ENDÜSTRİ MÜHENDİSLİĞİ ALANINDA AKADEMİK TARTIŞMALAR

Editör: Doç.Dr.Alper HAMZADAYI

yaz
yayınları

Endüstri Mühendisliği Alanında Akademik Tartışmalar

Editör

Doç.Dr. Alper HAMZADAYI

yaz yayınları

2026

Endüstri Mühendisliği Alanında Akademik Tartışmalar

Editör: Doç.Dr. Alper HAMZADAYI

E_ISBN 978-625-8996-67-8

Haziran 2026 – Afyonkarahisar

Dizgi/Mizanpaj: YAZ Yayınları

Kapak Tasarım: YAZ Yayınları

YAZ Yayınları. Yayıncı Sertifika No: 73086

M.İhtisas OSB Mah. 4A Cad. No:3/3
İscehisar/AFYONKARAHİSAR

www.yazyayinlari.com

yazyayinlari@gmail.com

İÇİNDEKİLER

U-SHAPED MIXED-MODEL ASSEMBLY LINE BALANCING PROBLEM CONSIDERING ERGONOMIC RISK

Seçil KULAÇ[1]

1. INTRODUCTION

Assembly lines are widely used in modern manufacturing systems, enabling high-volume production through the allocation of tasks to workstations. The assembly line balancing problem (ALBP) focuses on assigning tasks while satisfying precedence relations and improving performance measures such as cycle time and resource utilization (Boysen, Schulze, & Scholl, 2022). Although conventional ALBP models have improved production efficiency, they are generally based on time-oriented criteria and provide limited consideration of human-related factors.

Recent manufacturing research has increasingly considered human-centered and sustainable production principles associated with Industry 5.0 (Ghorbani, Keivanpour, Sekkay, & Imbeau, 2023). Despite advances in automation, manual operations remain important because of their flexibility. However, repetitive tasks, awkward postures, and physical demands may lead to work-related musculoskeletal disorders (WMSDs), negatively affecting both worker well-being and production performance (Zhang, Tang, Ruiz, & Zhang, 2020). Consequently, ergonomic considerations have received increasing attention in assembly line balancing studies.

[1] Öğr. Gör. Dr., Bursa Technical University, Quality Coordination Office, ORCID: 0000-0003-3432-0099.

In recent years, various ALBP studies have integrated ergonomic factors into optimization models. Among these approaches, energy expenditure (EE) has been widely used as an indicator of task-level physical workload and ergonomic risk. Previous studies demonstrated that EE can be incorporated into mathematical formulations to support workload evaluation and ergonomic decision-making (Stecke & Mokhtarzadeh, 2022). More recent studies combined ergonomic and operational objectives within integrated models. For example, EE has been evaluated together with noise exposure in mixed-model assembly environments to improve workload distribution and ergonomic conditions (Dalle Mura & Dini, 2023). Other studies considered ergonomic and economic objectives simultaneously, emphasizing trade-offs in assembly system design (Weckenborg, Thies, & Spengler, 2022). Fatigue- and recovery-based models have also been proposed to represent the temporal variation of ergonomic risk (Abdous, Delorme, Battini, Sgarbossa, & Berger-Douce, 2023). In addition, human–robot collaboration has been incorporated into optimization frameworks to address ergonomic risk and operational performance together (Huang, Sheng, Luo, Lu, Fu, & Yin, 2024). Cost-oriented formulations integrating ergonomic constraints and worker–cobot assignment decisions further demonstrate the relationship between ergonomic and economic objectives in assembly systems (Bekdemir & Taşan, 2024).

EE-based approaches estimate workload by considering the metabolic energy required during task execution, including dynamic movements and static postures (Zheng, Li, Zhang, Zhang, & Tang, 2025). Compared with qualitative ergonomic assessment methods, EE provides a continuous representation of workload intensity, allowing ergonomic constraints to be incorporated directly into mathematical models. In addition, EE-based measures support the evaluation of workload distribution

across workers and stations, enabling the simultaneous consideration of operational efficiency and ergonomic balance.

Assembly line configuration is another important factor affecting balancing performance. U-shaped assembly lines provide greater flexibility than straight-line systems by allowing tasks to be assigned to both forward and backward workstation sides. This structure enlarges the feasible solution space and improves workload redistribution possibilities, although it also increases the complexity of precedence management (Jiao, Cao, Li, Li, & Deng, 2022). Furthermore, many industrial systems operate under mixed-model production conditions, where different product types are assembled on the same line. Such environments increase variability in task times and complicate balancing decisions.

Despite these developments, several limitations remain in the literature. Many studies mainly focus on minimizing ergonomic risk without adequately considering workload distribution among workstations, which may result in uneven workload allocation. In addition, most formulations rely on a single ergonomic indicator and therefore do not simultaneously represent workload intensity and workload distribution. Practical factors such as zoning constraints are also frequently neglected. Furthermore, studies integrating ergonomic considerations into U-shaped mixed-model assembly lines remain limited, particularly within unified multi-objective optimization frameworks.

Motivated by these limitations, this study addresses the ergonomic U-shaped mixed-model assembly line balancing problem (Ergo-UALBP-2) using an integrated optimization framework. The proposed model uses EE to quantify task-level ergonomic workload and introduces two ergonomic measures: maximum ergonomic risk (ER) and ergonomic class imbalance

(EC). ER represents the highest cumulative workload assigned to a workstation, whereas EC evaluates workload distribution across predefined workload classes. The objective function integrates cycle time (CT), ER, and EC within a weighted framework. The formulation incorporates precedence relations together with positive and negative zoning constraints to represent practical task compatibility requirements. In addition, the U-shaped structure allows assignments on both forward and backward workstation sides, while mixed-model production is represented through model-dependent processing times.

2. LITERATURE REVIEW

The ALBP has been widely studied as an optimization problem focused on assigning tasks to workstations under precedence and capacity constraints. Classical formulations, known as simple assembly line balancing problems, generally aim to minimize the number of stations or cycle time while ensuring feasible task assignments (Boysen et al., 2022). Over time, these models have been extended to better represent practical production environments by incorporating structural flexibility, processing variability, and additional operational constraints.

Among these extensions, U-shaped assembly lines have attracted attention due to their flexibility and efficiency advantages. In these systems, tasks can be assigned to both forward and backward sides of the line, improving workstation utilization and workload distribution (Jiao et al., 2022). In addition, U-shaped configurations allow workers to operate across multiple stations, supporting flexible work organization (Zülch & Zülch, 2017). This flexibility has contributed to the development of parallel and hybrid U-shaped systems, which improve resource utilization and system capacity through shared

workstations and cross-line task assignments (Kucukkoc & Zhang, 2015; Chutima & Khotsaenlee, 2022).

Mixed-model assembly lines have also been extensively investigated to address production environments where multiple product variants are assembled on the same line. Such systems introduce model-dependent task times and increased workload variability, which complicate balancing decisions (Boysen, Fliedner, & Scholl, 2009; Hwang & Katayama, 2009). The integration of ergonomic considerations into these environments further increases model complexity by requiring the simultaneous consideration of operational and human-related factors (Zhang et al., 2020).

In recent years, ergonomic assembly line balancing has become an important research area. These studies incorporate human-related factors such as physical workload and fatigue into optimization models. Various ergonomic assessment methods, including REBA, RULA, OWAS, OCRA, and EAWS, have been used to quantify ergonomic risk, each addressing different aspects of physical strain (Otto & Battaïa, 2017). Ergonomic measures are generally incorporated either as constraints limiting allowable risk levels or as objective functions aimed at improving working conditions (Otto & Scholl, 2011; Battini, Calzavara, Otto, & Sgarbossa, 2016).

More recently, EE-based approaches have received increasing attention. EE-based methods estimate the metabolic energy required during task execution and provide a continuous representation of physical workload that can be directly integrated into mathematical models (Zheng et al., 2025). This representation supports workstation-level workload evaluation and facilitates the incorporation of ergonomic constraints into optimization frameworks. In addition, EE-based measures enable

the comparison of workload distribution across workstations, supporting more balanced assembly line designs.

The literature also shows increasing interest in multi-objective formulations that jointly consider operational efficiency and ergonomic performance measures such as cycle time and ergonomic risk (Otto & Scholl, 2011). In addition, practical factors such as zoning constraints and task incompatibilities have been incorporated to better represent real production settings.

Despite these developments, most studies model ergonomic risk primarily as a maximum allowable limit at the workstation level. Although this approach controls peak workload, it does not ensure balanced workload distribution among stations, particularly in U-shaped and mixed-model systems. Furthermore, many studies rely on a single aggregated ergonomic measure that reflects workload intensity but does not capture workload distribution across different risk categories. Therefore, there remains a need for integrated approaches that jointly consider workload intensity and balanced ergonomic risk distribution in U-shaped mixed-model assembly lines.

3. MATERIAL AND METHODS

3.1. Problem description

This study considers the Ergo UALBP-2 within a multi-criteria optimization framework. The problem aims to assign tasks to a fixed number of workstations while jointly considering production efficiency and ergonomic performance. The U-shaped line structure allows tasks to be assigned to either the forward or backward side of each workstation, increasing assignment flexibility and expanding the feasible solution space.

The objective function integrates cycle time, ergonomic risk, and ergonomic class imbalance within a weighted structure,

allowing different decision priorities to be represented. In this way, the model addresses the trade-off between production efficiency and ergonomic performance.

Task assignments are subject to operational and ergonomic constraints. Precedence relations ensure technological feasibility, while zoning constraints represent practical production requirements. Positive zoning requires specific tasks to be assigned to the same workstation, whereas negative zoning prevents incompatible task combinations. In addition, station-level ergonomic limits are considered to restrict excessive workload accumulation.

The problem is formulated under the following assumptions:

- Task processing times, precedence relations, and EE values are known in advance and remain constant.
- Each task is exclusively assigned to one station without division.
- The number of workstations is fixed, and each workstation is assigned at least one task.
- Precedence constraints must be satisfied for all task assignments.
- Tasks can be allocated to the forward or backward side of a workstation in the U-shaped configuration.
- The system operates as a mixed-model line, with model-dependent task processing times.

The proposed mathematical model was developed based on the framework introduced by Mokhtarzadeh, Rabbani, and Manavizadeh (2021). The notations used in the mathematical formulation are presented in Table 1.

Table 1. Notation of the proposed mathematical model

Indices and sets	
i, j	Index of tasks, $i, j \in \{1,2,\dots,NT\}$
s	Index of stations, $s \in \{1,2,\dots,NS\}$
m	Index of product models, $m \in \{1,2,\dots,NM\}$
P	Set of immediate precedence relations; $(i,j) \in P$ means task i precedes task j
Parameters	
NT	Number of assembly tasks
NS	Number of stations
NM	Number of product models
t_{im}	Processing time of task i of a product of model m
E_i	Ergonomic risk of task i
$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$	Represent the relative importance weights of the six workload levels, namely very light, light, moderate, heavy, very heavy, and extremely heavy, respectively.
w_1, w_2, w_3	Weights of the objective function
$\gamma_i, \lambda_i, \vartheta_i, \delta_i, \theta_i, \beta_i \in \{0,1\}$	indicate whether task i belongs to the very light, light, moderate, heavy, very heavy, or extremely heavy workload category, respectively.
$\rho_{ij} = \begin{cases} 1, \text{ if task } i \text{ and } j \text{ must be assigned to the same station;} \\ 0, \text{ otherwise} \end{cases}$	
$\eta_{ij} = \begin{cases} 1, \text{ if task } i \text{ and } j \text{ cannot be assigned to the same station;} \\ 0, \text{ otherwise} \end{cases}$	
Decision variables	
$x_{is} \in \{0,1\}$	1 if task i is assigned to station s on the forward side; 0 otherwise
$y_{is} \in \{0,1\}$	1 if task i is assigned to station s on the backward side; 0 otherwise
ER	Maximum ergonomic risk across all stations
EC	Maximum ergonomic class imbalance across all stations
CT	Cycle time

Objective function:

$$Minimize\ z = w_1 CT + w_2 ER + w_3 EC \tag{1}$$

Constraints:

$$\sum_{i=1}^{NT} (E_i(x_{is} + y_{is})) \leq ER \qquad \forall s \in \{1,2,\dots,\mathrm{NS}\} \tag{2}$$

$$\alpha_1 \sum_{i=1}^{NT} \gamma_i (x_{is} + y_{is}) + \alpha_2 \sum_{i=1}^{NT} \lambda_i (x_{is} + y_{is}) + \alpha_3 \sum_{i=1}^{NT} \vartheta_i (x_{is} + y_{is}) + \alpha_4 \sum_{i=1}^{NT} \delta_i (x_{is} + y_{is}) + \alpha_5 \sum_{i=1}^{NT} \theta_i (x_{is} + y_{is}) + \alpha_6 \sum_{i=1}^{NT} \beta_i (x_{is} + y_{is}) \leq EC \qquad \forall s \in \{1, 2, \dots, \mathrm{NS}\} \qquad (3)$$

$$\sum_{s=1}^{NS} (x_{is} + y_{is}) = 1 \qquad \forall i \in \{1, 2, \dots, \mathrm{NT}\} \qquad (4)$$

$$\sum_{i=1}^{NT} t_{im} (x_{is} + y_{is}) \leq CT \qquad \forall s \in \{1, 2, \dots, \mathrm{NS}\}, \forall m \in \{1, 2, \dots, NM\} \qquad (5)$$

$$\sum_{s=1}^{NS} (NS - s + 1)(x_{is} - x_{js}) \geq 0 \qquad \forall (i, j) \in P \qquad (6)$$

$$\sum_{s=1}^{NS} (NS - s + 1)(y_{js} - y_{is}) \geq 0 \qquad \forall (i, j) \in P \qquad (7)$$

$$\sum_{i=1}^{NT} (x_{is} + y_{is}) \geq 1 \qquad \forall s \in \{1, 2, \dots, \mathrm{NS}\} \qquad (8)$$

$$x_{is} = x_{js} \qquad \forall i, j \in \{1, 2, \dots, \mathrm{NT}\}, \forall s \in \{1, 2, \dots, \mathrm{NS}\}, \quad \rho_{ij} = 1 \qquad (9)$$

$$y_{is} = y_{js} \qquad \forall i, j \in \{1, 2, \dots, \mathrm{NT}\}, \forall s \in \{1, 2, \dots, \mathrm{NS}\}, \quad \rho_{ij} = 1 \qquad (10)$$

$$x_{is} + x_{js} \leq 1 \qquad \forall i, j \in \{1, 2, \dots, \mathrm{NT}\}, \forall s \in \{1, 2, \dots, \mathrm{NS}\}, \quad \eta_{ij} = 1 \tag{11}$$

$$y_{is} + y_{js} \leq 1 \qquad \forall i, j \in \{1, 2, \dots, \mathrm{NT}\}, \forall s \in \{1, 2, \dots, \mathrm{NS}\}, \quad \eta_{ij} = 1 \tag{12}$$

$$CT \geq 0, \mathrm{ER} \geq 0, EC \geq 0 \tag{13}$$

$$\mathrm{x_{is}}, y_{\mathrm{is}} \epsilon \{0,1\} \qquad \forall i \in \{1, 2, \dots, \mathrm{NT}\}, \forall s \in \{1, 2, \dots, \mathrm{NS}\} \tag{14}$$

Equation (1) formulates the problem as a multi-objective optimization model in which CT, ER, and EC are minimized through a weighted sum approach. ER and EC are defined as decision variables rather than exogenous parameters, allowing ergonomic performance to be optimized concurrently with production efficiency. Equation (2) introduces ER as an upper-bounding variable for station-level ergonomic risk. It ensures that the cumulative ergonomic load assigned to each station does not exceed ER. The minimization of ER reduces the maximum workload concentration and contributes to a more balanced distribution of ergonomic risk across stations. Equation (3) defines EC as an upper bound on the weighted accumulation of ergonomic workload categories at each station. This constraint regulates the distribution of tasks across workload intensity levels. The minimization of EC promotes a more uniform allocation of ergonomic classes and mitigates the concentration of high-intensity tasks within individual stations. According to Equation (4), each task is assigned uniquely to one station, considering both the forward and backward sides of the U-shaped line. Equation (5) introduces the cycle time restriction by requiring that the cumulative processing time of tasks at any station remains within the specified limit CT. Equations (6) and (7) enforce the precedence relations for tasks assigned to the forward and backward sides of the U-shaped line, respectively.

