\text{-}\lambda)\tau_A K_A \omega_A - (1\text{-}\lambda)c_A K_A \omega_A\}
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
F(y) = \frac{dy}{dt} = y(1\text{-}y)\{&x[R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B \omega_B \\
&- \eta p_A R_A + \eta s_A K_A \omega_A] - (1\text{-}\eta)\tau_B K_B \omega_B - (1\text{-}\eta)c_B K_B \omega_B\}
\end{aligned}
\tag{8}
$$

Eqs. 7 and 8 portray the dynamic game evolution mechanism of symmetric manufacturing firms' strategy choices in the resource-sharing model under fairness concerns.

4. An evolutionary game model of resource-sharing in symmetric enterprises

4.1 Stability analysis of game strategies of A-class manufacturing enterprises

From the theory of stability of differential equations, it is known that the point of stability of replicated dynamic equations is the point where the first order derivative at its zero is less than 0, it is necessary to satisfy the:

$$\left\{\frac{dF(x)}{dx} \leq 0 \middle| F(x) = 0, x \in [0,1]\right\} \tag{9}$$

Analyse the stability of the strategy evolution based on the replication dynamics equation of the sharing strategy adopted by the manufacturing enterprise in A-class, let:

$$y^* = \frac{[(1\text{-}\lambda)c_A K_A \omega_A - (1\text{-}\lambda)\tau_A K_A \omega_A]}{[R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B]} \tag{10}$$

When $y = y^*$, we have $F(x) = 0$, all x are in steady state.

When $y \neq y^*$, let $F(x) = 0$, it can be obtained that $x_1 = 0$ and $x_2 = 1$ are equilibrium points. From Eq. 7 we have:

$$\frac{dF(x)}{dx} = (1\text{-}2x)\{y[R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B$$
$$+ \lambda t_B R_B + \lambda s_B K_B \omega_B] + (1\text{-}\lambda)\tau_A K_A \omega_A - (1\text{-}\lambda)c_A K_A \omega_A\} \tag{11}$$

Local evolutionary stabilisation strategies are discussed below.

When $R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B < 0$, we have $y > y^*$. By $\frac{dF(x)}{dx}|_{x=0} < 0$, $\frac{dF(x)}{dx}|_{x=1} < 0$, $x_1 = 0$ is the evolutionary stability point.

When $R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B > 0$. There are two scenarios.

(1) If $(1\text{-}\lambda)\tau_A K_A \omega_A - (1\text{-}\lambda)c_A K_A \omega_A > R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A] + (1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B$, $y < y^*$. By $\frac{dF(x)}{dx}|_{x=0} < 0$, $\frac{dF(x)}{dx}|_{x=1} < 0$, $x_1 = 0$ is the evolutionary stability point.

(2) If $(1\text{-}\lambda)\tau_A K_A \omega_A - (1\text{-}\lambda)c_A K_A \omega_A < R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B$, we have $0 < y^* < 1$. At this point, the stabilisation strategy of A-class manufacturing enterprises depends on the size of y.

When $y > y^*$, We have $\frac{dF(x)}{dx}|_{x=0} < 0$, $x_2 = 1$ is the evolutionary stability point. But when $y < y^*$, We have $\frac{dF(x)}{dx}|_{x=1} < 0$, $x_1 = 0$ is the evolutionary stability point.

4.2 Stability analysis of game strategies of B-class manufacturing enterprises

Similarly, from the theory of stability of differential equations, the point of stability of the replicated dynamic equations for manufacturing enterprises of B-class is the point at which the first-order derivatives at its zeros are less than 0.

$$\left\{\frac{dF(y)}{dy} \leq 0 \middle| F(y) = 0, y \in [0,1]\right\} \tag{12}$$

Analyse the stability of the strategy evolution based on the replication dynamics equation of the sharing strategy adopted by the manufacturing enterprise in B-class, let:

$$x^* = \frac{[(1\text{-}\eta)\tau_B K_B \omega_B + (1\text{-}\eta)c_B K_B \omega_B]}{[R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A]} \tag{13}$$

When $x = x^*$, we have $F(y) = 0$, all y are in steady state.

When $x \neq x^*$, let $F(y) = 0$, it can be obtained that $y_1 = 0$ and $y_2 = 1$ are equilibrium points.

From Eq. 8 we have:

$$\frac{dF(y)}{dy} = (1\text{-}2y)\{x[R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B +(1\text{-}\eta)s_B K_B \omega_B$$
$$-\eta p_A R_A + \eta s_A K_A \omega_A] - (1\text{-}\eta)\tau_B K_B \omega_B - (1\text{-}\eta)c_B K_B \omega_B\} \qquad (14)$$

When $R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B +(1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A < 0$, we have $x > x^*$. By $\frac{dF(y)}{dy}|_{y=0} < 0$, $\frac{dF(y)}{dy}|_{y=1} > 0$, $y_1 = 0$ is the evolutionary stability point.

When $R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B +(1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A > 0$, There are two scenarios.

(1) If $(1\text{-}\eta)\tau_B K_B \omega_B + (1\text{-}\eta)c_B K_B \omega_B > R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B +(1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A$, we have $x < x^*$. By $\frac{dF(y)}{dy}|_{y=0} < 0$, $\frac{dF(y)}{dy}|_{y=1} > 0$, $y_1 = 0$ is the evolutionary stability point.

(2) If $(1\text{-}\eta)\tau_B K_B \omega_B + (1\text{-}\eta)c_B K_B \omega_B < R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B +(1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A$, we have $0 < x^* < 1$. At this point, the stabilisation strategy of B-class manufacturing enterprises depends on the size of x.

When $x > x^*$, We have $\frac{dF(y)}{dy}|_{y=1} < 0$, $y_2 = 1$ is the evolutionary stability point. But when $x < x^*$, We have $\frac{dF(y)}{dy}|_{y=0} < 0$, $y_1 = 0$ is the evolutionary stability point.

5. Stability analysis of the evolution of combinatorial strategies in gaming systems

From Eqs. 7 and 8, it can be seen that there are five evolutionary game equilibria in the system as $(0,0),(1,0),(0,1),(1,1)$ and (x^*, y^*). The calculation process is shown in Appendix A.

However, the equilibrium point obtained by copying the dynamic equation may not necessarily be a stable strategy for system evolution. The stable strategy for system evolution can be obtained through local stability analysis using Friedman's stability determination method for the Jacobian matrix proposed in 1991 [26]. The Jacobian matrix (denoted as J) takes the following form, see Appendix B for details:

$$J = \begin{bmatrix} \dfrac{\partial F(x)}{\partial x} & \dfrac{\partial F(x)}{\partial y} \\ \dfrac{\partial F(y)}{\partial x} & \dfrac{\partial F(y)}{\partial y} \end{bmatrix} \qquad (15)$$

It can be further obtained that $det(J)$ and $tr(J)$ as follows:

$$\begin{aligned} det(J) = &(1\text{-}2x)(1\text{-}2y)(y(R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A +(1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B \\ &+ \lambda s_B K_B \omega_B) + (1\text{-}\lambda)\tau_A K_A \omega_A - (1\text{-}\lambda)c_A K_A \omega_A)(x(R_0 - (1\text{-}\eta)p_B R_B \\ &+ (1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A) - (1\text{-}\eta)\tau_B K_B \omega_B \\ &- (1\text{-}\eta)c_B K_B \omega_B) - xy(1\text{-}x)(1\text{-}y)(R_0 - (1\text{-}\lambda)p_A R_A \\ &+ (1\text{-}\lambda)t_A R_A +(1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B)(R_0 \\ &- (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B +(1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A) \end{aligned} \qquad (16)$$

$$\begin{aligned} tr(J) = &(1\text{-}2x)(y(R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A +(1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B \\ &+ \lambda s_B K_B \omega_B) + (1\text{-}\lambda)\tau_A K_A \omega_A - (1\text{-}\lambda)c_A K_A \omega_A) + (1\text{-}2y)(x(R_0 \\ &- (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B +(1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A) \\ &- (1\text{-}\eta)\tau_B K_B \omega_B - (1\text{-}\eta)c_B K_B \omega_B) \end{aligned} \qquad (17)$$

When the equilibrium point is any local asymptotic stability point, it is the evolutionary stability strategy (ESS). The stability analysis of the equilibrium points is carried out according to the local stability analysis method of Jacobi matrix, and the results are shown in Table 3.

In situation 1, there are two evolutionary stabilisation points of the system, which are the points D(0,0), B(1,1) that satisfy the $det(J) > 0$ and $tr(J) < 0$ conditions, and their corresponding evolutionary stabilisation strategies are (non-sharing, non-sharing) and (sharing, sharing). The system evolutionary path in this situation is shown in Fig. 2(a). The evolutionary stable state at this time depends on $P(x^*, y^*)$, i.e. $0 < x^* < 1$, $0 < y^* < 1$, reflecting the existence of heterogeneity in the basic conditions and starting points of the two types of enterprises, may be due to the differences in production and manufacturing, showing different sensitivities, and therefore choose the strategic behaviour of (non-sharing, non-sharing), but with the presentation of the benefits of resource-sharing, the two types of enterprises will change their behaviour to choose the resource-sharing, and the formation of symbiotic win-win between the enterprises of the new business model.

In situation 2, there is an evolutionary stability point in the system, and its corresponding evolutionary stability strategy is (non-sharing, non-sharing). The system evolution path in this scenario is shown in Fig. 2(b). At this point, $0 < x^* < 1$, $y^* > 1$, the sharing behaviour of A-class manufacturing enterprises will be in a mixed state, there exists the probability of choosing to share and also choosing not to share, but as the dynamics of this system evolves to a stable period, the manufacturing enterprises will ultimately choose not to share resources. Reflecting the business operation, although the platform operates the resource-sharing system with some success and realises revenue spillovers for the enterprise, the enterprise will choose not to share resources as it has equity concerns at this point in time and is overly concerned with the revenue gap.

Table 2 Payoff matrix of evolutionary game for manufacturing resource-sharing

		Manufacturing enterprise B	
		Sharing	Non-Sharing
Manufacturing enterprise A	Sharing	$(1-\lambda)(R_A + IR_A - SC_A + TR_A + SR_A - PC_A + R_0)$ $+\lambda(R_B - IR_B - SC_B + TR_B + SR_B - PC_B + R_0)$	$(1-\lambda)(R_A + IR_A - SC_A) + \lambda R_B$
		$(1-\eta)(R_B - IR_B - SC_B + TR_B + SR_B - PC_B + R_0)$ $+\eta(R_A + IR_A - SC_A + TR_A + SR_A - PC_A + R_0)$	$(1-\eta)R_B + \eta(R_A + IR_A - SC_A)$
	Non-Sharing	$(1-\lambda)R_A + \lambda(R_B - IR_B - SC_B)$ $(1-\eta)(R_B - IR_B - SC_B + \eta R_A$	$(1-\lambda)R_A + \lambda R_B$ $(1-\eta)R_B + \eta R_A$

Table 3 Stability analysis results

Situation	Equilibrium point	$det(J)$ symbolic	$tr(J)$ symbolic	Condition
(1)	(0,0)	+	-	ESS
$0 < (1\text{-}\eta)\tau_B K_B\omega_B + (1\text{-}\eta)c_B K_B\omega_B < R_0 - (1\text{-}\eta)p_B R_B$ $+(1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B\omega_B - \eta p_A R_A + \eta s_A K_A\omega_A 0$	(1,0)	+	+	Instability point
$< (1\text{-}\lambda)c_A K_A\omega_A - (1\text{-}\lambda)\tau_A K_A\omega_A$ $< R_0 - (1\text{-}\lambda)p_A R_A$	(0,1)	+	+	Instability point
$+(1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A\omega_A - \lambda p_B R_B + \lambda t_B R_B$	(1,1)	+	-	ESS
$+ \lambda s_B K_B\omega_B$	(x^*, y^*)	-	0	centre point
(2)	(0,0)	+	-	ESS
$0 < (1\text{-}\eta)\tau_B K_B\omega_B + (1\text{-}\eta)c_B K_B\omega_B < R_0 - (1\text{-}\eta)p_B R_B$ $+(1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B\omega_B - \eta p_A R_A + \eta s_A K_A\omega_A$	(1,0)	+	+	Instability point
$(1\text{-}\lambda)c_A K_A\omega_A - (1\text{-}\lambda)\tau_A K_A\omega_A$	(0,1)	-	not sure	saddle point
$> R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A$ $+(1\text{-}\lambda)s_A K_A\omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B\omega_B > 0$	(1,1)	-	not sure	saddle point
(3)	(0,0)	+	-	ESS
$(1\text{-}\eta)\tau_B K_B\omega_B + (1\text{-}\eta)c_B K_B\omega_B > R_0 - (1\text{-}\eta)p_B R_B$ $+(1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B\omega_B - \eta p_A R_A + \eta s_A K_A\omega_A > 0$	(1,0)	-	not sure	saddle point
$0 < (1\text{-}\lambda)c_A K_A\omega_A - (1\text{-}\lambda)\tau_A K_A\omega_A < R_0 - (1\text{-}\lambda)p_A R_A$	(0,1)	+	+	Instability point
$+(1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A\omega_A - \lambda p_B R_B + \lambda t_B R_B$ $+ \lambda s_B K_B\omega_B$	(1,1)	-	not sure	saddle point
(4)	(0,0)	+	-	ESS
$(1\text{-}\eta)\tau_B K_B\omega_B + (1\text{-}\eta)c_B K_B\omega_B > R_0 - (1\text{-}\eta)p_B R_B$ $+(1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B\omega_B - \eta p_A R_A + \eta s_A K_A\omega_A > 0$	(1,0)	-	not sure	saddle point
$(1\text{-}\lambda)c_A K_A\omega_A - (1\text{-}\lambda)\tau_A K_A\omega_A$	(0,1)	-	not sure	saddle point
$> R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A$ $+(1\text{-}\lambda)s_A K_A\omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B\omega_B > 0$	(1,1)	+	+	Instability point

In situation 3, the system also has an evolutionary stability point, and its corresponding evolutionary stability strategy is (non-sharing, non-sharing). The system evolution path in this scenario is shown in Fig. 2(c). At this point, $x^* > 1$, $0 < y^* < 1$. The sharing behaviour of B-class manufacturing enterprises will be in a mixed state, due to the fact that the information technology level of B-class enterprises themselves is higher than that of A-class manufacturing enterprises, although through sharing resources to obtain new technologies and reduce manufacturing costs, but also the loss of their own advantages to increase the cost of risk, and coupled with fairness concerns make such manufacturing enterprises give up the sharing of resources, choosing to share resources not to share.

In situation 4, the system also has an evolutionary stability point, which corresponds to an evolutionary stability strategy of (non-sharing, non-sharing). The system evolution path in this scenario is shown in Fig. 2(d). At this point, $x^* > 1$, $y^* > 1$, both types of manufacturing enterprises adopt the non-sharing behaviour, the reason is that when the enterprise chooses to share resources, the direct and indirect benefits it obtains are less than the sum of the cost and risk cost of its data sharing, and due to too much attention to fairness, it does not satisfy the distribution of the value spillover brought by the data sharing, so it chooses to choose not to share resources.
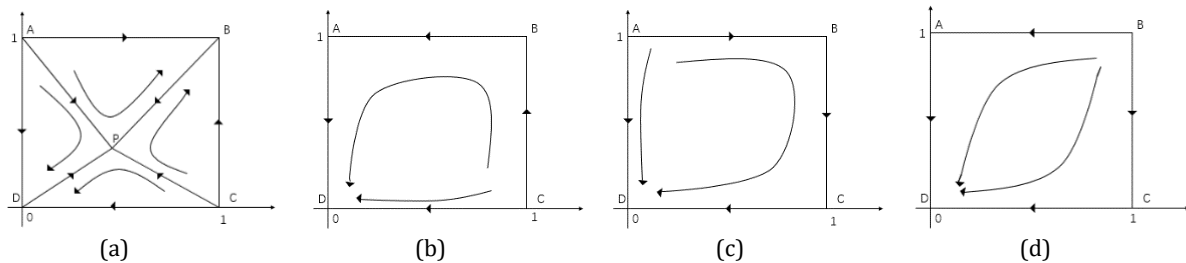

(a)   (b)   (c)   (d)

Fig. 2 System evolutionary path

6. Evolutionary game simulation analysis

To analyse the impact of various influencing factors on system stability, the initial values of the parameters are set as follows: $R_A = 0.2$, $R_B = 0.2$, $R_0 = 0.2$, $\omega_A = 0.1$, $\omega_B = 0.2$, $K_A = 3$, $K_B = 2$. With the initial values of most of the variables fixed, six influencing factors are analysed, including the system cost coefficient, the informatization level coefficient, the information loss coefficient, the resource transformation ability coefficient, the resource coordination ability coefficient, and the penalty factor, to determine how they affect the evolutionary results.

(1) The impact of changes in the system cost coefficients on the evolutionary results

When the initial value of the sharing probability for A-class and B-class enterprises is fixed at 0.2, the system cost coefficients are set to 0.2 and 0.7 respectively, and the evolutionary results are shown in Fig. 3.

When equity neutrality is present, it reduces enterprises' resource-sharing behaviour as the system cost coefficient increases, and there is a negative correlation between the two. But when there is fairness concern behaviour, on the contrary, it increases enterprises' sharing behaviour. However, if the fairness concern coefficient of B-class enterprises is higher than that of A-class enterprises, A-class enterprises will reduce sharing and may change their behaviour. And B-class enterprises change their behavioural strategies when the level of fairness concern is higher. At the same time the level of fairness concern coefficient of both types of enterprises does not affect the behaviour of the B-class enterprises themselves. The above situation indicates the heterogeneity of fairness concern behaviour. Due to the different nature of enterprises, A-class enterprises will pay more attention to the fairness concern of other enterprises while focusing on their profits, while B-class enterprises, due to their high level of informatization, only focus on their interests. Fairness concerns among enterprises will strengthen the impact of cost coefficients on sharing strategies.

(2) The impact of changes in the informatization level coefficient on the evolutionary results

The initial value of the sharing probability for A-class enterprises is fixed at 0.2, and the informatization level coefficients are set to 0.1, 0.5, and 0.9, respectively. The evolution results are shown in Fig. 4.

When fairness is neutral, as the coefficient of information technology level increases it strengthens the behaviour of firms in sharing resources and there is a positive correlation between the two. When A-class enterprises engage in fairness concern behaviour, and the enterprises gradually change their behaviour of sharing resources. At the same time, if the fairness concern behaviour of B-class enterprises is higher than that of A-class enterprises, A-class enterprises will exhibit behaviour of changing their strategies. When the level of informatization of A-class enterprises increases, it will not affect the initial sharing strategy. However, once there is a fairness concern behaviour, attention will be paid to its level and compared with B-class enterprises, thereby changing its strategy. This indicates that fairness concern behaviour exists in enterprise resource-sharing behaviour and is directional.

(3) The impact of changes in the information loss coefficient on the evolutionary results

The initial value of the sharing probability for B-class enterprises is fixed at 0.2, and the information loss coefficients are set to 0.2, 0.5, and 0.8, respectively. The evolution results are shown in Fig. 5.

When fairness is neutral, when the information loss coefficient is small, B-class enterprises tend to adopt a sharing strategy. As the information loss coefficient increases, it gradually changes to non sharing. When there is fairness concern behaviour, B- class enterprises will exhibit the behaviour of changing strategies and improving convergence speed, and the degree of its impact will become more apparent with the increase of the fairness concern coefficient. This indicates that for B-class enterprises, due to the nature of their shared resources, they are very concerned about the interests between enterprises, and fairness concern behaviour will increase the impact of the information loss coefficient on the sharing behaviour of B-class enterprises.

(4) The impact of changes in the resource conversion capacity coefficient on the evolutionary results

When the initial value of the sharing probability for A-class and B-class enterprises is fixed at 0.2, the resource conversion capacity coefficients are set to 0.2 and 0.7 respectively, and the evolution results are shown in Fig. 6.

When fairness is neutral, as the resource transformation ability coefficient increases it enhances the enterprise sharing strategy, and there is a positive correlation between the two. For A-class enterprises, the sharing strategy change behaviour occurs when the equity concern coefficient is large. Similarly, for B-class enterprises, their impact is higher than that of A-class enterprises, and their behaviour is more pronounced. This indicates that for any type of shared resource enterprise, once the shared resources are converted into their resources, they will pay more attention to the phenomenon of fairness. If there is a problem that is not conducive to overall coordination or their behaviour, they will immediately change their sharing strategy, and fairness concern behaviour has a reinforcing effect.

(5) The impact of changes in the resource collaboration capability coefficient on the evolutionary results

When the initial value of the sharing probability for A-class and B-class enterprises is fixed at 0.2, the resource collaboration capability coefficients are set to 0.2 and 0.7 respectively, and the evolutionary results are shown in Fig. 7.

When fairness is neutral, as the coefficient of resource synergy increases, A-class enterprises tend to engage in sharing behaviour, and there is a positive correlation between the two. Meanwhile, for A-class enterprises, when there is a fairness concern behaviour, their sharing behaviour will accelerate convergence speed with the increase of the fairness concern coefficient. For B-class enterprises, when fairness is neutral, as the coefficient of resource synergy increases, the convergence speed of their sharing probability to 1 accelerates. However, when there is a small fairness concern coefficient, the speed slows down, and during the growth process, B-class enterprises

change their sharing strategy. This indicates that for B-class enterprises, once they focus on fairness concerns between enterprises, they will reconsider whether the benefits they can gain from sharing soft manufacturing resources are beneficial to their development and change their behaviour.

(6) The impact of changes in the penalty factors on the evolutionary results

When the initial value of the sharing probability for A-class and B-class enterprises is fixed at 0.2, the resource collaboration capability coefficients are set to 0.2 and 0.7 respectively, and the evolutionary results are shown in Fig. 8.

When fairness is neutral, as the penalty factor increases, A-class enterprises will shift from sharing strategies to non sharing, with a negative correlation between the two. At the same time, for A-class enterprises, when there is fairness concern behaviour, their sharing behaviour will weaken the convergence speed with the increase of fairness concern coefficient. For B-class enterprises, when the penalty factor is small, they will choose a sharing strategy. As it continues to increase, B-class enterprises tend to choose a non-sharing strategy. The existence of fairness concern behaviour will slow down the convergence speed of corporate strategies, and when fairness concern behaviour is strong, companies tend to not share strategies. At this point, it indicates that companies will change their behaviour due to an increase in penalty factors, and different types of companies have opposite selection strategies when facing strong fairness concerns.
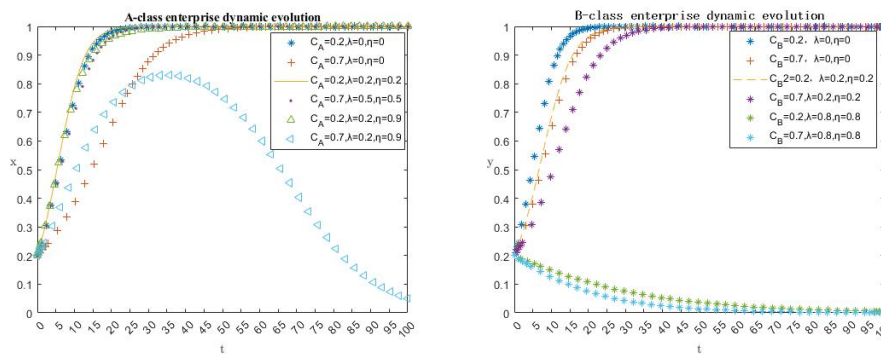

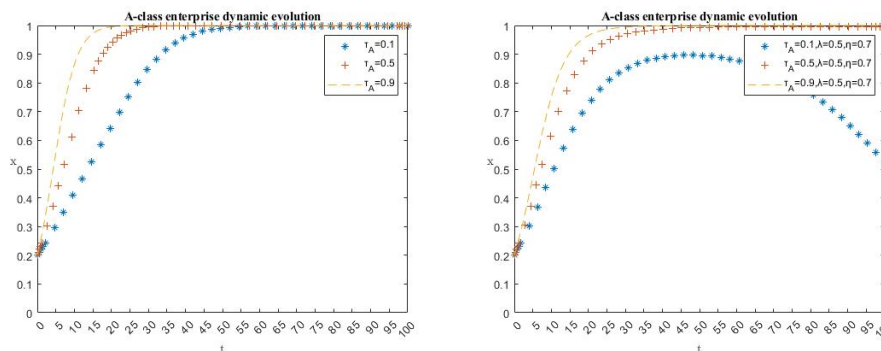**Fig. 3** The impact of changes in the system cost coefficient on the evolution results

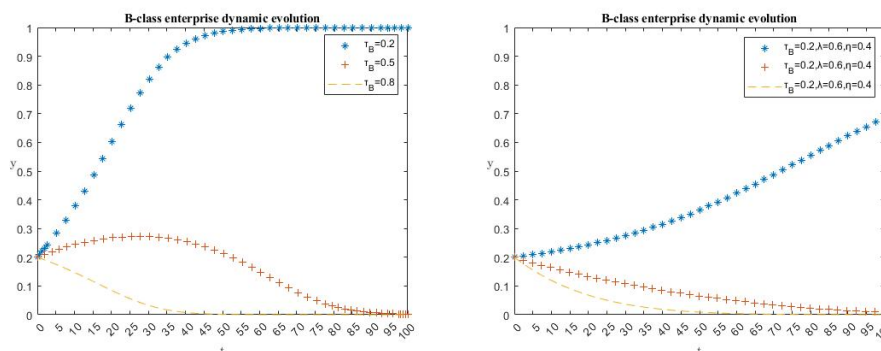**Fig. 4** The impact of changes in informatization level coefficient on evolutionary results

Fig. 5 The impact of changes in the information loss coefficient on the evolutionary results
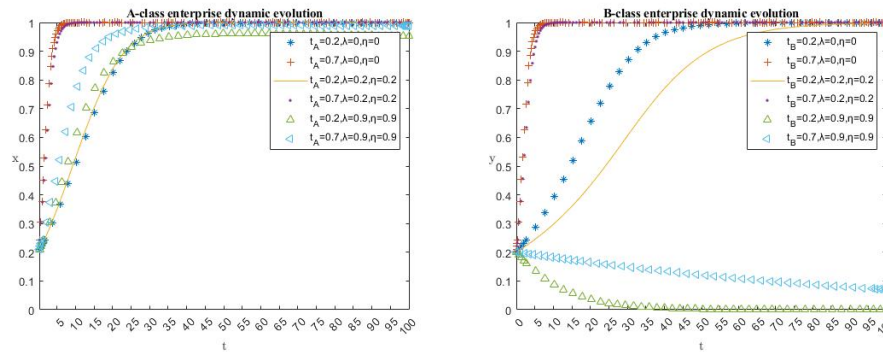
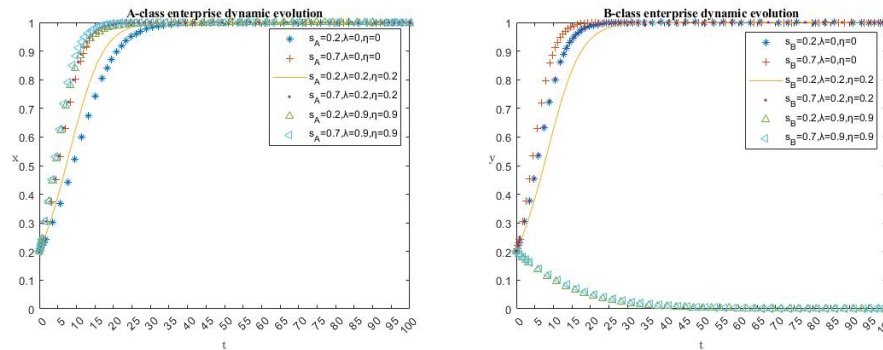**Fig. 6** The impact of changes in the resource conversion capacity coefficient on the evolutionary results

Fig. 7 The impact of changes in resource collaboration capability coefficients on evolutionary results
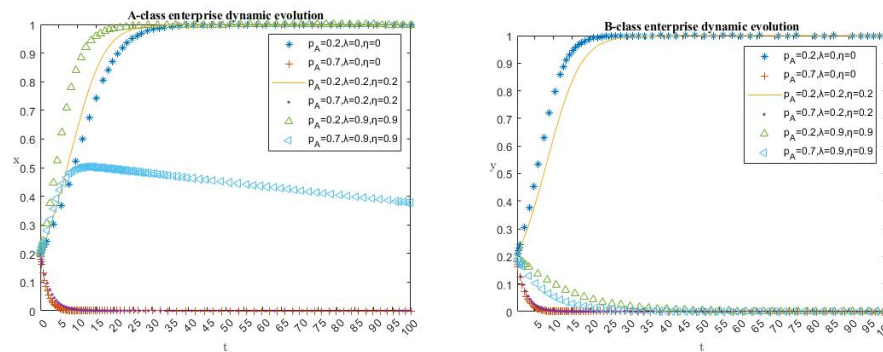

Fig. 8 The impact of penalty factor changes on evolutionary outcomes

7. Conclusion

This paper establishes a game model for the evolution of manufacturing resource-sharing in symmetric enterprises from the perspective of fairness concern, and uses the software MATLAB to simulate the system in dynamic simulation, comparing and analysing the impacts of different factors on the evolution results, and arriving at the following conclusions and recommendations.