These constraints ensure that each predecessor task is positioned before its corresponding successor according to the directional structure of the U-line. The term NS−s+1 represents the relative position of station s, allowing the precedence relations to be expressed consistently for both sides of the U-shaped configuration. Equation (8) guarantees that each station is assigned at least one task, thereby preventing idle stations and preserving the structural integrity of the line. Equations (9) and (10) impose positive zoning requirements by ensuring that specified task pairs are assigned to the same station and the same side of the U-shaped line. Equations (11) and (12) enforce negative zoning restrictions by preventing incompatible task pairs from being assigned to the same station on the same side of the line. Equation (13) defines the non-negativity conditions for the continuous decision variables CT, ER, and EC. Equation (14) specifies the binary nature of the assignment variables.

4. RESULTS AND DISCUSSION

4.1. Experimental settings

This study evaluates the Ergo UALBP-2 using benchmark instances obtained from the ALBP database (assembly-line-balancing.de). Since the original datasets do not include ergonomic information, EE values are assigned to each task in order to incorporate ergonomic considerations into the model. The assignment process follows the workload classification proposed by Cai, Wang, Luo, and Xu (2025), which categorizes physical workload into six levels, namely very light, light, moderate, heavy, very heavy, and extremely heavy work, each defined over specific energy expenditure intervals. This classification provides a structured basis for interpreting EE values within established workload levels. Accordingly, EE values are randomly generated within the corresponding intervals

for each task, ensuring variability while maintaining consistency with realistic workload conditions.

Workload intensity is represented in the mathematical model through workload class coefficients defined using the model notation as α_1, α_2, α_3, α_4, α_5, and α_6 corresponding to increasing workload levels from very light to extremely heavy. These coefficients are set to 0.02, 0.08, 0.15, 0.20, 0.25, and 0.30, respectively, and are incorporated into the model to capture the effect of ergonomic load on system performance.

The model is solved under different objective weight configurations to analyze the trade-off between productivity and ergonomic considerations. The weighted objective function is formulated using the parameters w_1, w_2, and w_3 which represent the relative importance of cycle time, ergonomic risk, and ergonomic class imbalance, respectively. Four different configurations are considered in the computational analysis. The first configuration focuses solely on cycle time minimization by setting $w_1 = 1$ and $w_2 = w_3 = 0$, representing a purely productivity-oriented scenario. The second and third configurations introduce small but non-zero weights for ergonomic components, with $w_2 = w_3 = 0.02$ and $w_2 = w_3 = 0.05$, enabling the evaluation of moderately balanced trade-offs between operational efficiency and ergonomic performance. The final configuration assigns equal importance to all objective components by setting $w_1 = w_2 = w_3 = 1$, reflecting a strongly multi-criteria decision-making perspective.

All problem instances are solved using IBM ILOG CPLEX Optimization Studio under default parameter settings. A time limit of 1000 seconds is imposed for each instance, and the best feasible solution obtained within this limit is reported. In contrast to earlier versions of the study, no heuristic or

metaheuristic solution approach is employed, and all results are derived directly from the MILP formulation. The performance of the model is evaluated in terms of CT, ER, EC, and objective value, providing a comprehensive assessment of system behavior under different weight configurations.

4.2. Computational results

The computational results obtained from the MILP model under different objective weight configurations are presented in Table 2. The analysis focuses on the impact of incorporating ergonomic considerations into the optimization framework, particularly in terms of CT, ER, and EC.

The results indicate that the baseline configuration, in which only cycle time is minimized ($w_1 = 1,\ w_2 = w_3 = 0$), yields the lowest CT values across all problem instances. However, this configuration is associated with relatively high levels of ergonomic risk and class imbalance, highlighting the limitations of purely productivity-oriented optimization. As ergonomic components are gradually introduced into the objective function, a clear trade-off emerges between operational efficiency and ergonomic performance.

When small weights are assigned to ergonomic criteria ($w_2 = w_3 = 0.02$ and $w_2 = w_3 = 0.05$), the model is able to significantly reduce both ER and EC values while maintaining nearly identical cycle time levels. In most problem instances, CT remains unchanged under these configurations, demonstrating that substantial ergonomic improvements can be achieved without compromising productivity. This behavior is particularly evident in medium- and large-scale instances such as P3, P6, and P8, where ER reductions exceed 30% and EC reductions reach over 50% in some cases.

A more pronounced effect is observed when all objective components are equally weighted ($w_1 = w_2 = w_3 = 1$). Under

this configuration, ergonomic performance is further improved, with ER and EC values reaching their lowest levels across all scenarios. However, this improvement is accompanied by a slight deterioration in cycle time for certain instances. For example, in problems such as P2 and P7, small increases in CT are observed, reflecting the stronger emphasis placed on ergonomic criteria. Despite this, the average change in CT across all instances remains negligible, with an overall variation of approximately −0.2%, indicating that the model maintains a stable level of productivity even under strong multi-criteria conditions.

From an aggregated perspective, the incorporation of ergonomic criteria leads to an average improvement of approximately 27% in ER and 38% in EC across all instances. These results demonstrate that the proposed model effectively balances operational and ergonomic objectives, achieving significant reductions in ergonomic risk and workload imbalance with minimal impact on cycle time. The findings further suggest that moderate weight configurations provide a favorable compromise, offering substantial ergonomic benefits without introducing performance degradation.

Overall, the results confirm that incorporating ergonomic considerations improves workload distribution across stations and reduces peak ergonomic loads, particularly in large-scale assembly systems.

Table 2. Performance of the MILP model under different weight scenarios

Problem	Weights			Number of Models	Number of Tasks	Number of Stations	MILP Results				Improvement (%)		
	w_1	w_2	w_3				Objective Value	CT	ER	EC	CT	ER	EC
P1	1	0	0	1	11	3	16.0	16	37.0	1.0			
	1	0.02	0.02				16.7	16	33.1	0.7	0.0%	10.5%	28.0%
	1	0.05	0.05				17.7	16	33.1	0.7	0.0%	10.5%	28.0%
	1	1	1				49.8	16	33.1	0.7	0.0%	10.5%	28.0%
P2	1	0	0	1	11	3	62.0	62	34.0	1.0			
	1	0.02	0.02				62.6	62	29.0	0.6	0.0%	14.9%	38.0%
	1	0.05	0.05				63.5	62	29.0	0.6	0.0%	14.9%	38.0%
	1	1	1				91.2	64	26.6	0.6	-3.2%	21.8%	40.0%
P3	1	0	0	3	19	3	157.0	157	93.0	3.0			
	1	0.02	0.02				158.3	157	61.6	1.3	0.0%	33.8%	55.7%
	1	0.05	0.05				160.1	157	61.6	1.3	0.0%	33.8%	55.7%
	1	1	1				219.9	157	61.6	1.3	0.0%	33.8%	55.7%
P4	1	0	0	1	30	6	54.0	54	41.0	1.0			
	1	0.02	0.02				54.7	54	36.2	0.8	0.0%	11.7%	16.0%
	1	0.05	0.05				55.9	54	36.2	0.8	0.0%	11.7%	16.0%
	1	1	1				91.1	54	36.2	0.8	0.0%	11.7%	16.0%
P5	1	0	0	1	35	5	97	97	78.0	2.0			
	1	0.02	0.02				98.1	97	55.0	1.3	0.0%	29.5%	34.5%
	1	0.05	0.05				99.8	97	55.0	1.3	0.0%	29.4%	36.5%
	1	1	1				153.3	97	55.0	1.3	0.0%	29.4%	36.5%
P6	1	0	0	1	45	6	92.0	92	126.0	3.0			
	1	0.02	0.02				93.3	92	64.0	1.5	0.0%	49.2%	51.3%
	1	0.05	0.05				95.3	92	63.9	1.4	0.0%	49.3%	52.3%
	1	1	1				157.3	92	63.8	1.5	0.0%	49.3%	51.7%
P7	1	0	0	8	50	8	847.0	847	74.0	2.0			
	1	0.02	0.02				846.4	845	68.3	1.5	0.2%	7.7%	26.0%
	1	0.05	0.05				853.0	850	58.1	1.3	-0.4%	21.4%	35.0%
	1	1	1				908.9	854	53.7	1.2	-0.8%	27.4%	40.0%
P8	1	0	0	1	70	6	585.0	585	166.0	5.0			
	1	0.02	0.02				587.0	585	98.0	2.6	0.0%	41.0%	48.2%
	1	0.05	0.05				589.9	585	96.3	2.6	0.0%	42.0%	48.4%
	1	1	1				683.9	585	96.2	2.7	0.0%	42.1%	45.6%
Average											-0.2%	27%	38%

5. CONCLUSIONS

This study addresses the Ergo-UALBP-2 within a unified mathematical programming framework, aligned with the principles of Industry 5.0, which emphasize human-centered and sustainable production systems. The proposed formulation integrates cycle time, ergonomic risk, and ergonomic class imbalance into a single weighted objective structure. In addition to classical assembly line balancing constraints, including precedence relations and zoning requirements, the model incorporates station-level ergonomic considerations that target

the minimization of maximum ergonomic risk across stations and the improvement of workload class balance.

EE is used to represent task-level physical workload and to derive ergonomic risk levels. Based on this representation, the model simultaneously reduces peak ergonomic exposure and improves the distribution of workload classes across stations, enabling a structured evaluation of ergonomic performance within the optimization framework.

Computational experiments are conducted using benchmark instances derived from the ALBP database. The model is solved using a MILP approach under multiple weight configurations to examine the interaction between productivity and ergonomic performance. The results show that cycle time-oriented configurations are associated with higher levels of ergonomic risk and imbalance. The inclusion of ergonomic components in the objective function leads to consistent improvements in both maximum ergonomic risk and workload class balance. Moderate weight configurations provide a balanced outcome, where ergonomic improvements are achieved while cycle time values remain stable. Configurations with equal weights further enhance ergonomic performance, with limited variation observed in cycle time across most instances.

The findings indicate that the proposed formulation provides a structured framework for analyzing trade-offs between operational efficiency and ergonomic considerations. The results show that improvements in maximum ergonomic risk and workload balance can be achieved within the same system configuration, particularly under moderate weighting strategies. This behavior becomes more pronounced as problem size increases, reflecting the importance of integrated ergonomic modeling in complex assembly systems.

The proposed framework can be extended in several directions. Different assembly line configurations, including straight lines, two-sided lines, parallel lines, and human–robot collaborative systems, may be incorporated to analyze structural differences in ergonomic load distribution. Alternative ergonomic assessment methods, such as RULA, REBA, and OWAS, may be integrated alongside EE-based measures to provide a broader evaluation of physical workload. From a methodological perspective, heuristic and metaheuristic solution approaches, including genetic algorithms, simulated annealing, particle swarm optimization, and hyper-heuristic frameworks, may be developed to address large-scale instances and complex constraint structures. Reinforcement learning-based adaptive strategies may also be explored for dynamic decision-making environments. In addition, multi-objective optimization approaches, such as Pareto-based evolutionary algorithms, may provide further insights into the relationship between productivity and ergonomic performance.

REFERENCES

Abdous, M. A., Delorme, X., Battini, D., Sgarbossa, F., & Berger-Douce, S. (2023). Assembly line balancing problem with ergonomics: a new fatigue and recovery model. International Journal of Production Research, 61(3). https://doi.org/10.1080/00207543.2021.2015081

Battini, D., Calzavara, M., Otto, A., & Sgarbossa, F. (2016). The Integrated Assembly Line Balancing and Parts Feeding Problem with Ergonomics Considerations. IFAC-PapersOnLine, 49(12). https://doi.org/10.1016/j.ifacol.2016.07.594

Bekdemir, P., & Taşan, S. Ö. (2024). Cost-based assembly line balancing and worker-cobot assignment problem under ergonomic constraints. Journal of the Faculty of Engineering and Architecture of Gazi University, 39(1). https://doi.org/10.17341/gazimmfd.1182311

Boysen, N., Fliedner, M., & Scholl, A. (2009). Sequencing mixed-model assembly lines: Survey, classification and model critique. In European Journal of Operational Research (Vol. 192, Issue 2). https://doi.org/10.1016/j.ejor.2007.09.013

Boysen, N., Schulze, P., & Scholl, A. (2022). Assembly line balancing: What happened in the last fifteen years? In European Journal of Operational Research (Vol. 301, Issue 3). https://doi.org/10.1016/j.ejor.2021.11.043

Cai, M., Wang, G., Luo, X., & Xu, X. (2025). Task allocation of human-robot collaborative assembly line considering assembly complexity and workload balance. International Journal of Production Research, 63(13), 4749–4775. https://doi.org/10.1080/00207543.2024.2442546

Chutima, P., & Khotsaenlee, A. (2022). Multi-objective parallel adjacent U-shaped assembly line balancing collaborated by robots and normal and disabled workers. Computers and Operations Research, 143. https://doi.org/10.1016/j.cor.2022.105775

Dalle Mura, M., & Dini, G. (2023). Improving ergonomics in mixed-model assembly lines balancing noise exposure and energy expenditure. CIRP Journal of Manufacturing Science and Technology, 40. https://doi.org/10.1016/j.cirpj.2022.11.005

Ghorbani, E., Keivanpour, S., Sekkay, F., & Imbeau, D. (2023). Ergonomic Assembly Line Balancing Problems Evolution and Future Trends with Insights into Industry 5.0 Paradigm. June, 1–34. https://www.researchgate.net/publication/372391019

Huang, Y., Sheng, B., Luo, R., Lu, Y., Fu, G., & Yin, X. (2024). Solving human-robot collaborative mixed-model two-sided assembly line balancing using multi-objective discrete artificial bee colony algorithm. Computers and Industrial Engineering, 187. https://doi.org/10.1016/j.cie.2023.109776

Hwang, R., & Katayama, H. (2009). A multi-decision genetic approach for workload balancing of mixed-model U-shaped assembly line systems. International Journal of Production Research, 47(14). https://doi.org/10.1080/00207540701851772

Jiao, Y., Cao, N., Li, J., Li, L., & Deng, X. (2022). Balancing a U-Shaped Assembly Line with a Heuristic Algorithm Based on a Comprehensive Rank Value. Sustainability (Switzerland), 14(2). https://doi.org/10.3390/su14020775

Kucukkoc, I., & Zhang, D. Z. (2015). Balancing of parallel U-shaped assembly lines. Computers and Operations Research, 64. https://doi.org/10.1016/j.cor.2015.05.014

Mokhtarzadeh, M., Rabbani, M., & Manavizadeh, N. (2021). A novel two-stage framework for reducing ergonomic risks of a mixed-model parallel U-shaped assembly-line. Applied Mathematical Modelling, 93. https://doi.org/10.1016/j.apm.2020.12.027

Stecke, K. E., & Mokhtarzadeh, M. (2022). Balancing collaborative human–robot assembly lines to optimise cycle time and ergonomic risk. International Journal of Production Research, 60(1). https://doi.org/10.1080/00207543.2021.1989077

Otto, A., & Battaïa, O. (2017). Reducing physical ergonomic risks at assembly lines by line balancing and job rotation: A survey. Computers and Industrial Engineering, 111. https://doi.org/10.1016/j.cie.2017.04.011

Otto, A., & Scholl, A. (2011). Incorporating ergonomic risks into assembly line balancing. European Journal of Operational Research, 212(2). https://doi.org/10.1016/j.ejor.2011.01.056

Weckenborg, C., Thies, C., & Spengler, T. S. (2022). Harmonizing ergonomics and economics of assembly lines using collaborative robots and exoskeletons. Journal of Manufacturing Systems, 62. https://doi.org/10.1016/j.jmsy.2022.02.005

Zhang, Z., Tang, Q. H., Ruiz, R., & Zhang, L. (2020). Ergonomic risk and cycle time minimization for the U-shaped worker assignment assembly line balancing problem: A multi-objective approach. Computers and Operations Research, 118. https://doi.org/10.1016/j.cor.2020.104905

Zheng, C., Li, Z., Zhang, Z., Zhang, L., & Tang, Q. (2025). Multi-objective restarted simulated annealing algorithm for assembly line balancing problem with collaborative robots considering ergonomics risks. In Flexible Services and Manufacturing Journal (Issue 0123456789). Springer US. https://doi.org/10.1007/s10696-025-09590-0

Zülch, M., & Zülch, G. (2017). Production logistics and ergonomic evaluation of U-shaped assembly systems. International Journal of Production Economics, 190. https://doi.org/10.1016/j.ijpe.2017.01.004

HYBRID ITERATED GREEDY AND DECODER-BASED METAHEURISTICS FOR UNRELATED PARALLEL MACHINE SCHEDULING

Alper HAMZADAYI[1]

1. INTRODUCTION

Large-scale scheduling problems frequently become computationally intractable when exact optimization approaches are directly applied (Pinedo, 2012). As the number of jobs, machines, and operational constraints increases, the search space grows combinatorially, making it difficult to obtain high-quality solutions within practical computational times. Consequently, advanced metaheuristic frameworks have become increasingly important for solving complex production scheduling environments encountered in modern manufacturing systems (Ruiz & Stützle, 2007).

Among modern optimization approaches, Iterated Greedy (IG) algorithms have attracted substantial attention because of their simplicity, robustness, and strong performance across different scheduling environments (Ruiz & Stützle, 2007). The main principle of IG algorithms relies on repeatedly destroying and reconstructing partial solutions in order to balance diversification and intensification throughout the search process. More recently, hybrid variants integrating local search procedures, adaptive reconstruction strategies, and decoder-guided evaluation mechanisms have demonstrated particularly

[1] Doç. Dr., Van Yüzüncü Yıl Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, ORCID: 0000-0003-4035-2775.

strong performance for large-scale combinatorial optimization problems (Feo & Resende, 1995; Mladenović & Hansen, 1997).

One of the most critical components of hybrid metaheuristics is the solution evaluation mechanism. In many scheduling environments, the effectiveness of the search process strongly depends on how candidate solutions are decoded and transformed into feasible production schedules. Decoder-based evaluation procedures dynamically coordinate sequencing decisions, timing interactions, setup operations, and resource availability constraints during schedule construction. Consequently, decoder design directly affects both search diversification capability and final solution quality.

This issue becomes particularly important in scheduling systems involving sequence-dependent setup operations and limited shared setup resources. In common-server environments, setup activities compete for server availability, creating strong synchronization dependencies among machines (Allahverdi, 2015). As a result, local sequencing decisions may generate highly complex global interactions throughout the production system. Therefore, the quality of the scheduling process depends not only on machine assignments but also on how setup operations are coordinated during decoding.

The scheduling environment becomes even more challenging when unrelated parallel machines are considered. In such systems, processing times vary according to machine–job combinations and some jobs may only be processed on eligible machines due to technological restrictions. Consequently, hybrid optimization approaches must simultaneously coordinate machine assignments, sequence-dependent setups, server operations, and timing decisions within highly constrained search spaces.

Although numerous heuristic and metaheuristic approaches have been proposed for scheduling problems with additional resources, most existing studies primarily focus on makespan-oriented objectives and classical dispatching strategies. Furthermore, decoder-guided search mechanisms remain relatively underexplored despite their significant influence on scheduling quality. Bektur and Saraç (2019) investigated the R_M, $S_1/ST_{sd}/\sum w_j \times T_j$ scheduling problem and proposed tabu search and simulated annealing approaches for minimizing total weighted tardiness. Their study demonstrated the computational difficulty of coordinating setup-server operations and machine scheduling decisions under weighted tardiness objectives. Candidate solutions were evaluated using a decoding mechanism based on minimum completion time decisions. However, the effect of alternative decoder strategies and advanced hybrid search mechanisms was not extensively investigated.

More recently, hybrid metaheuristic approaches integrating GRASP, Variable Neighborhood Search, Iterated Greedy, and adaptive local search procedures have demonstrated strong optimization capability for large-scale scheduling environments (Feo & Resende, 1995; Mladenović & Hansen, 1997; Ruiz & Stützle, 2007). Nevertheless, decoder-guided search strategies capable of dynamically coordinating setup-server operations during evaluation remain limited in the literature.

Motivated by these observations, this study proposes a hybrid Iterated Greedy framework, denoted as IG-HM, for unrelated parallel machine scheduling with sequence-dependent setup times under a common-server constraint. The proposed framework integrates GRASP-based construction, Variable Neighborhood Descent (VND), decoder-guided reconstruction,

and adaptive perturbation mechanisms within a unified search architecture.

Unlike conventional IG implementations relying on static evaluation rules, the proposed framework dynamically constructs server sequences during decoding using alternative evaluation strategies. Three decoder mechanisms are investigated: a minimum completion time decoder (MCT), a marginal weighted tardiness decoder (MinΔ), and a two-step look-ahead decoder (TSLA). The proposed TSLA structure explicitly considers future setup interactions during evaluation, allowing the search process to avoid locally attractive yet globally harmful sequencing decisions.

The proposed framework therefore combines diversification, adaptive decoding, and neighborhood-based intensification within a unified hybrid metaheuristic structure. Extensive computational experiments are conducted to evaluate the impact of decoder strategies, runtime settings, and neighborhood structures on overall scheduling performance.

The main contributions of this study can be summarized as follows:

1. A hybrid Iterated Greedy framework integrating GRASP-based construction and VND local search is proposed for unrelated parallel machine scheduling with sequence-dependent setup times under a common-server constraint.
2. Three alternative decoder-based evaluation mechanisms are developed and systematically analyzed within the proposed hybrid search framework.
3. A two-step look-ahead decoder (TSLA) is proposed to improve setup-server coordination during the decoding process.

4. Extensive computational experiments are conducted to analyze the impact of decoder strategies and runtime settings on scheduling performance.
5. The proposed IG-HM framework is statistically compared against the TS-I algorithm from the literature under multiple runtime configurations.

The remainder of the chapter is organized as follows. Section 2 presents the proposed IG-HM framework and its algorithmic components. Section 3 reports the computational experiments and comparative analyses. Finally, Section 4 concludes the chapter and discusses future research directions.

2. MATERIALS AND METHODS

2.1. An Iterated Greedy–Based Hybrid Algorithm for Solving the R_M, $S_1/ST_{sd}/\sum w_j \times T_j$ Problem

This section presents the proposed hybrid Iterated Greedy algorithm, denoted as IG-HM, developed to solve the R_M, $S_1/ST_{sd}/\sum w_j \times T_j$ problem. IG-HM adopts a machine–job-based solution representation and integrates GRASP-based construction (Feo & Resende, 1995), IG perturbation (Ruiz & Stützle, 2007), and Variable Neighborhood Descent (VND) for solution improvement within a unified framework (Mladenović & Hansen, 1997). The algorithm parameters are summarized in Table 1. The decoder configuration (decConf) defines the policy used to evaluate schedules, while α controls the size of the Restricted Candidate List in the GRASP phase. The destruction rate ρ determines the proportion of jobs removed during each IG perturbation, T_acc governs the probabilistic acceptance of non-improving solutions in a manner similar to Simulated Annealing (Kirkpatrick et al., 1983), and τ_max specifies the maximum runtime of a single algorithm execution.

Table 1. Parameters used in the proposed IG-HM algorithm.

Parameter	Description
decConf	The decoder configuration used to calculate fitness value of a given schedule.
α	GRASP restricted candidate list (RCL) parameter.
ρ	Destruction rate in the Iterated Greedy (IG) procedure.
T_acc	Acceptance temperature, analogous to Simulated Annealing.
τ *max*	Maximum runtime allowed for a single IG execution.

Algorithm 1 summarizes the experimental procedure used in this study. For each instance in the dataset, the proposed hybrid algorithm is independently executed for NR replications. In each replication, a single run of the IG algorithm is performed and the resulting objective value and schedule are recorded. After completing all replications, per-instance summary statistics, including the minimum, maximum, and average solution values as well as the best schedule, are computed and used for performance comparison and analysis.

To reduce stochastic bias and ensure fair comparisons, all experiments are conducted under a common random numbers (CRN) framework (Law & Kelton, 2000). For each instance–replication pair (ins, rep), a deterministic random seed is generated according to Eq. (1).

$$seed_{ins,rep} = base\_seed + ins \times K + rep \quad (1)$$

Here, base_seed is a fixed integer used to initialize the pseudo-random number generator, K is a sufficiently large constant that separates seed ranges across instances (K≫NR), and NR denotes the number of replications. This construction assigns each instance a distinct block of seed values and ensures a one-to-one mapping between each instance–replication pair (ins, rep) and the corresponding $seed_{ins,rep}$. As a result, all algorithms are executed with identical random seeds for the same instance–replication pair, so that all stochastic decisions are driven by the same pseudo-random number stream. This CRN design reduces

variance in performance comparisons and improves fairness and reproducibility, allowing observed differences to be attributed to algorithmic mechanisms rather than stochastic effects.

Algorithm 2 outlines the main procedure of the IG-HM algorithm. An initial solution is first generated using the GRASP construction method (Feo & Resende, 1995) and immediately improved by a VND phase (Mladenović & Hansen, 1997), yielding the initial incumbent solution. At each iteration, a candidate solution is produced through an IG-style destruction–reconstruction process (Ruiz & Stützle, 2007), in which a subset of jobs is removed according to the destruction rate ρ and reinserted using a decoder-guided greedy strategy.

Algorithm 1. CRN-based experimental procedure.

For each instance file in Data **do**

  Read instance *ins*

  For *rep* ← 1 to *NR* **do**

    seed ← apply Eq. (1) to determine the seed

    rep_obj, rep_schedule ← **IG-HM (***ins, decConf, α, ρ, T_acc, τ_max, seed***)**

    Record *rep_obj, rep_schedule*

  End For

  Compute summary statistics (minimum solution value, maximum solution value, average solution value, best schedule) for instance *ins*

End For

Save per-instance summary and best schedule

The reconstructed solution is then further refined by VND. Improving solutions are accepted deterministically, while non-improving solutions may be accepted with a probability defined in Eq. (2).

$$p\_acc = e^{-\frac{\Delta}{T\_acc}} \qquad (2)$$

In Eq. (2), Δ denotes the deterioration in solution quality. The best solution is tracked throughout the search, and the algorithm returns this solution when the time limit τ_max is reached.

2.1. Initialization

In IG-HM, a machine–job-based representation is used, and the initial solution is generated through a GRASP-based construction procedure (Algorithm 3). Jobs are iteratively assigned to eligible machines using a randomized greedy rule. For each unscheduled job–machine pair, the marginal contribution to total weighted tardiness is evaluated, and a Restricted Candidate List (RCL) is formed accordingly. A pair is then randomly selected from the RCL and appended to the corresponding machine schedule. This process continues until all jobs are scheduled.

Algorithm 2. Overview of the proposed IG-HM algorithm.

Algorithm IG-HM (*ins, decConf, α, ρ, T_acc, τ_max, seed***)**

1. *t_start* ← Get the current CPU time

2. (*Schedule_cur*, *Fitness_cur*) ← **GRASP_Construct (***α, decConf***)** ▷ Algorithm 3

3. (*Schedule_cur*, *Fitness_cur*) ← **VND_FirstImprovement (***Schedule_cur*, *Fitness_cur*, *decConf***)** ▷ Algorithm 4

4. *Schedule_best* ← *Schedule_cur*

5. *Fitness_best* ← *Fitness_cur*

6. While (current CPU time − *t_start* < *τ_max***) do**

7. *Schedule_cand* ← *Schedule_cur*

8. *removedJobs* ← **IG_Destroy (***Schedule_cand*, *ρ***)** ▷ Algorithm 5

9. (*Schedule_cand*, *Fitness_cand*) ← **IG_Reconstruct (***Schedule_cand*, *decConf*, *removedJobs***)** ▷ Algorithm 6

10. (*Schedule_ cand*, *Fitness_ cand*) ← **VND_FirstImprovement (***Schedule_cand*, *Fitness_cand*, *decConf***)**

11. If *Fitness_cand* ≤ *Fitness_cur* **then**

12. *Schedule_cur* ← *Schedule_ cand*

```
13.     Fitness_cur ← Fitness_cand
14.   Else
15.     Δ ← Fitness_cand - Fitness_cur
16.     p_acc ← exp (−Δ / T_acc)
17.     If rand (0,1) < p_acc then
18.       Schedule_cur ← Schedule_ cand
19.       Fitness_cur ← Fitness_cand
20.     End If
21.   End If
22.   If Fitness_cur < Fitness_best then
23.     Schedule_best ← Schedule_cur
24.     Fitness_best ← Fitness_cur
25.   End If
26. End While
27. Return (Schedule_best, Fitness_best)
```

2.2. Local Search

Local improvement is carried out using a Variable Neighborhood Descent (VND) strategy (Mladenović & Hansen, 1997) with a first-improvement rule, as outlined in Algorithm 4. Neighborhoods are explored sequentially, and the search restarts from the first neighborhood whenever an improvement is found. The procedure terminates when no improving move exists in any neighborhood. Five neighborhood operators are embedded in the VND framework (Algorithms 4.1–4.5). Intra-machine neighborhoods include IntraInsert, IntraSwap, and Intra2Opt, which perform insertion, swap, and subsequence reversal moves within the same machine sequence. These operators correspond to classical permutation-based moves commonly used in local search (Feo & Resende, 1995; Ruiz & Stützle, 2007), with the 2-opt move adapted from Lin and Kernighan (1973). Inter-machine neighborhoods consist of InterRelocate and InterSwap, which relocate or exchange jobs between machines, allowing

simultaneous changes in assignment and sequencing decisions. All candidate solutions are evaluated using the decoder specified by the configuration *decConf*.

Algorithm 3. Initial solution construction procedure.

Algorithm GRASP_Construct (*α, decConf***)**
1. For each machine $m \leftarrow 1$ to M **do**
2. $Schedule[m] \leftarrow []$ ▷ schedule for the machine m
3. $t[m] \leftarrow 0$ ▷ completion time on machine m
4. $last[m] \leftarrow -1$ ▷ last job processed on machine m
5. End For
6. $CST \leftarrow 0$ ▷ global server time
7. $used\_jobs \leftarrow \emptyset$
8. While ($|used\_jobs| < J$**) do**
//Build candidate contributions
9. $D \leftarrow []$ ▷ list of tuples (m, j, d (*contribution*))
10. For each *unscheduled* job j **do**
11. For each machine m that can process j **do**
12. $A \leftarrow$ **max (**$t[m]$, CST**)**
13. $setup \leftarrow (last[m] == -1 \ ? \ 0 : SD_{last[m],j})$
14. $C \leftarrow A + setup + PR_{j,m}$
15. $d \leftarrow w_j \times$ **max (**$0, C - DD_j$**)** ▷ weighted tardiness contribution
16. **append (**m, j, d**)** to D
17. End For
18. End For
//Build the Restricted Candidate List (*RCL*)
19. $minD \leftarrow$ **min (**d in D**)**
20. maxD ← **max** (d in D)
21. $thr \leftarrow minD + \alpha \times (maxD - minD)$
22. $RCL \leftarrow \{(m, j)$ in $D \mid d \leq thr\}$
//Randomly select from *RCL* and update sequences
23. pick random (*mSel, jSel***)** from *RCL*
24. append (*jSel* to *Schedule*[*mSel*]**)**
25. $A \leftarrow$ **max (**t[*mSel*], *CST***)**
26. $setup \leftarrow$ (last[*mSel*] == -1) ? 0 : $SD_{last[mSel], jSel}$
27. $t[mSel] \leftarrow A + setup + PR_{jSel, mSel}$ ▷ machine completion time
28. $last[mSel] \leftarrow jSel$
29. $CST \leftarrow A + setup$ ▷ server completion time
30. add *jSel* to *used_jobs*
31. End While
32. *Schedule_Fitness* ← **Decode (***Schedule, decConf***)**
33. Return (*Schedule, Schedule_Fitness***)**

Algorithm 4. Local search with first-improvement strategy.