(1) The system cost coefficient has a significant negative impact on the enterprise manufacturing resource-sharing strategy, when the system cost coefficient is large, the cost of entering the platform to share resources far exceeds the benefits it brings, and the increase of sharing cost will reduce the subject's willingness to cooperate. Then the more advanced manufacturing technology will not let the enterprise change the production mode, so the low cost and high efficiency should be made the advantage of the sharing platform. Information loss coefficient is also an important coefficient that affects the sharing behaviour of enterprises, the sharing platform must strengthen the resource risk control and improve the reliability of resource interoperability and mutual sharing, so the setting of the penalty factor is an important guarantee for the long-term development of the enterprise resource-sharing model. Manufacturing enterprise's own information technology level and resource transformation, resource synergy ability to improve, increase the main resource-sharing willingness, the resulting indirect benefits for enterprises to choose resource-

sharing provides a strong impetus. Therefore, for manufacturing enterprises must enhance their own degree of information technology, thereby promoting the depth of resource-sharing.

(2) Manufacturing enterprise data sharing strategy selection is a dynamic game process, the subject of the two strategies selection will be affected by the size of the gain and fairness concern behaviour, and the fairness perception effect on the manufacturing enterprise data sharing strategy selection of the impact of the existence of a threshold. When the subject's fairness concern coefficient is smaller than its own strategy threshold, the system evolves to a stable state of sharing, but with the enhancement of the influence of the fairness perception effect, the sharing party is more and more dissatisfied with the current distribution of benefits, which will destroy the sharing behaviour and affect the stability of the system. When fairness is neutral, the willingness of manufacturing enterprises to share resources gradually increases with the improvement of the level of synergy and resource transformation capability. However, under the fairness concern, the increase of shared resource revenue will not enhance the subject's willingness to share.

Based on the above research conclusions, and combined with the actual situation of business operations, this paper puts forward the following suggestions:

As a hard manufacturing resource enterprise, it should continuously improve its own system level, update and upgrade its equipment in time to better connect with the sharing platform, and also abide by the contract or agreement to constrain its own behaviour. With the continuous development of shared resources, enterprises must improve their own degree of information technology, and continuously increase their investment in research and development and innovation. As a soft manufacturing resource providers should be clear about the benefits of sharing platforms, should not only focus on their own development, but to jointly maintain the collective development, will be more favourable to themselves. Due to their own information advantages, they should try to explore the value of shared resources, focus on research and development, and collaborate with hard manufacturing enterprises to develop products for mutual benefit. At the same time, due to the existence of fair concern behaviour, can be set for manufacturing enterprises transparent profit distribution mechanism to achieve the effect of resource-sharing strategy stability. Through the formulation of reasonable ex ante contracts or the construction of an effective reward and punishment mechanism, it can enhance the revenue guarantee of resource-sharing behaviour of manufacturing enterprises, strengthen the fairness perception of the sharing subject, and improve the satisfaction of both enterprises with their own revenue distribution. To control the cost of manufacturing enterprises participating in shared manufacturing, a certain threshold can be set to protect the willingness of enterprises to participate. In addition, the enterprises involved in resource-sharing have the same goal, supervise each other, effectively reduce the risks faced in the process of information collaboration and transformation, and guarantee the information security of manufacturing enterprises, thus avoiding the destruction of inter-enterprise sharing cooperation due to the enhancement of the fairness concern, and jointly promoting the in-depth development of resource-sharing.

At the same time to take active measures to develop the digital economy, the use of digital dividends and efficient development of the platform economy. With the help of blockchain and other emerging information technologies, the authentication and storage of shared resource data are completed in the blockchain system. The decentralised nature of blockchain makes the data no longer stored in a single centralised server, thus reducing the risk of tampering with or loss of the shared data, and only the authorised manufacturing enterprises can access and modify the shared data to safeguard the data security of the shared resources. All enterprises involved in shared manufacturing can access the same data, improving information transparency and reducing information asymmetry. Blockchain can record and track every link of shared resources, realising the traceability of shared resources and clarifying the rights and obligations of shared enterprises. The joint participation of multiple parties enables manufacturing enterprises to respond more flexibly and efficiently to market changes in the modern competitive environment, fuelling the stable and positive development of resource-sharing.

Acknowledgement

References

[1] Li, R., Rao, J., Wan, L. (2022). The digital economy, enterprise digital transformation, and enterprise innovation, *Managerial and Decision Economics,* Vol. 43, No. 7, 2875-2886, doi: 10.1002/mde.3569.

[2] Qi, Y., Niu, Y., Zhou, Z. (2023). Digital economy empowering the development level of chinese manufacturing industry, *Economic Computation and Economic Cybernetics Studies and Research,* Vol. 57, No. 4, 243-258, doi: 10.24818/18423264/57.4.23.15.

[3] Benoit, S., Baker, T.L., Bolton, R.N., Gruber, T., Kandampully, J. (2017). A triadic framework for collaborative consumption (CC): Motives, activities and resources & capabilities of actors, *Journal of Business Research,* Vol. 79, 219-227, doi: 10.1016/j.jbusres.2017.05.004.

[4] Krzeczkowska, M., Slabon, A. (2024). Science and society – A new era for science communication in the context of sustainable development, *Chemistry Didactics Ecology Metrology,* Vol. 28, No. 1-2, 121-134, doi: 10.2478/cdem-2023-0007.

[5] Fehr, E., Sehmidt, K.M. (1999). A theory of fairness, competition, and cooperation, *The Quarterly Journal of Economics,* Vol. 114, No. 3, 817-868, doi: 10.1162/003355399556151.

[6] Mair, J., Reischauer, G. (2017). Capturing the dynamics of the sharing economy: Institutional research on the plural forms and practices of sharing economy organizations, *Technological Forecasting and Social Change*, Vol. 125, 11-20, doi: 10.1016/j.techfore.2017.05.023.

[7] Kazancoglu, Y., Kazancoglu, I., Sagnak, M. (2018). A new holistic conceptual framework for green supply chain management performance assessment based on circular economy, *Journal of Cleaner Production*, Vol. 195, No. 10, 1282-1299, doi: 10.1016/j.jclepro.2018.06.015.

[8] Jiang, P., Li, P. (2020). Shared factory: A new production node for social manufacturing in the context of sharing economy, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* Vol. 234, No. 1-2, 285-294, doi: 10.1177/0954405419863220.

[9] Yu, C., Jiang, X., Yu, S., Yang, C. (2020). Blockchain-based shared manufacturing in support of cyber physical systems: concept, framework, and operation, *Robotics and Computer-Integrated Manufacturing,* Vol. 64, Article No. 101931, doi: 10.1016/j.rcim.2019.101931.

[10] Billo, R.E., Dearborn, F., Hostick, C.J., Spanner, G.E., Stahlman, E.J., Aurand, S.S. (2007). A group technology model to assess consolidation and reconfiguration of multiple industrial operations—A shared manufacturing solution, *International Journal of Computer Integrated Manufacturing*, Vol. 6, No. 5, 311-322, doi: 10.1080/09511929308944583.

[11] Khoa, B.T., Huynh, L.T., Nguyen, M.H. (2020). The relationship between perceived value and peer engagement in sharing economy: A case study of ridesharing services, *Journal of System and Management Sciences*, Vol. 10, No. 4, 149-172, doi: 10.33168/JSMS.2020.0410.

[12] Liu, P., Liu, C., Wei, X. (2021). Optimal allocation of shared manufacturing resources based on bilevel programming, *Discrete Dynamics in Nature and Society,* Vol. 2021, No. 1, Article No. 6474241, doi: 10.1155/2021/6474241.

[13] Ho, T.H., Zhang, J.J. (2008). Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Management Science,* Vol. 54, No. 4, 686-700, doi: 10.1287/mnsc.1070.0788.

[14] Wang, Y.-L., Yin, X.-M., Zeng, X.-Y., Chen, W. (2023). Supply chain decision-making considering green technology effort: Effect on random output and retail price with fairness concerns, *Economic Computation and Economic Cybernetics Studies and Research,* Vol. 57, No. 1, 103-120, doi: 10.24818/18423264/57.1.23.07.

[15] Zhang, H.Y., Zhang, Z., Pu, X.J., Li, Y.H. (2019). Green manufacturing strategy considering retailers' fairness concerns, *Sustainability,* Vol. 11, No. 17, Article No. 4646, doi: 10.3390/su11174646.

[16] Gong, G., Li, K., Zha, Q. (2023). A maximum fairness consensus model with limited cost in group decision making, *Computers & Industrial Engineering*, Vol. 175, Article No. 108891, doi: 10.1016/j.cie.2022.108891.

[17] Du, J., Liu, S., Liu, Y. (2022). A limited cost consensus approach with fairness concern and its application, *European Journal of Operational Research*, Vol. 298, No. 1, 261-275, doi: 10.1016/j.ejor.2021.06.039.

[18] Wang, T.Y., Zhang, H. (2023). Blockchain-based tripartite evolutionary game study of manufacturing capacity sharing, *Advances in Production Engineering & Management,* Vol. 18, No. 2, 225-236, doi: 10.14743/apem2023.2.469.

[19] Cao, P., Duan, G., Tu, J., Jiang, Q., Yang, X., Li, C. (2023). Blockchain-based process quality data sharing platform for aviation suppliers, *IEEE Access,* Vol. 11, 19007-19023, doi: 10.1109/ACCESS.2023.3246984.

[20] Liu, P., Wei, X.L., Liu, C.Y. (2022). Tripartite evolutionary game analysis of shared manufacturing by manufacturing companies under government regulation mechanism, *Discrete Dynamics in Nature and Society*, Vol. 2022, No. 1, Article No. 7706727, doi: 10.1155/2022/7706727.

[21] Zhu, P.P., Shen, J., Xu, M. (2023). Study on the evolution of information sharing strategy for users of online patient community, *Personal and Ubiquitous Computing,* Vol. 27, 689-695, doi: 10.1007/s00779-020-01464-6.

[22] Zhang, L., Lu, Q., Huang, R., Chen, S., Yang, Q., Gu, J. (2023). A dynamic incentive mechanism for smart grid data sharing based on evolutionary game theory, *Energies*, Vol. 16, No. 24, Article No. 8125, doi: 10.3390/en16248125.

[23] Xiao, M., Tian, Z.Y. (2022). Evolutionary game analysis of company collaborative strategy in cloud manufacturing platform environment, *Advances in Production Engineering & Management*, Vol. 17, No. 3, 295-310, doi: 10.14743/apem2022.3.437.

[24] Liu, W., Long, S., Wei, S., Xie, D., Wang, J., Liu, X. (2021). Smart logistics ecological cooperation with data sharing and platform empowerment: An examination with evolutionary game model, *International Journal of Production Research*, Vol. 60, No. 13, 4295-4315, doi: 10.1080/00207543.2021.1925173.

[25] Liu, W., Wu, C.-H., Tsai, S.-B., Shao, X.-F., Wacławek, M. (2023). Corporate environmental management in the context of digital transformation, *Ecological Chemistry and Engineering S,* Vol. 30, No. 1, 91-92, doi: 10.2478/eces-2023-0006.

[26] Friedman, D. (1991). Evolutionary games in economics, *Econometrica*, Vol. 59, No. 3, 637-666, doi: 10.2307/2938222.

Appendix A

Let $F(x) = 0$, we can get: $x_1 = 0$, $x_2 = 1$, $y^* = \frac{[(1-\lambda)c_A K_A \omega_A - (1-\lambda)\tau_A K_A \omega_A]}{[R_0 - (1-\lambda)p_A R_A + (1-\lambda)t_A R_A + (1-\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B]}$.

Let $F(y) = 0$, we can get: $y_1 = 0$, $y_2 = 1$, $x^* = \frac{[(1-\eta)\tau_B K_B \omega_B + (1-\eta)c_B K_B \omega_B]}{[R_0 - (1-\eta)p_B R_B + (1-\eta)t_B R_B + (1-\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A]}$.

Appendix B

Which: $\frac{\partial F(x)}{\partial x} = (1\text{-}2x)\{y[R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B] + (1\text{-}\lambda)\tau_A K_A \omega_A - (1\text{-}\lambda)c_A K_A \omega_A\}$, $\frac{\partial F(x)}{\partial y} = x(1\text{-}x)[R_0 - (1\text{-}\lambda)p_A R_A + (1\text{-}\lambda)t_A R_A + (1\text{-}\lambda)s_A K_A \omega_A - \lambda p_B R_B + \lambda t_B R_B + \lambda s_B K_B \omega_B]$,

$\frac{\partial F(y)}{\partial x} = y(1\text{-}y)[R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A]$,

$\frac{\partial F(y)}{\partial y} = (1\text{-}2y)\{x[R_0 - (1\text{-}\eta)p_B R_B + (1\text{-}\eta)t_B R_B + (1\text{-}\eta)s_B K_B \omega_B - \eta p_A R_A + \eta s_A K_A \omega_A] - (1\text{-}\eta)\tau_B K_B \omega_B - (1\text{-}\eta)c_B K_B \omega_B\}$

Current state and production characteristics of the Polish tanning industry: A case study

Bielak, E.[a,*], **Zakrzewska, M.**[b]

[a]Department of Non-food Product Quality and Safety, Krakow University of Economics, Krakow, Poland
[b]Department of Management Process, Krakow University of Economics, Krakow, Poland

A B S T R A C T

This article presents the results of a study on the current state and evaluation of the Polish tanning industry, focusing on its production characteristics. The research sample included 220 companies that were contacted to gather information about their operations. Some of these companies have been suspended, liquidated, or have changed their business profiles. Approximately 30 % confirmed that they are still active in leather manufacturing, indicating that the Polish tanning industry is experiencing a process of deindustrialization. Surveys conducted in 20 companies revealed that Polish tanneries operate on national, European, and global scales. Most of them are micro or small enterprises with annual revenues of up to PLN 5 million. The primary factor defining their competitiveness is the high quality of the products and services they offer. They mainly process calfskin and cowhides sourced from Poland and abroad, primarily for the footwear sector. Polish tanneries are aware of global trends, including the industry's shift towards ecological practices, the adoption of modern technologies, and the introduction of innovations. Given the current challenges facing the Polish tanning industry, it is essential to take action to improve the health of this sector of the economy.

1. Introduction

Leather is a valuable raw material used in the production of a variety of products that are considered luxury by consumers and have been eagerly purchased for years, regardless of the prevailing fashion. Leather manufacturing has a long history and tradition. However, over the years, the tannery sector has changed due to technological progress, innovative solutions, and, above all, the introduction of new regulations or the modification of existing ones. It estimates that the European tanning sector consists of almost 1,600 companies, over 34,000 employees, and generates a turnover of up to €7.4 billion. In recent years, a gradual concentration has been observed. As recently as 2000, tanneries employed an average of 24 workers, today it is 21. European tanneries range from small and medium-sized enterprises, continuing long family traditions, to large companies operating internationally [1]. The modern leather industry in Europe is a strategic area of the manufacturing sector, resulting from a combination of tradition and con-

tinuous innovation. What is more the European market of the leather sector leads, among others, in terms of quality, technology, innovation, and sustainable development [2].

Three countries were identified as the largest producers of leather in 2022, i.e. Brazil, the United States, and Turkey. Their combined share of global production is 33 %. Countries playing a significant role in global leather export include Italy, Brazil, and China [3]. According to estimates by the Food and Agriculture Organization of the United Nations (FAO), more than 50 % of the leather produced worldwide is used in the clothing industry. Only about one-fifth is produced in European tanneries, which have high environmental and safety standards, and require sustainable processes and chemicals. The result is a high-quality, strength material whose intricate production steps justify calling it a luxury product [4]. It is used to produce various goods also referred to as luxury [5]. Analysing the destination of leathers supplied by European tanneries, this raw material is mainly used in footwear (41 %), leather goods/accessories (19 %), and furniture (17 %), but also automotive industry (13 %) and clothing (8 %) [6].

Data from the report Global Leather Goods Market – Industry Trends and Forecast to 2030 [7] confirms the ever-increasing demand for leather goods, projecting that the global market for these products will continue to grow in the coming years, reaching USD 699,906.77 million by 2030. Premium and high-quality luxury products made from natural leather are expected to be popular. New and innovative solutions are expected from the global leather industry in terms of raw material properties and design. However, the sector may be challenged by, inter alia, problems with technology, equipment, and the lack of adequate leather processing skills. Among the factors that may have a negative impact on the development of the global leather market are problems with the availability of raw materials or competition from other materials that are synthetic substitutes for natural leather. An important aspect, according to Cafasso [8], is to educate consumers about the origin of the leather products they use, the methods and ways in which they are produced, and the associated environmental impact. It is worth mentioning that many researchers concerned the ecology issues and the various aspects of sustainability for example according to the footwear industry [9-11], which is closely associated with the tanning/leather sector.

The Polish economy is currently undergoing the process of reindustrialization and is the 7th largest industrial country in the European Union and 21st in the world, supplying many important products. However, when it comes to the leather industry, the decline of this industry is observed in Poland. Data from the Central Statistical Office show that the structure of sold production of the clothing, textile, and leather industries in Poland in 2020 decreased by about 9 percentage points compared to 1985 and amounted to 2 % [12]. In the case of the leather and leather products industry, the sales value for 2021 was PLN 3,836.2 million [13]. In recent years, there has been a gradual decline in employment in the sector, but a positive trend is an increase in wages [14]. The most recent data shows that in Poland at the end of December 2023, there were 310 leather manufacturing and tanning-related entities (PKD 15.11.Z), as well as 2623 footwear manufacturers (PKD 15.20.Z) and 1558 manufacturers of bag and saddlery products (PKD 15.12.Z) [15]. The Polish leather-footwear industry uses mainly domestic raw materials for production. Entrepreneurs operating in this sector cooperate closely with the vocational education system. This is because the work done in their factories is a combination of traditional craftsmanship, the art of design and construction, as well as the ability to operate increasingly advanced production machinery [16]. Even though demand for leather products remains high and, as mentioned above, a further increase in the value of the global leather market is predicted, more and more tanneries are closing down in Poland.

This study aimed to present the current state and make an assessment of the Polish tanning industry, taking into consideration the characteristics of the production process carried out in tanneries located in Poland. Once a database of domestic tanneries had been established, a survey questionnaire was designed and sent to companies via email. Respondents were also allowed to take part in the survey by answering it during a telephone interview. This allowed the verification of the previously developed database of tanneries, the selection of companies that are still active in production, and the extraction of information regarding their leather manufac-

turing processes. The following research questions were formulated and verified in the course of the study:

- RQ1: What is the current overall state of the tanning industry in Poland?
- RQ2: What most determines competitiveness in the tanning industry in Poland?
- RQ3: Do the tanneries in Poland know current global business trends concerning their sector of the economy?
- RQ4: What raw materials are processed in tanneries in Poland, using what tanning methods, and what goods are produced from Polish leather?

The remainder of this article is organized as follows. The second section describes the research procedure and method used for data collection. This is followed by the presentation of the results of the research obtained with elements of discussion. The research was then summarized and conclusions were drawn while answering the research questions posed. The final part of the article includes the formulation of recommendations for further research along with a discussion of the limitations encountered due to the specifics of the research topic and methods adopted.

2. Research methodology

2.1 Procedure of the research

The research procedure to achieve the objective of the study is presented in Fig. 1. The research was conducted in three phases:

(1) RESEARCH PREPARATION PHASE: Following the procedure, the study started by formulating the research problem in the form of a general question: What is the current state of the tanning industry in Poland and what are the characteristics of the production process carried out in polish tanneries? Then, to define the scope of the study, an attempt was made to analyse scientific texts on the issue under investigation. A search for information and materials on scientific studies carried out on the Polish tanning industry has resulted in the conclusion that there is a research gap in this area. Based on the information collected, research questions were formulated, as mentioned in the Introduction.

(2) RESEARCH EXECUTION PHASE: A survey approach was selected and used to collect data to enable further analysis and answers to the research questions. The characteristics of the research tool are described in detail in section 2.2. The survey was sent to tanneries located all over Poland. The data collected through the questionnaire was analysed.

(3) RESEARCH SUMMARY PHASE: Based on the interpretation of the research results obtained, the research questions were answered and conclusions were drawn. In addition, recommendations were formulated in connection with the study.
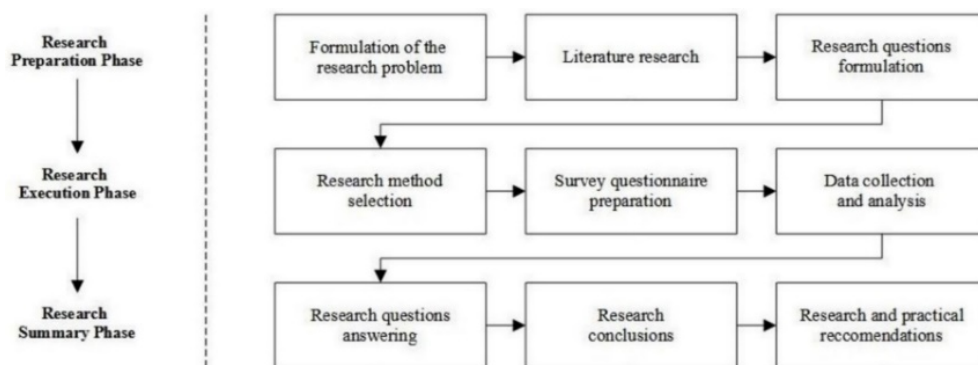

Fig. 1 Diagram showing the procedure for the research process (Source: Own elaboration)

The authors declare that the research was conducted in accordance with ethical principles. The article presents non-interventional studies. All interviewees gave informed consent to participate in the survey.

2.2 Survey questionnaire design

Based on the review of current industry literature, the operationalization of the research problem, and having regard to the formulated research questions, a survey questionnaire was designed. In the authors' opinion, a survey was the most appropriate tool for substantive and organizational reasons. The survey questionnaire used in the study was created in the Microsoft Forms application.

The survey questionnaire was divided into three sections: formal, substantive, and final. The formal section was the introduction to the survey. It included a return to the addressees of the survey, an introduction to the research problem, a presentation of the aim and authors of the research, assurances of the anonymity of the answers provided, and information on how the data collected in the research will be used. The substantive part is the part that includes questions to respondents on the issues of the research problem. The final part of the questionnaire was devoted to thanking for participation in the research and information regarding the possibility of obtaining the research report.

The questionnaire of the survey consisted of 23 questions, which took the form of both closed and open questions. The order of questions in the substantive part of the questionnaire was related to the division of variables into the following categories: company characteristics, market position, information on the production process, and management information. This publication presents the data obtained through the respondents' answers to questions 1 to 14, i.e. those relating to the characteristics of the company, its market position, and the production process it follows. Of the 14 questions analysed, 3 were open-ended – respondents were asked to provide their own answers. The remaining 11 questions were closed: 6 single-choice questions, 4 multiple-choice questions, and 1 question for putting the given answers in the right order. The information gathered through the respondents' answers to the remaining 9 of the 24 questions of the survey will be included by the authors in the next article, which will focus on management in Polish tanneries.

2.3 Development of a database of Polish tanneries and data collection

To present the current state and characteristics of the Polish tanning industry, a database of tanneries was developed. The initial list of tanneries was drawn up in cooperation with the Polish Chamber of Shoe and Leather Industry [17], an industry organization that has been active in the leather industry since 1989 and whose mission is to create an economically and organizationally strong leather industry in Poland and to support its activities internationally. The list was further supplemented with additional tanneries, based on information obtained from 8 online company and/or service search facilities shown in Table 1.

Table 1 Websites used in the development of the tannery database (Source: Own elaboration)

No.	Website name	Website address
1	Az-Polska.com	https://www.az-polska.com/
2	Baza firm pkt.pl	https://www.pkt.pl/
3	BiznesFinder.pl	https://www.biznesfinder.pl/
4	Cylex Szukaj Lokalnie Polska	https://www.cylex-polska.pl/
5	Firmy.net	https://www.firmy.net/
6	Ohmycraft.pl	https://ohmycraft.pl/
7	Panorama Firm	https://panoramafirm.pl/
8	Polski Katalog Firm	https://pkf.org.pl/

To gather information on Polish tanneries, the following search terms were entered into search engines on the websites listed in Table 1: '*tanning*', '*tannery*', and '*leather manufacturing*'. The result of applying the described search method was the creation of a database containing, for most companies, data such as company name, postal address (in some cases with exact territorial location shown on a map of the country), contact phone number(s), email address, web address.

Data collection in the study of the state and production process characteristics in Polish tanneries was carried out using:

- an online survey to which a link was sent via email – this was completed by a representative of the tannery, or
- a telephone survey in which a representative of the tannery answered questions in an online survey, the answers being marked by an interviewer.

The timeframe for the development of the tannery database and survey data collection is shown in Fig. 2. The list of tanneries located in Poland created at the end of 2022 based on information obtained from the Polish Chamber of the Footwear and Leather Industry and company search engines, included 220 companies. The email address provided online was held by 139 of them, so in the first instance, an email was sent to these enterprises in January 2023 with information about the survey, a link to the survey, and a request to complete it. Of the companies reached via email, only 5 chose to complete the survey questionnaire in the first quarter of 2023.

The database developed contained between 1 and 3 contact numbers (landline and/or mobile) for 219 companies. As a next step, a telephone contact was therefore attempted, which proved successful for 89 companies. From the information obtained from the callers, it appeared that 2 of the 89 companies had changed their business profile, while 67 tanneries (about 30 % of the 220) confirmed that their core business is related to leather manufacturing. However, 7 of the 67 companies indicated that they were currently in the process of reducing production and phasing out. A few respondents also indicated that the enterprises they represent are small, 1- or 2-person businesses. During the interviews, it became apparent that 20 companies (about 9 % of the 220) were no longer engaged in leather manufacturing, had suspended or liquidated, or were in liquidation bankruptcy. The remaining 130 companies (about 59 % of the 220), with which it was not possible to establish telephone contact despite three attempts made over six months (April to September 2023), are likely in a similar situation. After dialling the telephone numbers to these tanneries, found in the companies' online directories, the voicemail turned on or the answering machine said *"The number dialled does not exist", "There is no such number"* or *"The number dialled is not answering at the moment"*, or even though the call tone could be heard in the handset, no one answered the phone.

During a telephone interview with 67 tanneries that confirmed their active leather production, several respondents requested that we resend an email containing a link to the survey. Notably, nine callers provided alternative email addresses—either company or personal—different from those listed in our database. Email addresses were also obtained for 4 enterprises where contact via this route had previously been impossible, due to the lack of availability of an electronic address on the Internet. Only 4 of the 67 individuals representing actively operating tanneries were willing to complete the survey during the telephone interview, while 21 callers definitively refused to participate in the survey, either by telephone or email. In the end, 20 completed survey questionnaires were obtained.
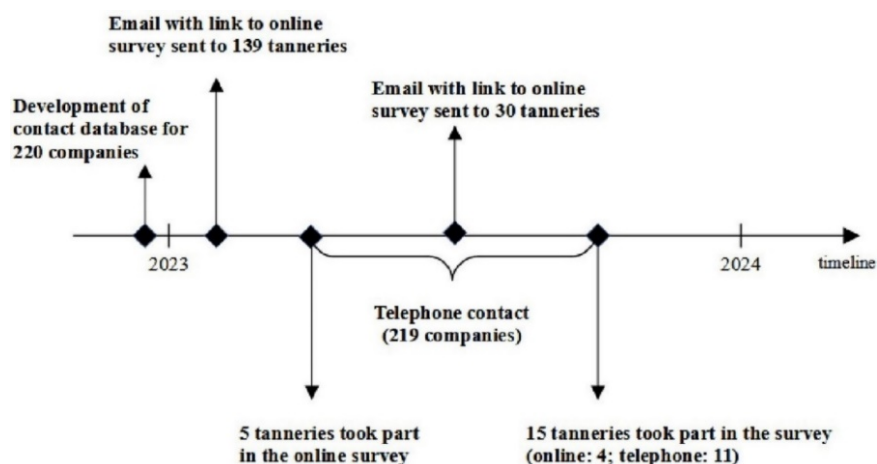

Fig. 2 Timeline of the tannery database development and survey data collection (Source: Own elaboration)

2.4 Evaluation of the manoeuvrability in the conducted studies

To assess the research methodology used and the quality of the results obtained, in the context of their representativeness, return rates were determined (Table 2). According to Kopyść [18], their estimation allows, among other things, to verify the effectiveness of the applied data collection methodology and informs whether the sample was properly selected.