Algorithm VND_FirstImprovement(*Schedule, Fitness, decConf***)**

1. *Nlist* ← [IntraInsert, IntraSwap, Intra2Opt, InterRelocate, InterSwap]
2. *improved* ← true
3. **While** *improved* **do**
4. *improved* ← false
5. **For** *k* ← 1 to |*Nlist*| **do**
6. (*Schedule, Fitness, improved*) ← *Nlist*[*k*](*Schedule, Fitness, decConf*)
7. **If** *improved* **then**
8. Break // first improvement → restart from first neighborhood
9. **End If**
10. **End For**
11. End While
12. Return (*Schedule, Fitness*)

Algorithm 4.1. Intra-Machine Insertion Move (IntraInsert).

Procedure IntraInsert (*Schedule, Fitness, decConf***)**

1. **For** each machine *m* in machines **do**
2. **For** each pair of positions (*a, b*) in *Schedule*[*m*] with $a \neq b$ **do**
3. *S'* ← *Schedule*
4. remove the job at position *a* in *S'*[*m*], then insert this job at position *b* in *S'*[*m*]
5. *F'* ← **Decode (***S', decConf***)**
6. **If** *F'* < *Fitness* **then**
7. **Return (***S', F',* true**)**
8. **End If**
9. **End For**
10. End For
11. Return (*Schedule, Fitness,* false)

Algorithm 4.2. Intra-Machine Swap Move (IntraSwap).

Procedure IntraSwap (*Schedule, Fitness, decConf***)**

1. For each machine *m* in machines **do**
2. **For** each pair of positions (*a*, *b*) in *Schedule*[*m*] with *a* < *b* **do**
3. *S'* ← *Schedule*
4. swap the jobs at positions *a* and *b* in *S'*[*m*]
5. *F'* ← **Decode (***S', decConf***)**
6. **If** *F'* < *Fitness* **then**
7. **Return (***S', F',* true**)**
8. **End If**
9. **End For**
10. End For
11. Return (*Schedule, Fitness,* false**)**

Algorithm 4.3. Intra-Machine 2-opt Reversal Move (Intra2Opt).

Procedure Intra2Opt (*Schedule, Fitness, decConf***)**

1. For each machine *m* in machines **do**
2. **For** each pair of indices (*a*, *b*) in *Schedule*[*m*] with *a* < *b* **do**
3. *S'* ← *Schedule*
4. reverse the subsequence of *S'*[*m*] between positions *a* and *b* (inclusive)
5. *F'* ← Decode (*S', decConf*)
6. **If** *F'* < *Fitness* **then**
7. **Return (***S', F',* true**)**
8. **End If**
9. **End For**
10. End For
11. Return (*Schedule, Fitness,* false**)**

Algorithm 4.4. Inter-Machine Relocation Move (InterRelocate).

Procedure InterRelocate (*Schedule, Fitness, decConf***)**

1. For each ordered pair of machines (m, k) with $m \neq k$ **do**
2. For each job position a in *Schedule*[m] **do**
3. For each insertion position *pos* in a selected subset of positions of *Schedule*[k] **do**
4. $S' \leftarrow$ *Schedule*
5. remove the job at position a in $S'[m]$
6. insert this job into $S'[k]$ at position *pos*
7. $F' \leftarrow$ **Decode (***S', decConf***)**
8. If $F' <$ *Fitness* **then**
9. Return (*S', F',* true**)**
10. End If
11. End For
12. End For
13. End For
14. Return (*Schedule, Fitness,* false**)**

Algorithm 4.5. Inter-Machine Swap Move (InterSwap).

Procedure InterSwap (*Schedule, Fitness, decConf***)**

1. For each unordered pair of machines (m, k) with $m < k$ **do**
2. For each job position a in *Schedule*[m] **do**
3. For each job position b in *Schedule*[k] **do**
4. $S' \leftarrow$ *Schedule*
5. swap the jobs at positions a and b between $S'[m]$ and $S'[k]$
6. $F' \leftarrow$ **Decode (***S', decConf***)**
7. If $F' <$ *Fitness* **then**
8. Return (*S', F',* true**)**
9. End If
10. End For
11. End For
12. End For
13. Return (*Schedule, Fitness,* false**)**

2.3. Solution Destruction and Reconstruction

The perturbation mechanism of IG-HM follows the classical destroy–reconstruct principle of Iterated Greedy algorithms (Ruiz & Stützle, 2007). At each iteration, a subset of jobs is removed from the current schedule to promote diversification, and the resulting partial solution is rebuilt using a greedy, decoder-guided reconstruction strategy. The destruction and reconstruction procedures are detailed in Algorithms 5 and 6. In the destruction phase (Algorithm 5), a fraction ρ of jobs is randomly removed from each machine sequence, where the number of removed jobs is proportional to the sequence length. All removed jobs are collected in a set and temporarily excluded from the schedule. In the reconstruction phase (Algorithm 6), jobs are reinserted one by one by evaluating all feasible insertion positions on eligible machines using the selected decoder configuration (decConf). Each job is placed in the position yielding the lowest objective value; if no improving insertion exists, it is assigned to a randomly selected eligible machine. After all jobs are reinserted, the schedule is decoded and its objective value is computed.

Algorithm 5. Destruction phase of the Hybrid IG algorithm.

Algorithm **IG_Destroy (***Schedule_cand, ρ***)**

1. *removedJobs* ← ∅

2. **For** each machine *m* ← *M* **do**

3. *len* ← length of *Schedule_cand*[*m*]

4. *remCount* ← ⌊*ρ* × *len*⌋

5. **Repeat** *remCount* **times**

6. **If** *Schedule_cand*[*m*] is empty **then**

7. **break**

8. **End If**

9. *pos* ← random integer in [1, length (*Schedule_cand*[*m*])]

10. *job* ← *Schedule_cand*[*m*][*pos*]

11. **append** job to *removedJobs*

12. **remove** job from *Schedule_cand*[*m*] at position *pos*

13. **End Repeat**

14. End for

15. Return *removedJobs*

Algorithm 6. Decoder-guided reconstruction procedure.

Algorithm IG_Reconstruct (*Schedule_cand, decConf, removedJobs***)**

1. **For** job $j \in$ *removedJobs* **do**

2. *bestF* ← +∞; *bestMach* ← −1; *bestPos* ← −1

3. **For** each machine $m \leftarrow M$ **do**

4. **If** job *j* is not eligible on machine *m* ($B_{j,m} = 0$) **then**

5. **continue**

6. **End If**

7. **For** *pos* ← 1 to length(*Schedule_cand*[*m*]-1) **do**

8. *S'* ← *Schedule_cand*

9. Place job *j* into *Schedule_cand*[*m*] at position *pos*

10. *F'* ← **Decode (***S'*, *decConf***)**

11. **If** *F'* < *bestF* **then**

12. *bestF* ← *F'*; *bestMach* ← *m*; *bestPos* ← *pos*

13. **End If**

14. **End For**

15. **End For**

16. **If** *bestMach* = −1 **then**

17. choose any machine *m* with $B_{j,m}$ =1 and insert *j* at the end of *Schedule_cand*[*m*]

18. **Else**

19. Place job *j* into *Schedule_cand*[*bestMach*] at position *bestPos*

20. **End If**

21. **End For**

22. *Fitness* ← **Decode (***Schedule_cand, decConf***)**

23. **Return (***Schedule_cand, Fitness***)**

2.4. Solution Evaluation

The solution evaluation strategy is a key component of the proposed algorithm, as it transforms a machine-level schedule into a time-feasible solution under the common-server constraint. It determines the server setup sequence and computes job start times, completion times, and tardiness through a step-by-step decoding procedure that explicitly coordinates machine operations with the common server. Machine selection at each step is governed by a predefined rule. The overall decoding process is summarized in Algorithm 7, with alternative machine-selection strategies described in Algorithms 7.1–7.3. Table 2 lists the main state variables maintained during decoding, including machine availability, server completion times, job completion times, and marginal tardiness contributions. At each iteration, the function pick_machine selects a machine with unscheduled jobs, and the earliest feasible start time of the corresponding job is computed as the maximum of the machine completion time and the current server time. After applying the sequence-dependent setup, processing begins and all time-related variables are updated. This procedure continues until all jobs are scheduled, after which total weighted tardiness is computed as the fitness value. Three alternative decoding strategies are implemented via the pick_machine function.

Table 2. Notation for the decoder-based evaluation procedures.

Parameter	Description
$A[j]$	Machine-available time before server for job j
$L[j]$	Server completion time of job j
$C[j]$	Completion time for job j
$T[j]$	Tardiness value for job j
$M[j]$	Latest assigned job j to the related machine
$pos[j]$	Position of job j on machine $M[j]$
m_sel	Index of machine whose next job will be processed by the server
$\Delta[m]$	Marginal increase in total weighted tardiness if the next job on machine m is processed next by the server

Algorithm 7. The main logic of the decoding procedure.

Algorithm Decode (*Schedule, decConf***)**

For $j \leftarrow 1$ to J **do**
  $A[j] \leftarrow 0$; $L[j] \leftarrow 0$; $C[j] \leftarrow 0$; $T[j] \leftarrow 0$; $M[j] \leftarrow -1$; $pos[j] \leftarrow -1$
End For
For $m \leftarrow 1$ to M **do**
  $idx[m] \leftarrow 1$ ▷ next unscheduled position in $Schedule[m][\cdot]$
  $last[m] \leftarrow -1$ ▷ last job processed on machine m
  $t[m] \leftarrow 0$ ▷ completion time on machine m
End For
$CST \leftarrow 0$ ▷ global server time
$done \leftarrow 0$ ▷ number of decoded jobs
While $done < J$ **do**
  $m\_sel \leftarrow$ **pick_machine (***Schedule, idx, last, t, CST, decConf***)**
  If $m\_sel == -1$ **then**
    break ▷ safety; should not occur for a feasible schedule
  End If
  $j \leftarrow Schedule[m\_sel][idx[m\_sel]]$; $h \leftarrow last[m\_sel]$
  $A[j] \leftarrow$ **max (** $t[m\_sel]$, CST **)**
  $setup \leftarrow (h == -1\ ?\ 0 : SD_{h,j})$
  $L[j] \leftarrow A[j] + setup$
  $C[j] \leftarrow L[j] + PR_{j,m\_sel}$
  $T[j] \leftarrow$ **max (**0, $C[j] - DD_j$**)**
  $M[j] \leftarrow m\_sel$
  $pos[j] \leftarrow idx[m\_sel]$
  $t[m\_sel] \leftarrow C[j]$ ▷ update machine completion time
  $CST \leftarrow L[j]$ ▷ update global server time
  $last[m\_sel] \leftarrow j$
  $idx[m\_sel] \leftarrow idx[m\_sel] + 1$
  $done \leftarrow done + 1$
End While
$Fitness \leftarrow 0$
For $j \leftarrow 1$ to J do
  $Fitness \leftarrow Fitness + w_j \times T[j]$
End For
Return (*Fitness***)**

The minimum completion time (MCT) rule of Bektur and Saraç (2019) is implemented as one of the decoding procedures. It selects the machine whose next job yields the smallest completion time, favoring early machine availability; ties are broken using the ratio of job weight to processing-plus-setup time to prioritize more urgent jobs. The procedure is detailed in Algorithm 7.1. The marginal-tardiness procedure (MinΔ) evaluates the marginal increase in total weighted tardiness associated with scheduling each candidate job next and selects the machine with the smallest penalty. This procedure is described in Algorithms 7.1 and 7.2. The two-step look-ahead procedure (TSLA) extends this approach by considering two consecutive decisions. Candidate machines are first ranked by their immediate marginal penalties, after which the two best candidates are evaluated under both possible execution orders using a short simulation. The machine yielding the smaller combined tardiness increase is selected. TSLA is presented in Algorithms 7.1 and 7.3.

Algorithm 7.1. Machine selection policy (three decoder rules).

Algorithm pick_machine (*Schedule, idx, last, t, CST, decConf***)**

1. *M_active* ← list of machines *m* such that *idx*[*m*] ≤ |*Schedule*[*m*]| ▷ Identify all machines that still contain at least one unscheduled job

2. **Switch** *decConf* **do**

3. **Case** *MCT*:

4. *bestM* ← −1; *bestT* ← +∞; *bestScore* ← −∞

5. **For** each *m* in *M_active* **do**

6. *j* ← *Schedule*[*m*][*idx*[*m*]]; *h* ← *last*[*m*]

7. *setup* ← (*h* == −1 ? 0 : $SD_{h,j}$)

8. P[*m*] ← $PR_{j,m}$ + *setup*

9. *score* ← w_j / **max (**1, P[*m*]**)**

10. **If** (*t*[*m*] < *bestT*) or (*t*[*m*] == *bestT* and *score* > *bestScore***) then**

11. *bestT* ← *t*[*m*]; *bestScore* ← *score; bestM* ← *m*

12. **End If**

13. **End For**

14. **Return** *bestM*

15. **Case** MinΔ:
16. *bestM* ← −1; *bestΔ* ← +∞
17. **For** each *m* in *M_active* **do**
18. *Δ*[*m*] ← **Delta (***m, Schedule, idx, last, t, CST***)**
19. **If** *Δ*[*m*] < *bestΔ* **then** *bestΔ* ← *Δ*[*m*]; *bestM* ← *m* **End If**
20. **End For**
21. **Return** *bestM*
22. **Case** TSLA:
//Step 1: build candidate list with single-step deltas
23. *Cand* ← ∅ ▷ list of pairs (*m*, *Δ*[*m*])
24. **For** each *m* in *M_active* **do**
25. *Δ*[*m*] ← **Delta (***m, Schedule, idx, last, t, CST***)**
26. **append** (*m*, *Δ*[*m*]) to *Cand*
27. **End For**
28. **If** length(*Cand*) == 1 **then**
29. *bestM* ← *Cand*[1] ▷ only one candidate
30. **Return** *bestM*
31. **End If**
//Step 2: select two machines with the smallest *Δ*[*m*] values
32. Let *m_A* and *m_B* be the two machines in *Cand* with the smallest *Δ*[*m*]
33. *cost_AB* ← **Simulate_two (***m_A, m_B, Schedule, idx, last, t, CST***)**
35. *cost_BA* ← **Simulate_two (***m_B, m_A, Schedule, idx, last, t, CST***)**
35. **If** *cost_AB* ≤ *cost_BA* **then Return** *m_A* **Else Return** *m_B* **End If**
36. **Switch End**

Algorithm 7.2. Single-step marginal contribution Δ[m].

Algorithm Delta (*m, Schedule, idx, last, t, CST***)**
1. *j* ← *Schedule*[*m*][*idx*[*m*]]
2. *h* ← *last*[*m*]
3. *A*[*j*] ← **max (***t*[*m*], *CST***)**
4. *setup* ← (*h* == −1 ? 0 : $SD_{h,j}$)
5. *C*[*j*] ← *A*[*j*] + *setup* + $PR_{j,m}$
6. *tard* ← **max (**0, *C*[*j*] − DD_j**)**
7. *Δ* ← w_j × *tard*
8. **Return** *Δ*

Algorithm 7.3. Two-step look-ahead procedure.