Table 2 Return rates in the survey (Source: Own elaboration based on Kopyść [18])

Rates	Equation (%)	Equation components	Values of components	Values of rates (%)
Total return (r_o)	$r_o = \dfrac{p}{n} \; x \; 100$	p – number of fully completed questionnaires received	$p = 20$	9
Return rate of all questionnaires (r_w)	$r_w = \dfrac{p+c}{n} \; x \; 100$			9
Actual return of questionnaires (r_r)	$r_r = \dfrac{p}{n-z} \; x \; 100$	n – number of entities in the sample	$n = 220$	30
Rate of survey participation (r_u)	$r_u = \dfrac{p}{k} \; x \; 100$	c – number of partially completed questionnaires received	$z = 153$	30
		z – number of outdated addresses	$k = 67$	
Rate of refusal (r_x)	$r_x = \dfrac{o}{k} \; x \; 100$	k – number of entities contacted	$o = 21$	31
		o – number of entities that refused to participate in the survey	$c = 0$	

The total return rate (r_o), i.e. the rate indicating the percentage of respondents who answered taking into account all selected respondents, was estimated at around 9 %. Such a low total return rate is because in the case of 20 companies, information about their liquidation was obtained, while the vast majority (131 companies) could not be contacted in general, despite repeated attempts, which may also indicate that their activities have ceased. Another indicator estimated was the return rate of all questionnaires (r_w), reporting the percentage of completed questionnaires obtained, regardless of whether respondents answered all the questions in the questionnaire. As each of the questionnaires returned by respondents was filled in its entirety, the return rate of all questionnaires equalled the total return rate, i.e. 9 %. This demonstrates, among other things, the good quality of the survey questionnaire that was designed, as respondents wanted to answer the questions asked and had no trouble doing so.

Taking into account outdated addresses, the actual return of questionnaires rate (r_r) allowed for a more meaningful estimate of the answers to the surveys. Outdated addresses were considered to be 131 companies that could not be contacted in general (neither by email nor by telephone), 20 tanneries that confirmed that they had ceased operations, and 2 companies that reported a change in their business profile. The actual return rate was approximately 30 %. The rate of survey participation (r_u), taking into account the number of actively operating tanneries that could be contacted by telephone, was approximately 30 %. It allowed to show a relationship between the number of enterprises contacted with no answer. The rate of refusal (r_x), was around 31 %. Its estimation allowed to obtain information on respondents who were explicitly unwilling to participate in the surveys. This could have been due to, among other things, a lack of interest in the research topic, concerns about sharing company data, or other factors.

3. Results and findings

3.1 General characteristics of Polish tanneries using 20 companies as examples

Of the 20 tanneries that took part in the survey, 10 (50 %) are located in the Mazowieckie Voivodship and 5 (25 %) are located in the Małopolskie Voivodship. One company from the Dolnośląskie Voivodship, Lubelskie Voivodship, Lubuskie Voivodship, Pomorskie Voivodship,

and Śląskie Voivodship (5 % respectively) also decided to complete the survey. The territorial distribution of these tanneries is shown in Fig. 3. Considering the settlement units, the majority—14 tanneries (70 %)—are located in rural areas, while the remaining 6 are situated in smaller or larger cities (4 tanneries, or 20 %, in cities with up to 50,000 inhabitants, and 2 tanneries, or 10 %, in cities with 51,000 to 100,000 inhabitants).
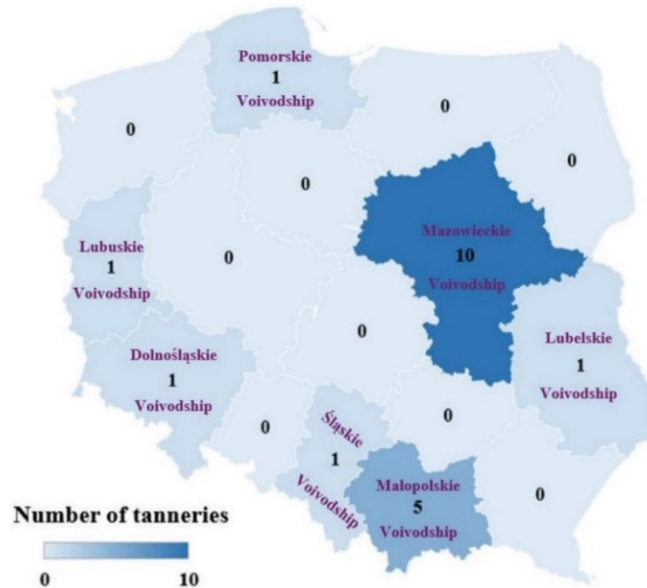

Fig. 3 Territorial location of the tanneries that took part in the study
(Source: Own elaboration using Bing © GeoNames, Microsoft, TomTom)

When characterising the geographical range of their activities, only 5 % of the tanneries indicated a local range (village/town area), while 15 % a regional range (voivodship area). The companies in question mainly operate nationally and internationally within Europe (45 % and 35 % respectively), with a select few (about 1/3) also indicating a global reach. Seven of the 20 (35 %) tanneries that participated in the study are micro-enterprises with less than 10 employees. The largest number, i.e. 9 (45 %), are small enterprises employing between 10 and 49 people. 15 % of those surveyed (3 companies), with a workforce of 50-249, are classified as medium-sized enterprises. One of the tanneries surveyed is a large company with more than 250 employees. This confirms the information contained in the *Social & Environmental Report 2020: The European Leather Industry* [1], according to which European tanneries are mainly small and medium-sized enterprises, but also large companies with an international reach. Taking into consideration the average annual net turnover of the last 10 years, in the case of 6 (30 %) tanneries it was up to PLN 1 million, while 8 companies (40 %) admitted that it was in the range of PLN 1-5 million. A revenue structure of 5-10 million PLN was achieved by 2 (10 %) tanneries, while large profits, exceeding an annual average of 10 million PLN, were indicated by 4 (20 %) of the companies participating in the study.

When asked about their market position, about 2/3 (65 %) of the tanneries considered themselves to be one of the many leather production companies in Poland. The remaining 7 companies (35 %) position themselves as a leader in the tanning sector on the domestic market. The vast majority, i.e. 14 (70 %) of the surveyed companies claim that the most important factor defining their competitiveness is the high quality of the products and/or services they offer (Fig. 4). According to a few (20 %) tanneries, their advantage is mainly determined by the price of products and/or services. One of the companies indicated an option related to technological innovation, and one also admitted that it relies on modern product design (5 % of the respondents respectively).

A lot of interesting information was obtained in the answers to an open question, in which the companies surveyed were asked to identify the most important global business trends in the tanning industry. To clarify and ensure the correct interpretation of the question, the questionnaire included an explanation that business trends are understood as, among other things, activ-

ities related to management, brand promotion, and networking. Responses to the question indicated were varied, with broad and precise answers in the case of 3 tanneries:

- *'Western markets and access to them through green certifications (origin of raw materials, eco-friendly, low-carbon, etc.). First and foremost LWG certification. Promotion of leather to consumers as a renewable, natural and strength, and therefore ecological material, which is a recycled waste from the meat industry and an answer to the 'fast-fashion' trend dominated by oil-based cheap synthetic materials. In addition, automation is an answer to the lack of skilled labor. Reduction of energy consumption in processing'*,

- *'wide-ranging certification of the enterprise and products, modern technologies (also ecological), modern machinery, promotion on the world market, cooperation with other companies of the leather industry in terms of new trends of innovative products'*,

- *'as of today, the most important business trend is to manage production in the most economically viable way in order not to lose financial liquidity due to the constantly increasing costs of services and semi-finished products used in production. Current increases in the chemicals used in production are up to 70 % by 2021. In such turmoil and stress, it is difficult to find funds for brand promotion.'*
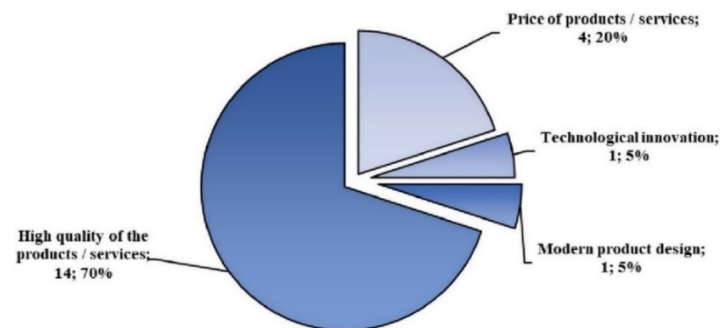

Fig. 4 Determinants of competitiveness in Polish tanneries (Source: Own elaboration)

The majority of respondents to the question on business trends gave a short, slogan-like answer, i.e. *'brand promotion', 'brand promotion, quality, customer acquisition', 'price, product promotion', 'price and quality play the most important role', 'constant contact with customers', 'social media', 'cooperation with the environment', 'establishing cooperation', 'creation of new products', 'innovation, availability of high-quality products'*. One of the tanneries indicated that *'it is necessary to re-brand, production of wellingtons, corrugated packaging, certainly not diversification of production'*.

3.2 Analysis of issues related to leather production carried out in Polish tanneries

As shown in Fig. 5, the raw materials most frequently processed in Polish tanneries are calfskin and cowhide. These hides are processed in 17 (85 %) of the 20 companies surveyed. In addition, tanneries located in Poland also use, but to a lesser extent, sheep, goat, and pig skins (options indicated by 5 of the tanneries respectively, 25 % each). 2 (10 %) of the enterprises surveyed chose the answer 'other skins'. These data are in line with European statistics, according to which tanneries located in this part of the world mainly produce bovine leather, with a share of more than 80 % [2]. This is because this type of leather can be used to manufacture a wide variety of products [19].

Continuing on the theme of raw material characteristics, respondents in the next two open questions were asked to indicate the countries of origin of the skin/hides processed in their tanneries (when enterprises source raw material from more than one country, average percentages had to be added) and to indicate the average number of tonnes of raw material processed per year. The responses obtained are shown in Table 3. Approximately half (45 %) of the tanneries surveyed purchase their raw material exclusively in Poland. Skin and hides processed in the remaining tanneries are sourced from domestic suppliers and imported from both European and non-European countries. The countries of origin of skin and hides mentioned by the respondents

are shown in Fig. 6. The country from which Polish tanneries most often import skins and hides is Germany. Analysing the responses regarding the average number of tonnes of raw material processed per year by Polish tanneries, a significant discrepancy was found between the companies. 5 (25 %) of the 20 tanneries surveyed process less than 100 tons of skin/hides per year, while 10 (50 %) process between 100 and 1000 tons. 5 (25 %) are producing on a much larger scale, i.e. processing 20,000 tonnes or more of raw material per year.
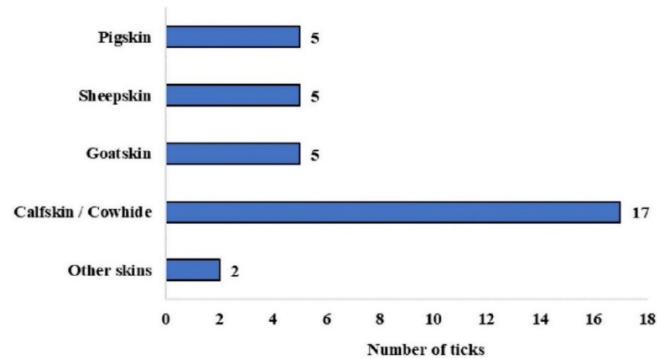

Fig. 5 Type of raw material processed in Polish tanneries (Source: Own elaboration)

Table 3 Countries of origin and volume of raw material processed in Polish tanneries (Source: Own elaboration)

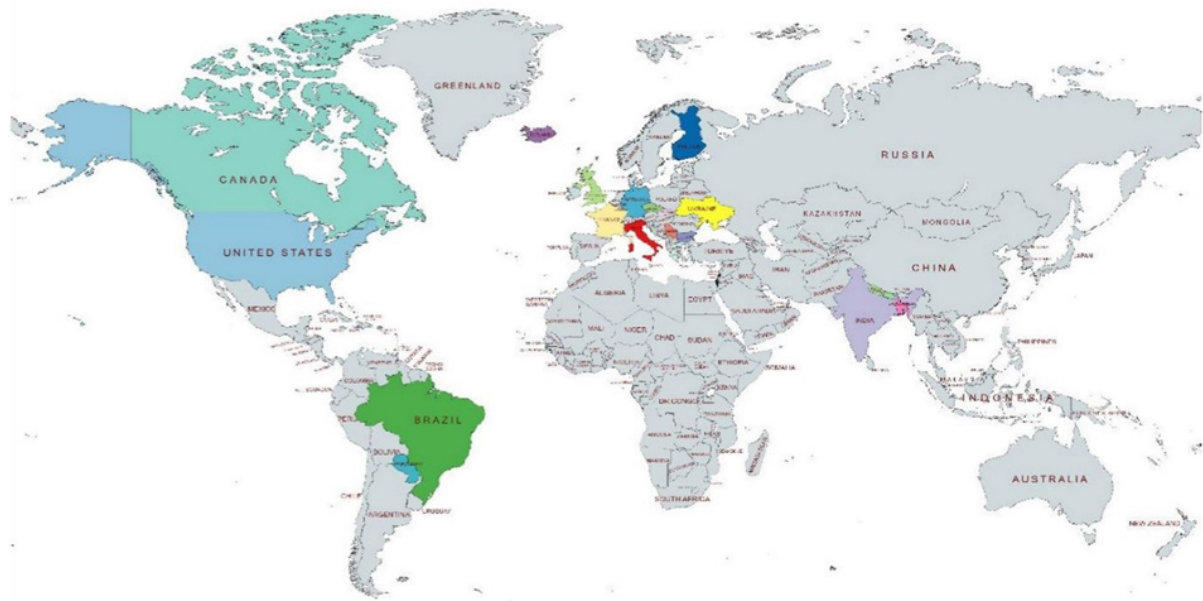| Tannery no.	Countries of origin of the skins/hides	Average share (%)	The average number of tonnes of raw material processed per year
1	Poland	50	350
	Greece	15	
	Bulgaria	15	
	Ukraine	5	
	Bangladesh, India, Brazil, Italy, Germany	15	
2	Poland	100	150
3	Poland	65-70	20 000
	Germany, USA, Canada, other countries	30-35	
4	Poland	100	1
5	Poland	40	700
	Bulgaria	60	
6	Poland	100	180
7	Poland	30	1 000
	Germany	30	
	other countries	40	
8	Poland	100	50
9	Iceland	90	9
	Germany	10	
10	Poland	100	1 700 000
11	Poland	100	100
12	Poland	100	220 000
13	the whole world, from Paraguay to Nepal	100	30
14	Poland	50	3 300 000
	Germany	50	
15	Poland	80	200
	France	20	
16	Poland	100	1
17	Finland	30	500
	Israel, Greece, France, Serbia, Germany	70	
18	Poland	30	115
	Slovakia	50	
	Czech Republic	20	
19	Germany	60	30 000
	England	20	
	USA	10	
	other countries	10	
20	Poland	100	300

Fig. 6 Countries from which raw material processed in Polish tanneries is imported
(Source: Own elaboration using mapchart.net)

Most (65 %) of the tanneries surveyed carry out a full 3-stage leather manufacturing process [20], i.e., they carry out the phase of 1) preparing the skins/hides for tanning (beamhouse), 2) proper tanning, and 3) finishing. The tanning itself (phase 2) is carried out by 1 (5 %) of the enterprises, while 3 (15 %) only carry out the final phase, i.e. finishing the material they previously acquired in the form of a tanned semi-finished product. One (5 %) of the tanneries is engaged in preparing the skins/hides for tanning (phase 1) and proper tanning (phase 2), with no further finishing. Two of the companies purchase skins/hides prepared to receive tanning, known as pelt, and subject it to tanning (phase 2) and finishing (phase 3).

The tanning method carried out in the 17 tanneries surveyed is shown in Table 4. Tanning with trivalent chromium (Cr3+) salts is carried out in 13 (65 %) enterprises. This method is widely used in leather production worldwide, as the most popular tanning method, allowing to obtain soft leathers, characterised by high strength and good elasticity useful for the production of, among others, footwear, clothing, furniture, and automotive upholstery [21]. According to estimates, about 75 % of leather made today, is produced using this tanning method [22]. Eight (40 %) of the tanneries surveyed use vegetable tannins. This the oldest tanning method makes it possible to obtain hard and resistant leathers, which are used in the production of shoe soles, bags, belts, furnishings components [23], or saddlery products such as saddles and harnesses [24]. As can be seen from Table 4, other tanning methods are also used in tanneries located in Poland, in addition to the two mentioned. It should be noted that 10 of the tanneries that completed the survey rely on a one tanning method, 5 tanneries use two tanning methods, while the remaining 2 enterprises use at least three different types of tanning agents to obtain leather with specific properties desired by the customer.

In the last question, respondents were asked to indicate the destination of the leather produced at the tanneries they represent. When completing the questionnaire, the sample types of goods given (footwear, gallantry, clothing, upholstery, etc.) had to be ranked in descending order of production volume (from the largest to the smallest share). The majority of the respondents, i.e. 12 (60 %) of the tanneries indicated that they primarily produce leather for the footwear industry. This is in line with European statistics, according to which most of the leather supplied by tanneries located in Europe is used for footwear production [6]. Two enterprises (10 % each, respectively) produce leather primarily for clothing and gallantry, while 1 (5 %) tannery oriented its production primarily for upholstery leather. Three (15 %) of the 20 companies surveyed primarily produce leather for uses other than those previously mentioned.

Table 4 Types of tanning method carried out in Polish tanneries (Source: Own elaboration)

Tannery no.	Tanning methods		
	chrome	vegetable	other
1*			
2	x		
3	x		x
4	x		
5	x	x	
6		x	
7*			
8	x		
9		x	
10	x		
11	x		
12*			
13			x
14	x		
15	x	x	
16		x	
17	x	x	x
18	x		x
19	x	x	x
20	x	x	
Total	13	8	5

*/ Tanneries where the tanning process is not carried out.

4. Discussion

According to the information obtained at the beginning of the study, approximately 220 tanneries have been located in Poland in recent years. As part of the ongoing research, the compiled database of tanneries was updated. In the process of data collection, it was confirmed that 20 companies with this profile (about 9 %) have been suspended or liquidated, or the enterprises are in liquidation bankruptcy. Another 130 companies (about 59 %) are likely to be in a similar situation and have not been contacted via email and/or telephone, despite several attempts. Additionally, 67 tanneries (about 30 %) confirmed that they are still actively engaged in leather manufacturing, but 7 of them admitted that they are currently in the process of reducing production and phasing out. Two of the 220 companies, which until recently were tanneries, have changed their business profile. This indicates that the tanning industry in Poland has undergone a process of deindustrialization in recent years, especially during the COVID-19 pandemic, which is the answer to research question RQ1.

The characterisation of the production process carried out in Polish tanneries was based on the example of 20 tanneries that participated in the survey. These enterprises are primarily located in villages and small towns (with populations up to 50,000) within the Mazowieckie and Małopolskie Voivodships. They operate not only domestically but also in Europe, with some indicating a global reach for their activities. The surveyed tanneries are usually micro or small enterprises with up to 49 employees and annual revenues of up to PLN 5 million net. About 2/3 (65 %) of the companies indicated that they were one of many leather producers in the domestic market, while the others placed themselves in a leadership role. According to the majority of the surveyed tanneries, the most important factor defining their competitiveness in the market is the high quality of the products and services they offer. This answers the research question RQ2.

Polish tanneries are aware of the relevance of global trends related to the industry's transformation towards ecology and sustainability. They point out, among other things, the need to reduce energy consumption in processing, the introduction of modern, environmentally friendly production technologies, and the need for certification confirming that their operations are environmentally focused. Current business trends for Polish tanneries also include automation of production and investment in modern machinery. An important aspect is brand promotion in domestic and global markets, contact with customers, including using social media, and cooperation with the environment. The tanneries point to the need to develop customer-appealing, in-

novative products, characterized by good quality and affordable prices. This answers the research question RQ3.

Regarding production characteristics, information was obtained that the surveyed Polish tanneries process mainly calfskin and cowhides, purchased domestically and/or imported from various European and non-European countries, in quantities ranging from 1 to as much as 3 300 000 tons per year. In most enterprises, a full 3-stage process of leather manufacturing is carried out, i.e. preparation of skins/hides for tanning, proper tanning, and finishing. Selected companies limit production to one or two of these stages. The most common tanning method used in the surveyed enterprises is chrome tanning; vegetable tanning is used less frequently. Polish tanneries are mainly oriented toward the production of raw material for the footwear industry, and to a lesser extent for clothing or gallantry. This answers the research question RQ4.

Skins and hides, which are a major by-product of the food industry, will be supplied to markets as long as meat and dairy products are consumed. The most appropriate management of them is related to the use of technological processes, thanks to which a natural material is obtained—a raw material with unique properties that cannot be replaced by any synthetic substitute [25]. Given this, as well as because of the current situation of the Polish tanning industry, it is necessary to take measures to improve the condition of this sector of the economy.

According to the authors, entrepreneurs in the tanning industry in Poland should consciously analyse and adapt to the most important global trends in the sector, which, according to the survey results, are well known to them. To increase the competitiveness of the Polish tanning industry, action should be taken aimed at improving production processes in tanneries, with the transformation of the industry towards more ecological and sustainable production. The focus should be on eco-friendly tanning methods, recycling of waste and used leather goods, and thus minimizing the negative environmental impact. What is more, it is necessary to track and adapt new technologies, consumer patterns, and changes in their preferences. Action should focus on finding alternative sources of skins and hides obtaining and manufacturing leather with new and/or better properties to increase the flexibility and diversity of offerings.

The state and characteristics of the Polish tanning industry, as presented in the article, are issues that have not been subjected to wider analysis and studies by researchers for years. This undoubtedly demonstrates the existing research gap in this area. Thus, the scientific research designed and carried out made it possible to obtain up-to-date knowledge regarding the situation of this important branch of the national industry, the directions of its changes and the most important problem, which is the progressive deindustrialization of the industry. Similar topics have also not been addressed in the current literature with regard to other European and non-European countries. Therefore, the studies presented in the article may inspire to perform similar analyses with regard to other countries. Based on such studies, it would be possible to draw broader conclusions and assess the situation of the European and global tanning industry, point out the challenges and problems it faces, and propose concrete solutions to them.

5. Research recommendations and limitations

Further research into the structural changes and needs of the Polish tanning industry in the context of the deindustrialization process is recommended. The research should include an analysis of the causes and consequences of this transformation and the search for new development opportunities in this sector. In addition, the authors also suggest research into the development of leather manufacturing technologies in a more environmentally friendly direction to create sustainable products with better properties and higher quality. An example is the research that has been carried out for several years, also in Poland, on the use of natural biocides - essential oils in the production of leather for the internal parts of footwear, to obtain raw material with antimicrobial properties [26, 27]. In the context of the development of the tanning industry, not only in Poland, but around the world, it would also be worth noting the possibility of applying machine learning solutions. Support in this regard could contribute, among other things, to increasing production efficiency, reducing costs, improving product quality and reducing the negative impact of the industry in question on the environment. The application of machine learning in

manufacturing processes and beyond is a topical and frequently addressed issue in publications that have appeared in recent years [28-29].

When planning further research, the limitations encountered by the article's authors should be borne in mind. Conducting survey research may face various constraints, both general and specific to the research topic. Limitations when conducting survey research to assess the current state of the tanning industry in Poland and characteristics of the production process include several aspects:

- For survey results to be reliable, the sample should be representative of the entire population surveyed. In the case of a survey on the state of the tanning industry, difficulties may arise, for example, from not having access to a complete list of all companies in the sector.
- Another limitation proves to be the low response rate to the survey and the low involvement of respondents. This may affect the representativeness of the sample and the reliability of the results. Some information about the tanning industry in Poland may be confidential or protected by law, which may make it difficult to collect a complete picture.
- It is also worth bearing in mind the subjectivity of responses. Respondents may have opinions on the state of the tanning industry that may be formed by their own experiences or prejudices. This can result in the introduction of errors or distortion of results.

To effectively carry out further research in the area of the Polish tanning industry, it is recommended to increase the motivation of respondents. In addition, the authors of the study recommend making study visits to tanneries and conducting in-depth interviews with tannery entrepreneurs.

Acknowledgement

Funding

References

[1] Confederation of National Associations of Tanners and Dressers of the European Community. Social & environmental report 2020, The European leather industry, from *https://euroleather.com/doc/SER/European%20Leather%20Industry%20-%20Social%20and%20Environmental%20Report%202020%20-%20EN%20web.pdf*, accessed March 29, 2024.
[2] Skills 4 Smart TCLF 2030. European leather industry, Leather sector, from *https://s4tclfblueprint.eu/project/tclf-sectors/european-leather-industry/*, accessed April 19, 2024.
[3] IndexBox. World - leather - market analysis, forecast, size, trends and insights, from *https://www.indexbox.io/store/world-leather-market-report-analysis-and-forecast-to-2020/*, April 21, 2024.
[4] Lanxess (2019). Przemysł pełen kontrowersji: Produkcja skór [An industry full of controversy: Leather manufacturing], *Chemia i Biznes*, Vol. 6, No. 57, 28-30.
[5] Kim, H.Y., Kwon, Y.J. (2017). Blurring production-consumption boundaries: Making my own luxury bag, *Journal of Business Research*, Vol. 74, 120-125, doi: 10.1016/j.jbusres.2016.10.022.
[6] Directorate General for Internal Market, Industry, Entrepreneurship, European Commission. Textiles, leather and fur industries, from *https://single-market-economy.ec.europa.eu/sectors/textiles-ecosystem/textiles-leather-fur_en*, accessed April 12, 2024.
[7] Data Bridge Market Research. Global leather goods market – Industry trends and forecast to 2030, from *https://www.databridgemarketresearch.com/reports/global-leather-goods-market*, accessed March 29, 2024.

[8] Cafasso, A.M. (2023). From hide to handbag: A holistic review of the leather industry and its environmental impact, *GLO* 100 *Introduction to global studies, Policy analysis papers,* Salve Regina University, from *https://digitalcommons.salve.edu/glo100/3*, accessed April 29, 2024.

[9] Karaosman, H., Perry, P., Brun, A., Morales-Alonso, G. (2020). Behind the runway: Extending sustainability in luxury fashion supply chains, *Journal of Business Research*, Vol. 117, 652-663, doi: 10.1016/j.jbusres. 2018.09.017.

[10] Marconi, M., Marilungo, E., Papetti, A., Germani, M. (2017). Traceability as a means to investigate supply chain sustainability: The real case of a leather shoe supply chain, *International Journal of Production Research*, Vol. 55, No. 22, 6638-6652, doi: 10.1080/00207543.2017.1332437.

[11] Staikos, T., Rahimifard, S. (2007). A decision-making model for waste management in the footwear industry, *International Journal of Production Research* Vol. 45, No. 18-19, 4403-4422 doi: 10.1080/00207540701450187.

[12] Kulas, B. (2022). Przemysł w Polsce [Industry in Poland], Geografia24.pl, from *https://geografia24.pl/przemysl-w-polsce/*, accessed April 25, 2024.

[13] NESTOR Doradztwo Finansowe i Gospodarcze Ireneusz Jabłoński. Program optymalnego cyklu szkoleń dla nowo przyjmowanych osób do sektora przemysłu mody i innowacyjnych tekstyliów zgodny z zapotrzebowaniem pracodawców [Program of the optimal training cycle for newcomers to the fashion and innovative textile industry conforming to employer demands], PARP Grupa PFR, Sektorowa Rada ds. Kompetencji Moda i Innowacyjne Tekstylia.

[14] Polish Chamber of Shoe and Leather Industry. Ogólnopolskie badanie kondycji polskiej branży obuwniczo-skórzanej. Wyniki ankiety. [National study of the condition of the Polish footwear and leather industry. Survey results], from *https://www.pips.pl/ogolnopolskie-badanie-kondycji-polskiej-branzy-obuwniczo-skorzanej-wyniki-ankiety/*, accessed October 21, 2024.

[15] Statistics Poland. Kwartalna informacja o podmiotach gospodarki narodowej w rejestrze regon rok 2023 [Quarterly information on entities of the national economy in the regon register for 2023], from *https://stat.gov.pl/obszary-tematyczne/podmioty-gospodarcze-wyniki-finansowe-zmiany-strukturalne-grup-podmiotow/kwartalna-informacja-o-podmiotach-gospodarki-narodowej-w-rejestrze-regon-rok-2023,7,13.html*, accessed April 20, 2024.

[16] Sektorowa Rada ds. Kompetencji Moda i Innowacyjne Tekstylia. Poradnik: Sektor skórzany [Handbook: Leather sector], from *http://www.modakompetencje.pl/wp-content/uploads/2019/11/Poradnik_sektor_sk%C3%B3rzany.pdf*, accessed April 19, 2024.

[17] Polish Chamber of Shoe and Leather Industry. Polska Izba Przemysłu Skórzanego: Wejdź w naszą skórę [Polish Chamber of Shoe and Leather Industry: Go into our skin], from *https://www.pips.pl/*, accessed April 10, 2024.

[18] Kopyść, P., Jak liczyć zwrotność w badaniach marketingowych [How to count maneuverability in marketing research], Kraina Biznesu, from *https://krainabiznesu.pl/projektowanie-biznesu/jak-liczyc-zwrotnosc-w-badaniach-marketingowych/*, accessed April 22, 2024.