Algorithm Simulate_two (*m_first, m_second, Schedule, idx, last, t, CST***)**

//First job: machine *m_first*

1. $jA \leftarrow Schedule[m\_first][idx[m\_first]]$

2. $hA \leftarrow last[m\_first]$

3. $AA \leftarrow \max(t[m\_first], CST)$

4. $sA \leftarrow (hA == -1 \ ? \ 0 : SD_{hA, jA})$

5. $CA \leftarrow AA + sA + PR_{jA, m\_first}$

6. $tardA \leftarrow \mathbf{max}(0, CA - DD_{jA})$

7. $dA \leftarrow w_{jA} \times tardA$

8. $CST2 \leftarrow AA + sA$ ▸ server time after completing job *jA*

▸ Second job: machine *m_second*

9. $jB \leftarrow Schedule[m\_second][idx[m\_second]]$

10. $hB \leftarrow last[m\_second]$

11. $AB \leftarrow \mathbf{max}(t[m\_second], CST2)$

12. $sB \leftarrow (hB == -1 \ ? \ 0 : SD_{hB, jB})$

13. $CB \leftarrow AB + sB + PR_{jB, m\_second}$

14. $tardB \leftarrow \mathbf{max}(0, CB - DD_{jB})$

15. $dB \leftarrow w_{jB} \times tardB$

16. $cost \leftarrow dA + dB$

17. Return *cost*

2.5. Time Complexity Analysis of the Algorithm

The computational complexity of the proposed IG-HM algorithm can be summarized as follows. Decoding a complete schedule requires O(N×M) time. The GRASP construction phase performs N placement steps, each evaluating all eligible job–machine pairs, resulting in $O(N^2 \times M)$ time. During the IG phase, approximately $\rho \times N$ jobs are reinserted, and all feasible insertion positions are evaluated; since each insertion triggers a full decoding, the reconstruction phase requires $O(N^3 \times M)$ time. The VND local search explores $O(N^2)$ neighborhood moves, each

evaluated through decoding, yielding $O(N^3 \times M)$ time per VND pass. Over *iter* destroy–reconstruct–VND iterations, the overall complexity of one IG run is $O(iter \times N^3 \times M)$. Although this worst-case bound is polynomial in N and M, the cubic dependence on N implies a rapidly increasing computational burden for large-scale instances.

2.6. Numerical Example for the Evaluation Procedures

This section presents a numerical example to illustrate the behavior of the decoder-based evaluation procedures and to compare the implemented decoding strategies. Table 3 reports the parameters of the illustrative instance, including processing times, machine eligibility, job weights, due dates, and sequence-dependent setup times. The example is used to analyze how different decoders construct the global server sequence and how these choices affect total weighted tardiness.

Table 3. Parameters of the illustrative example.

$PR_{j,m}$		$B_{j,m}$		w_j	DD_j	$SD_{i,j}$						
M1	**M2**	**M1**	**M2**			**jobs**	**1**	**2**	**3**	**4**	**5**	**6**
4	3	1	1	4	16	**1**	0	7	9	1	5	5
10	5	1	1	3	23	**2**	4	0	8	9	7	7
9	4	1	1	1	26	**3**	7	3	0	7	10	0
10	5	1	1	1	14	**4**	8	4	6	0	2	10
8	2	1	1	1	18	**5**	9	4	10	5	0	0
7	10	1	1	1	23	**6**	9	6	6	8	8	0

To ensure a fair comparison, the machine-level job sequences are fixed as (5, 1, 6) on Machine 1 and (2, 3, 4) on Machine 2 for all decoders. Consequently, any performance differences arise solely from the construction of the global server sequence, which is generated dynamically according to each decoder's selection rule. At each decoding step, the next job on each machine is evaluated, and the selected job is appended to the server sequence while system times are updated. Under the MCT-based decoder, the job with the smallest completion time is selected. Table 4 summarizes the corresponding decoding decisions and the evolution of machine and server times.

Table 4. *MCT* procedure-based decoder step-by-step evaluation.

Step	Candidates (*m*, *j*)	*A*[*j*]	*L*[*j*]	*t*[*m*]	Selected Job	Updated (*t*[1], *t*[2], *CST*)	T[*j*]
1	(1,5), (2,2)	0, 0	0, 0	8, 5	5 (M1)	(8, 0, 0)	0
2	(1,1), (2,2)	8, 0	17, 0	21, 5	2 (M2)	(8, 5, 0)	0
3	(1,1), (2,3)	8, 5	17, 13	21, 17	3 (M2)	(8, 17, 13)	0
4	(1,1), (2,4)	13, 17	22, 24	26, 29	1 (M1)	(26, 17, 22)	10
5	(1,6), (2,4)	26, 22	31, 29	38, 34	4 (M2)	(26, 34, 29)	20
6	(1,6)	29	34	41	6 (M1)	(41, 34, 34)	18

Using the MCT-based decoder, the resulting server sequence is [5, 2, 3, 1, 4, 6], yielding a total weighted tardiness of 78. The corresponding schedule is shown in Figure 1.

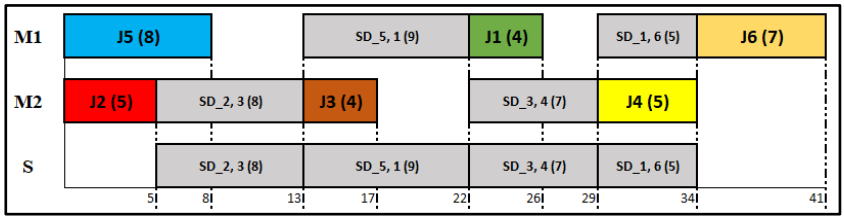


Figure 1. Schedule for the *MCT* procedure-based decoder.

The MinΔ-based decoder selects the job with the smallest marginal weighted tardiness, $\Delta = w_j \times T_j$, thereby accounting for job priorities and due dates in a reactive manner. The corresponding decoding decisions are summarized in Table 5.

Table 5. MinΔ procedure-based decoder step-by-step evaluation.

Step	Candidates (*m*, *j*)	Δ[*m* = 1]	Δ[*m* = 2]	Selected Job	Updated (*t*[1], *t*[2], *CST*)	T[j]
1	(1,5), (2,2)	0	0	5 (M1)	(8, 0, 0)	0
2	(1,1), (2,2)	20	0	2 (M2)	(8, 5, 0)	0
3	(1,1), (2,3)	20	0	3 (M2)	(8, 17, 13)	0
4	(1,1), (2,4)	40	15	4 (M2)	(8, 29, 24)	15
5	(1,1 only)	84	-	1 (M1)	(37, 29, 33)	21
6	(1,6 only)	26	-	6 (M1)	(49, 29, 42)	26

Using the MinΔ-based decoder, the resulting server sequence is [5, 2, 3, 4, 1, 6], yielding a total weighted tardiness of 125. The corresponding schedule is shown in Figure 2.

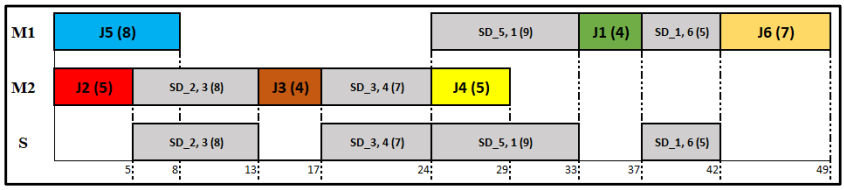


Figure 2. Schedule for the MinΔ procedure-based decoder.

The TSLA-based decoder applies a two-step look-ahead mechanism that compares alternative serving orders, enabling it to avoid short-sighted decisions that may be costly in later stages. The corresponding evaluations and decisions are reported in Table 6.

Table 6. TSLA procedure-based decoder step-by-step evaluation.

Step	Candidates (m, j)	Δ[1]	Δ[2]	Δ[AB]	Δ[BA]	Selected Job	Updated (t[1], t[2], CST)	T[j]
1	(1, 5), (2, 2)	0	0	0	0	5 (M1)	(8, 0, 0)	0
2	(1, 1), (2, 2)	20	0	20	20	2 (M2)	(8, 5, 0)	0
3	(1, 1), (2, 3)	20	0	23	40	1 (M1)	(21, 5, 17)	5
4	(1, 6), (2, 3)	10	3	17	22	3 (M2)	(21, 29, 25)	3
5	(1, 6), (2, 4)	14	27	42	52	6 (M1)	(37, 29, 30)	14
6	(2, 4)	-	28	-	-	4 (M2)	(37, 42, 37)	28

Using the TSLA-based decoder, the resulting server sequence is [5, 2, 1, 3, 6, 4], yielding a total weighted tardiness of 65. The corresponding schedule is shown in Figure 3.

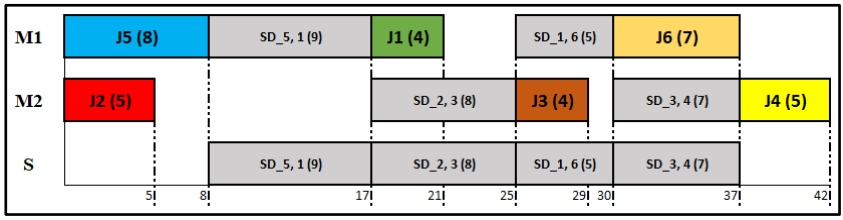


Figure 3. Schedule for the TSLA procedure-based decoder.

3. RESULTS AND DISCUSSION

The proposed methods are evaluated on a large set of benchmark instances under several computational settings. The experimental study focuses on the heuristic effectiveness. In particular, the analysis includes: (i) generation of benchmark instances, (ii) parameter calibration of the proposed hybrid IG algorithm, and (iii) assessment of the proposed approximation strategies under alternative evaluation procedures.

Among the methods reported in Bektur and Saraç (2019), the TS-I algorithm demonstrated the strongest overall performance. Therefore, TS-I is employed as the reference metaheuristic in this study. To avoid implementation-related bias, all compared algorithms were developed in the same software environment and tested under identical computational conditions. The proposed IG-based procedures are investigated using four different solution-evaluation policies. The implementations were carried out in Visual C++, and Gurobi 9.0.3 was used for solving the mathematical models. All experiments were executed on a workstation with an Intel Core i7-8565U CPU and 16 GB memory.

3.1. Generation of Test Problems

A new benchmark data set was constructed since the original data set introduced by Bektur and Saraç (2019) is not publicly accessible. The generated instances follow the same structural assumptions used in the original study so that comparable difficulty levels can be preserved. The experimental factors were determined by varying the number of machines and jobs. Five different machine settings were examined: 2, 3, 5, 7, and 10 machines. For each setting, two job-size categories were considered, where the total number of jobs equals either five or ten times the number of machines. Machine-dependent processing times were sampled from the interval [10,100] using a uniform distribution. Job priority weights were also uniformly generated between 1 and 10. Three different setup-time environments were created to represent varying levels of sequence dependency. These environments correspond to low, medium, and high setup variability, generated from the intervals [5,25], [5,50], and [25,50], respectively. Machine eligibility was examined under two alternative scenarios. In the unrestricted case, every job can be processed on all machines. In the restricted case, eligibility is determined probabilistically with an

assignment probability of 0.7. Due dates were produced using the same due-date generation mechanism adopted by Bektur and Saraç (2019), where the tardiness factors are set to 0.5 and 0.8 and the randomness parameter is fixed at 0.2. Overall, the combination of all parameter levels yields 120 experimental settings. By generating two random replications for each setting, the final benchmark set contains 240 instances.

3.2. Parameter calibration of the proposed algorithm

The calibration of the hybrid IG method (IG-HM) focuses on three key parameters: the GRASP restricted candidate list parameter α, the destruction ratio ρ, and the acceptance temperature T_acc. Parameter tuning was conducted using a two-stage experimental protocol on a representative subset of benchmark instances. This subset includes four machine settings (M ∈ {3, 5, 7, 10}) and two workload levels (J = 5×M and J = 10×M), yielding eight instance classes. This design captures algorithmic behavior across different problem scales and workload intensities.

In the first phase, a screening experiment based on a fractional factorial design was conducted. The parameters α ∈ {0.1, 0.6}, ρ ∈ {0.2, 0.6}, and T_acc ∈ {5, 20} were examined. All configurations were evaluated under a common random numbers (CRN) scheme with NR=10 replications to reduce stochastic variability. To ensure comparable computational effort, the runtime limit was scaled as τ_max = M×J×10 milliseconds. Algorithmic performance was measured using the relative percentage deviation (RPD) defined in Eq. (3).

$$RPD_{i,r} = (Some_{sol} - Best_{sol})/Best_{sol} \times 100 \qquad (3)$$

where $Sol_{i,r}$ denotes the objective value obtained in replication r for instance i, and Sol_i^* represents the best-known solution for that instance. The overall performance of each

parameter configuration was evaluated by the average RPD (ARDP) computed as in Eq. (4).

$$ARPD = \frac{1}{|\,I\,| \times \mathrm{NR}} \sum_{i \in I} \sum_{r=1}^{NR} RPD_{i,r} \qquad (4)$$

where |I| is the training instance set (|I|=12) and NR = 10 is the number of replications. The results of this screening experiment are reported in Table 7, while the statistical comparison of parameter levels is summarized in Table 8.

Table 7. Screening design results (Phase–1).

α	*ρ*	*T_acc*	*ARPD* (%)	Std. Dev.
0.1	0.2	5	21.5	3.9
0.1	0.6	5	24.2	4.5
0.6	0.2	20	13.6	2.4
0.6	0.6	20	10.8	1.8

Table 8. ANOVA results for screening phase.

Factor	F-Value	p-Value
α	12.8	0.002
ρ	9.4	0.008
T_acc	15.2	0.001

As shown in Table 7, configurations with small α values and zero acceptance temperature resulted in clearly inferior performance, whereas moderate values of α and ρ together with a positive T_acc led to substantial improvements in solution quality. In order to statistically validate these findings, a one-way analysis of variance (ANOVA) test was applied to the screening

results, followed by Tukey's honestly significant difference (HSD) procedure. The ANOVA results in Table 8 indicate that all three parameters have a statistically significant effect on solution quality ($p < 0.01$), while the post-hoc analysis confirms that intermediate parameter settings significantly outperform extreme configurations.

In the second stage, a finer grid search was carried out in the neighbourhood of the most promising configurations, testing $\alpha \in \{0.3, 0.4, 0.5\}$, $\rho \in \{0.3, 0.4, 0.5\}$, and $T\_acc \in \{5, 10, 15\}$ using the same training instances and experimental protocol. The aggregated results of this phase are presented in Table 9. Among all tested combinations, the parameter setting $(\alpha^*, \rho^*, T\_acc^*) = (0.4, 0.4, 10)$ achieved the lowest ARPD and exhibited the smallest standard deviation across instances, indicating both high solution quality and robust performance. Consequently, this configuration was selected as the final parameter setting and adopted in all subsequent computational experiments.

Table 9. Fine tuning results (Phase–2).

α	***ρ***	***T_acc***	***ARPD***	**Std. Dev.**
0.3	0.3	5	9.9	1.9
0.4	0.4	10	**7.8**	**1.3**
0.5	0.5	15	10.6	2.1

Finally, Figure 4 illustrates the superiority of the selected parameter configuration over alternative settings in terms of ARPD.

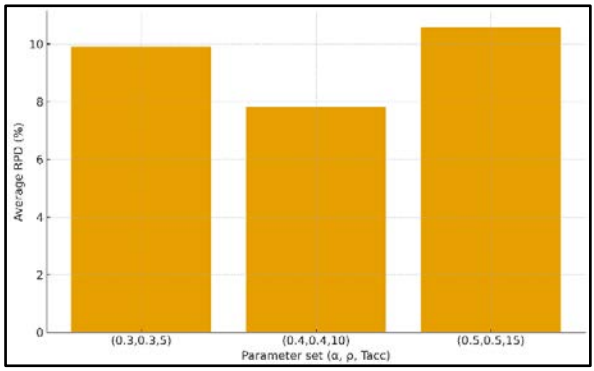


Figure 4. Final validation plot comparing the selected parameter configuration with alternative settings in terms of ARPD.