[19] Kopitar, D., Bosnjak, F.Z. Akalovic, J., Skenderi, Z. (2022). Thermophysiological properties of bovine leather in dependence on the sampling point, tanning and finishing agents, *Journal of Industrial Textiles,* Vol. 51, 5 suppl., 8906S-8924S, doi: 10.1177/15280837221077048.

[20] National Institute of Industrial Research, Board of Consultants and Engineers (2011). *Leather processing & Tanning technology: Handbook*, National Institute of Industrial Research Project Consultancy Services, Dehli, India.

[21] G-Star Raw C.V. Guideline for leather manufacturers, from *https://img2.g-star.com/image/upload/v01/CSR/PDF/Guide_for_Leather_Manufacturers_Version_1.0.pdf*, accessed April 21, 2024.

[22] Leather Naturally. Summary of the different types of tanning, Manufacturing, leather naturally, from *https://www.leathernaturally.org/resources/fact-sheets/summary-of-the-different-types-of-tanning/*, accessed April 10, 2024.

[23] Cámara Argentina de Productores de Extractos de Quebracho, The vegetable tanning process, from https://www.tannins.org/the-vegetable-tanning-process/ accessed March 29, 2024.

[24] Gillan, K., Murray, J. (2019*). Comprehensive guide to leather repair and restoration: Leather repair training manual,* Advanced Leather Solutions, Anderson, California, USA.

[25] Textile Exchange. Preferred fiber & Materials, Market report 2021, from *https://textileexchange.org/app/uploads/2021/08/Textile-Exchange_Preferred-Fiber-and-Materials-Market-Report_2021.pdf*, accessed April 21, 2024.

[26] Bielak, E., Marcinkowska, E., Syguła-Cholewińska, J. (2020). Investigation of finishing of leather for inside parts of the shoes with a natural biocide, *Scientific Reports*, Vol. 10, No. 3467, 1-10, doi: 10.1038/s41598-020-60285-y.

[27] Bielak, E., Sawoszczuk, T., Syguła-Cholewińska, J. (2023). Application of chromatographic and microbiological analyses to identify and assess the durability of antimicrobial properties of innovative materials for the footwear industry – Leather modified with Origanum vulgare oil, *Industrial & Engineering Chemistry Research*, Vol. 62, No. 14, 5864-5876, doi: 10.1021/acs.iecr.2c04527.

[28] Jankovič, D., Šimic, M., Heraković, N. (2024). A comparative study of machine learning regression models for production systems condition monitoring, *Advances in Production Engineering & Management*, Vol. 19, No. 1, 78-92, doi: 10.14743/apem2024.1.494.

[29] Deng, G.F. (2023). Dynamic price competition market for retailers in the context of consumer learning behavior and supplier competition: Machine learning-enhanced agent-based modeling and simulation, *Advances in Production Engineering & Management*, Vol. 18, No. 4, 434-446, doi: 10.14743/apem2023.4.483.

APEM *journal*

Optimization of reverse logistics network for end-of-life vehicles: A Shanghai case study

Yao, J.[a,*]

[a]China Iron & Steel Research Institute Group, Beijing, P.R. China

ABSTRACT

With the surge in car ownership, the end-of-life vehicles recycling market has shown enormous development potential. As the reverse logistics network for recycling end-of-life vehicles suffers from high operating costs and low recycling rates, there is an urgent need to upgrade the actual recycling measures for end-of-life vehicles to an operable level. This article first uses the OGM (1, N) model to predict the number of end-of-life vehicles in the coming years. At the same time, a reverse logistics network model with the second-hand car market as the recycling centre was constructed with the goal of minimizing the total cost of the end-of-life vehicles reverse logistics network. The network model was simulated using mixed integer programming (MILP), and the optimal solution was solved through LINGO 12 programming. Through an example analysis of Shanghai, it is found that the market of end-of-life vehicles will embrace growth, and it is verified that the optimized reverse logistics network can effectively reduce the operation cost and logistics cost of recycling centres, and can effectively improve the actual recycling rate of end-of-life vehicles. Finally, the optimized site selection results are obtained, and a specific traffic distribution scheme is proposed, which is crucial for promoting cars that meet scrap standards to be recycled through formal channels and reducing logistics costs for recycling enterprises.

1. Introduction

With the continuous development of the Chinese automobile market, China became the world's largest producer and seller of automobiles in 2009. As of 2020, China's car ownership has reached 281 million units, and the end-of-life car recycling market has huge potential. According to statistics, the scrap rate of vehicles in China is about 3 %, while the scrap rate in developed countries reaches 6 % to 8 %. At the same time, the recycling and dismantling rate in China is relatively low, only 1 % to 1.5 % of the car ownership, while the recycling and dismantling rate in developed countries is 5 % to 6 %. In the face of the contradiction between the rapid growth of car ownership and the low recovery rate of end-of-life vehicles, in the context of a low-carbon and digital economy [1, 2], China's reverse logistics services led by third-party end-of-life vehicles recycling and dismantling enterprises are facing great challenges, for which the Chinese government has also formulated relevant policies and regulations [3]. For example, the Implementation Rules for the Management of End-of-Life Motor Vehicles Recycling (2020) proposed the goal of market-oriented, specialized, and intensive development of the end-of-life motor vehicle recycling and dis-

mantling industry. However, China's end-of-life vehicles recycling companies still face some problems. Scattered resources, insufficient capital investment, sloppy business management, outdated dismantling technology and a lack of standardisation and normalisation of recycling and dismantling operations are among their main challenges. Moreover, dismantling enterprises still focus on the sale of scrap steel as their main profit target, neglecting the added value of components, and the reuse rate of components is very low. There are even some illegal and non-standard recycling enterprises, which not only disrupt the normal order of the end-of-life vehicles dismantling and recycling market, but also have a negative impact on environmental protection and resource utilization. In addition, according to the lifecycle of automotive products, the peak period of mandatory scrapping standards for automobiles has not yet been reached, and the number of end-of-life vehicles is continuously increasing at a rate of over 15 % per year. The future surge in the end-of-life vehicles market also poses greater challenges to environmental protection and resource utilization. In this regard, in-depth study of automotive reverse logistics can achieve the rational use of resources, reduce environmental pollution and enhance the economic benefits of enterprises, solve the problems faced by the industry and achieve sustainable development, which is of great significance to improve the overall level and competitiveness of China's end-of-life vehicles recycling and dismantling industry.

2. Literature review

Scholars have conducted research using different research methods in predicting the number of end-of-life vehicles and car ownership [4]. Among them, Hao *et al.* proposed a combined prediction model composed of grey model, exponential smoothing method and artificial neural network optimized by Particle swarm optimization (PSO) algorithm to solve the nonlinear characteristics and uncertainty problems of the number of end-of-life vehicles recovered, and verified the accuracy of the prediction through an example [5]. Mohammadali-Ali *et al.* considered vehicle life (EOL factor) to improve the effectiveness of end-of-life vehicles recycling when establishing a system dynamics model [6]. Zhang *et al.* used singular spectrum analysis (SSA) of univariate time series models and vector autoregressive (VAR) models of multivariate models to predict car ownership [7]. Dargay and Gately, from an economic perspective, linked car ownership with per capita income and combined it with dynamic econometric models to establish a prediction model for car ownership [8].

At present, research on the optimization of reverse logistics networks for end-of-life vehicles has achieved many results, and the objective function used in existing research is mainly the lowest total cost. Demirel and Gokcen proposed a mixed integer Linear programming model for the reverse logistics network design of used and end-of-life vehicles under the participation of different roles, which successfully reduced the cost of network transportation of end-of-life vehicles and related materials [9]. Lin *et al.* established a FLAERN mathematical model with the goal of cost minimization to solve the facility location allocation problem in the end-of-life vehicles recycling network, and proposed an algorithm based on artificial bee colony for optimization [10]. Govindan and Gholizadeh comprehensively investigated the relationship between the elastic sustainable RLs and the variable and flexible capabilities of facilities in electric vehicles, and adopted a robust optimized cross-entropy hybrid solution method to calculate the total cost of the elastic sustainable RLN with the main goal of minimizing the total cost of the RLN [11]. Seval and Ozturk developed a mathematical planning model for reverse logistics network design to manage a reverse flow network of end-of-life vehicles in the context of a manufacturer being responsible for the entire lifecycle of its products, which was used to determine the number and location of network facilities as well as the volume of material flows [12]. Amin *et al.* conducted a Case study on the key participants in North American ELVs collection for the reverse logistics path problem of recycling used vehicles, and proposed a two-stage heuristic solution to maximize the optimal distribution of dealers between internal fleets and between internal fleets and external operators [13]. Min *et al.* proposed a mixed integer nonlinear programming model and genetic algorithm aimed at solving a spatially and temporally integrated reverse logistics problem involving returned products [14]. Marin and Pelegrin analysed the Return to Plant Location Problem (RPLP)

based on the heuristic and exact solution method of Lagrange decomposition [15]. Du and Evans established a double objective MIP optimization model for the reverse logistics network problem of Third-party logistics companies providing logistics services for after-sales service networks, and designed a solution consisting of discrete search method, dual Simplex algorithm method and constraint method [16].

Similar to the reverse logistics network optimization problem of end-of-life vehicles, facilities location [17], production Linear programming [18, 19], supply chain performance evaluation [20], production and procurement planning are also combinatorial optimization problems [21, 22], and the modelling of combinatorial optimisation problems as mixed integer programming models is widely used in the solution of these problems [23, 24]. And the global optimization software LINGO is used for rapid solution, which has achieved good results in the examples. For example, Yong and Jing used LINGO to establish a logistics network optimization design model, select a carbon emission distribution centre demand matching model, a distribution path optimization model and a multi-objective optimization model aimed at minimizing carbon emissions, and optimize the model [25]. Xiao *et al.* constructed a four level reverse logistics network model including the origin of end-of-life vehicles, the recycling centre, the remanufacturing centre and the disassembler, and used LINGO to solve the mixed integer Linear programming mathematical model established, pointing out that the location of the dismantling centre and the capability rating strategy have an important impact on the total cost of the logistics network [26, 27].

Based on this, this article first predicts the future prospects of the end-of-life vehicles market using the OGM (1, N) model based on grey theory. Further, the reverse logistics network model with the second-hand automobile market as the recovery and transit centre is constructed by fully considering the factors such as the dismantling capacity of the dismantling point, the logistics volume balance of each network node, and the Fixed cost. Select appropriate decision variables, construct a linear objective function with the minimum total cost, model the reverse network problem of end-of-life vehicles as a Mixed Integer Programming model, and then use LINGO software to obtain the optimal solution. Finally, a case study was conducted to verify the effectiveness of the model.

3. Research methodology

3.1 Prediction model for the number of end-of-life vehicles

The specific modelling process of OGM (1, N) model is as follows.

(1) Grey sequence generation

Set $X_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n))$ as the dependent variable sequence. Sequence $X_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n))$ is an independent variable sequence with high correlation with sequence $X_1^{(0)}$.

(2) Perform a cumulative generation (1-AGO) on $X_i^{(0)}$ to obtain the sequence $X_1^{(1)}$:

$$X_i^{(1)} = \left(x_i^{(1)}(1), x_i^{(1)}(2), \dots, x_i^{(1)}(n) \right) \tag{1}$$

Among them: $x_i^{(1)}(k) = \sum_{i=1}^{k} x_i^{(0)}(i), k = 1,2,\dots,n$.

(3) Generate the nearest neighbour mean sequence $Z_1^{(1)}$ from $X_1^{(1)}$:

$$Z_1^{(1)} = \left(z_1^{(1)}(2), z_1^{(1)}(3), \dots, z_1^{(1)}(m) \right) \tag{2}$$

Among them: $z_1^{(1)}(k) = [x_1^{(1)}(k) + x_1^{(1)}(k-1)]/2, k = 2,3,\dots,m$.

(4) Build OGM (1, N) model:

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{i=2}^{k} b_i x_i^{(1)}(k) + h_1(k-1) + h_2 \tag{3}$$

The $h_1(k-1)$ and h_2 in the equation are referred to as the linear correction term and grey action of the OGM (1, N) model, respectively, a is referred to as the development coefficient, $b_i x_i^{(1)}(k)$ is called the driving term, b_i is called the driving coefficient.

(5) Parameter Estimation of OGM (1, N) model

The Least Squares Satisfaction of Parameter Column $\hat{a} = [b_1, b_2, \ldots, b_N, a, h_1, h_2]^T$ in OGM (1, N) model:

$$\hat{a} = (B^T B)^{-1} B^T Y \tag{4}$$

Among them:

$$B = \begin{bmatrix} x_2^{(1)}(2) & x_3^{(1)}(2) & \cdots & x_N^{(1)}(2) & -z_2^{(1)}(2) & 1 & 1 \\ x_2^{(1)}(3) & x_3^{(1)}(3) & & x_N^{(1)}(3) & -z_2^{(1)}(3) & 2 & 1 \\ \vdots & \vdots & \ddots & \vdots & & & \vdots \\ x_2^{(1)}(m) & x_3^{(1)}(m) & \cdots & x_N^{(1)}(m) & -z_2^{(1)}(2) & m-1 & 1 \end{bmatrix} \tag{5}$$

$$Y = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(m) \end{bmatrix} \tag{6}$$

So $\hat{x}_1^{(0)}(k) = \sum_{i=2}^{k} b_i x_i^{(1)}(k) - a z_1^{(1)}(k) + h_2 + h_1(k-1)$ is the differential model of OGM (1, N).

$$\hat{x}_1^{(0)}(k) = \sum_{t=1}^{k-1} \left[u_1 \sum_{i=2}^{N} u_2^{t-1} b_i x_i^{(1)}(k-t+1) \right] + u_2^{k-1} \hat{x}_1^{(1)}(1) + \sum_{j=0}^{k-2} u_2^j \left[(k-j)u_3 + u_4 \right] \tag{7}$$

In the Eq. 7, $k = 2,3,\ldots$.

$$u_1 = \frac{1}{1+0.5a}, u_2 = \frac{1-0.5a}{1+0.5a}, u_3 = \frac{h_1}{1+0.5a}, u_4 = \frac{h_2-h_1}{1+0.5a}$$

(6) Obtaining predicted values

The following is the cumulative reduction formula of OGM (1, N) model:

$$\hat{x}_1^{(0)}(k) = \hat{x}_1^{(1)}(k) - \hat{x}_1^{(1)}(k-1), k = 2,3,\ldots,m \tag{8}$$

3.2 Construction of reverse logistics network model for end-of-life vehicles

This article constructs a reverse logistics network model with the second-hand car market as the recycling centre, based on the two key issues of reverse logistics network design and low recycling rate in the optimization of recycling end-of-life vehicles, with the goal of minimizing the total cost of the reverse logistics network for end-of-life vehicles. At the same time, constraints such as the conservation of logistics volume between network nodes, limitations on the processing capacity of recycling centres, limitations on the dismantling capacity of dismantling centres, and the number of established logistics network recycling centres are set to ensure overall coordination among relevant organizations and fully complete recycling tasks.

Parameter settings

To convert the reverse logistics in the recycling process of end-of-life vehicles into a specific mathematical model, some parameters need to be annotated and explained. The parameter symbols and definitions used in this study are shown in Table 1.

<div align="center">Table 1 Definition of parameter symbols</div>

Category	Symbol	Definition
Set	$KHQ(i)$	Customer area address set ($i = 1,2,3, \dots, n, \forall i \in K$)
	$HSZX(j)$	Recycling centre addresses set ($j = 1,2,3, \dots, m, \forall j \in HSZX$)
	$CJZX(z)$	Dismantling centre address set ($z = 1,2,3, \dots, p, \forall z \in CJZX$)
	$AZ(w)$	Address of remanufacturing centre or landfill site set ($w = 1,2, \dots, t, \forall w \in AZ$)
	$U(i,j)$	Relevant indicator set from customer area i to recycling centre j ($i = 1,2,3, \dots, n, j = 1,2,3, \dots, m, \forall i \in K$，$\forall j \in HSZX$)
	$T(j,z)$	Relevant indicator set from recovery centre j to dismantling centre z ($j = 1,2,3, \dots, m, z = 1,2,3, \dots, p, \forall j \in HSZX$，$\forall z \in CJZX$)
	$R(z,w)$	Relevant indicator set from dismantling centre z to remanufacturing centre or landfill plant w ($z = 1,2,3, \dots, p, w = 1,2, \dots, t, \forall z \in CJZX, \forall w \in AZ$)
Parameter	$CH1(j)$	Fixed investment cost of recycling centre j
	$CH2(j)$	Operating costs amortized per vehicle recovered by the recycling centre j
	$CC(z)$	The dismantling cost of dismantling a single end-of-life vehicles
	$LKH(i,j)$	Distance from recycling centre i to dismantling centre j
	$QK(i)$	Total number of end-of-life vehicles in customer area i
	$N(j)$	Maximum recycling capacity of recycling centre j
	$E(z)$	Maximum dismantling capacity of dismantling centre z
	$M(z)$	Recycling quantity of dismantling centre
	$V1$	Unit freight from customer area to recycling centre (RMB/vehicle · kilometre)
	$V2$	Unit shipping cost from recycling centre to dismantling centre (RMB/vehicle · kilometre)
	A	Unit transportation from dismantling centre to component remanufacturing centre (RMB/vehicle · kilometre)
	B	Unit transportation from dismantling centre to steel enterprise (RMB/vehicle · kilometre)
	C	Unit transportation from dismantling centre to landfill site (RMB/vehicle · kilometre)
Decision variables	$Y(j)$	If a new recycling centre is set up at h, take 1; otherwise, take 0
	$QKH(i,j)$	The number of end-of-life motor vehicles transported from customer area i to recycling centre j
	$QHC(j,z)$	The number of end-of-life motor vehicles from recycling centre j to dismantling centre z
	$QCA(z,w)$	The number of vehicles transporting motor vehicle waste from the dismantling centre j to the waste landfill w
	$QCZ(z,w)$	The number of automotive engines transported from dismantling centre j to remanufacturing centre w
	$QCG(j)$	The number of end-of-life vehicles and scrap metal transported from the dismantling centre to the steel plant

Model specification

- This article studies the logistics network of a single product and cycle, where the weight of a single vehicle represents the overall quality.
- The transportation distance between each logistics node is known, and the transportation cost per unit distance is known, and the transportation cost is positively correlated with the transportation distance.
- Material and component losses during transportation are not taken into account, including the scrap disposal rate during dismantling.
- The alternative solutions for the construction or expansion of facilities in the automotive reverse logistics network are known, including their alternative locations and number, location selection near the second-hand car market in a certain area, and setting their recycling capacity in advance.
- The transportation from the recycling centre to the dismantling centre will be estimated based on transporting multiple end-of-life vehicles at once.
- Ignoring the uncertainty of time and the dynamism of the supply chain, the cost of inventory backlog is included in the unit operating cost.
- The third-party logistics company can reach cooperation with them, and the cooperation status will not change in a short time.

Objective function

The objective function of establishing a model with the lowest total cost of reverse logistics network for end-of-life vehicles:

$$min\ TC = TC1 + TC2 + TC3 + TC4 + TC5 + TC6 + TC7 + TC8 \tag{9}$$

- Fixed investment cost of network recycling transit centres ($TC1$)

$TC1$ is the sum of the products of $CH1(j)$ and $Y(j)$:

$$TC1 = \sum_{j=1}^{n} CH1(j)\ Y(j) \tag{10}$$

- Operating costs of network recycling transit centres ($TC2$)

$TC2$ is the sum of the products of $QKH(i,j)$ and $CH2\ (j)$:

$$TC2 = \sum_{i=1}^{m} \sum_{j=1}^{n} QKH(i,j)\ CH2(j) \tag{11}$$

- The total dismantling cost of the dismantling company ($TC3$)

$TC3$ is the sum of the products of $QHC(j,z)$ and $CC\ (z)$:

$$TC3 = \sum_{j=1}^{n} \sum_{z=1}^{p} QHC(j,z)\ CC(z) \tag{12}$$

- The total logistics cost of network transportation

The calculation of the total transportation cost is based on the sum of the products of the distance between each network node, the freight cost per kilometre, and the quantity of recycling between network nodes.

The total transportation cost includes the shipping cost from the recycling centre to the dismantling centre ($TC4$):

$$TC4 = \sum_{i=1}^{m} \sum_{j=1}^{n} LKH(i,j)\ QKH(i,j)\ V1 \tag{13}$$

The total cost of shipping between the recycling centre and the dismantling centre ($TC5$):

$$TC5 = \sum_{j=1}^{n} \sum_{z=1}^{p} LHC(j,z)\ QHC(j,z)\ V2 \tag{14}$$

Freight cost from dismantling centre to remanufacturing company ($TC6$), garbage factory ($TC7$), steel factory ($TC8$) are shown as follows:

$$TC6 = \sum_{z=1}^{p} \sum_{w=1}^{t} LCA(z,w)\ QKH(z,w)\ A \tag{15}$$

$$TC7 = \sum_{z=1}^{p} \sum_{w=1}^{t} LCZ(z,w)\ QKH(z,w)\ F \tag{16}$$

$$TC8 = \sum_{z=1}^{p} LCG(j)\ QCG(j)\ G \tag{17}$$

Restraint condition

Based on the current situation of the reverse logistics network for end-of-life vehicles, this study establishes constraints from aspects such as the conservation of logistics volume between net-

work nodes, limitations on the processing capacity of recycling centres, limitations on the dismantling capacity of dismantling centres, the number of established in the logistics network recycling centres, and integer constraints.

(1) Constraints on conservation of logistics volume

Based on the principle of material flow conservation between various nodes in the logistics network, the following constraint conditions for material flow conservation are established.

Balance of total material flow of end-of-life vehicles:

$$\sum_{i=1}^{m} QK(i) = \sum_{i=1}^{m}\sum_{j=1}^{n} QKH(i,j) = \sum_{j=1}^{n}\sum_{z=1}^{p} QHC(j,z) =$$
$$= \sum_{z=1}^{p}\sum_{w=1}^{t} QCA(z,w) = \sum_{z=1}^{p}\sum_{w=1}^{t} QCA(z,w) = \sum_{Z=1}^{p} QCG(z) \tag{18}$$

Conservation of material flow at node i in each customer area:

$$QK(i) = \sum_{j=1}^{n} QKH(i,j) \qquad (i = 1,2,3,\dots,n) \tag{19}$$

Conservation of material flow at various network recycling and transfer centre nodes:

$$\sum_{i=1}^{n} QKH(i,j) = \sum_{z=1}^{p} QHC(j,z) \, (j = 1,2,3,\dots,m) \tag{20}$$

Conservation of material flow at nodes from steel mills, garbage dumps, remanufacturing companies to dismantling enterprises:

$$\sum_{j=1}^{m} QHC(j,z) = \sum_{w=1}^{t} QCA(z,w) \, (z = 1,2,3,\dots,p) \tag{21}$$

$$\sum_{j=1}^{m} QHC(j,z) = \sum_{w=1}^{t} QCA(z,w) \, (z = 1,2,3,\dots,p) \tag{22}$$

$$\sum_{j=1}^{m} QHC(j,z) = QCG(j) \qquad (z = 1,2,3,\dots,p) \tag{23}$$

(2) Limitations on processing capacity of recycling centres

According to the recycling capacity limit of the recycling centre, the 0-1 variable $Y(j)$ of the recycling centre meets the minimum and maximum processing quantity limits. The value of $Y(j)$ is determined by the conservation constraint of the recycling centre's processing capacity, thereby determining the location:

$$\sum_{i=1}^{n} QKH(i,j) \leq N(j)\,Y(j) \quad (j = 1,2,3,\dots,m) \tag{24}$$

$$\sum_{z=1}^{p} QHC(j,z) \leq N(j)\,Y(j) \quad (j = 1,2,3,\dots,m) \tag{25}$$

$$\sum_{i=1}^{n} QKH(i,j) \geq P(j)\,Y(j) \quad (j = 1,2,3,\dots,m) \tag{26}$$

$$\sum_{z=1}^{p} QHC(j,z) \geq N(j)\,Y(j) \quad (j = 1,2,3,\dots,m) \tag{27}$$

$$\sum_{j=1}^{m} QHC(j,z) \leq M(z) \quad\quad (z = 1,2,3,\dots,p) \tag{28}$$

(3) Limitations on the dismantling capacity of the dismantling centre

Establish constraint conditions based on the dismantling capacity limitations of each dismantling centre to meet the law of conservation of logistics volume between each node:

$$\sum_{j=1}^{m} QHC(j,z) \geq E(z) \quad (z = 1,2,3,\dots,p) \tag{29}$$

$$\sum_{w=1}^{t} QCA(z,w) \geq E(z) \quad (z = 1,2,3,\dots,p) \tag{30}$$

$$\sum_{w=1}^{t} QCA(z,w) \leq M(z) \quad (z = 1,2,3,\dots,p) \tag{31}$$

$$\sum_{w=1}^{t} QCZ(z,w) \geq E(z) \quad (z = 1,2,3,\dots,p) \tag{32}$$

$$\sum_{w=1}^{t} QCZ(z,w) \leq M(z) \quad (z = 1,2,3,\dots,p) \tag{33}$$

(4) Limitation on the number of newly established recycling points:

$$\sum_{j=1}^{3} Y(j) = 3 \tag{34}$$

$$3 \leq \sum_{j=4}^{n} Y(j) \leq 5 \tag{35}$$

(5) Integer constraint:

$$Y(j) \in \{0，1\} \quad (j = 1,2,3,\dots,n) \tag{36}$$

4. Results of the study

4.1 Example description and data source

This article takes Shanghai, the fifth largest city in China in terms of car ownership, as a research example to analyse and verify the effectiveness of the model. Select participants closely related to the reverse logistics of end-of-life vehicles for on-site research and interviews, and obtain the required data from government official websites and relevant literature.

(1) According to Google Maps data, Fig. 1 shows the distribution of end-of-life vehicle network nodes in Shanghai.

Fig. 1 Distribution of scrap vehicle network nodes in Shanghai

(2) Distances between nodes were computed using Google Maps, with the following specifics: the distance between the customer area and the recycling centre; the distance between the recycling centre and the dismantling enterprise; the distance between the dismantling enterprise and the landfill plant; the distance between the dismantling centre and Shanghai Baosteel Group Corporation; and the distance between the dismantling enterprise and the remanufacturing company. Here, only the distance data between the 16 customer areas and the selected recycling centres for end-of-life vehicles in Shanghai are shown in Table 2.

Table 2 Distance between customer area and recycling centre

Distance (km)	Wuning Road Service Branch	Boyuan Road Service Branch	Hulan Road Service Branch	Shanghai Used Car Trade Market	Shanghai Port second-hand car trading market	Shanghai Xinzhuang Old Motor Vehicle Trading Market	Shanghai Pudong Used Motor Vehicle Trading Market	Shanghai Chongming Old Motor Vehicle Trading Market	Shanghai Bailian United Used Motor Vehicle Trading Market
Pudong New Area	23	45	23	54	17	29	13	90	18
Huangpu District	9.4	31	17	45	22	23	20	103	11
Xuhui District	25	8	23	36	48	16	21	109	18
Changning District	3.3	18	20	33	31	19	23	104	13
Jing'an District	5.4	26	17	40	30	20	21	102	11
Putuo District	3.7	19	16	26	30	22	31	103	10
Hongkou District	11	34	15	38	16	28	19	90	8.9
Yangpu District	16	35	22	43	16	29	17	89	16
Minhang District	19	23	33	36	46	5.7	21	115	27
Baoshan District	24	38	10	51	22	43	37	98	19
Jiading District	32	17	26	20	43	32	53	117	24
Jinshan District	70	67	86	73	102	54	71	168	77
Songjiang District	40	38	54	44	67	21	41	137	48
Qingpu District	38	36	55	25	62	28	52	143	50
Fengxian District	43	48	57	62	67	31	42	139	51
Chongming District	101	118	98	125	81	120	95	2	93

(3) The recycling of end-of-life vehicles is generally carried out through the recycling and dismantling company's doorstep transportation or by customers driving the vehicles that are about to be end-of-life to the recycling company. Based on actual research, this study believes that the transportation unit fuel consumption during vehicle recycling is 0.2 liters/kilometre, and the freight cost between each node is shown in Table 3.

Table 3 Freight costs between nodes

	V1	V2	A	B	C
Freight cost (RMB/kilometre · vehicle)	2.8	0.8	0.16	0.4	0.9

4.2 Prediction results of end-of-life vehicles in Shanghai

The grey correlation analysis was carried out on the relevant indicators selected for the prediction of the number of 16 end-of-life vehicles from 2011 to 2018. Finally, with 0.83 as the threshold, Gross regional product, per capita gross product, total social consumer goods, disposable income of residents, and car ownership were taken as the main factors of the grey correlation prediction. The final prediction indicates that the number of cars scrapped in Shanghai from 2019 to 2021 is 110,853, 130,054, and 140,509, respectively, as calculated using MATLAB.

4.3 Analysis of the results of the reverse logistics network model for end-of-life vehicles in Shanghai

Effectively reducing the cost of reverse logistics network

Referring to the reverse logistics network data of end-of-life vehicles in Shanghai, a mixed integer programming (MILP) simulation network model was used to simulate the operation, and LINGO 12 programming was used to solve the optimal solution. Integrate resources between various nodes to find the optimal traffic allocation result. And select a location for the recycling centre. The solution report is displayed as follows:

> Global optimal solution found.
> Objective value: 0.1302113E + 09
> Objective bound: 0.1302113E + 09
> Infeasibilities: 0.000000
> Extended solver steps: 0
> Total solver iterations: 149

The current program has a total of 283 variables, with 149 iterations. The optimal value of the objective function min $TC = 0.1302113E + 09$, which means the logistics cost is 130429.9 million RMB.

> Recovery centre cost: $TC1 + TC2 = 183800 + 811424 = 995224$
> Logistics freight cost: $TC4 + TC5 + TC6 + TC7 + TC8 = 2160869 + 658096.4 + 639380.5 + 1389350 + 822120.8 = 5669816.7$

At present, there are about 50 recycling points in the existing reverse logistics network for end-of-life vehicles, and the average operating cost of each recycling point is about 300000 RMB. The total operating cost exceeds the sum of logistics freight cost and the operating cost of the recycling centre.

Choosing the second-hand car market as the recycling and transfer centre and establishing a reverse logistics network system for end-of-life vehicles can effectively reduce the logistics and operational costs of reverse logistics for end-of-life vehicles. Allowing profits to end-of-life vehicles dismantling companies increases the profit margin of the dismantling centre, causing a large number of end-of-life vehicles flowing into informal channels such as the underground black market to flow into formal channels for scrapping.

Effectively improving the recycling rate of reverse logistics networks

Table 4 Scrap vehicle recycling data from 2015 to 2016

Time	Car ownership (10000 units)	Theoretical scrap volume (10000 units)	Actual cancellation volume (10000 units)	Actual recycling volume (10000 units)
2015	332.35	13.2	8.5	4.01
2016	359.48	14.37	9.7	4.2

The theoretical scrap quantity is generally 4 % of the total inventory, taking the historical data in Table 8 for 2016 as an example:

$$\text{Actual scrap rate} = \frac{\text{Actual cancellation volume}}{\text{Theoretical scrap volume}} \cdot 100\,\% = \frac{9.7}{14.37} \cdot 100\,\% = 67.5\,\%$$

$$\text{Actual recovery rate} = \frac{\text{Actual recycling volume}}{\text{Actual cancellation volume}} \cdot 100\,\% = \frac{4.2}{9.7} \cdot 100\,\% = 43.2\,\%$$

According to industry calculations, generally 40 % enter formal dismantling channels, 30 % are illegally dismantled and dismantled, and 30 % are re launched through second-hand modifications. Half of the 46700 end-of-life vehicles that have not been deregistered can be recovered and deregistered through inspection of the second-hand car market.

$$\text{Actual scrap rate} = \frac{\text{Actual cancellation volume}}{\text{Theoretical scrap volume}} \cdot 100\,\% = \frac{12.04}{14.37} \cdot 100\,\% = 83.8\,\%$$

$$\text{Actual recovery rate} = \frac{\text{Actual recycling volume}}{\text{Actual cancellation volume}} \cdot 100\,\% = \frac{6.535}{12.035} \cdot 100\,\% = 54.3\,\%$$

Using data from 2016, it is expected that the actual scrap rate will increase from 67.5 % to 83.8 %, and the actual recovery rate will increase from 43.2 % to 54.3 %.

Optimization of site selection for network recycling centres

According to the results of $Y(j)$, the following six second-hand car trading markets will be designated as the recycling transfer points for the reverse logistics network of end-of-life vehicles in Shanghai in the future, as shown in Table 5.

Table 5 Location of network recycling centres using the used car market as a transit point

Wuning Road Service Branch	2907 Zhongshan North Road, Putuo District, Shanghai
Boyuan Road Service Branch	Certification Hall, Floor 1, No. 969, Boyuan Road, Jiangqiao Town, Jiading District, Shanghai
Hulan Road Service Branch	Room 108, Building 4, No. 525 Hulan Road, Baoshan District, Shanghai
Shanghai Xinzhuang Old Motor Vehicle Trading Market	3318 Gudai Road, Shanghai
Shanghai Pudong Used Motor Vehicle Trading Market	1441 Yuqiao Road, Pudong New Area, Shanghai
Shanghai Bailian United Used Motor Vehicle Trading Market	3550 Gonghexin Road, Jing'an District, Shanghai

Optimized allocation of network resources

Based on the above site selection optimization results, the following resource optimization allocation suggestions are proposed to effectively reduce the total operating cost of the entire reverse logistics network for end-of-life vehicles.

(1) The flow distribution between the customer area and the recycling area is shown in Table 6.

Table 6 Flow Distribution of customer area – Recycling area (Unit: Vehicles)

Distance (km)	Wuning Road Service Branch	Boyuan Road Service Branch	Hulan Road Service Branch	Shanghai Used Car Trade Market	Shanghai Port second-hand car trading market	Shanghai Xinzhuang Old Motor Vehicle Trading Market	Shanghai Pudong Used Motor Vehicle Trading Market	Shanghai Chongming Old Motor Vehicle Trading Market	Shanghai Bailian United Used Motor Vehicle Trading Market
Pudong New Area	0	0	0	0	0	0	0	0	1236
Huangpu District	0	0	0	0	0	1110	7608	0	0
Xuhui District	0	5341	0	0	0	0	0	0	0
Changning District	694	1150	0	0	0	3043	0	0	0
Jing'an District	0	1586	3968	0	0	0	0	0	2218
Putuo District	6306	0	0	0	0	0	0	0	0
Hongkou District	0	0	0	0	0	0	0	0	9154

Table 6 (Continuation)

Yangpu District	0	0	0	0	0	0	0	0	5830
Minhang District	0	0	0	0	0	1850	0	0	0
Baoshan District	0	0	2032	0	0	0	0	0	0
Jiading District	0	923	0	0	0	0	0	0	0
Jinshan District	0	0	0	0	0	370	0	0	0
Songjiang district	0	0	0	0	0	785	0	0	0
Qingpu District	0	0	0	0	0	490	0	0	0
Fengxian District	0	0	0	0	0	452	0	0	0
Chongming District	0	0	0	0	0	0	156	0	0

(2) To reduce costs, the following flow distribution plan is proposed for the recycling centre to the dismantling centre, as shown in Table 7.

Table 7 Flow distribution of recycling centre – Dismantling centre (Unit: Vehicles)

Distance (km)	Xinguang Renewable Resources (Shanghai) Co., Ltd	Oulubao (Baosteel) Dismantling Co., Ltd	Shanghai Huadong Dismantling Co., Ltd	Shanghai Huajian Dismantling Co., Ltd	Shanghai Motor Vehicle Recycling Service Centre	Shanghai Xinzhuang Dismantling Co., Ltd	Shanghai Jiaoyun Bus Dismantling Co., Ltd
Wuning Road Service Branch	0	0	0	0	0	0	7000
Boyuan Road Service Branch	8000	0	0	0	0	0	1000
Hulan Road Service Branch	0	6000	0	0	0	0	0
Shanghai Used Car Trade Market	0	0	0	0	0	0	0
Shanghai Port second-hand car trading Market	0	0	0	0	0	0	0
Shanghai Xinzhuang Old Motor Vehicle Trading Market	0	0	0	0	0	8000	0
Shanghai Pudong Used Motor Vehicle Trading Market	0	9000	0	0	0	0	0
Shanghai Chongming Old Motor Vehicle Trading Market	0	0	0	0	0	0	0
Shanghai Bailian United Used Motor Vehicle Trading Market	0	5000	3000	2500	2800	1902	2000

(3) The article proposes that Xinguang Renewable Resources (Shanghai) Co., Ltd. will transport the dismantled waste to Chongming Garbage Dump. Within the upper limit of the treatment capacity of the Laogang Garbage Dump, the dismantling company will try its best to transport the waste to the Laogang Garbage Dump for treatment, as shown in Table 8.

Table 8 Flow distribution of dismantling centre – Landfill plant (Unit: Vehicles)

Distance (km)	Xinguang Renewable Resources (Shanghai) Co., Ltd	Oulubao (Baosteel) Dismantling Co., Ltd	Shanghai Huadong Dismantling Co., Ltd	Shanghai Huajian Dismantling Co., Ltd	Shanghai Motor Vehicle Recycling Service Centre	Shanghai Xinzhuang Dismantling Co., Ltd	Shanghai Jiaoyun Bus Dismantling Co., Ltd
Chongming Garbage Landfill Plant	8000	0	0	0	0	0	0
Laogang Solid Waste Comprehensive Utilization Base	0	20000	3000	2500	2800	9902	10000

(4) The flow distribution between the dismantling centre and the steel plant is shown in Table 9.

Table 9 Flow distribution of dismantling centre – Steel plant (Unit: Vehicle)

Distance (kilometre)	Xinguang Renewable Resources (Shanghai) Co., Ltd	Oulubao (Baosteel) Dismantling Co., Ltd	Shanghai Huadong Dismantling Co., Ltd	Shanghai Huajian Dismantling Co., Ltd	Shanghai Motor Vehicle Recycling Service Centre	Shanghai Xinzhuang Dismantling Co., Ltd	Shanghai Jiaoyun Bus Dismantling Co., Ltd
Shanghai Baosteel Group Corporation	8000	20000	3000	2500	2800	9902	10000

(5) The article puts forward the opinion that Oulubao (Baosteel) Vehicle Dismantling Co., Ltd. and Shanghai Jiaoyun Bus Dismantling Co., Ltd. will transport the five major assemblies of the recovered vehicles to SAIC Motor, and other dismantling companies in Shanghai will transport the five major assemblies of the recovered vehicles to Shanghai Fumei Auto Automatic Transmission Technical Service Co., Ltd, as shown in Table 10.

Table 10 Flow distribution of dismantling centre – Remanufacturing company (Unit: Vehicle)

Distance (km)	Xinguang Renewable Resources (Shanghai) Co., Ltd	Oulubao (Baosteel) Dismantling Co., Ltd	Shanghai Huadong Dismantling Co., Ltd	Shanghai Huajian Dismantling Co., Ltd	Shanghai Motor Vehicle Recycling Service Centre	Shanghai Xinzhuang Dismantling Co., Ltd	Shanghai Jiaoyun Bus Dismantling Co., Ltd
Shanghai Xingfu Rebede Powertrain Co., Ltd. (SAIC Motor)	0	20000	0	0	0	0	10000
Shanghai Fumei Auto Automatic Transmission Technical Service Co., Ltd. (authorized by Dongfeng Peugeot-Citroën and Great Wall Motor)	8000	0	3000	2500	2800	9902	0

5. Discussion

In terms of theoretical contributions, this study mainly analyses and predicts the prospects of Shanghai's automobile market first, then breaks the tradition that most scrap automobile dismantling companies in Shanghai set up recycling centres only to collect materials rather than serve as transfer points, and establishes a tripartite logistics network system with the second-hand car trading market as the recycling centre to realize the optimal layout of full coverage of recycling points and distribution systems. The most efficient implementation of transport.

At the same time, this study also has great practical significance to the society. First of all, the automobile manufacturing industry consumes a lot of resources, and its development speed is accelerating, and the demand for resources is also increasing. Research on the reverse logistics network of end-of-life vehicles is conducive to improving the utilization rate of resources, promoting the development of circular economy [28], and promoting the recycling of raw materials such as steel, rubber, non-ferrous metals and the five assemblies of automobiles. In addition, the exhaust gas emitted by end-of-life vehicles on the road will be three times that of ordinary cars, and the unreasonable disposal of end-of-life vehicles will cause serious potential environmental pollution problems. This study focuses on the control of reverse logistics of automobiles, which is helpful to reduce environmental pollution and reduce the harm to the environment caused by the use of end-of-life vehicles. At the enterprise level, in the recycling process, more than 95 % of the raw materials of an old automobile can be reused. Realizing the recycling of end-of-life vehicles parts and five assemblies is conducive to reducing the production cost of enterprises, improving the utilization rate of raw materials in the automotive manufacturing field, reducing the cost of raw materials, and creating profits for the forward logistics process of automobiles.

6. Conclusion

This study proposes a logistics network location planning and flow allocation proposal with the second-hand car market as the recycling centre. Based on the historical number of end-of-life vehicles in Shanghai, the future quantity is predicted and used as input data for the model to simulate network operation. The grey prediction results show that the recycling volume of end-of-life vehicles in Shanghai will reach 140000 in 2021, and the prospects for the end-of-life vehicles recycling industry are very optimistic. To solve the optimization problem of reverse logistics network for end-of-life vehicles, this study used LINGO 12 software for model solving. The obtained resource optimization allocation plan effectively improves the actual recovery rate and reduces the operational and logistics costs of the logistics network.

The reverse logistics network of end-of-life vehicles is a complex and uncertain system. However, the research object of this paper is only the reverse logistics network of end-of-life vehicles with single-variety, single-cycle, recycling manufacturing and utilization, which is complex and multi-product in reality. The detailed study of multi-product situation is more suitable for practical problems. The construction of mixed integer model considering environmental protection, economy, resource utilization and other benefits can provide a comprehensive framework for scrap vehicle reverse logistics network. The follow-up research will also focus on the dynamic prediction of the amount of recycling, which will be based on the changes of many factors such as market demand, technological development, policies and regulations, in order to realize the dynamic real-time allocation of resources and adapt to the changing environment and market demand, so as to achieve the innovation breakthrough of reverse logistics network.

References

[1] Iruthayaraj, D.L., Arockiam, R.R.P., Subbaian, J. (2023). Real-time indoor environment quality monitoring for vehicle cabin, *Environmental Engineering and Management Journal,* Vol 22, No. 11, 1801-1811, doi: 10.30638/eemj.2023.156.

[2] Qi, Y., Niu, Y., Zhou, Z. (2023). Digital economy empowering the development level of Chinese manufacturing industry, *Economic Computation and Economic Cybernetics Studies and Research,* Vol. 57 No. 4, 243-258, doi: 10.24818/18423264/57.4.23.15.

[3] Wang, L., Chen, X.Y., Zhang, H. (2021). Joint distribution models in fast-moving consumer goods wholesale enterprise: Comparative analysis and a case study, *Advances in Production Engineering & Management*, Vol. 16, No. 2, 212-222, doi: 10.14743/apem2021.2.395.

[4] Simic, V. (2013). End-of-life vehicle recycling - A review of the state-of-the-art, *Tehnički Vjesnik – Technical Gazette,* Vol. 20, No. 2, 371-380.

[5] Hao, H., Zhang, Q., Wang, Z., Zhang, J. (2018). Forecasting the number of end-of-life vehicles using a hybrid model based on grey model and artificial neural network, *Journal of Cleaner Production,* Vol. 202, 684-696, doi: 10.1016/j.jclepro.2018.08.176.

[6] Mohamad-Ali, N., Ghazilla, R.A.R., Abdul-Rashid, S.H., Sakundarini, N., Ahmad-Yazid, A., Stephenie, L. (2017). Development of an end-of-life vehicle recovery model using system dynamics and future research needs, *IOP Conference Series: Materials Science and Engineering,* Vol. 210 No. 1, Article No. 012075, doi: 10.1088/1757-899X/210/1/012075.

[7] Zhang, Y., Zhong, M., Geng, N., Jiang, Y. (2017). Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China, *Plos One,* Vol. 12, No. 5, 1-15, doi: 10.1371/journal.pone.0176729.

[8] Dargay, J., Gately, D. (1999). Income's effect on car and vehicle ownership, worldwide: 1960-2015, *Transportation research, Part A: Policy and Practice,* Vol. 33, No. 2, 101-138, doi: 10.1016/S0965-8564(98)00026-3.

[9] Demirel, E., Demirel, N., Gökçen, H. (2016). A mixed integer linear programming model to optimize reverse logistics activities of end-of-life vehicles in Turkey, *Journal of Cleaner Production,* Vol. 112, Part 3, 2101-2113, doi: 10.1016/j.jclepro.2014.10.079.

[10] Lin, Y., Jia, H., Yang, Y., Tian, G., Tao, F., Ling, L. (2018). An improved artificial bee colony for facility location allocation problem of end-of-life vehicles recovery network, *Journal of Cleaner Production,* Vol. 205, 134-144, doi: 10.1016/j.jclepro.2018.09.086.

[11] Govindan, K., Gholizadeh, H. (2021). Robust network design for sustainable-resilient reverse logistics network using big data: A case study of end-of-life vehicles, *Transportation Research Part E: Logistics and Transportation Review,* Vol. 149, Article No. 102279, doi: 10.1016/j.tre.2021.102279.

[12] Ene, S., Öztürk, N. (2015), Network modeling for reverse flows of end-of-life vehicles, *Waste Management,* Vol. 38 284-296, doi: 10.1016/j.wasman.2015.01.007.

[13] Chaabane, A., Montecinos, J., Ouhimmou, M., Khabou, A. (2021). Vehicle routing problem for reverse logistics of End-of-Life Vehicles (ELVs), *Waste Management*, Vol. 120, 209-220, doi: 10.1016/j.wasman.2020.11.008.

[14] Min, H., Ko, C.S., Ko, H.J. (2006). The spatial and temporal consolidation of returned products in a closed-loop supply chain network, *Computers & Industrial Engineering,* Vol. 51, No. 2, 309-320, doi: 10.1016/j.cie.2006.02.010.

[15] Marin, A., Pelegrin, B. (1998). The return plant location problem: Modelling and resolution, *European Journal of Operational Research,* Vol. 104, No. 2, 375-392, doi: 10.1016/S0377-2217(97)00192-6.

[16] Du, F., Evans, G.W. (2008). A bi-objective reverse logistics network analysis for post-sale service, *Computers & Operations Research,* Vol. 35, No. 8, 2617-2634, doi: 10.1016/j.cor.2006.12.020.

[17] Kheirabadi, M., Naderi, B., Arshadikhamseh, A., Roshanaei, V. (2019). A mixed-integer program and a Lagrangian-based decomposition algorithm for the supply chain network design with quantity discount and transportation modes, *Expert Systems with Applications,* Vol. 137, 504-516, doi: 10.1016/j.eswa.2019.07.004.

[18] Steinrücke, M. (2015). Integrated production, distribution and scheduling in the aluminium industry: A continuous-time MILP model and decomposition method, *International Journal of Production Research,* Vol. 53, No. 19, 5912-5930, doi: 10.1080/00207543.2015.1023401.

[19] Tan, Y., Takakuwa, S. (2011). Use of simulation in a factory for business continuity planning, *International Journal of Simulation Modelling,* Vol. 10, No. 1, 17-26, doi: 10.2507/IJSIMM10(1)2.172.

[20] Li, Y., Abtahi, A.-R., Seyedan, M. (2019). Supply chain performance evaluation using fuzzy network data envelopment analysis: A case study in automotive industry, *Annals of Operations Research,* Vol. 275, No. 2, 461-484, doi: 10.1007/s10479-018-3027-4.

[21] Guzmán, E., Poler, R., Andres, B. (2023). A matheuristic approach combining genetic algorithm and mixed integer linear programming model for production and distribution planning in the supply chain, *Advances in Production Engineering and Management,* Vol. 18, No. 1, 19-31, doi: 10.14743/apem2023.1.454.

[22] Ding, F., Liu, S., Li, X. (2022). Pareto optimality of centralized procurement based on genetic algorithm, *Tehnički Vjesnik – Technical Gazette,* Vol. 29, No. 6, 2058-2066, doi: 10.17559/TV-20220723180901.

[23] Jünger, M., Liebling, T.M., Naddef, D., Nemhauser, G.L., Pulleyblank, W.R., Reinelt, G., Rinaldi, G., Wolsey, L.A. (2010). *50 Years of integer programming 1958-2008,* Springer, Berlin, Germany, doi: 10.1007/978-3-540-68279-0.

[24] Niu, X.Y., Liu, S.F., Huang, Q.L. (2022). End-of-line delivery vehicle routing optimization based on large-scale neighbourhood search algorithms considering customer-consumer delivery location preferences, *Advances in Production Engineering & Management*, Vol. 17, No. 4, 439-454, doi: 10.14743/apem2022.4.447.

[25] Ye, Y., Wang, J. (2017). Study of logistics network optimization model considering carbon emissions*, International Journal of System Assurance Engineering and Management,* Vol. 8, Suppl. 2, 1102-1108, doi: 10.1007/s13198-017-0576-x.

[26] Xiao, Z., Sun, J., Shu, W., Wang, T. (2019). Location-allocation problem of reverse logistics for end-of-life vehicles based on the measurement of carbon emissions, *Computers & Industrial Engineering,* Vol. 127, 169-181, doi: 10.1016/j.cie.2018.12.012.

[27] Ocampo, L.A., Himang, C.M., Kumar, A., Brezocnik, M. (2019). A novel multiple criteria decision-making approach based on fuzzy DEMATEL, fuzzy ANP and fuzzy AHP for mapping collection and distribution centers in reverse logistics, *Advances in Production Engineering & Management,* Vol. 14, No. 3, 297-322, doi: 10.14743/apem2019.3.329.

[28] Cheng, Y., Xu, J. (2021). Model of environmental management science based on circular economy theory, *Ecological Chemistry and Engineering S,* Vol. 28, No. 4, 513-524, doi: 10.2478/eces-2021-0034.

Simulation analysis of dual-end queuing ride-hailing system considering driver-side queue management

Tang, M.C.[a,b,*], Cao, J.[a], Gong, D.Q.[c], Xue, G.[c], Khoa, B.T.[b]

[a]Xuzhou University of Technology, P.R. China
[b]Industrial University of Ho Chih Minh City, Vietnam
[c]International Center for Informatics Research, Beijing Jiaotong University, P.R. China

A B S T R A C T

Ride-hailing services have transformed urban transportation through convenience yet introduced new complexities around efficiency and traffic management. This study investigates the dual-queuing problem in ride-hailing from the driver perspective using a multi-agent simulation approach. The focus is dissecting the dynamics between driver search times and passenger wait times, which critically influence operational efficiency especially during peak demand. Exploring these interactions aims to uncover insights that could improve service efficiency and customer satisfaction. Addressing such ride-hailing challenges is vital not just for individual providers but also for advancing sustainable mobility across rapidly growing metropolitan regions. Enhanced efficiency connects to broader urban development narratives around livability, accessibility, and responsible mobility ecosystems.

1. Introduction

As China swiftly emerges as a global leader in on-demand mobility, ride-hailing services have profoundly transformed urban transportation paradigms. Leveraging mobile technologies, these innovative platforms efficiently match passengers and drivers, catering to escalating mobility demands within metropolitan expanses [1, 2]. Having altered commuting habits, ride-sharing now constitutes an integral mobility component for contemporary urban living [3, 4].

As an emergent mode of transport, ride-hailing presents service characteristics unprecedented in traditional travel methods. It has altered commuting habits and has had a substantial impact on the urban transportation landscape. With the rapid development of the ride-hailing market and fierce competition, enhancing service efficiency and customer satisfaction has become pivotal to the industry's growth [5]. A major challenge faced by the sector, particularly during peak times and within congested city environments, is the effective management of the matching process between drivers and passengers to reduce waiting times and improve operational efficiency. A notable operational issue within the ride-hailing industry is the often inefficient process of drivers seeking passengers, which not only diminishes drivers' work efficiency but also extends passengers' waiting periods [6]. This issue becomes especially pronounced during urban peak hours. The

efficiency of the match between drivers and passengers directly affects the entire service system's efficiency, making the effective balance of driver search time and passenger wait time a pressing problem to resolve [7].

The necessity of this study lies in exploring and addressing the dual-queue problem on the driver's end within the ride-hailing industry, which is crucial for enhancing the service efficiency and customer satisfaction of the entire industry. This research aims to delve into the processes of driver search and passenger wait through in-depth analysis and simulation to identify strategies that elevate service efficiency, thereby providing effective operational and decision-making support for ride-hailing platforms. Moreover, by studying the driver-side queuing model, this research also aims to offer insights beneficial to urban traffic management and service optimization [8].

This study faces multiple challenges. First is the technical challenge of constructing a simulation model that accurately mirrors real traffic and service conditions, requiring precise modeling of complex urban traffic environments and dynamic passenger demands [9]. Secondly, processing and analyzing the vast amount of data generated by the simulation model to extract meaningful insights and strategies is non-trivial. Additionally, designing efficient algorithms to optimize driver search routes and reduce passenger waiting times is key to our research [10].

To achieve these objectives, this study employs a multi-agent simulation model. This method, by simulating the interactive behavior of drivers and passengers within a virtual environment, allows us to replicate and analyze complex urban traffic environments and ride-hailing service processes under controlled conditions. The simulation model is capable of emulating various traffic conditions and passenger demand patterns, providing an effective tool for analyzing driver search times and passenger waiting times. The crux of this approach is its ability to reproduce the complex scenarios of real life and provide real-time feedback on strategies.

The primary contribution of this research is the development of an innovative simulation model that effectively simulates the dual-queue system from the driver's perspective, analyzing its impact on the overall efficiency of ride-hailing services. Through this model, we can propose specific strategies to reduce driver search times and passenger wait times, thereby enhancing the operational efficiency of ride-hailing platforms. Additionally, this research offers valuable insights into ride-hailing services, particularly on how to effectively manage the match between drivers and passengers during peak periods. These contributions are significant not only for the practical operations of the ride-hailing industry but also for providing theoretical support for future urban traffic management and service optimization. Through comprehensive and in-depth research, this article aims to contribute significantly to the academic field of ride-hailing services and provide practical and feasible recommendations for business operations.

Tabel 1 Reminder of this paper

Section	Description	Key points
Abstract	Overview of the study focusing on ride-hailing services and dual-end queuing problems.	Highlights the significance of driver search times and passenger waiting times in ride-hailing efficiency.
Introduction	Contextualizes the rise of ride-hailing services in urban transportation, especially in China.	Emphasizes the impact of ride-hailing on urban mobility and the importance of managing driver-passenger matching.
Related works	Review of existing literature on ride-hailing system efficiency and multi-agent simulation studies.	Discusses previous findings on operational efficiency, user behavior, and dispatch strategies in ride-hailing services.
Methodology	Details the approach for analyzing the dual-end queuing problem using a multi-agent simulation model.	Describes research hypothesis, ride-hailing vehicle arrivals analysis, dual-end queuing model, and parameter settings.
Simulation results	Presents the findings from the simulation, focusing on passenger accumulation and efficiency metrics.	Analyzes the correlation between passenger numbers, driver search times, and system efficiency.
Discussion	Theoretical and practical implications of the research, along with limitations and future directions.	Reflects on the study's contribution to urban transportation and ride-hailing service management.
Conclusions	Summarizes the study's findings and their implications for optimizing ride-hailing services.	Emphasizes the importance of managing driver search and passenger wait times for service efficiency.

Table 1 provides a structured summary this paper. It outlines the key sections of the research paper, including the Abstract, Introduction, Related Works, Methodology, Simulation Results, Discussion, Conclusions, and References. Each section is concisely described, emphasizing the study's focus on the complexities of ride-hailing services, particularly the efficiency of driver-passenger matching and the impact of driver search and passenger wait times on overall service efficiency. The table encapsulates the research's theoretical and practical implications, methodological approaches, and key findings, offering a coherent overview of the paper's content and contributions to the field of urban transportation and ride-hailing service management.

2. Related works

2.1 Ride-hailing system operational efficiency

Within the contemporary transportation ecosystem, the operational efficiency, service quality, and user satisfaction of ride-hailing services constitute key research areas. Multidimensional studies in this domain have indicated that Feng *et al.* [4] engaged in a comprehensive systemic analysis to explore avenues for enhancing the operational efficiency and service quality of ride-hailing services. Additionally, Xu *et al.* [5] centered their research on the supply curve of ride-hailing systems, revealing how market conditions affect the balance of supply and demand for these services.

Further advancing the field, another study by Xu *et al.* [6] focused on optimizing vehicle dispatch within ride-hailing systems, demonstrating the application of technological means to increase service efficiency. Complementarily, Li *et al.* [7] conducted an in-depth exploration of user behavior within ride-hailing systems, particularly the decision-making processes of boundedly rational users. On another front, Lu *et al.* [8] sought to balance efficiency and fairness in ride-hailing services, especially in the context of carpooling design during specific times such as late-night hours. Meanwhile, Schlenther *et al.* [9] emphasized the issue of spatial equity in the provision of ride-hailing services.

Collectively, these studies offer a comprehensive understanding of the operations of ride-hailing systems, covering aspects such as service efficiency, user behavior, dispatch strategies, and equity. Not only do they provide new theoretical perspectives for the academic sphere, but they also offer practical guidance and strategy recommendations for operational practices.

Building on the foundation of existing research, our study delves further into the dual-queue problem at the driver's end within ride-hailing systems. We pay special attention to the relationship between driver search behavior and passenger wait times and how this relationship impacts the efficiency of the entire service system. Through the development of an integrated simulation model, we plan to simulate and analyze driver and passenger behaviors under varying conditions, seeking novel pathways to enhance service efficiency. Simultaneously, we will explore how to maintain system fairness and sustainability while ensuring efficient service, providing new theoretical insights into the ride-hailing service domain and offering feasible solutions to meet market demands and urban transportation challenges.