3.3. Performance Comparison of Approximation Methods

This section evaluates the proposed approximation-based solution methods under different runtime limits and decoder configurations. The experimental analysis is conducted in three stages. First, the performance of the proposed IG-HM algorithm is compared with the TS-I method of Bektur and Saraç (2019). Second, the impact of alternative decoding strategies on solution quality is examined.

3.3.1. Performance Comparison of IG-HM with TS-I

Since TS-I was originally implemented with the MCT decoding strategy in Bektur and Saraç (2019), the same decoder is used here to ensure methodological consistency in the first experimental phase. Three runtime budgets are considered, with each run limited to $\tau\_max = M \times J \times 10$ milliseconds for $q \in \{1, 2, 3\}$. Each algorithm is tested on the full benchmark set of 240 instances, with 10 independent runs per instance, resulting in $3 \times 240 \times 10 = 7.200$ times executions per algorithm. Table 10 reports the results for the smallest runtime setting (q=1). Despite the tight time limit, IG-HM outperforms TS-I across nearly all problem sizes. TS-I identifies 81 best solutions, whereas IG-HM

identifies 195, indicating a substantially higher frequency of superior solutions. The average ARPD of IG-HM is 6.986, compared to 14.648 for TS-I, demonstrating that IG-HM maintains good solution quality even under short runtimes.

For small instances (M=2), both algorithms show comparable behavior. As problem size increases (M≥5), the performance of TS-I deteriorates markedly, with average ARPD values often exceeding 20%, while IG-HM consistently remains below 10%. For the largest instances (M=7 and M=10), IG-HM consistently delivers better solutions than TS-I, confirming its robustness on large-scale problems.

Table 10. The average ARPD and RBS with q = 1.

(M, J)	TS-I with *MCT decoder*		IG-HM with *MCT decoder*	
	RBS	**Avg. ARDP**	**RBS**	**Avg. ARDP**
(2, 10)	24	2.003	24	1.593
(2, 20)	16	7.871	11	8.391
Average	0.833 (40)	4.937	0.729 (35)	4.992
(3, 15)	24	2.941	9	6.931
(3, 30)	2	20.107	22	7.234
Average	0.541 (26)	11.524	0.645 (31)	7.083
(5, 25)	8	13.420	16	9.158
(5, 50)	1	35.307	23	8.947
Average	0.187 (9)	24.364	0.812 (39)	9.052
(7, 35)	2	16.362	22	7.633
(7, 70)	1	24.171	23	7.908
Average	0.062 (3)	20.267	0.937 (45)	7.771
(10, 50)	0	13.758	24	6.002
(10, 100)	3	10.537	21	6.060
Average	0.062 (3)	12.147	0.937 (45)	6.031
All Average	0.337 (81)	14.648	0.812 (195)	6.986

Table 11 reports the performance of TS-I and IG-HM under the medium runtime setting (q=2). The results show a clear advantage of IG-HM across all problem classes. While TS-I remains competitive for the smallest instances (M=2), its performance rapidly deteriorates as problem size increases. In contrast, IG-HM maintains stable solution quality for all configurations. The difference is especially pronounced for medium and large instances. For M≥5, the average ARPD of TS-I exceeds 18%, whereas IG-HM consistently keeps this value below 0.08%. In addition, IG-HM attains 191 best solutions compared to 91 achieved by TS-I, indicating a substantially higher frequency of superior solutions.

Table 12 reports the results for q=3. Although TS-I benefits from the longer runtime, its performance remains clearly inferior to that of IG-HM. Under this setting, IG-HM achieves an average ARPD of 5.273 and identifies 202 best solutions, whereas TS-I records an average ARPD of 11.427 with only 72 best solutions. These results indicate that IG-HM is able to exploit additional runtime more effectively and deliver substantially higher-quality solutions. The effect of problem size on average ARPD of TS-I and IG-HM algorithms and the effect of problem size on RBS of TS-I and IG-HM algorithms are given in Figure 5 and 6, respectively.

Table 11. The average ARPD and RBS with q = 2.

(M, J)	TS-I with *MCT decoder*		IG-HM with *MCT decoder*	
	RBS	**Avg. ARDP**	**RBS**	**Avg. ARDP**
(2, 10)	24	1.879	23	0.015
(2, 20)	20	4.412	10	0.073
Average	0.916 (44)	3.145	0.687 (33)	0.044
(3, 15)	22	1.449	14	0.045
(3, 30)	6	14.388	18	0.073
Average	0.583 (28)	7.919	0.666 (32)	0.059
(5, 25)	11	8.591	14	0.074
(5, 50)	1	31.400	23	0.084
Average	0.250 (12)	19.995	0.770 (37)	0.079
(7, 35)	3	12.409	21	0.070
(7, 70)	1	25.016	23	0.077
Average	0.083 (4)	18.713	0.916 (44)	0.074
(10, 50)	0	10.935	24	0.064
(10, 100)	3	11.247	21	0.058
Average	0.062 (3)	11.091	0.937 (45)	0.061
All Average	0.379 (91)	12.173	0.795 (191)	0.063

Table 12. The average ARPD and RBS with q = 3.

(M, J)	TS-I with *MCT decoder*		IG-HM with *MCT decoder*	
	RBS	**Avg. ARDP**	**RBS**	**Avg. ARDP**
(2, 10)	13	1.752	24	0.000
(2, 20)	12	5.310	21	1.383
Average	0.520 (25)	3.531	0.937 (45)	0.691
(3, 15)	14	4.614	22	1.222
(3, 30)	6	11.785	18	7.675
Average	0.416 (20)	8.200	0.833 (40)	4.448
(5, 25)	10	7.456	14	6.484
(5, 50)	2	27.184	22	8.504
Average	0.250 (12)	17.320	0.750 (36)	7.494
(7, 35)	7	11.345	17	7.233
(7, 70)	1	23.909	23	7.774
Average	0.166 (8)	17.627	0.833 (40)	7.504
(10, 50)	4	9.669	20	6.618
(10, 100)	3	11.247	21	5.840
Average	0.145 (7)	10.458	0.854 (41)	6.229
All Average	0.300 (72)	11.427	0.841 (202)	5.273

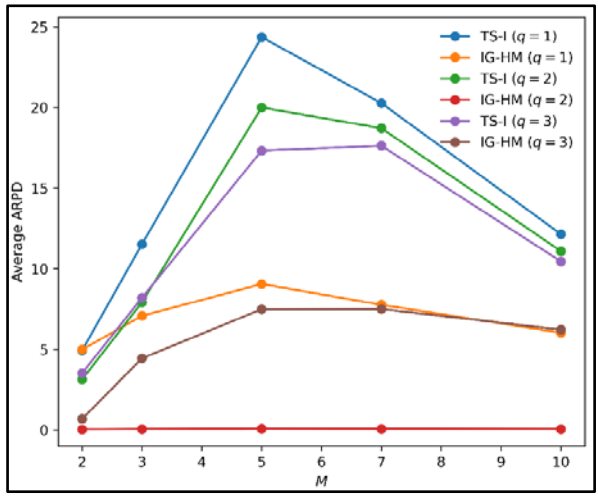


Figure 5. Effect of problem size on average ARPD of TS-I and IG-HM algorithms.

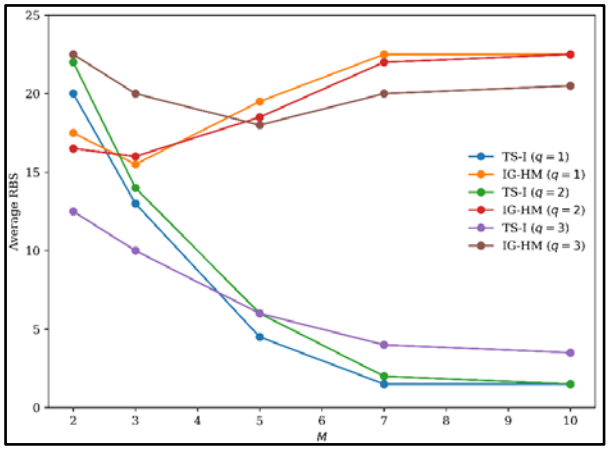


Figure 6. Effect of problem size on RBS of TS-I and IG-HM algorithms.

To examine the statistical significance of the performance differences between IG-HM and TS-I, a paired non-parametric Wilcoxon signed-rank test was applied to the RPD values from all experimental runs. This test was selected because it supports paired comparisons and does not require normality assumptions

(Derrac et al., 2011; García et al., 2009). For each runtime setting $q \in \{1,2,3\}$, the test was performed using all paired results across instances and replications. The null hypothesis assumes no difference between the algorithms, while the alternative hypothesis states that IG-HM performs better. As reported in Table 13, the null hypothesis is rejected for all three runtime settings, indicating that IG-HM statistically outperforms TS-I at the 0.5% significance level ($\alpha=0.005$).

Table 13. Wilcoxon signed-rank test results.

q	TS mean	IG mean	Wilcoxon W	Significance
1	14.648	6.986	518106.5	$p < 10^{-10}$
2	12.173	0.063	24371.0	$p < 10^{-12}$
3	11.427	5.273	530678.5	$p < 10^{-8}$

3.3.2. Impact of decoding strategies on IG-HM performance

The second experimental study examines the effect of decoder selection on the performance of the proposed IG-HM algorithm. Three decoding strategies, MCT, MinΔ, and TSLA, are evaluated using two metrics: the number of best solutions (RBS) and the average ARDP. Comparative results are reported in Table 14. The findings show that decoder choice substantially affects both the behavior and effectiveness of IG-HM. Among the tested strategies, the MCT decoder delivers the strongest performance, achieving the lowest Avg. ARDP values in almost all problem groups and the highest number of best solutions. Across the experiments, MCT attains an Avg. ARDP of 36.662 and identifies 142 best solutions, outperforming the alternatives. The MinΔ decoder exhibits the weakest performance. It fails to produce any best solution across all test groups and yields significantly higher deviation values, with an Avg. ARDP of

133.143, indicating limited guidance for the IG-HM search process under the considered settings. Although TSLA does not surpass MCT in average solution quality, it consistently outperforms MinΔ and achieves an Avg. ARDP of 50.789 with 109 best solutions. For several medium and large instances, TSLA approaches the performance of MCT, suggesting competitive behavior in more challenging cases.

Table 14. Performance comparison of different decoders within the IG-HM algorithm.

(M, J)	IG-HM with *MCT decoder*		IG-HM with MinΔ *decoder*		IG-HM with TSLA *decoder*	
	RBS	**Avg. ARDP**	**RBS**	**Avg. ARDP**	**RBS**	**Avg. ARDP**
(2, 10)	22	2.230	0	42.545	12	12.752
(2, 20)	5	54.228	0	145.669	19	7.860
Average	0.562 (27)	28.229	0 (0)	94.107	0.645 (31)	10.306
(3, 15)	7	20.988	0	70.835	18	8.172
(3, 30)	9	54.155	0	217.440	15	76.300
Average	0.333 (16)	37.571	0 (0)	144.138	0.687 (33)	42.236
(5, 25)	14	44.770	0	105.927	10	25.413
(5, 50)	24	10.869	0	196.596	0	135.086
Average	0.791 (38)	27.820	0 (0)	151.262	0.208 (10)	80.250
(7, 35)	13	22.351	0	112.418	11	59.742
(7, 70)	13	91.072	0	180.052	11	48.592
Average	0.541 (26)	56.711	0 (0)	146.235	0.458 (22)	54.167
(10, 50)	19	9.258	0	98.862	5	69.298
(10, 100)	16	56.697	0	161.090	8	64.678
Average	0.729 (35)	32.978	0 (0)	129.976	0.270 (13)	66.988
All Average	0.591 (142)	36.662	0 (0)	133.143	0.454 (109)	50.789

4. CONCLUSION

This study addresses an unrelated parallel machine scheduling problem with sequence-dependent setup times, machine eligibility constraints, and a common setup server, aiming to minimize total weighted tardiness. The problem reflects manufacturing environments with heterogeneous machines and a shared setup resource, where strong interactions between machine operations and server availability significantly increase scheduling complexity.

For larger instances, an Iterated Greedy–based algorithms are developed. The first approach, IG-HM, combines GRASP-based construction, decoder-guided evaluation, and Variable Neighborhood Descent within an Iterated Greedy framework. Three decoding strategies are examined to guide server sequence construction: Minimum Completion Time (MCT), Marginal Weighted Tardiness (MinΔ), and Two-Step Lookahead (TSLA). Among these, MCT proves to be consistently effective, while TSLA offers complementary benefits by considering limited foresight, which becomes advantageous in certain problem settings.

Further improvements may also be achieved by embedding adaptive learning mechanisms into IG-HM, allowing decoding strategies to adjust dynamically during the search.

REFERENCES

Allahverdi, A. (2015). The third comprehensive survey on scheduling problems with setup times/costs. European Journal of Operational Research, 246(2), 345–378. https://doi.org/10.1016/j.ejor.2015.04.004.

Bektur, G., & Saraç, T. (2019). A mathematical model and heuristic algorithms for an unrelated parallel machine scheduling problem with sequence-dependent setup times, machine eligibility restrictions and a common server. Computers & Operations Research, 103, 46–63. https://doi.org/10.1016/j.cor.2018.10.010.

Derrac, J., García, S., Molina, D., & Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm and Evolutionary Computation, 1(1), 3-18, https://doi.org/10.1016/j.swevo.2011.02.002.

Feo, T. A., & Resende. M. G. C. (1995). Greedy randomized adaptive search procedures. Journal of Global Optimization, 6(2), 109–133. https://doi.org/10.1007/BF01096763.

García, S., Fernández, A., Luengo, J., & Herrera, F. (2009). A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft Computing, 13, 959–977, https://doi.org/10.1007/s00500-008-0392-y.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. Science, 220(4598), 671–680. https://doi.org/10.1126/science.220.4598.671.

Law, A. M., & Kelton, W. D. (2000). Simulation modeling and analysis (3rd ed.). McGraw–Hill.

Lin, S., & Kernighan, B. W. (1973). An effective heuristic algorithm for the traveling-salesman problem. Operations Research, 21(2), 498–516. https://www.jstor.org/stable/169020.

Mladenović, N., & Hansen, P. (1997). Variable neighborhood search. Computers & Operations Research, 24(11), 1097–1100. https://doi.org/10.1016/S0305-0548(97)00031-2.

Pinedo, M. (2012). Scheduling: Theory, algorithms and systems. Springer, New York. A.B.D.

Ruiz, R., & Stützle, T. (2007). A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem. European Journal of Operational Research, 177(3), 2033–2049. https://doi.org/10.1016/j.ejor.2005.12.009.

AN ARC-BASED MATH-HEURISTIC FOR UNRELATED PARALLEL MACHINE SCHEDULING WITH SEQUENCE-DEPENDENT SETUPS AND A COMMON SERVER

Alper HAMZADAYI[1]

1. INTRODUCTION

Modern manufacturing systems require production environments capable of handling product variety, customized demands, and short delivery times. In such systems, production efficiency depends not only on machine utilization but also on the coordination of auxiliary resources such as setup operators, robotic systems, and transportation units (Pinedo, 2012; Potts & Strusevich, 2009). Among these resources, setup operations play a critical role, particularly when setup durations depend on job sequencing decisions, resulting in sequence-dependent setup times that significantly increase scheduling complexity (Allahverdi et al., 1999; Allahverdi et al., 2008). Such setups commonly arise in semiconductor manufacturing, flexible machining systems, and automated production environments, where inefficient sequencing may cause setup congestion, machine idle times, and production delays (Rabadi et al., 2006; Lin & Ying, 2014; Bitar et al., 2016).

Additional complexity emerges when setup activities require a shared setup server. In common-server scheduling systems, setup operations performed by a single server create strong synchronization dependencies among machines,

[1] Doç. Dr., Van Yüzüncü Yıl Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, ORCID: 0000-0003-4035-2775.

preventing independent machine scheduling (Allahverdi, 2015; Cheng et al., 2017; Hasani et al., 2014). The problem becomes even more challenging in unrelated parallel machine environments, where processing times vary according to machine–job combinations and machine eligibility restrictions may exist.