2.2 Multi-agent simulation studies

In the pursuit of improving shared mobility services, numerous studies leveraging multi-agent simulation models have provided significant insights. Inturri *et al.* [10] utilized a multi-agent simulation approach to plan and design new shared mobility services. Their research team, by simulating diverse traffic and shared mobility scenarios, discussed how to effectively implement shared mobility services in urban environments. This work underscored the potential applications of multi-agent systems in understanding and optimizing shared mobility services. Li *et al.* [11] employed mean field multi-agent reinforcement learning for the efficient allocation of ride-hailing orders, showcasing how advanced machine learning algorithms can be used to optimize ride-hailing order distribution, thus enhancing overall service efficiency and reducing passenger waiting times. Ke *et al.* [12] used a multi-agent deep reinforcement learning framework to study delay strategies within ride-sourcing systems, focusing on how to effectively dispatch ride-

sourcing vehicles under high-demand conditions to balance supply and demand and improve system efficiency. Galland *et al.* [13] simulated individual mobility behavior in carpooling using a multi-agent simulation, providing insights on how to optimize carpooling services to reduce traffic congestion and environmental impact. Mei *et al.* [14] explored multi-modal travel policies to improve park-and-ride efficiency through multi-agent simulation, demonstrating how policy interventions can optimize the connection between public transportation and private car usage, enhancing overall travel efficiency.

These studies demonstrate the tremendous potential of multi-agent simulations in understanding and optimizing shared mobility services. By simulating complex traffic environments and user behaviors, they provide valuable insights for the planning and design of shared mobility services. Particularly with the use of advanced machine learning techniques to optimize service distribution and scheduling, these studies illustrate how data-driven approaches can enhance service efficiency and user satisfaction. Additionally, these studies consider environmental sustainability and policy impacts while improving travel efficiency, highlighting the importance of integrated approaches in solving transportation issues.

Building upon these foundations, our research further explores the application of multi-agent simulation in optimizing ride-hailing services. We focus on how simulation models can better understand and optimize the interactions between drivers and passengers and how these interactions affect service efficiency and user experience. Additionally, our research will investigate how to reduce environmental impacts and consider policy factors while ensuring efficient service. In this way, our study aims to provide new theoretical insights and practical strategies for the development of shared mobility services.

3. Methodology

3.1 Research hypothesis

In the domain of urban transportation, the dynamics of taxi services can be conceptualized as a queuing system where both the taxis and the passengers are involved in a bidirectional waiting mechanism—waiting to board and to pick up, respectively. Within this system, the metric of efficiency in boarding is intrinsically linked to the length of the queue and the associated waiting time. Theoretically, the waiting time is directly proportional to the queue length; an increase in the number of waiting passengers or taxis invariably leads to longer waiting periods.

Optimizing the efficiency of such a taxi-boarding system necessitates a strategic determination of the quantity of boarding points. An optimal number of boarding points can theoretically minimize the queue length on both ends of the system, thereby reducing the waiting time and enhancing the overall efficiency of the service. For instance, a singular boarding point would result in a bottleneck scenario where all taxis and passengers converge, leading to extended queues and increased waiting times, thereby diminishing the system's efficiency. Conversely, the introduction of multiple boarding points could distribute the demand, subsequently shortening the queue lengths and reducing waiting times, culminating in an elevated efficiency of the boarding process.

Therefore, viewing the process of taxi boarding as a queuing system sheds light on the pivotal role of queuing lengths and waiting times as critical indicators of efficiency. It emphasizes the necessity of a judicious allocation of boarding points to minimize queuing lengths, thereby optimizing the efficacy of the taxi-boarding system. This theoretical framework can guide the structuring of urban taxi services to achieve maximum operational efficiency.

3.2 Analysis of ride-hailing vehicle arrivals

Assuming that the vehicle flow is relatively low when ride-hailing cars reach their destination, the Poisson distribution is applied to model the arrival pattern of these vehicles. The number of ride-hailing taxis at the destination can thus be represented by the following equations:

$$X \sim P(\lambda) \tag{1}$$

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0,1,2,\cdots) \tag{2}$$

To further refine our model, it is essential to estimate the parameter λ within the Poisson distribution, which represents the mean rate of arrivals. Given that the overall distribution is known and randomness is adequately accounted for, it is prudent to employ the method of maximum likelihood estimation (MLE) over moment estimation or other point estimation techniques. MLE is particularly advantageous in this context due to its efficacy in estimating parameters for well-defined probability distributions like the Poisson distribution.

The specific steps to solve for the parameter λ involve setting up the likelihood function based on the Poisson probability mass function, taking its natural logarithm, and then finding the value of λ that maximizes this log-likelihood function. This optimization process typically involves taking the derivative of the log-likelihood function with respect to λ, setting it equal to zero, and solving for λ.

$$\dot{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \tag{3}$$

Ultimately, the estimate of λ is obtained, as shown in Eq. 3. This estimate provides the expected number of ride-hailing cars arriving at the destination within a given timeframe, allowing for more effective management and allocation of resources to match service capacity with customer demand.

3.3 Dual-end queuing model

In the realm of ride-hailing services, a bi-directional queuing model encapsulates the interactions between drivers and passengers, each seeking "service" in terms of securing a ride or a fare, respectively. This model can be typified as a multi-server queuing system, where the average queue length is a critical measure of operational efficiency.

The formulas for calculating average queue length is:

$$L_q = \frac{p_0 \rho^s \rho_s}{s!(1-\rho_s)^2} \tag{4}$$

The probability of no customers in the system is

$$p_0 = \left[\sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s}{s!(1-\rho_s)}\right]^{-1}, \tag{5}$$

and the probability of the service station being busy

$$\rho = \frac{\lambda}{\mu}, \tag{6}$$

$$\rho_s = \frac{\rho}{s} = \frac{\lambda}{s\mu}, \tag{7}$$

The average number of customers in the system is:

$$L = L_q + \rho, \tag{8}$$

L can be derived from the average queue length. This, in turn, informs the average time customers spend in the system

$$W = \frac{L_q}{\lambda} + \frac{1}{\mu}, \tag{9}$$

and their average waiting time:

$$W_q = \frac{L_q}{\lambda}. \tag{10}$$

Additionally, the probability of k vehicles waiting in the system is:

$$P_k = \begin{cases} \frac{\rho^k}{k!}p_0, \ k < s \\ \frac{\rho^k}{s!s^{k-s}}p_0, \ k \geq s \end{cases}, \tag{11}$$

which serves as a significant indicator of service capacity and customer demand alignment.

These metrics – average waiting time, average number of customers, and average time spent in the system – collectively gauge the performance and quality of the service system. They reflect

the system's operational tempo and efficiency, providing insights into resource allocation and customer behavior patterns. Importantly, the queuing probability when k vehicles are present aids in evaluating the system's capability to meet customer service requirements and manage busy periods.

In essence, these metrics are more than mathematical expressions; they are vital tools for assessing system performance and service quality. By understanding and applying these indicators, business decision-makers can enhance service processes, optimize resource distribution, and ultimately, elevate customer satisfaction.

In summary, the dual-queue model in ride-hailing systems, represented by a series of equations, offers a robust framework for analyzing and improving service efficiency. By refining these mathematical models and focusing on the core indicators, researchers and practitioners alike can better navigate the complexities of ride-hailing services, ensuring that operational decisions are informed, strategic, and customer-centric.

3.4 Parameter settings

To conduct the simulation for a ride-hailing service system, a set of parameters is input into a designated Input folder. These parameters are critical in shaping the simulation's framework and include:

N_{pass} (100000): This denotes the total number of passengers within the simulation, representing the demand side of the service.

N_{cars} (50): The total number of cars available, indicating the supply side of the ride-hailing system.

N_m (10000): The total simulation time, which reflects the operational time frame for the service analysis.

N_x (200) and N_y (200): These parameters define the number of grid points in the x and y directions, respectively, forming a grid that simulates the geographical area over which the service operates.

L_{block} (5): This specifies the number of points in a block, providing a measure for the simulation's spatial resolution.

$Scan_d$ (5): The scope of taxi service, which could potentially be used to define the radius within which a taxi searches for passengers.

max customer (10): The maximum number of customers that a taxi can serve before it's considered full and unable to accept new passengers.

$t_{customer}$ (100): The time step interval at which new passengers are generated in the simulation, dictating the influx of service requests.

The output data derived from these inputs are organized into three folders: pass_sum, search_sum, and wait_time. The pass_sum folder contains a log of timestamps and the corresponding cumulative number of passengers waiting for or currently on a ride. The search_sum keeps a record of all the times taken by taxis to find passengers, reflecting the efficiency of the service in matching supply with demand. Lastly, the wait_time folder tracks the duration that passengers wait before boarding, a direct measure of service responsiveness. In the pass_sum dataset, the first column lists the timestamps starting from zero, while the fourth column details the accumulated number of passengers at each timestamp. The search times and waiting times are illustrated in the third column of their respective datasets.

These parameters and the resulting data provide an intricate picture of the ride-hailing system's dynamics, allowing for an assessment of its efficiency, capacity to meet demand, and overall service quality.

Table 2 Simulation process of this paper

Stage	Description	Key considerations
Collection of input data	Gathering essential parameters such as passenger demand and driver availability.	Accurate data collection is crucial for a realistic simulation.
Simulation modeling	Simulating interactions between drivers and passengers based on the collected data.	This phase replicates real-world scenarios to understand the dynamics of the ride-hailing service.
Analysis of output data	Examining key metrics like system efficiency, driver search times, and passenger waiting times.	The analysis provides insights into the operational effectiveness and areas needing optimization.

Table 2 captures the simulation process of a dual-end queuing ride-hailing system through a detailed flow. This flowchart methodically outlines each step involved in the simulation, providing a clear and sequential depiction of the entire process. The simulation begins with the collection of input data, which includes crucial parameters such as passenger demand and driver availability. This initial stage sets the groundwork for the simulation by establishing the fundamental variables that will influence the system's dynamics. Following data collection, the process advances to the simulation modeling phase. Here, the interactions between drivers and passengers are simulated, considering the various factors collected in the initial stage. This modeling is pivotal as it replicates the real-world scenarios of a ride-hailing service, thereby enabling a comprehensive analysis of the system. The final stage of the simulation process involves analyzing the output data. Key metrics such as system efficiency, driver search times, and passenger waiting times are examined. This analysis provides valuable insights into the operational effectiveness of the ride-hailing system and highlights areas that require optimization.

Overall, Figure 1 effectively illustrates the structured approach of the simulation process, from the gathering of initial data to the final analysis of the system's performance. This step-by-step representation is crucial for understanding the methodology behind the simulation and the resulting conclusions drawn about the ride-hailing system's efficiency.

3.5 Evaluation of the dual-end queuing model simulation

In the evaluation of dual-end queuing models for ride-hailing simulations, two critical performance metrics are employed: the average waiting time for passengers and the average search time for drivers. The average waiting time, is calculated as the ratio of the total waiting time experienced by all customers to the total number of passengers served within the system, as shown in Eq. 12:

$$T_{\text{wait}} = \frac{T_{\text{wait\_total}}}{N_{\text{total}}} \tag{12}$$

Similarly, the average search time for drivers, is determined by the ratio of the total search time for all ride-hailing vehicles to the number of vehicles in the system, indicated in Eq. 13:

$$T_{\text{search}} = \frac{T_{\text{search\_totd}}}{N\_\text{cars}} \tag{13}$$

The operational excellence of a ride-hailing system is inversely proportional to both the average search time and the average waiting time—the lower these values, the more optimized the system is considered to be. A comprehensive score, denoted as G is utilized to assess the overall efficiency of the ride-hailing operation, which is expressed in Eq. 14:

$$G = -\theta_1 T_{\text{wait}} - \theta_2 \tag{14}$$

Here, G represents the total score of the system, while θ_1 and θ_2 symbolize the respective weights assigned to the average waiting time and the average search time. The maximization of G is equivalent to finding the optimal weights θ_1 and θ_1, such that the conditions for the equality in the arithmetic and geometric means inequality, represented by Eqs. 15 and 16:

$$G_n = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 x_2 \, \text{L} x_n} \tag{15}$$

$$A_n = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \text{L} + x_n}{n} \tag{16}$$

are satisfied. When the equality holds true, we have:

$$\frac{\theta_1}{\theta_2} = \frac{T_{\text{search}}}{T_{\text{wait}}} \tag{17}$$

Ultimately, as Eq. 18 suggests:

$$\frac{\theta_1}{\theta_2} = \frac{T_{\text{search\_totd}}}{T_{\text{wait\_totd}}} \cdot \frac{N_{\text{total}}}{N_{\_}\text{cars}} \tag{18}$$

when the input data for the number of passengers and other system inputs are fixed. Hence, the ratio of weights for the average waiting time and the average search time in the efficiency score solely depends above parameters.

By analyzing these performance metrics, it is possible to strategically manage ride-hailing operations to improve overall service quality. The implications of this model extend beyond mere operational statistics; they encompass strategic decision-making regarding resource allocation, service policy adjustments, and enhancements to customer experience. The model provides a quantitative framework to dissect and ameliorate the complex dynamics of ride-hailing services, ultimately guiding improvements in system design and management for enhanced service delivery.

4. Simulation results

In the pass_sum dataset of the ride-hailing simulation, the data represents a correlation between specific timestamps and the cumulative number of passengers at each of these moments, thus defining a relationship between the timestamp and N_{total} – the total accumulated passenger count. The ratio of θ_1 to θ_2 is positively correlated with N_{total}. This means that as time progresses, if N_{total} increases, so does the ratio of θ_1 to θ_2; if N_{total} remains constant, the ratio stays the same; and if N_{total} decreases, the ratio diminishes.

A scatter plot (Fig. 1) is generated using 100 data sets from pass_sum, with the horizontal axis representing timestamps (chosen at every 500th moment from a total of 50,000) and the vertical axis showing the cumulative number of passengers at those times. This scatter plot illustrates the general distribution of accumulated passenger numbers across 50,000 timestamps. From Figure 1, it's observed that peak passenger accumulation occurs near timestamps 10,000, 28,000, 40,000, and 44,000, while the lowest passenger numbers are near 0, 35,000, and 50,000.
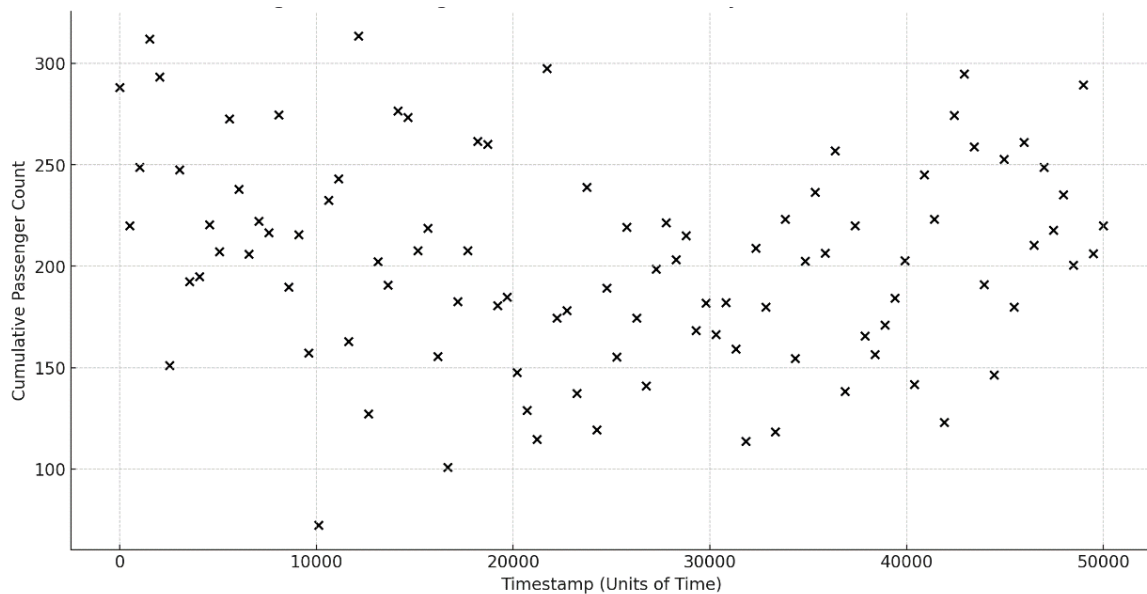
Fig. 1 Passengers accumulated in the system

The analysis of this data and the resulting plot indicates that overall, N_{total} tends to increase over time, though there are moments of decline and troughs. The positive correlation between the ratio of θ_1 and θ_2 and N_{total} suggests that the change in the ratio can approximate the trend of

accumulated passenger numbers in the system. Hence, at the point where N_{total} is at its maximum, the ratio of θ_1 to θ_2 also peaks, indicating that compared to the search time, the passenger waiting time has the most significant impact on system efficiency at those moments, and vice versa. Particularly around timestamp 40,000, where the ratio is highest, the passenger waiting time most significantly impacts system efficiency; at timestamp 0, where the ratio is lowest, the search time has the greatest impact on efficiency.

In practical terms, this simulation provides insights into operational dynamics of ride-hailing services. It highlights the importance of managing passenger wait times and driver search times to optimize system efficiency. In real-world scenarios, when a system begins operation, an absence of generated passengers equates to no customer waiting time, suggesting that such extreme cases should be excluded from analysis. As the system accumulates more passengers, search times for drivers tend to increase. In contrast, when there are enough drivers waiting, passenger waiting times do not significantly increase. This reflects real-life situations and suggests that reducing driver search times might be more crucial than reducing passenger wait times when passenger numbers are high; the opposite might be true when there are fewer passengers.

The implications of these findings are significant for the practical operation of ride-hailing services. They emphasize the need for dynamic management strategies that adjust to varying demand levels throughout the day. For example, during peak hours with high passenger numbers, ride-hailing platforms might focus more on deploying drivers efficiently to reduce their search times, thereby ensuring that passengers are picked up promptly and the overall system efficiency is maintained. Conversely, during off-peak hours with fewer passengers, the focus could shift to ensuring that passengers have shorter waiting times, perhaps by predicting demand patterns and pre-positioning vehicles in strategic locations.

Moreover, the simulation results can guide policy-making in urban transport management, where balancing the need for efficient ride-hailing services with broader concerns like traffic congestion and environmental impact is crucial. For instance, during times of maximum passenger accumulation, city planners and ride-hailing platforms could collaborate to implement measures that ease traffic congestion, such as opening additional pick-up points or providing incentives for shared rides.

In summary, the simulation results offer a comprehensive understanding of how passenger accumulation trends affect the efficiency of ride-hailing systems. By analyzing these trends, ride-hailing services can optimize their operations, improving customer satisfaction and contributing to more efficient urban transport systems. This research not only provides a model for analyzing and enhancing ride-hailing services but also serves as a template for understanding similar demand-responsive transport systems, potentially leading to innovations in urban mobility solutions.

Table 3 encapsulates the simulation results, highlighting the relationship between different timestamps, the total number of passengers at these points, and the ratio, a metric indicating the system efficiency. It shows that at peak times (e.g., 10,000, 28,000, 40,000, and 44,000 timestamps), the efficiency is most affected by passenger waiting times, indicating the need for efficient passenger handling during high-demand periods. Conversely, during low passenger accumulation (e.g., timestamps 0, 35,000, and 50,000), driver search times have a more significant impact on efficiency, suggesting a focus on optimizing driver deployment during these periods. This table serves as a concise summary of the simulation's findings, guiding operational strategies for ride-hailing services.

Fig. 2 presents a simulated bivariate line graph, which vividly captures the relationship between the average search time for drivers and the average waiting time for passengers in a ride-hailing scenario. As time progresses, both lines exhibit a slight inverse trend, suggesting a degree of interaction between the two. The fluctuations in average search time appear to correspond to changes in passenger waiting time, particularly near the points where the lines intersect, which may indicate a direct link between the efficiency of matching passengers to vehicles and passenger satisfaction at certain times. These insights offer strategic implications for operational tactics, especially in balancing resource allocation during periods of high demand and low activity.

Table 3 The simulation results

Timestamp (Units)	Cumulative Passenger Count	Ratio	Implications for System Efficiency
0	Low/Minimum	Low	Greater impact of driver search time on efficiency
10,000	High/Peak	High	Passenger waiting time significantly impacts efficiency
28,000	High/Peak	High	Passenger waiting time significantly impacts efficiency
35,000	Low	Lower	Driver search time becomes more crucial for efficiency
40,000	High/Peak	Highest	Passenger waiting time most significantly impacts efficiency
44,000	High/Peak	High	Passenger waiting time significantly impacts efficiency
50,000	Low/Minimum	Low	Greater impact of driver search time on efficiency
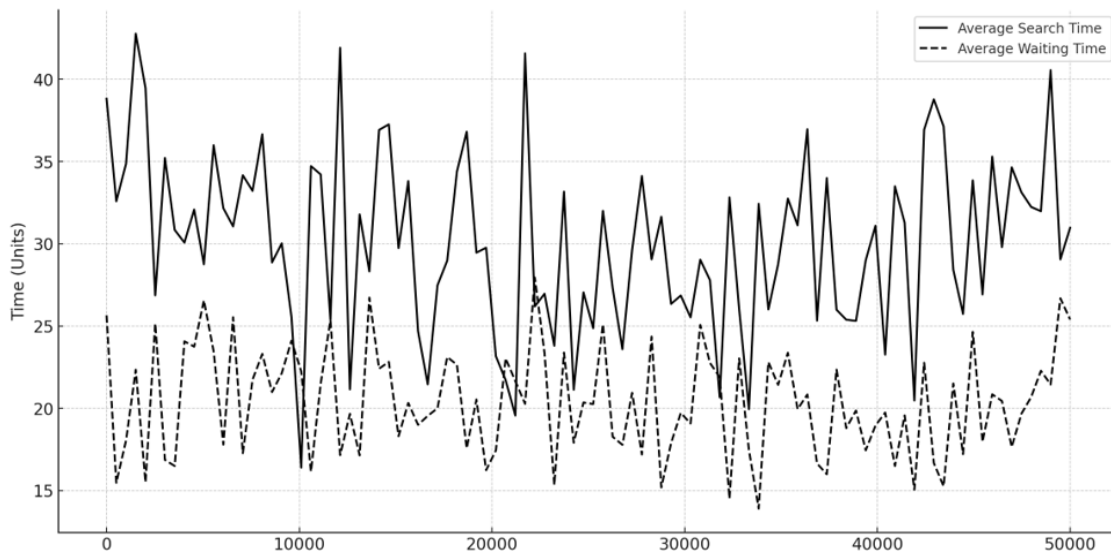
Fig. 2 Simulation of average search and waiting times in ride-hailing
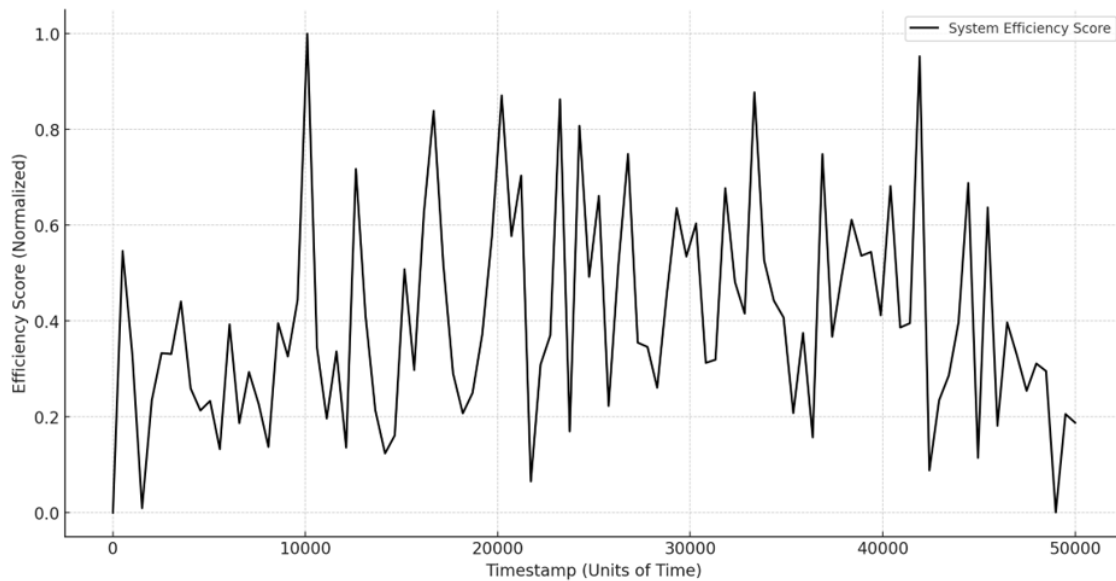
Fig. 3 Simulation of system efficiency score in ride-hailing

Fig. 3 illustrates the normalized System Efficiency Score in a ride-hailing service simulation, traced over time. The efficiency score, normalized between 0 and 1 for clarity, exhibits notable variability, with peaks suggesting instances of heightened efficiency and troughs indicating dips

in performance. This oscillation reflects the dynamic nature of ride-hailing operations, where efficiency is impacted by various factors, such as driver availability and passenger demand. The graph serves as a crucial analytical tool, indicating that strategic adjustments are essential to sustain high-efficiency levels and to address the periods of lower performance that could affect customer satisfaction and operational success.

5. Discussion

5.1 Theoretical implications

This research puts forth valuable theoretical insights for the realms of urban transportation and ride-hailing service optimization. At its core, the work illustrates the potency of deploying a dual-end queuing approach to closely examine and enhance ride-hailing efficacy. This framework grants more nuanced comprehension of the complex interplay between passenger wait times and driver vacant times, alongside the collective impact on system-wide efficiency.

The revelations spotlight the intricate nature of responsive transport networks, underscoring the necessity of integrating real-time tracking and predictive analytics into flexible operational strategies. By unpacking the dynamics between various performance variables, the study enriches prevailing theoretical models surrounding ride-share platforms. The findings affirmatively highlight the need for adaptive, data-driven management practices.

Additionally, the research contributes to the evolving discourse on urban mobility ecosystems, suggesting efficient ride-hailing mechanisms can constitute a pivotal mobility component within modern cities. The work emphasizes the value of bridging theoretical constructs with practical implementations, paving inroads for innovative solutions that synthesize consumer and societal transportation needs.

In summary, the study yields multi-faceted theoretical insights around structuring sophisticated ride-hailing models to empower decision-makers in enhancing system-level and customer outcomes. The research harbors valuable potential to inform existing knowledge as well as future platforms at the intersection of transportation technology and urban advancement.

5.2 Practical implications

The practical implications of this research on ride-hailing services are wide-ranging and impactful, proffering valuable insights for operators, regulators, and urban designers alike.

Foremost, the study spotlights the potency of dynamic resource allocation in ride-hailing stewardship. By comprehending passenger volume patterns and driver availability fluxes, providers can optimize fleet positioning to slash wait times during peak-demand. This adaptive approach enhances overall customer experiences. Additionally, the findings underline the merits of demand-responsive tactics to address volatility. Platforms can harness predictive data analytics to forecast spikes and pre-deploy supply to strategic zones accordingly. Such proactive maneuvers help balance stability and efficiency.

Moreover, the work endorses greater cooperation between ride-hailing entities and metropolitan authorities to ease congestion. By collaborating to share insights on ridership and performance, these stakeholders may co-create innovative solutions such as designated access points or carpool incentives. These symbiotic measures stand to simultaneously improve services, customer satisfaction, and environmental sustainability.

In summary, this research puts forth an insightful optimization framework primed to shape decision-making within the rapidly changing urban mobility arena. The illuminated dynamics and partnerships can empower operators to unlock enhanced strategies, authorities to encourage responsible innovation, and cities to progress towards accessibility and sustainability goals [15, 16].

5.3 Limitations and future directions

While this research proffers valuable insights into ride-hailing ecosystems, the study simultaneously highlights avenues for prospective exploration given certain inherent limitations.

Primarily, the simulation model hinges on defined assumptions that cannot encapsulate the full dynamism of real-world systems. Future works could address this constraint by integrating more diverse, live data covering traffic flows and passenger behavioral variances to enhance experiential representativeness. Additionally, supplemental research factors including regulatory policies, economic climates, and sustainability impacts warrant investigation to construct more holistic perspectives.

As intelligent technologies progress, opportunities abound to cultivate sophisticated algorithms that sharpen predictive capacities and operational optimization. Comparative studies across diverse urban geographies would also elucidate how localized elements like layouts and cultural attitudes influence ride-hailing efficacy. Such international juxtapositions could unveil tailored strategies for assorted metropolitan contexts.

In summary, while constituting an informative foundation, this research indicates ample avenues to surmount current limitations as the intelligent transportation domain continues advancing. Exploring human dynamics and technological possibilities can lead to increasingly realistic and hyper-responsive urban ride systems [17, 18].