Although scheduling problems with setup considerations have been widely studied, most existing works either simplify server coordination or focus mainly on makespan minimization. For clarity, the notation used in this study is summarized in Table 1, related studies are presented in Table 2, and algorithm abbreviations are given in Table 3. As shown in Table 2, studies simultaneously considering sequence-dependent setup times, common setup servers, and unrelated parallel machines remain relatively limited, particularly for weighted tardiness minimization problems.

Table 1. Problem Notation and Definitions.

Notation	Description
R_M, S_1	Unrelated parallel machine environment with a common setup server
ST	Sequence independent setup time
ST_{sd}	Sequence dependent setup time
C_{max}	Makespan, corresponding to the completion time of the last scheduled job
T_{max}	Maximum tardiness
ΣT_j	Total tardiness
Throughput	Production volume / output rate
$\Sigma w_j \times C_j$	Total weighted completion time
ΣEC	Total energy consumption
$\Sigma w_j \times F_j$	The total flow time weighted
$\Sigma w_j \times T_j$	Total tardiness weighted
ΣD	Variation in machine workloads
ΣC_j	Total completion time
ΣF_j	Total flow time
$\Sigma Loadvar$	Total workload variation
$\Sigma n_j \times \theta_j$	Number of completed products
$\Sigma moves$	Number of additional resource movements
$w_j \times L_j$	Total weighted lateness
TEC	Total Energy Consumption

Table 2. Studies on scheduling with additional resources.

Authors	Year	Additional Resource Usage	Setup Time	Objective	Method
García-de-Alba et al.	2025	S_1	ST_{sd}	C_{max}	MILP, Hybrid (IG+VND)
Şaştım and Hasgül	2024	S_1, res·11	ST_{sd}	C_{max}	MILP, SA, RS
Zhu et al.	2024	res·1	ST_{sd}	C_{max}, *Energy*	DABC
Hadhbi et al.	2023	S_1	ST_{sd}	$\Sigma w_j \times C_j$	MILP, Branch-and-Cut
Lopez-Esteve et al.	2023	res··	ST_{sd}	C_{max}	GRASP
Raboudi et al.	2022	S_1	STsd	$\Sigma w_i \times C_i$	MILP
Yunusoglu and Yildiz	2022	res⋯	ST_{sd}	C_{max}	CP
Li et al.	2022	res·1	ST_{sd}	C_{max}	ABC-AC
Lei and He	2022	res·1	ST_{sd}	C_{max}	AABC
Zhang et al.	2021	S>1	ST_{sd}	C_{max}, TEC	CEA
Al-Harkan et al.	2021	res⋯	ST	C_{max}	MHS
Bitar et al.	2021	res·11	ST_{sd}	$\Sigma w_j \times C_j$; $\Sigma n_j \times \theta_j$; $\Sigma moves$	ILP
Su et al.	2021	res1··	–	C_{max}	MILP, GA, Hybrid GA
Fanjul-Peyro	2020	res1··	ST_{sd}	C_{max}	MILP, Heuristic
Yepes-Borrero et al.	2020	res1··	ST	C_{max}	MILP, GRASP
Al-Harkan and Qamhan	2019	res⋯	ST_{sd}	C_{max}	MILP, Hybrid (VNS, SA)
Bektur and Saraç	2019	S_1	ST_{sd}	$\Sigma w_i \times T_i$	MILP, TS, SA
Qamhan and Alharkan	2019	res1·1	–	C_{max}	MILP
Zhu and Tianyu	2019	res··1	ST_{sd}	$\Sigma w_j \times C_j + \Sigma EC$	MILP, Multi-objective AIS
Afzalirad and Shafipour	2018	res⋯	ST_{sd}	C_{max}	ILP, GA, Hybrid GA
Fleszar and Hindi	2018	res1··	–	C_{max}	MILP, CP-based Heuristic
Villa et al.	2018	res1··	–	C_{max}	Heuristic
Fanjul-Peyro et al.	2017	res1··	–	C_{max}	MILP, Matheuristic
Manupati et al.	2017	res·11	ST	ΣC_j, ΣF_j, ΣT_j, $\Sigma Loadvar$	MINLP, AI-NSGA-II
Afzalirad and Rezaeian	2016	res⋯	ST_{sd}	C_{max}	ILP, GA, AIS
Bitar et al.	2016	res·11	ST_{sd}	$\Sigma w_i \times C_i$	MA
Özpeynirci et al.	2016	res··1	–	C_{max}	MILP, TS
Zheng and Wang	2016	res1·1	–	C_{max}	MILP, Two-stage FOA
Kerkhove and Vanhoucke	2014	S_1	ST	$\Sigma(w_j \times T_j + w_i \times L_j)$	MILP, Hybrid (TS, GA)
Lozano and Medaglia	2014	res··1	ST_{sd}	C_{max}, ΣT_j	Two-stage Heuristic (MILP + GRASP)
Torabi et al.	2013	res·11	–	$\Sigma w_j \times F_j + \Sigma w_j \times T_j + \Sigma D$	MINLP, MOPSO
Zhang et al.	2011	res··1	ST_{sd}	*Throughput*	RL
Chen	2006	res··1	ST	T_{max}	Heuristic
Chen and Wu	2006	res··1	ST	ΣT_i	Heuristic
Chen	2005	res··1	–	C_{max}	Heuristic

Bektur and Saraç (2019) studied the R_M, $S_1/ST_{sd}/\sum w_j T_j$ scheduling problem and proposed a mixed-integer linear programming model together with tabu search and simulated annealing algorithms. Their study highlighted the complexity of coordinating machine schedules and setup-server operations

under weighted tardiness objectives. However, server coordination was modeled indirectly, limiting computational efficiency for larger instances.

Table 3. Description of algorithms used in Table 2.

Abbreviation	Full Name
ILP	Integer Linear Programming
MILP	Mixed Integer Linear Programming
MINLP	Mixed Integer Nonlinear Programming
CP	Constraint Programming
GA	Genetic Algorithm
MHS	Modified Harmony Search
ABC-AC	Artificial Bee Colony with Adaptive Competition
AABC	Adaptive Artificial Bee Colony
DABC	Dynamical Artificial Bee Colony
AIS	Artificial Immune System
MA	Memetic Algorithm
FOA	Fruit Fly Optimization Algorithm
MOPSO	Multi-Objective Particle Swarm Optimization
RL	Reinforcement Learning
NSGA-II	Non-dominated Sorting Genetic Algorithm II
SA	Simulated annealing
TS	Tabu search
GRASP	Greedy Randomized Adaptive Search Procedure
VND	Variable Neighborhood Descent
IG	Iterated Greedy
RS	Random Descent Search
CEA	Combinatorial Evolutionary Algorithm

Recently, arc-based mixed-integer linear programming (AB-MILP) formulations have gained attention due to their ability to explicitly represent sequencing and synchronization relations. In common-server environments, arc-based structures enable direct integration of server precedence decisions into the mathematical model (Hamzadayı & Arvas, 2026). Nevertheless, the integration of arc-based formulations with math-heuristic evaluation mechanisms remains relatively unexplored for unrelated parallel machine scheduling problems with common setup servers.

Motivated by this gap, this study proposes a novel AB-MILP formulation for unrelated parallel machine scheduling with sequence-dependent setup times under a common-server

constraint. The proposed framework explicitly coordinates machine sequencing, server operations, and machine eligibility restrictions through a global server sequence. In addition, a reduced mathematical evaluation mechanism is developed by fixing server permutations during optimization, improving computational tractability.

The main contributions of this study are summarized as follows:

- A novel AB-MILP formulation is proposed for unrelated parallel machine scheduling with sequence-dependent setup times and a common setup server.
- A reduced mathematical evaluation mechanism based on server-sequence fixing is developed.
- A math-heuristic Iterated Greedy framework (IG-MILP) integrating the reduced AB-MILP formulation is proposed.
- Extensive computational experiments are conducted to evaluate the proposed approaches.

The remainder of the chapter is organized as follows. Section 2 presents the scheduling problem and the AB-MILP formulation, and describes the proposed IG-MILP framework. Section 3 reports computational results, and Section 4 concludes the study.

2. MATERIALS AND METHODS

2.1. AB-MILP formulation of the Problem

This section presents the scheduling environment and the proposed arc-based mixed-integer linear programming (AB-MILP) formulation. The problem involves unrelated parallel machines with machine eligibility constraints and sequence-dependent setup times performed by a common setup server.

Since both the machine and the server are occupied during setup operations, scheduling decisions must jointly coordinate machine sequences and server operations. Each job is associated with a due date and a priority weight, and the objective is to minimize total weighted tardiness. The sets, parameters, and decision variables are summarized in Tables 4–6, followed by the objective function and model constraints.

Table 4. Sets and indices of the model.

Sets and indices	Description
$i, j, k \in \{1, 2, \ldots, J\}$	Set of jobs, with indices *i*, *j*, *k*.
$m \in \{1, 2, \ldots, M\}$	Set of unrelated parallel machines, with index *m*.

Table 5. Model parameters.

Parameters	Description
$PR_{i,m}$	Processing time of job *i* on machine *m*.
$B_{i,m} \in \{0,1\}$	Eligibility parameter; equals 1 if job *i* can be processed on machine *m*, and 0 otherwise.
$SD_{i,j}$	Sequence-dependent setup time required when job *j* follows job *i*.
DD_i	Due date of job *i*.
w_i	Tardiness weight of job *i*.
$\emptyset$	A sufficiently large positive constant (Big-M).

Table 6. Model Variables.

Binary decision variables	
$Y_{i,m}$	Indicates assignment of job *i* to machine *m*.
$X_{i,j,m}$	Takes value 1 when job *i* immediately precedes job *j* on machine *m*.
$XS_{i,m}$	Identifies job *i* as the first job processed on machine *m*.
$XE_{i,m}$	Identifies job *i* as the last job processed on machine *m*.
$Z_{i,j}$	Defines the ordering of setup tasks on the common server, where job *i* precedes job *j*.
Continuous decision variables	
S_i	The processing starting time of job *i*.
C_i	The completion time associated with job *i*.
T_i	The tardiness value of job *i*.
SS_i	Start time of the setup for job *i*.
DV_i	Total setup duration accumulated before the processing of job *i*.
$ord_{i,m}$	Positional order of job *i* on machine *m*.

The proposed AB-MILP model is presented below.

$$\text{Minimize} \sum_{i=1}^{J} w_i \times T_i \quad (1)$$

$$\sum_{m=1}^{M} Y_{i,m} = 1 \ \forall i \in J \quad (2)$$

$$Y_{i,m} \leq B_{i,m} \ \forall i \in J; \ \forall m \in M \quad (3)$$

$$XS_{i,m} + \sum_{j=1:j\neq i}^{J} X_{j,i,m} = Y_{i,m} \ \forall i \in J; \ \forall m \in M \quad (4)$$

$$XE_{i,m} + \sum_{j=1:j\neq i}^{J} X_{i,j,m} = Y_{i,m} \ \forall i \in J; \ \forall m \in M \quad (5)$$

$$X_{i,i,m} = 0 \ \forall i \in J; \ \forall m \in M \quad (6)$$

$$\sum_{i=1}^{J} XS_{i,m} \leq 1 \ \forall m \in M \quad (7)$$

$$\sum_{i=1}^{J} XE_{i,m} \leq 1 \ \forall m \in M \quad (8)$$

$$\sum_{i=1}^{J} XS_{i,m} = \sum_{i=1}^{J} XE_{i,m} \ \forall m \in M \quad (9)$$

$$ord_{j,m} \geq ord_{i,m} + 1 - |J| \times \left(1 - X_{i,j,m}\right) \forall i,j \in J: i \neq j; \ \forall m \in M \quad (10)$$

$$ord_{i,m} \leq |J| \times Y_{i,m} \ \forall i \in J; \ \forall m \in M \quad (11)$$

$$ord_{i,m} \geq XS_{i,m}\ \forall i\ \in\ J;\ \forall m \in M \tag{12}$$

$$Z_{i,j} + Z_{j,i} = 1\ \ \forall i,j \in J: i \neq j \tag{13}$$

$$Z_{i,j} \geq Z_{i,k} + Z_{k,j} - 1\ \ \forall i,j,k \in J: \mathrm{i} \neq k, \mathrm{k} \neq j, \mathrm{i} \neq j \tag{14}$$

$$Z_{i,j} \geq \sum_{m=1}^{M} X_{i,j,m}\ \ \forall i,j \in J: i \neq j \tag{15}$$

$$Z_{i,i} = 0\ \ \forall i\ \in\ J \tag{16}$$

$$C_i = S_i + \sum_{m=1}^{M} PR_{i,m} \times Y_{i,m}\quad \forall i\ \in\ J \tag{17}$$

$$S_j \geq S_i + PR_{i,m} + SD_{i,j} - \emptyset\ \times \left(1 - X_{i,j,m}\right) \forall i,j \in J: i \neq j;\ \forall m \in M \tag{18}$$

$$DV_i = \sum_{j=1:j\neq i}^{J} \sum_{m=1}^{M} SD_{j,i} \times X_{j,i,m}\ \ \forall i\ \in\ J \tag{19}$$

$$SS_i = S_i - \ DV_i\ \ \forall i\ \in\ J \tag{20}$$

$$SS_j \geq SS_i + DV_i - \emptyset\ \times \left(1 - Z_{i,j}\right) \forall i,j \in J: i \neq j \tag{21}$$

$$T_i \geq C_i - DD_i\, \forall i\ \in\ J \tag{22}$$

$$Y_{i,m} \in \{0,1\}\ \forall i\ \in\ J, \forall m\ \in\ M;\ X_{i,j,m}\ \in \{0,1\}\ \forall i,j \in J, \forall m \in M;\ XS_{i,m}\ \in \{0,1\}\ \forall i\ \in\ J, \forall m\ \in\ M;\ XE_{i,m} \in \{0,1\}\ \forall i\ \in\ J, \forall m\ \in\ M;\ Z_{i,j}\ \in \{0,1\} \forall i,j \in J;\ S_i \geq 0\ \forall i\ \in\ J;\ C_i \geq 0\ \forall i\ \in\ J;\ T_i\ \geq 0 \forall i\ \in\ J;\ St_i \geq 0\ \forall i\ \in\ J;\ DV_i \geq 0\ \forall i\ \in\ J; ord_{i,m} \geq 0\ \forall i\ \in\ J, \forall m\ \in\ M \tag{23}$$

Constraint set (1) minimizes total weighted tardiness. Constraint sets (2) and (3) ensure that each job is assigned to

exactly one eligible machine. Constraint sets (4)–(9) establish feasible machine-level sequencing structures by defining predecessor–successor relationships and preventing infeasible assignments. Constraint sets (10)–(12) apply MTZ-based ordering constraints to maintain consistent processing orders on each machine. Constraint sets (13)–(16) regulate the sequencing of setup operations performed by the common server while ensuring consistency with machine-level precedence relations. Constraint sets (17)–(20) define job completion times, setup durations, and the synchronization between server operations and job processing. Constraint set (21) enforces the single-capacity restriction of the common server by preventing overlapping setup operations. Constraint set (22) calculates tardiness values, while constraint set (23) defines the domain and integrality conditions of the decision variables.

To ensure valid temporal precedence relations, the Big-M constant must be sufficiently large to dominate all feasible time differences in the schedule. Instead of using an arbitrary value, a problem-specific upper bound is constructed based on the processing and setup characteristics of each instance. This instance-dependent $\emptyset$ value is defined through constraint sets (24)–(27).

$$SD_{max} = \max_{\forall i,j \in J : i \neq j} SD_{i,j} \tag{24}$$

$$PR_i^{max} = \max_{\forall m \in M : B_{i,m}=1} PR_{i,m} \tag{25}$$

$$PR_{sum}^{max} = \sum_{i=1}^{J} PR_i^{max} \tag{26}$$

$$\emptyset = PR_{sum}^{max} + |J| \times SD_{max} \tag{27}$$

Constraint set (24) defines the maximum sequence-dependent setup time, denoted by SD_{max}. Constraint set (25) determines the maximum feasible processing time of each job across eligible machines, while constraint set (26) aggregates these values to obtain an upper bound on the total processing workload. Finally, constraint set (27) defines the Ω constant by combining upper bounds on processing and setup times. This value is used to relax precedence constraints when no sequencing relationship exists between jobs.