6. Conclusion

This research utilized an innovative multi-agent simulation to provide new perspectives on the dual-queuing conundrum in ride-hailing services. By meticulously tracking driver search times versus passenger wait times, the study revealed crucial insights around managing operational efficiency. As the simulation indicates, these time-based factors have an asymmetrical, proportional impact on overall system performance. When passenger demand surges during peak periods, the influence of driver search time becomes more pronounced.

Consequently, a strategic priority emerges for ride-hailing firms seeking to optimize efficiency, especially under high congestion scenarios. The focus should shift to policies and mechanisms that reduce driver vacant times between rides. Targeting search time not only directly improves system-level performance but also alleviates passenger wait time, thereby enhancing service quality and satisfaction.

As urban transportation networks evolve, such evidence-based findings can guide the sustainable growth of ride-sharing. The multi-agent simulation approach could be extended to determine optimization thresholds and evaluate the impact of interventions like driver incentives or demand-responsive pricing. Accounting for human dynamics is vital towards mobility ecosystems that balance convenience, profitability, and responsibility.

References

[1]   Sun, J., Liu, S.F., Zhang, X.H., Gong, D.Q. (2022). Simulation-based modelling of the impact of ridesharing on urban system, *International Journal of Simulation Modelling*, Vol. 21, No. 1, 148-159, doi: 10.2507/IJSIMM21-1-CO2.

[2]   Huang, Q.L., Wang, W.J., Liang, X.J., Xu, L., Niu, X.Y., Yang, X.Y. (2022). Last-mile delivery optimization considering the demand of market distribution methods: A case studies using adaptive large neighborhood search algorithm, *Advances in Production Engineering & Management*, Vol. 17, No. 3, 350-366, doi: 10.14743/apem2022.3.441.

[3]   Han, X., Zhao, P.X., Kong, D.X. (2022). A bi-objective optimization of airport ferry vehicle scheduling based on heuristic algorithm: A real data case study, *Advances in Production Engineering & Management*, Vol. 17, No. 2, 183-192, doi: 10.14743/apem2022.2.429.

[4]   Feng, G., Kong, G., Wang, Z. (2021). We are on the way: Analysis of on-demand ride-hailing systems, *Manufacturing & Service Operations Management*, Vol. 23, No. 5, 1237-1256, doi: 10.1287/msom.2020.0880.

[5]   Xu, Z., Yin, Y., Ye, J. (2020). On the supply curve of ride-hailing systems, *Transportation Research Part B: Methodological*, Vol. 132, 29-43, doi: 10.1016/j.trb.2019.02.011.

[6]   Xu, Y., Wang, W., Xiong, G., Liu, X., Wu, W., Liu, K. (2022). Network-flow-based efficient vehicle dispatch for city-scale ride-hailing systems, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 6, 5526-5538, doi: 10.1109/TITS.2021.3054893.

[7]   Li, Y., Liu, Y. (2021). Optimizing flexible one-to-two matching in ride-hailing systems with boundedly rational users, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 150, Article No. 102329, doi: 10.1016/j.tre.2021.102329.

[8] Lu, C., Wu, J., Wu, C., Qin, Y., Li, Q., Ma, N. (2021). Efficiency or fairness? Carpooling design for online ride-hailing platform in transport hubs at midnight, In: *Proceedings of the 29th International Conference on Advances in Geographic Information Systems,* Beijing, China*,* 244-255, doi: 10.1145/3474717.3483953.

[9] Schlenther, T., Leich, G., Maciejewski, M., Nagel, K. (2023). Addressing spatial service provision equity for pooled ride-hailing services through rebalancing, *IET Intelligent Transport Systems*, Vol. 17, No. 3, 547-556, doi: 10.1049/itr2.12279.

[10] Inturri, G., Le Pira, M., Giuffrida, N., Ignaccolo, M., Pluchino, A., Rapisarda, A., D'Angelo, R. (2019). Multi-agent simulation for planning and designing new shared mobility services, *Research in Transportation Economics*, Vol. 73, 34-44, doi: 10.1016/j.retrec.2018.11.009.

[11] Li, M., Qin, Z., Jiao, Y., Yang, Y., Wang, J., Wang, C., Wu, G., Ye, J. (2019). Efficient ridesharing order dispatching with mean field multi-agent reinforcement learning, In: *Proceedings of WWW '19: The Web Conference*, San Francisco, California, USA*,* 983-994, doi: 10.1145/3308558.3313433.

[12] Ke, J., Xiao, F., Yang, H., Ye, J. (2022). Learning to delay in ride-sourcing systems: A multi-agent deep reinforcement learning framework, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 5, 2280-2292, doi: 10.1109/TKDE.2020.3006084.

[13] Galland, S., Knapen, L., Yasar, A.-U.-H., Gaud, N., Janssens, D., Lamotte, O., Koukam, A., Wets, G. (2014). Multi-agent simulation of individual mobility behavior in carpooling, *Transportation Research Part C: Emerging Technologies*, Vol. 45, 83-98, doi: 10.1016/j.trc.2013.12.012.

[14] Mei, Z., Wei, D., Ding, W., Wang, D., Ma, D. (2023). Multi-agent simulation for multi-mode travel policy to improve park and ride efficiency, *Computers & Industrial Engineering*, Vol. 185, Article No. 109660, doi: 10.1016/j.cie.2023.109660.

[15] Dragan, D., Šinko, S., Keshavarzsaleh, A., Rosi, M. (2022). Road freight transport forecasting: A fuzzy Monte-Carlo simulation-based model selection approach, *Tehnički Vjesnik – Technical Gazette,* Vol. 29, No. 1, 81-91, doi: 10.17559/TV-20210110140112.

[16] Gao, X., Kong, C., Wang, H., Dong, B., Ma, Z., Ren, D. (2022). Calculation of stability limit displacement of surrounding rock of deep-buried soft rock tunnel construction based on fuzzy logic matching algorithm, *Tehnički Vjesnik – Technical Gazette,* Vol. 29, No. 2, 441-448, doi: 10.17559/TV-20210607100722.

[17] Kim, Y.-J., Jung, K.-H. (2022). Social data anlysis on the perception of emergency simulation education of nursing college students using the q method, *Journal of Logistics, Informatics and Service Science*, Vol. 9, No. 1, 291-306, doi: 10.33168/LISS.2022.0118.

[18] Wei, G., Sarman, A.M., Li, M., Shen, L. (2023). A comprehensive approach for thermal comfort analysis in green intelligent buildings using BIM technology, *Journal of System and Management Sciences*, Vol. 13, No. 2, 515-528, doi: 10.33168/JSMS.2023.0235.

Characterizing the effects of SiC and Al₂O₃ on the mechanical properties of Al6082 hybrid metal matrix composites: An experimental and neural network approach

Masood, A.A.[a], **Ali, A.**[a], **Madhu, P.**[b], **Yashas Gowda, T.G.**[b,*], **Jeevan, T.P.**[b], **Sharath, B.N.**[b,*]

[a]Department of Industrial Engineering, College of Engineering, Prince Sattam bin Abdulaziz University, Alkharj, Saudi Arabia
[b]Department of Mechanical Engineering, Malnad College of Engineering, Hassan, affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India

A B S T R A C T

The use of advanced materials in the field of aerospace and automotive applications has led to use of metal matrix composites (MMC's) due to their excellent mechanical properties. Aluminium metal matrix composite is one of the materials which can be strengthened by reinforcing it with hard ceramic particles. In the current work Al6082 matrix hybrid composites reinforced with silicon carbide (SiC) and aluminium oxide (Al2O3) was developed by using stir casting technique. The weight percentage of SiC was varied from 0 wt.% to 8 wt.% and keeping 3 wt.% Al2O3 constants. The tensile, hardness, density and impact tests were conducted, and the results obtained revealed that the addition of silicon carbide and Al2O3 particles in Al6082 enhances the mechanical properties of the prepared hybrid composites. The artificial neural network (ANN) model, which was trained using a dataset consisting of experimental results, has effectively captured the correlation between the weight percentage (wt.%) of silicon carbide (SiC) and the mechanical properties of the composite material. Through the examination of this model, valuable insights can be obtained regarding the distinct contributions of SiC to the mechanical properties of Al6082.

1. Introduction

MMCs such as space structures, deck plates, automobile and railway have critical applications in the industrial sector and invited the attention of the brake disc. About 5496 tons to 80000 tons during the period 2012 to 2019 is expected to increase global demand for MMCs, and it is constantly growing. The important feature of MMCs is the properties such as low density, better mechanical properties, thermal expansion and conductivity and high stiffness which are tailor made which enhances the performances. The reinforced particulate composites are getting much attention due to the reduction of weight for many important applications [1-4]. Metal matrix composites

(MMCs) are a type of material that merges the favourable attributes of metals with the augmented mechanical and physical features of reinforcing materials. Metal Matrix Composites (MMCs) are known to exhibit enhanced mechanical properties such as increased strength, stiffness, wear resistance, and thermal stability when ceramic reinforcements are integrated into the metal matrix. This contrasts with conventional metals which lack these desirable attributes. In recent times particulate reinforced metal matrix composites have emerged as important materials due to its characteristics and the low-cost benefits. The important reinforcements used are, mainly the carbides, oxides and nitrides. The various processing techniques have been employed to optimize the properties of the materials required for different applications [5-8]. Mixture of the metals and ceramics grow the required feature in evolving the metal matrix materials. The accumulation of the materials with high modulus and ductile property induces the material with the transitional property between matrix alloy and ceramic reinforcement. When the reinforcement phase in the composites is comprised the transfers of load to them takes place. This transfer depends on the bonding force in between matrix and the reinforcement phase and employed method of fabrication. Stir casting is one of the manufacturing techniques which provide good adhesion between reinforcement and matrix. Al6082 is one of the predominant metals in which selection of right alloy for a given application considering the various properties such as density, tensile strength, ductility, formidability, weldability, wear and corrosion resistance. The variety of ceramics are used as reinforcements to aluminium alloy in there MMC's such as Al_2O_3, TiC, TiB_2, Al_3Zr, SiC, B_4C [9-13]. The embedded reinforcement particles in hybrid matrix composites will have better properties such as higher stiffness, better toughness lighter weight better toughness and higher wear resistance [14-18]. A frequently researched composite involves the utilization of aluminium alloy, as the matrix material, which is strengthened by the inclusion of silicon carbide (SiC) and aluminium oxide (Al_2O_3) particles. Silicon carbide (SiC) and aluminium oxide (Al_2O_3) are commonly utilized as reinforcement materials owing to their exceptional mechanical characteristics, elevated melting temperatures, and chemical inertness.

Silicon carbide (SiC) is a ceramic material that exhibits remarkable properties such as exceptional hardness, high strength, and excellent thermal conductivity. It is known for its hardness and brittleness. Additionally, it exhibits a diminished coefficient of thermal expansion, rendering it suitable for employment in conjunction with aluminium alloys. Conversely, Al_2O_3, which is commonly referred to as alumina, exhibits elevated levels of hardness, commendable wear resistance, and exceptional electrical insulation characteristics. The material displays excellent chemical resistance and is capable of enduring elevated temperatures. Incorporating SiC and Al_2O_3 reinforcements into the Al matrix has been observed to substantially augment the composite's mechanical properties. The reinforcing elements function as agents that enhance the strength and stiffness of the composite by hindering the movement of dislocations and obstructing the propagation of cracks. In addition, the inclusion of SiC and Al_2O_3 particles enhances the composite's wear resistance and tribological characteristics, rendering it appropriate for utilization in scenarios necessitating elevated endurance and reduced friction. In addition, the inclusion of SiC and Al_2O_3 particles has the potential to impact the thermal characteristics of the composite material. The thermal conductivity of these reinforcements is exceptional, thereby facilitating the efficient dissipation of heat. The attribute in question holds significant importance in scenarios that necessitate efficient regulation of heat, such as in the context of electronic gadgets, heat sinks, and automotive constituents. The incorporation of SiC and Al_2O_3 reinforcements within Al hybrid metal matrix composites has been observed to confer a variety of advantages, such as heightened mechanical potency. The aforementioned characteristics render them appealing for diverse engineering implementations, encompassing aerospace, automotive, electronics, and thermal regulation systems. The performance attributes and impact of the composite material are subject to various factors such as the distribution, particle size, reinforcement content, and fabrication techniques. These factors require meticulous optimization to attain the intended properties [19, 20]. Artificial Neural Networks (ANNs) have been widely employed in the characterization of material properties owing to their capacity to acquire knowledge of intricate patterns and relationships from empirical data. Artificial neural network (ANN) methodologies have demonstrated their efficacy in the analysis of diverse materials, encompassing metals, polymers, ceramics, composites,

and other such substances. Researchers endeavour to develop a predictive model utilizing Artificial Neural Networks (ANNs) to effectively characterize material properties. The primary objective is to create a model that can accurately estimate material properties by considering input parameters or features [21, 22].

The current investigation centres on the examination of the mechanical characteristics of Al6082 hybrid metal matrix composites. These composites are reinforced with different wt.% of silicon carbide (SiC), in addition to a constant 3 wt.% of aluminium oxide (Al_2O_3). The composite's overall strength and ductility will be evaluated through the implementation of tensile testing. The utilization of hardness measurements can offer valuable insights into a material's capacity to withstand indentation and surface wear. The determination of a composite's stiffness and capacity to endure deformation under load can be facilitated through the utilization of tensile test measurements. Apart from the mechanical properties, the microstructure analysis will be conducted to evaluate other characteristics, such as the distribution of particles and the interfacial bonding between the matrix and reinforcements. The present analysis aims to offer valuable insights into the structural integrity of the composite material and the efficacy of the reinforcement distribution.

The primary objective of this investigation is to enhance the mechanical characteristics of a hybrid metal matrix composite by optimizing the weight proportions of SiC and Al_2O_3 in the Al6082 matrix. The outcomes of this study will enhance the comprehension of the impact of SiC and Al_2O_3 reinforcements on the efficacy of Al6082 composites. This will furnish significant insights for potential applications in industries where augmented mechanical properties hold paramount importance.

2. Materials and methods

2.1 Al 6082

The rationale behind the selection of Al6082 as the matrix material is attributed to its extensive utilization across diverse industrial sectors, owing to its exceptional amalgamation of robustness, anti-corrosive properties, and ease of weld ability. The incorporation of SiC and Al_2O_3 into Al6082 is anticipated to enhance its mechanical characteristics, including but not limited to strength, hardness, stiffness, and wear resistance. Al6082 is an alloy with manganese as the main constituent with good corrosion resistance. Al6082 produces tight coils of chips, when chip breakers are used during machining, this provides good machinability. Table 1 shows the composition of Aluminium 6082.

Table 1 Chemical composition of Al 6082 alloy [23]

Al 6082	Cr	Cu	Fe	Mg	Mn	Si	Ti	Zn	Al
% Composition	0.25	0.1	0.5	1.2	1.0	1.3	0.1	0.2	95.2-98.3

2.2 Aluminium oxide

Al_2O_3 is the budget-friendly and generally used material in the family of engineering ceramics. Its substantial practice is in the manufacture of aluminium metal, even though it is also used as an abrasive on account of its hardness and as a refractory material because of its high melting point. In contrast, aluminium oxide (Al_2O_3) exhibits notable characteristics such as elevated hardness, resistance to wear, and exceptional electrical insulation capabilities. Incorporating Al_2O_3 particles into the composite material can augment its wear resistance and durability. Additionally, aluminium oxide exhibits favourable thermal stability, rendering it appropriate for implementation in scenarios necessitating elevated temperature endurance [24].

2.3 Silicon carbide

Silicon carbide (SiC) is a widely recognized material for reinforcement purposes due to its exceptional mechanical characteristics, which encompass elevated strength, hardness, and thermal conductivity. The properties have the potential to augment the structural integrity and rigidity of the composite material. Moreover, SiC exhibits a diminished coefficient of thermal expansion, rendering it congruous with the aluminium alloy matrix and diminishing the probability of failures

induced by thermal mismatch. The rationale behind selecting Al6082 as the matrix material stems from its extensive application in diverse industrial sectors owing to its exceptional amalgamation of strength, resistance to corrosion. The anticipated outcome of reinforcing Al6082 with SiC and Al_2O_3 is an enhancement in its mechanical characteristics, including but not limited to strength, hardness, stiffness, and wear resistance [25].

2.4 Preparation of Al 6082 – Al2O3 – SiC hybrid composites

Stir casting is employed in the present investigation to fabricate aluminium hybrid metal matrix composite with Al6082 as base metal, SiC and Al2O3 as reinforcement material (Table 2). Initially, the particles are preheated at 300 °C – 400 °C for 2 hours to eradicate the volatile substances and the particle temperature is maintained nearer to melting temperature of about 750 °C. The pre-heating give rise to the artificial oxidation of the particle surface forming SiO2 layer in SiC and this SiO2 layer thus formed in turn helps in refining the wettability of the particles. Next, the Al6082 ingots were charged into the furnace and melting was allowed to continue till an even temperature of 700 °C is attained, subsequently a degassing agent (Hexachloroethane) was supplemented into the molten metal for elimination of dissolved gas to control the porosity in the cast. The melt was then permitted to cool till 600 °C to a semi-solid state. At this point, the SiC and Al2O3 mixture was added to the melt and manual stirring of the slurry was performed for 20 min. After the manual stirring, the composite slurry was reheated and maintained at a temperature of 700 °C ± 10 °C and then mechanical stirring was performed. The stirring was done for 10 min at an average stirring rate of 350 rpm. Finally, the molten metal at a temperature of 710 °C was then poured into the mould of diameter 25 mm and height 250 mm, which was preheated at 300 °C and then permitted to solidify [26, 27].

Table 2 Wt.% of reinforcement in Al6082

Samples	Wt.% of Al	Wt.% of SiC	Wt.% of Al_2O_3
M	100	-	-
M1	95	2	3
M2	93	4	3
M3	91	6	3
M4	89	8	3

2.5 Testing of composites

The testing of composites for density, tensile strength, hardness and impact were carried out as per ASTM standards. The density of material is evaluated using Archimedes principle. Rockwell hardness tester was employed for determining hardness. The tensile strength of the specimen prepared were evaluated using UTM of 50-ton capacity. A notched sample of Al 6082 was used to calculate the impact strength. The scanning electron microscope with a resolution of 3.0 nm at 30kV and magnification of 5X to 1000X was used to study the clear microstructure of the specimens. The specimen was highly polished by using bluffing machine and high-quality emery papers [28-34].

3. Results and discussions

3.1 Microstructural studies

The SEM micrographs of the analysed composites are depicted in Fig. 1. The micrographs presented herein offer a comprehensive depiction of the cross-sectional morphology of SiC and Al_2O_3 particles within the Al6082 matrix. The images depict microstructures that exhibit the ceramic phase of silicon carbide, displaying a range of luminosity in shades of grey. Upon analysis of the micrographs, it is evident that the microstructure layer is impacted by the composition of the composites. The microstructure layer undergoes a shift as a result of the incremental incorporation of reinforcement materials, namely SiC and Al_2O_3, which leads to modifications in their respective percentages. The micrographs provide valuable information regarding the dispersion, shape, and interfacial properties of SiC and Al_2O_3 particles in the matrix.
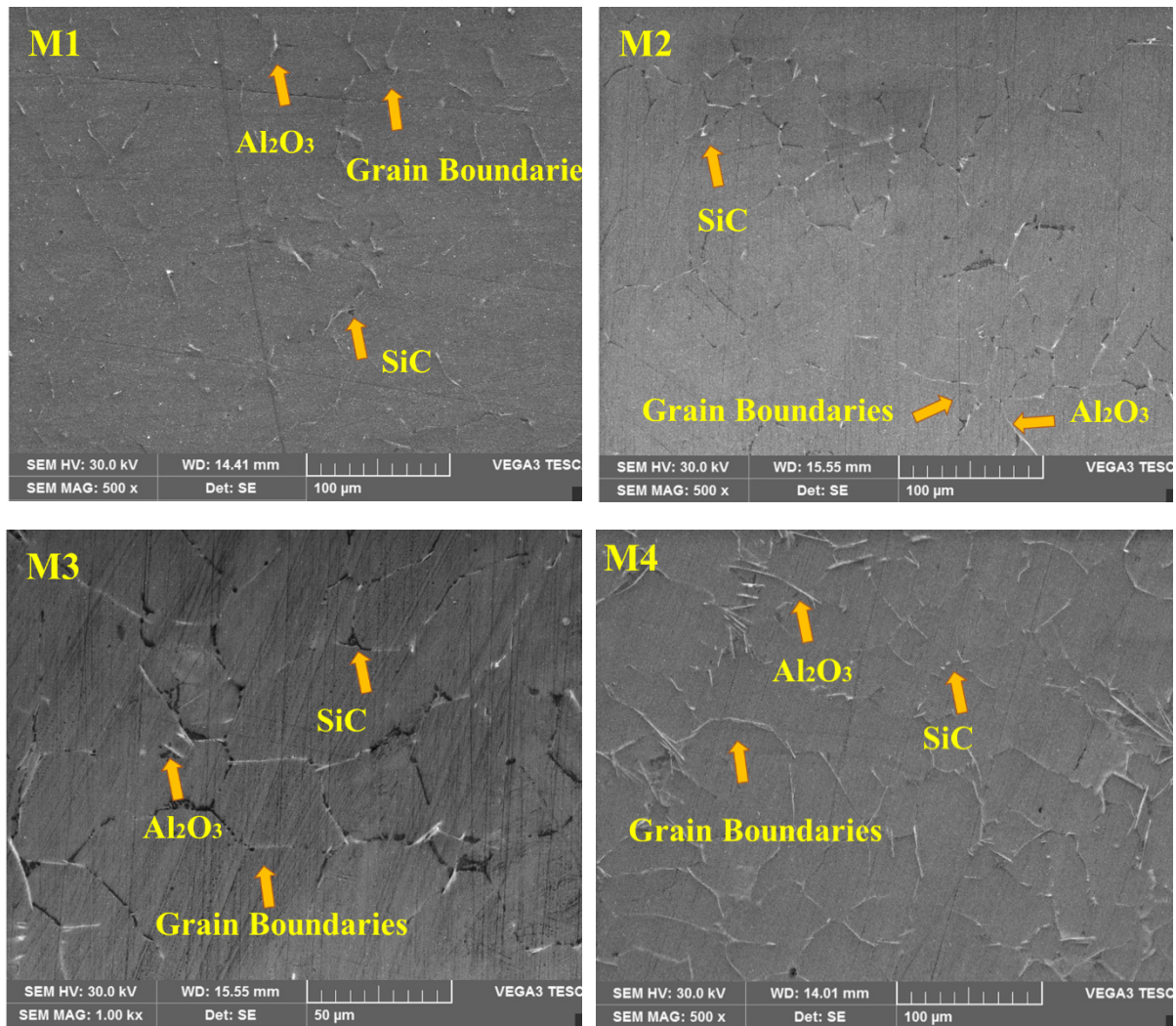
Fig. 1 SEM micrographs of the hybridized composites

3.2 Density measurement

The parameter of density holds significant importance in the examination of composite materials, as it offers valuable insights into their composition and overall performance. The present investigation aimed to examine the density fluctuations among various reinforcement material compositions (namely SiC and Al2O3) within the Al6082 matrix. The graphical representation depicted in Fig. 2 showcases the correlation between the density of the composites and the proportion of SiC reinforcement. The plot's results suggest that the density of the composite materials exhibits a positive correlation with the percentage of SiC reinforcement. As the silicon carbide content within the composite material is augmented, the overall density of the material is correspondingly elevated. The observation is in accordance with the established fact that SiC (3.1 g/cm3) and Al2O3 (3.95 g/cm3) possess a higher density in comparison to the unadulterated Al6082 matrix (2.71 g/cm3). The elevated densities of the hybrid composite specimens, in contrast to those of pure Al6082, can be ascribed to the incorporation of SiC and Al2O3 reinforcements. The incorporation of reinforcements with higher densities than the matrix material results in a notable elevation of the composite's overall density. This occurrence is anticipated and represents a prevalent attribute of composite materials.
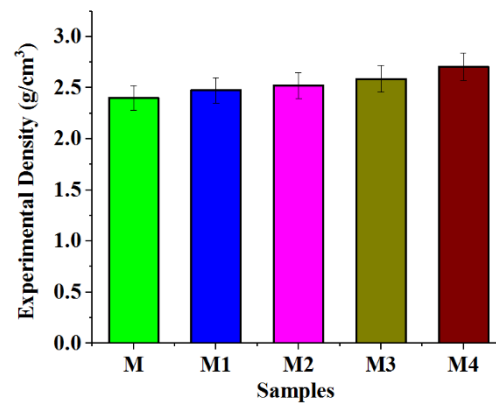
Fig. 2 Density measurement

3.3 Tensile strength

The findings depicted in Fig. 3 demonstrate that the Al6082 hybrid composite (M4) containing 3 wt.% Al2O3 and 8 wt.% SiC exhibits greater tensile strength in comparison to the remaining compositions. The observed enhancement in the tensile strength can be ascribed to the ceramic properties of the SiC particles and their interplay with the matrix substance. The incorporation of ceramic SiC particles is a critical factor in enhancing the strength of the composite material, because of their intrinsic high strength and hardness properties. The inclusion of SiC within the matrix of a composite material serves as a reinforcing agent, thereby augmenting its strength. The incorporation of SiC particles serves to impede the movement of dislocations and hinder the propagation of cracks, leading to an improvement in the composite's tensile strength as its entirety. The findings indicate that the Al-SiC chemical reaction leads to a slight change in silicon concentration, which in turn enhances the interfacial bonding between the matrix and SiC particles. The enhancement of interfacial bonding is a significant factor in facilitating stress transfer between the matrix and the reinforcement, thereby resulting in a rise in tensile strength. Furthermore, Fig. 4 illustrates the fluctuation of percentage elongation in relation to the SiC content. The data indicates that the percentage of elongation exhibits an initial increase, rising from 10.54 % at 0 wt.% SiC to 12.45 % at 4 wt.% SiC composition. The observed rise in elongation values suggests an enhancement in the composite's ductility. The inclusion of SiC particles has the potential to enhance the plastic deformation characteristics of a material, thereby impeding premature failure by furnishing additional energy-absorbing locations within the material.
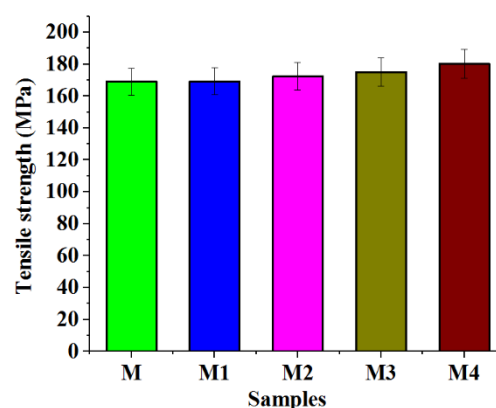
Fig. 3 Tensile strengths of samples

Nevertheless, when the SiC content is elevated to 8 wt.%, the percentage of elongation experiences a reduction to 8.16 % (M4). The observed decline in elongation indicates a reduction in the material's ductility as the concentration of SiC increases (Fig. 4). The presence of a greater amount of SiC may result in an increased number of stress concentration sites, thereby restricting the material's capacity to undergo plastic deformation and elongation. The incorporation of SiC and Al2O3 reinforcements into the Al6082 matrix results in an enhancement of the material's tensile

strength. The utilization of SiC particles in ceramic form leads to an improvement in the interfacial bonding with the matrix, thereby causing a rise in the tensile strength. The percentage of elongation exhibits an initial increase as the SiC content is augmented, thereby signifying an enhancement in ductility. However, at elevated SiC concentrations, the percentage of elongation experiences a decline. The discoveries underscore the intricate correlation among reinforcement content, interfacial bonding, and mechanical characteristics, underscoring the necessity for meticulous optimization of composite composition to achieve the intended performance.
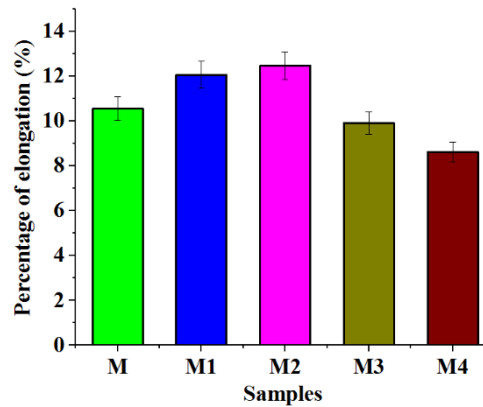
Fig. 4 Percentage of elongation of samples

3.4 Impact Strength

The findings illustrated in Fig. 5 indicate that the impact strength of Al6082 hybrid composites exhibited fluctuations in response to the varying proportions of SiC and Al2O3 reinforcements. The impact strength values pertaining to various compositions are presented as follows: The interfacial shear strength values for SiC concentrations of 2 wt.%, 4 wt.%, 6 wt.%, and 8 wt.% were found to be 30.2 N/m, 36.5 N/m, 40.3 N/m, and 46.9 N/m, respectively. These values were obtained while maintaining a constant 3 % Al2O3 content. The graphical representation unambiguously demonstrates that the impact strength of the composite material reinforced with 8 wt.% silicon carbide and 3 wt.% aluminium oxide was superior in comparison to the composite material reinforced with 2 wt.% silicon carbide and 3 wt.% aluminium oxide. The observation implies that the augmentation of SiC content has a beneficial effect on the impact strength of the hybrid composite. The enhanced impact strength of Al6082 can be ascribed to the incorporation of SiC and Al2O3 reinforcements into its matrix. Ceramic materials such as SiC and Al2O3 exhibit exceptional hardness and strength characteristics. Upon integration into the matrix, these entities function as fortifying agents and facilitate the prevention of crack propagation under conditions of impact loading.
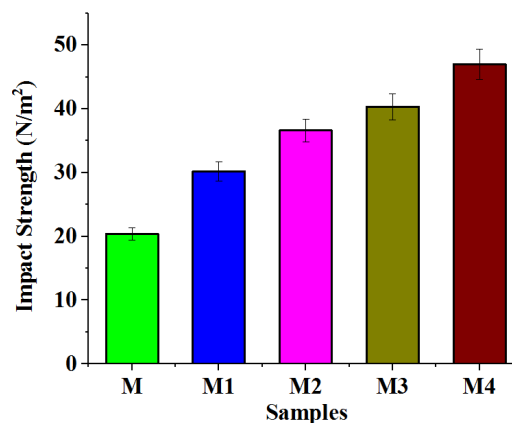
Fig. 5 Impact Strength of samples

The enhanced impact strength of the composite containing 8 weight percent SiC and 3 wt.% of Al2O3, in comparison to the composite comprising 2 wt.% of SiC and 3 wt.% of Al2O3, can be ascribed to the amplified volume fraction and dispersion of the reinforcing particles. An increase in the SiC content leads to a greater number of reinforcing particles that hinder the propagation of cracks, thereby enhancing the material's impact resistance. Furthermore, it is noteworthy that the hybrid composite specimens exhibit greater impact strength in comparison to the pure Al6082. The incorporation of SiC and Al2O3 reinforcements into the composite material presents a noteworthy benefit in terms of enhancing its impact strength. The data suggests that the hybrid composite material exhibits enhanced energy absorption and impact load resistance, rendering it a viable option for scenarios where impact resistance is of paramount importance.

3.5 Hardness test

As depicted in Fig. 6, it is evident that the hardness value of pure Al6082 is comparatively low. The present study reveals that there is an increase in the hardness value of hybrid composites with varying compositions of SiC and Al2O3. The composite containing 8 wt.% SiC and 3 % Al2O3 exhibits a greater hardness value in comparison to the composite containing 2 wt.% SiC and 3 % Al2O3. The observed trend suggests that the incorporation of SiC and Al2O3 reinforcements results in an augmentation of the hardness of the Al6082 matrix. The inclusion of SiC and Al2O3, which are classified as ceramic materials, exerts a substantial impact on the hardness characteristics of the composite. The composite material containing 8 wt.% of silicon carbide (SiC) and 3 wt.% of aluminium oxide (Al2O3) exhibits greater hardness in comparison to the composite material containing 2 wt.% of SiC and 3 wt.% of Al2O3. The present study's results indicate that an elevated SiC reinforcement content correlates with a rise in the hardness of the hybrid composite. The augmented level of hardness can be ascribed to the existence of the ceramic phase within the matrix alloy. The introduction of SiC and Al2O3 reinforcements results in a modification of the microstructure of the composite material. The ceramic particles function as reinforcing agents by impeding the movement of dislocations and offering supplementary resistance to plastic deformation. This phenomenon is a contributing factor to the overall augmentation in hardness. The composite material comprising Al6082 metal exhibits enhanced strength characteristics when reinforced with 8 wt.% SiC and 3 wt.% Al2O3. The observed phenomenon can be ascribed to the rising proportion of the ceramic phase present in the matrix alloy. The composite material's augmented hardness and strength render it appropriate for utilization in scenarios that necessitate enhanced mechanical properties and resilience against deformation.
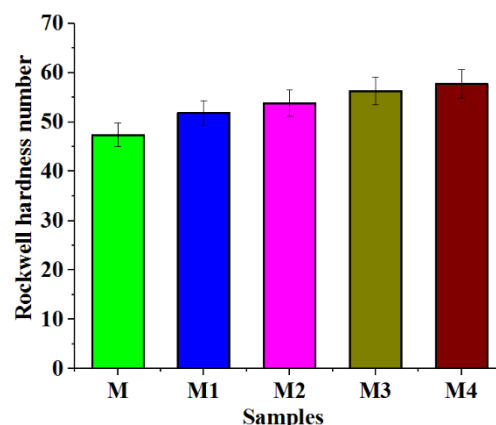
Fig. 6 Rockwell hardness number of samples

4. Mechanical property characterization of hybrid composites using artificial neural networks (ANN)

Artificial Neural Networks (ANNs) are a computational model that draws inspiration from the structural and functional characteristics of the human brain. Artificial neural networks (ANNs) are extensively employed in the field of machine learning and have demonstrated their efficacy in

addressing intricate challenges across diverse domains such as image and speech recognition, natural language processing, and data analysis. Artificial neural networks (Fig. 7) consist of interconnected nodes referred to as artificial neurons or perceptrons. The neurons within the system are arranged in a hierarchical structure, consisting of distinct layers, namely the input layer, one or more hidden layers, and the output layer. The input layer is responsible for receiving the input data, which undergoes processing and transformation as it traverses through the hidden layers. Ultimately, the output layer generates the intended output or prediction. Every individual neuron within an Artificial Neural Network (ANN) is responsible for receiving input signals, undergoing a specific transformation process, and subsequently generating an output signal. In the typical scenario, the inputs are subjected to a process of weighting, wherein each input is multiplied by an associated weight value. Subsequently, the neuron performs a computation by aggregating the weighted sum of its inputs and subsequently applying an activation function to ascertain its output. The utilization of an activation function in the model introduces non-linearity, thereby facilitating the acquisition of intricate patterns and relationships within the dataset. Training an artificial neural network (ANN) entails the utilization of a technique known as backpropagation. This technique facilitates the adjustment of neuron weights by considering the discrepancies between the predicted output and the observed output. The iterative process persists until the performance of the model attains a level that can be considered satisfactory. Artificial neural networks (ANNs) offer a robust framework for constructing intelligent systems capable of emulating specific facets of human cognition. This capability empowers machines to execute intricate tasks and generate precise predictions by leveraging available data.

The experimental and artificial neural network (ANN) prediction plots of all samples with density and UTS are depicted in Fig. 8 indicates that there is a reasonable level of concurrence between the outcomes obtained from the experiment and those derived from the artificial neural network (ANN). The Artificial Neural Network (ANN) was utilized to forecast the density and UTS phenomenon via comparison with experimental outcomes. The results indicated a favourable match with a regression coefficient of $R^2 = 0.974$ & 0.988 (as shown in Fig. 9) and a relative error of 0.003.
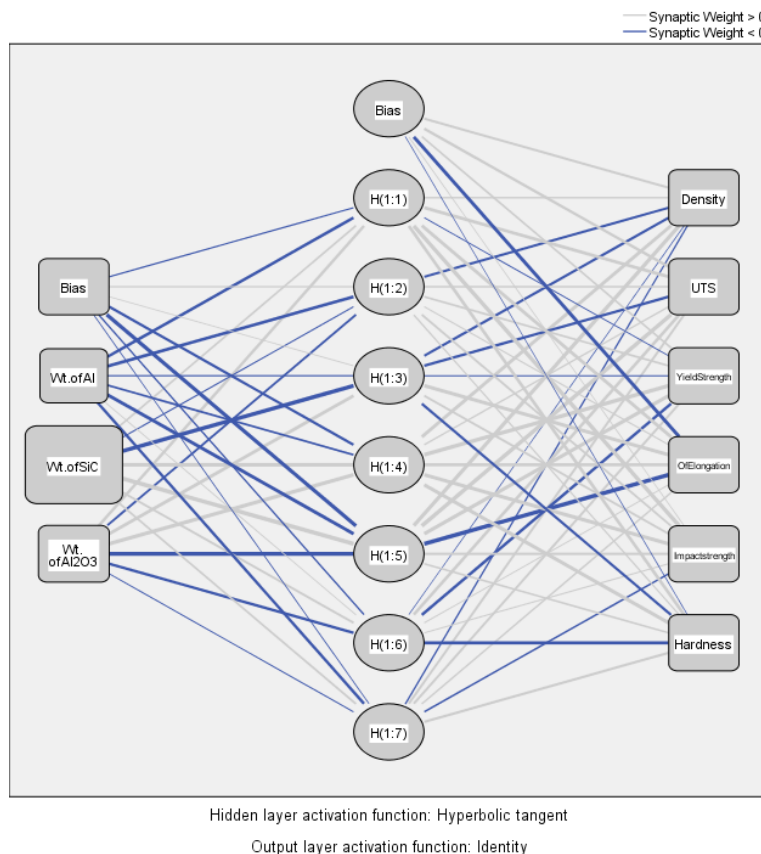
Fig. 7 ANN architecture

The experiment and artificial neural network (ANN) prediction plots of all sample with impact strength and hardness are depicted in Fig. 9. illustrates that the empirical findings are in concurrence with the artificial neural network (ANN) outcomes. The Artificial Neural Network (ANN) was utilized to forecast the impact strength and hardness phenomenon by means of a comparative analysis with experimental outcomes. The results of this analysis demonstrated a strong correlation with a regression coefficient of $R2 = 0.941$, as depicted in Fig. 10 and a relative error of 0.050.

Based on the findings presented in Fig. 10, it can be concluded that the wt.% of SiC holds the highest degree of influence among the composition. The Artificial Neural Network (ANN) was utilized to forecast the mechanical properties phenomenon through a comparative analysis with experimental outcomes.
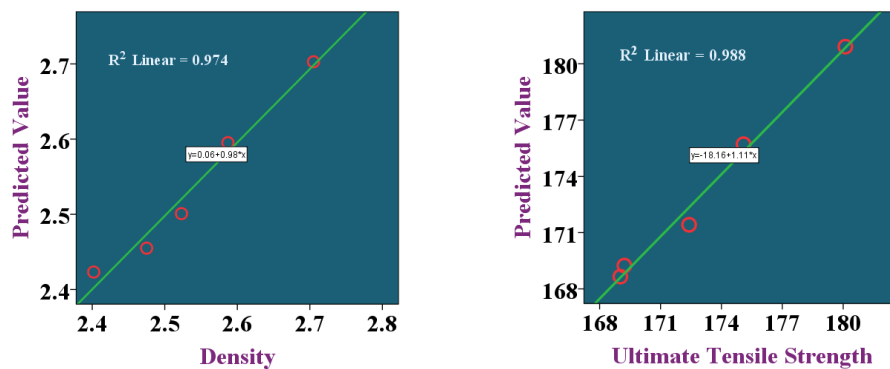
Fig. 8 Experimental vs. predicted values of the samples.
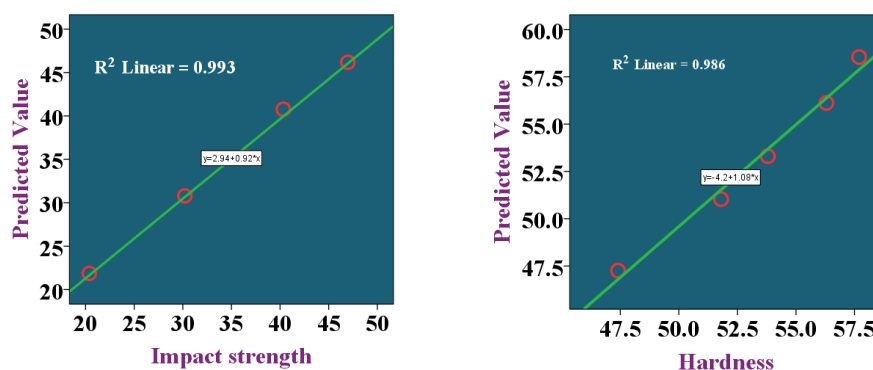
Fig. 9 Experimental vs. predicted values of the samples
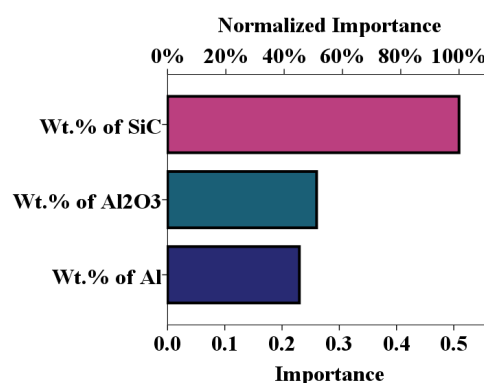
Fig. 10 Normalized factor importance

5. Conclusions

In the current study, Al6082/SiC/Al2O3 nanocomposites obtained by the method of stir-casting and the impact of nano-Al2O3 and nano-SiC particle on mechanical properties of the MMCs were investigated. The conclusions could be concised as follows.

- The experimental density values improved with the rise in proportion of SiC particles as reinforcement.
- The increase in wt. % of SiC enhanced tensile and impact strength of all the casted composite materials.
- The increment in weight percentage of SiC by keeping the load constant of 100 kg improved the hardness. The extreme hardness value was observed with the composites having 8 % wt. of SiC, while the minimum was seen in the composites having 2 % wt. of SiC.
- From the SEM analysis we observed that SiC and Alumina was uniformly distributed in Al6082 metal matrix.
- The implementation of Artificial Neural Network (ANN) analysis has provided confirmation that the weight percentage (wt.%) of silicon carbide (SiC) plays a substantial role in the observed improvement of mechanical properties in Al6082 alloy.

Thus, the experimental results clearly indicate that the Al6082 properties gets better with rise in the weight percentage of SiC reinforcement. Therefore, the integration of SiC and Al2O3 in Al6082 produces improved outcomes than pure Al6082.

Acknowledgement

References

[1] Chawla, K.K. (2012). Metal matrix composites, In: Chawla, K.K. (ed.), *Composite Materials,* Springer, New York, USA, 197-248, doi: 10.1007/978-0-387-74365-3_6.

[2] Rino, J.J., Chandramohan, D., Sucitharan, K.S. (2012). An overview on development of aluminium metal matrix composites with hybrid reinforcement, *International Journal of Science and Research*, Vol. 1, No. 3, 196-203, doi: 10.21275/IJSR13010104.

[3] Nicholls, C.J., Boswell, B., Davies, I.J., Islam, M.N. (2017). Review of machining metal matrix composites, *The International Journal of Advanced Manufacturing Technology,* Vol. 90, 2429-2441, doi: 10.1007/s00170-016-9558-4.

[4] Sharath, B.N., Venkatesh, C.V. (2021). Study on effect of boron carbide, aluminium oxide and graphite on dry sliding wear behaviour of aluminium based metal matrix composite at different temperature, *Tribologia – Finnish Journal of Tribology*,Vol. 38, No. 1-2, 35-46, doi: 10.30678/fjt.99931.

[5] Partheeban, C.M.A., Rajendran, M., Vettivel, S.C., Suresh, S., Moorthi, N.S.V. (2015). Mechanical behavior and failure analysis using online acoustic emission on nano-graphite reinforced Al6061–10TiB2 hybrid composite using powder metallurgy, *Materials Science and Engineering: A,* Vol. 632, 1-13, doi: 10.1016/j.msea.2015.02.064.

[6] Sharath, B.N., Madhu, K.S., Venkatesh, C.V. (2019). Experimental study on dry sliding wear behaviour of Al-B4C-Gr metal matrix composite at different temperatures, *Applied Mechanics and Materials,* Vol. 895, 96-101, doi: 10.4028/www.scientific.net/AMM.895.96.

[7] Imran, M., Khan, A.R.A. (2019). Characterization of Al-7075 metal matrix composites: A review, *Journal of Materials Research and Technology,* Vol. 8, No. 3, 3347-3356, doi: 10.1016/j.jmrt.2017.10.012.

[8] Sharath, B.N., Jeevan, T.P., Baig, M.A.A., Ashrith, H.S., Afzal, A., Reddy, A.R. (2021). Machinability studies on boron carbide and graphite reinforced aluminium hybrid composites, *Materials Today: Proceedings,* Vol. 46, Part 17, 8734-8741, doi: 10.1016/j.matpr.2021.04.036.

[9] Madhu, K.S., Venkatesh, C.V., Sharath, B.N., Karthik, S. (2021). Effect of boron carbide on wear resistance of graphite containing Al7029 based hybrid composites and its dry sliding wear characterization through experimental, response surface method and ANOVA, *Tribologia – Finnish Journal of Tribology,* Vol. 38, No. 3-4, 48-60, doi: 10.30678/fjt.111905.

[10] Sharath, B.N., Madhu, K.S., Pradeep, D.G., Venkatesh, C.V. (2021). Tribological suitability of aluminium hybrid composite above atmospheric temperature, *IOP Conference Series: Materials Science and Engineering,* Vol. 1189, Article No. 012018, doi: 10.1088/1757-899X/1189/1/012018.

[11] Singh, R., Shadab, M., Dash, A., Rai, R.N. (2019). Characterization of dry sliding wear mechanisms of AA5083/B4C metal matrix composite, *Journal of the Brazilian Society of Mechanical Sciences and Engineering,* Vol. 41, Article No. 98, doi: 10.1007/s40430-019-1593-2.

[12] Madhu, K.S., Sharath, B.N., Venkatesh, C.V., Pradeep, D.G. (2021). Evaluation of mechanical properties of ceramic reinforced Aluminium-7029 hybrid composite, *IOP Conference Series: Materials Science and Engineering,* Vol. 1189, No. 1, Article No. 012019, doi: 10.1088/1757-899X/1189/1/012019.

[13] Chikkegouda, S.P., Gurudath, B., Sharath, B.N., Karthik, S., Mahale, R.S. (2022). Mechanical and tribological characteristics of Aluminium 2618 matrix composite reinforced with boron carbide, *Biointerface Research in Applied Chemistry,* Vol. 12, No. 4, 4544-4556, doi: 10.33263/BRIAC124.45444556.

[14] Madhu, K.S., Venkatesh, C.V., Sharath, B.N., Karthik, S. (2023). Characterization and evaluation of mechanical properties of Al-Zn based hybrid metal matrix composites, *Applied Science and Engineering Progress*, Vol. 16, No. 1, 5804-5804, doi: 10.14416/j.asep.2022.03.008.

[15] Reddy, B.R., Srinivas, C. (2018). Fabrication and characterization of silicon carbide and fly ash reinforced aluminium metal matrix hybrid composites, *Materials Today: Proceedings,* Vol. 5, No. 2, Part 2, 8374-8381, doi: 10.1016/j.matpr.2017.11.531.

[16] Sharath, B.N., Karthik, S., Pradeep, D.G., Madhu, K.S., Venkatesh, C.V. (2022). Machinability studies on boron carbide and graphite reinforced Al7029-based hybrid composites, In: Natarajan, E., Vinodh, S., Rajkumar, V. (eds.), *Materials, design and manufacturing for sustainable environment, Lecture notes in mechanical engineering,* Springer, Singapore, 511-522, doi: 10.1007/978-981-19-3053-9_38.

[17] Ballupete Nagaraju, S., Kodigarahalli Somashekara, M., Puttegowda, M., Manjulaiah, H., Kini, C.R., Channarayapattana Venkataramaiah, V. (2022). Effect of B4C/Gr on hardness and wear behavior of Al2618 based hybrid composites through Taguchi and artificial neural network analysis, *Catalysts,* Vol. 12, No. 12, Article No. 1654, doi: 10.3390/catal12121654.

[18] Budapanahalli, S.H., Mallur, S.B., Patil, A.Y., Alosaimi, A.M., Khan, A., Hussein, M.A., Asiri, A.M. (2022). A tribological study on the effect of reinforcing SiC and Al$_2$O$_3$ in Al7075: Applications for spur gears, *Metals,* Vol. 12, No. 6, 1028, doi: 10.3390/met12061028.

[19] Aybarc, U., Dispinar, D., Seydibeyoglu, M.O. (2018). Aluminum metal matrix composites with SiC, Al 2 O 3 and graphene – Review, *Archives of Foundry Engineering,* Vol. 18, No. 2, 5-10, doi: 10.24425/122493.

[20] Krishna, S.G., Subrahmanyam, B.V., Lakshmi, B. (2019). Mechanical characteristics of Al-Al$_2$O$_3$-SiC metal matrix composites made by stir casting technique, *IOP Conference Series: Materials Science and Engineering,* Vol. 653, Article No. 012005, doi: 10.1088/1757-899X/653/1/012005.

[21] Nagaraju, S.B., Puttegowda, M., Somashekara, M.K., Girijappa, Y.G.T., Govindaswamy, P.D., Sathyanarayana, K. (2024). Advancing the performance of ceramic-reinforced Aluminum hybrid composites: A comprehensive review and future perspectives, *Applied Science and Engineering Progress,* Vol. 17, No. 2, 7034-7034, doi: 10.14416/j.asep.2023.10.001.

[22] Sharath, B.N., Venkatesh, C.V., Afzal, A., Aslfattahi, N., Aabid, A., Baig, M., Saleh, B. (2021). Multi ceramic particles inclusion in the aluminium matrix and wear characterization through experimental and response surface-artificial neural networks, *Materials,* Vol. 14, No. 11, Article No. 2895, doi: 10.3390/ma14112895.

[23] Sharma, V.K., Kumar, V., Joshi, R.S., Sharma, D. (2020). Experimental analysis and characterization of SiC and RE oxides reinforced Al-6063 alloy based hybrid composites, *The International Journal of Advanced Manufacturing Technology,* Vol. 108, 1173-1187, doi: 10.1007/s00170-020-05228-7.

[24] Khatkar, S.K., Suri, N.M., Kant, S., Pankaj (2018). A review on mechanical and tribological properties of graphite reinforced self lubricating hybrid metal matrix composites, *Reviews on Advanced Materials Science,* Vol. 56, No. 1-20, doi: 10.1515/rams-2018-0036.

[25] Jayashree, P.K., Gowri Shankar, M.C., Kini, A., Sharma, S.S., Shetty, R. (2013). Review on effect of silicon carbide (SiC) on stir cast aluminium metal matrix composites, *International Journal of Current Engineering and Technology,* Vol. 3, No. 3, 1061-1071.

[26] Sahu, M.K., Sahu, R.K. (2020). Experimental investigation, modeling, and optimization of wear parameters of B4C and fly-ash reinforced aluminum hybrid composite, *Frontiers in Physics,* Vol. 8, Article No. 219, doi: 10.3389/fphy.2020.00219.

[27] Anil Kumar, K.S., Murigendrappa, S.M., Kumar, H. (2019). Experimental investigation on effects of varying volume fractions of SiC nanoparticle reinforcement on microstructure and mechanical properties in friction-stir-welded dissimilar joints of AA2024-T351 and AA7075-T651, *Journal of Materials Research,* Vol. 34, No. 7, 1229-1247, doi: 10.1557/jmr.2018.445.

[28] Thirumoorthy, A., Arjunan, T.V., Senthil Kumar, K.L. (2019). Experimental investigation on mechanical properties of reinforced Al6061 composites and its prediction using KNN-ALO algorithms, *International Journal of Rapid Manufacturing,* Vol. 8, No. 3, 161-177, doi: 10.1504/IJRAPIDM.2019.100498.

[29] Kaushik, N.C., Rao, R.N. (2017). Influence of applied load on abrasive wear depth of hybrid Gr/SiC/Al–Mg–Si composites in a two-body condition, *Journal of Tribology,* Vol. 139, No. 6, Article No. 061601, doi: 10.1115/1.4035779.

[30] Allien, V.J., Kumar, H., Desai, V. (2019). Dynamic analysis and optimization of SiC reinforced Al6082 and Al7075 MMCs, *Materials Research Express,* Vol. 6, No. 5, Article No. 056528, doi: 10.1088/2053-1591/ab038e.

[31] Umer, U., Mohammed, M.K., Abidi, M.H., Alkhalefah, H., Kishawy, H.A. (2022). Machinability analysis and multi-response optimization using NGSA-II algorithm for particle reinforced aluminum based metal matrix composites, *Advances in Production Engineering & Management,* Vol. 17, No. 2, 205-218, doi: 10.14743/apem2022.2.431.

[32] Amin, M., Rathore, M.F., Ahmed, A.A., Saleem, W., Li, Q., Israr, A. (2023). A feed direction cutting force prediction model and analysis for ceramic matrix composites C/SiC based on rotary ultrasonic profile milling, *Advances in Production Engineering & Management,* Vol. 18, No. 3, 288-302, doi: 10.14743/apem2023.3.473.

[33] Stojanović, B., Ivanović, L. (2015). Application of aluminium hybrid composites in automotive industry, *Tehnički Vjesnik – Technical Gazette,* Vol. 22, No. 1, 247-251, doi: 10.17559/TV-20130905094303.

[34] Prince, M., Kumar, A.V., Kumar, G.M. (2022). Investigation on mechanical properties of aluminum 8011 metal matrix compositewith titanium carbide particulate reinforcement, *Tehnički Vjesnik – Technical Gazette,* Vol. 29, No. 6, 2105-2110, doi: 10.17559/TV-20220217173159.

Calendar of events

- North American Manufacturing Research Conference (NAMRC) 52, June 17-21, 2024, Knoxville, TN, USA.

- 18th International Conference on Industrial and Manufacturing Systems Engineering, August 9-10, 2024, Lagos, Nigeria.

- IMTS 2024, International Manufacturing Technology Show, September 9-14, 2024, Chicago, USA.

- 27TH European Conference on Artificial Intelligence, October 19-24, 2024, Santiago de Compostela, Spain.

- 16th EAI International Conference on Simulation Tools and Techniques, December 9-10, 2024, Bratislava, Slovakia.

- International Conference on Smart Materials & Structures, March 27-28, 2025, Berlin, Germany.

- 14th Spring World Congress on Engineering and Technology (SCET 2025), April 19-21, 2025, Guilin, China.

- International Connect & Expo on Material Science and Engineering, April 28-20, 2025, Rome, Italy.

- 2025 Annual Modeling and Simulation Conference (ANNSIM'25), May 26-29, 2025, Madrid, Spain.

- Fifth International Conference on Simulation for Additive Manufacturing (Sim-AM 2025), September 9-11, 2025, Pavia, Italy.

This page intentionally left blank.

Notes for contributors

General

Articles submitted to the *APEM journal* should be original and unpublished contributions and should not be under consideration for any other publication at the same time. Manuscript should be written in English. Responsibility for the contents of the paper rests upon the authors and not upon the editors or the publisher. The content from published paper in the *APEM journal* may be used under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0). For most up-to-date information please see the APEM journal homepage *apem-journal.org*.

Submission of papers

A submission must include the corresponding author's complete name, affiliation, address, phone and fax numbers, and e-mail address. All papers for consideration by *Advances in Production Engineering & Management* should be submitted by e-mail to the journal Editor-in-Chief:

Miran Brezocnik, Editor-in-Chief
UNIVERSITY OF MARIBOR
Faculty of Mechanical Engineering
Chair of Production Engineering
Smetanova ulica 17, SI – 2000 Maribor
Slovenia, European Union
E-mail: editor@apem-journal.org

Manuscript preparation

Manuscript should be prepared in *Microsoft Word 2010* (or higher version) word processor. *Word .docx* format is required. Papers on A4 format, single-spaced, typed in one column, using body text font size of 11 pt, should not exceed 12 pages, including abstract, keywords, body text, figures, tables, acknowledgements (if any), references, and appendices (if any). The title of the paper, authors' names, affiliations and headings of the body text should be in *Calibri* font. Body text, figures and tables captions have to be written in *Cambria* font. Mathematical equations and expressions must be set in *Microsoft Word Equation Editor* and written in *Cambria Math* font. For detail instructions on manuscript preparation please see instruction for authors in the *APEM journal* homepage *apem-journal.org*.

The review process

Every manuscript submitted for possible publication in the *APEM journal* is first briefly reviewed by the editor for general suitability for the journal. Notification of successful submission is sent. After initial screening, and checking by a special plagiarism detection tool, the manuscript is passed on to at least two referees. A double-blind peer review process ensures the content's validity and relevance. Optionally, authors are invited to suggest up to three well-respected experts in the field discussed in the article who might act as reviewers. The review process can take up to eight weeks on average. Based on the comments of the referees, the editor will take a decision about the paper. The following decisions can be made: accepting the paper, reconsidering the paper after changes, or rejecting the paper. Accepted papers may not be offered elsewhere for publication. The editor may, in some circumstances, vary this process at his discretion.

Proofs

Proofs will be sent to the corresponding author and should be returned within 3 days of receipt. Corrections should be restricted to typesetting errors and minor changes.

Offprints

An e-offprint, i.e., a PDF version of the published article, will be sent by e-mail to the corresponding author. Additionally, one complete copy of the journal will be sent free of charge to the corresponding author of the published article.

APEM
journal

Contents