2.2. IG-MILP: A Math-Heuristic Algorithm

This section presents an IG-based algorithm in which candidate solutions are evaluated using the proposed AB-MILP model. IG-MILP follows the standard structure of the classical IG algorithm, with its main steps summarized in Algorithm 1. The local search, AB-MILP-based evaluation, and destruction–construction operators of IG-MILP are described in the following subsections.

2.2.1. Local search

In IG-MILP, solution refinement is performed using a pairwise swap neighborhood (Algorithm 2) with a first-accept improvement strategy. The current solution is updated immediately upon finding a better neighbor, and the search continues until no further improvement is possible. This lightweight local search is applied at each IG iteration to enhance solution quality and maintain search stability.

2.2.2. Destruction and reconstruction

After initialization, IG-MILP iteratively updates the current solution through successive destruction and reconstruction phases. During destruction, a subset of jobs is randomly removed from the sequence to promote diversification. In reconstruction, the removed jobs are reinserted one by one

using a greedy insertion criterion that selects the most favorable position for each job. The corresponding procedures are detailed in Algorithm 3.

Algorithm 1. Algorithmic structure of the IG-MILP.

Algorithm IG-MILP (ρ, *T_acc*, τ*_max*, *ins*, *seed***)**
1. *t_start* ← Get the current CPU time
2. *Seq_cur* ← A job sequence of length *J* initialized using random ordering
3. *Fitness_cur* ← **MILP_Fitness (***Seq_cur***)**
4. (*Seq_cur*, *Fitness_cur*) ← **LocalSearch (***Seq_cur*, *Fitness_cur***)**
5. *Seq_best* ← *Seq_cur*
6. *Fitness_best* ← *Fitness_cur*
7. **While (**current CPU time − *t_start* < τ*_max***) do**
8. *Seq_cand* ← *Seq_cur*
9. (*Seq_cand*, *Fitness_cand*) ← **Destruction_Construction_IG (***Seq_cand*, ρ**)**
10. (*Seq_cand*, *Fitness_cand*) ← **LocalSearch (***Seq_cand*, *Fitness_cand***)**
11. **If** *Fitness_cand* ≤ *Fitness_cur* **then**
12. *Seq_cur* ← *Seq_ cand*
13. *Fitness_cur* ← *Fitness_cand*
14. **Else**
15. Δ ← *Fitness_cand - Fitness_cur*
16. *p_acc* ← **exp (**−Δ / *T_acc***)**
17. **If rand (**0,1**)** < *p_acc* **then**
18. *Seq_cur* ← *Seq_ cand*
19. *Fitness_cur* ← *Fitness_cand*
20. **End If**
21. **End If**
22. **If** *Fitness_cur* < *Fitness_best* **then**
23. *Seq_best* ← *Seq_cur*
24. *Fitness_best* ← *Fitness_cur*
25. **End If**
26. **End While**
27. **Return (***Seq_best*, *Fitness_best***)**

Algorithm 2. Local improvement procedure.

Algorithm LocalSearch (*Seq*, *Fitness***)**
1. *len* ← length of *Seq*
2. *b_Seq* ← *Seq*
3. *b_Fitness* ← *Fitness*
4. *improvment* ← True
5. **While** *improvment* **do**
6. *improvment* ← False
7. **For** *i* = 1 through *len* − 1 **do**
8. **For** *j* = *i* + 1 through *len* **do**
9. *N_Seq* ← swap positions *i* and *j* in *Seq*
10. *N_Fitness* ← **MILP_Fitness (***N_Seq***)**
11. **If** *N_Fitness* < *b_Fitness* **then**

12. *b_Seq* ← *N_Seq*
13. *b_Fitness* ← *N_Fitness*
14. *improvment* ← True
15. **End If**
16. **End For**
17. **End For**
18. End While
19. Return *b_Seq and b_Fitness*

Algorithm 3. Solution destruction and reconstruction.

Input: *Seq*: current job sequence, ρ: destruction ratio
Output: Updated sequence *Seq*, Corresponding fitness value
1. Let *len* denote the length of sequence *Seq*
2. Compute the number of jobs to be removed as $remCount = \lfloor \rho \times len \rfloor$
3. Randomly select *remCount* jobs from *Seq* and store them in the set *Removed*
4. Remove all jobs in *Removed* from *Seq*
5. For each removed job *j* **do**:
a. Set *bestPos* ← −1
b. Initialize *bestFit* ← +∞
c. For each possible insertion position $p = 1, \ldots, |Seq| + 1$ **do**:
i. Construct a temporary sequence *TempSeq* by inserting job *j* at position *p* in *Seq*
ii. Evaluate *TempSeq* using the MILP-based fitness function
iii. If the obtained fitness value is smaller than *bestFit* **then**:
- Update *bestFit* with the new fitness value
- Update *bestPos* with position *p*
d. Insert job *j* into *Seq* at position *bestPos*
6. Compute the final fitness of the reconstructed sequence *Seq* using the MILP evaluation
7. Return the updated sequence *Seq* and its fitness value

2.2.3. Mathematical Model Based Solution Evaluation

When solutions are evaluated using the proposed MILP model, candidate solutions generated by the IG algorithm are assessed through the AB-MILP formulation. To embed the structural information provided by the metaheuristic, a global precedence pattern is extracted from the IG solution. As summarized in Table 7, the permutation *Z* defines the setup order of the common server, while the derived priority vector

rank_array assigns each job a global rank consistent with this order, with smaller ranks indicating higher priority.

Table 7. Server Sequence and Job Ranking in the Model.

Location	1	2	3	4	5	6	7	8	9	10
Z	8	2	9	6	5	7	1	4	10	3
Job ID	1	2	3	4	5	6	7	8	9	10
rank_array	7	2	10	8	5	4	6	1	3	9

The proposed AB-MILP model-based solution evaluation is presented below.

Constraint sets (1)-(12), (17)-(20), and (22)

$$X_{i,j,m} = 0 \quad \forall i,j \in J, i \neq j: \text{rank\_array}(i) > \text{rank\_array(j)} \tag{28}$$

$$SS_{Z[k+1]} \geq SS_{Z[k]} + DV_{Z[k]} \quad \forall k \in 1,2,\dots,J-1 \tag{29}$$

$$Y_{i,m} \in \{0,1\}\ \forall i \in J, \forall m \in M;\ X_{i,j,m} \in \{0,1\}\ \forall i,j \in J, \forall m \in M;\ XS_{i,m} \in \{0,1\}\ \forall i \in J, \forall m \in M;\ XE_{i,m} \in \{0,1\}\ \forall i \in J, \forall m \in M;\ S_i \geq 0\ \forall i \in J;\ C_i \geq 0\ \forall i \in J;\ T_i \geq 0\ \forall i \in J;\ St_i \geq 0\ \forall i \in J;\ DV_i \geq 0\ \forall i \in J; ord_{i,m} \geq 0\ \forall i \in J, \forall m \in M \tag{30}$$

Constraint set (28) restricts machine-level sequencing decisions according to the rank order obtained from the metaheuristic, preventing precedence relations that contradict the global permutation. This limits the feasible region to high-quality solution structures. Since the server sequence is fixed by permutation *Z*, the original server-ordering constraints become unnecessary. Instead, constraint set (29) directly enforces the fixed server order by ensuring that the setup of job *Z*[*k*+1] starts only after the completion of the setup of job *Z*[k]. Consequently, server-order binary variables and related transitivity constraints are eliminated. Finally, constraint set (30) preserves only the

domain restrictions associated with the remaining active variables. By combining rank-based restrictions with a fixed server sequence, the reduced formulation significantly decreases the search space while maintaining temporal optimization capability within the math-heuristic framework.

3. RESULTS AND DISCUSSION

This section investigates the effectiveness of the proposed approaches through a comprehensive computational study. The experiments are organized into three parts. First, a benchmark data set is generated to reproduce the characteristics of the scheduling environment considered in the literature. Second, the proposed arc-based MILP formulation is compared with the BS-MILP model introduced by Bektur and Saraç (2019). Finally, the IG-MILP algorithm is analyzed with respect to both computational efficiency and solution quality. All algorithms were coded in Visual C++, while exact optimization models were solved using Gurobi 9.0.3. Computational experiments were conducted on a computer equipped with an Intel Core i7-8565U processor and 16 GB RAM.

3.1. Benchmark Instance Construction

Since the original benchmark instances of Bektur and Saraç (2019) are unavailable, a new benchmark set was generated following the same experimental design principles to ensure comparable problem characteristics. The experiments consider five machine settings, $M \in \{2, 3\}$, with the number of jobs defined as J=5×M. Processing times were generated from a discrete uniform distribution over [10,100], while job weights were sampled from [1,10]. To analyze setup-time variability, three setup intervals were considered: low [5,25], medium [5,50], and high [25,50]. Two machine-eligibility environments were examined: full eligibility and partial eligibility, where each job is

eligible for a machine with probability 0.7. Due dates were generated using the procedure of Bektur and Saraç (2019) with tardiness tightness parameters of 0.5 and 0.8 and a randomness factor of 0.2. Overall, 120 parameter configurations were obtained, and two independent instances were generated for each configuration, resulting in a total of 240 benchmark instances.

3.2. Comparison of Mathematical Formulations

This subsection compares the BS-MILP model of Bektur and Saraç (2019) with the proposed AB-MILP formulation under a common time limit of 3600 seconds. Solution quality, optimality gaps, and CPU times are reported in Table 8 and 9, where "NFS" denotes instances for which no feasible solution was obtained.

For 2-machine instances, AB-MILP clearly outperforms BS-MILP in terms of feasibility, solution quality, and computational efficiency. While BS-MILP reaches optimality in only 12 instances, AB-MILP obtains 23 optimal solutions and feasible results for all cases. Moreover, the average optimality gap is significantly reduced. For 3-machine instances, the problem becomes substantially harder. BS-MILP fails to obtain optimal solutions and frequently returns infeasible results or very large optimality gaps. In contrast, AB-MILP produces feasible solutions for all instances and reaches optimality in several cases. Due to their computational difficulty, larger 3-machine instances were excluded from further analysis.

Figures 1 and 2 compare both formulations in terms of optimality gap and CPU time. Overall, the results demonstrate that AB-MILP provides superior scalability and solution performance compared with BS-MILP.

Table 8. Computational results for 2-machine instances.

Instances involving 2 machines with either 10 or 20 jobs						
Ins.	**BS-MILP**			**AB-MILP**		
	SV	**Gap%**	**CPU**	**SV**	**Gap%**	**CPU**
1	398	100.00	3600.06	*398*	*0.00*	764.36
2	367	100.00	3600.02	*367*	*0.00*	413.00
3	2580	95.35	3600.04	*2580*	*0.00*	269.10
4	2481	96.74	3600.03	*2316*	*0.00*	265.52
5	*1300*	*0.00*	245.34	*1300*	*0.00*	41.70
6	*1702*	*0.00*	44.89	*1702*	*0.00*	40.66
7	*2233*	*0.00*	619.99	*2233*	*0.00*	18.59
8	*2989*	*0.00*	3237.62	*2989*	*0.00*	808.50
9	1692	100.00	3600.04	*1485*	*0.00*	1811.88
10	2002	100.00	3600.04	*1905*	*0.00*	2099.17
11	2632	94.87	3600.05	*2632*	*0.00*	97.23
12	2370	97.08	3600.06	*2370*	*0.00*	171.16
13	*920*	*0.00*	3061.51	*920*	*0.00*	263.73
14	*1055*	*0.00*	158.02	*1055*	*0.00*	48.49
15	*4080*	*0.00*	2560.50	*4080*	*0.00*	123.29
16	*3040*	*0.00*	1445.78	*3040*	*0.00*	119.79
17	1959	100.00	3600.06	*1959*	*0.00*	1473.69
18	2118	100.00	3600.05	*1912*	77.78	3600.01
19	6655	96.06	3600.07	*6655*	*0.00*	3193.88
20	5310	98.98	3600.05	*5310*	*0.00*	883.75
21	*2942*	*0.00*	2287.19	*2942*	*0.00*	2006.26
22	*4528*	*0.00*	143.00	*4528*	*0.00*	55.48
23	*6550*	*0.00*	649.18	*6550*	*0.00*	142.62
24	*7889*	*0.00*	929.70	*7889*	*0.00*	123.25
25	NFS	-	3600.13	*1751*	100.00	3600.04
26	NFS	-	3600.12	*1303*	100.00	3600.02
27	NFS	-	3600.16	*10618*	98.87	3600.02
28	NFS	-	3600.11	*8063*	100.00	3600.05
29	6115	100.00	3600.17	*4419*	100.00	3600.03
30	6120	100.00	3600.15	*2972*	100.00	3600.02
31	27603	100.00	3600.47	*14102*	97.42	3600.08
32	15526	100.00	3600.14	*12782*	99.01	3600.03
33	NFS	-	3600.19	*2337*	100.00	3600.03
34	NFS	-	3600.16	*2484*	100.00	3600.03
35	NFS	-	3600.16	*12591*	99.32	3600.04
36	NFS	-	3600.22	*10614*	99.97	3600.02
37	NFS	-	3600.17	*3110*	100.00	3600.02
38	10911	100.00	3600.20	*5433*	100.00	3600.02
39	NFS	-	3600.13	*18326*	96.38	3600.07
40	21789	100.00	3600.10	*14486*	93.53	3600.09
41	NFS	-	3600.15	*9128*	100.00	3600.03
42	NFS	-	3600.18	*8981*	100.00	3600.06
43	NFS	-	3600.21	*26263*	96.92	3600.06
44	NFS	-	3600.12	*16554*	96.65	3600.03
45	20946	100.00	3600.24	*14354*	100.00	3600.03
46	NFS	-	3600.21	*8257*	100.00	3600.03
47	22019	100.00	3600.11	*19253*	93.95	3600.06
48	33911	100.00	3600.14	*20467*	94.94	3600.05

Table 9. Computational results for 3-machine instances.

Instances involving 2 machines with either 10 or 20 jobs						
Ins.	**BS-MILP**			**AB-MILP**		
	SV	**Gap%**	**CPU**	**SV**	**Gap%**	**CPU**
49	2824	100.00	3600.19	*1636*	100.00	3600.03
50	3280	100.00	3600.16	*1336*	100.00	3600.03
51	4178	100.00	3600.11	*2667*	53.75	3600.03
52	4476	100.00	3600.17	*3369*	51.30	3600.10
53	2110	100.00	3600.11	*1419*	100.00	3600.04
54	2584	100.00	3600.18	*2094*	70.05	3600.32
55	8648	100.00	3600.09	*7643*	53.58	3600.01
56	8682	100.00	3600.11	*6780*	50.02	3600.10
57	NFS	-	3600.43	*1886*	100.00	3600.04
58	3882	100.00	3600.16	*1165*	100.00	3600.58
59	8240	100.00	3600.14	*3839*	49.05	3600.11
60	12631	100.00	3600.14	*6377*	54.46	3600.01
61	3158	100.00	3600.10	*2393*	86.80	3600.06
62	3856	100.00	3600.12	*3201*	99.72	3600.01
63	6288	100.00	3600.09	*4769*	42.03	3600.25
64	10514	100.00	3600.11	*8526*	44.92	3600.48
65	5833	100.00	3600.39	*4203*	99.98	3600.81
66	8713	100.00	3600.14	*6397*	*0.00*	512.102
67	10005	100.00	3600.18	*9054*	65.08	3600.21
68	13703	100.00	3600.25	*7155*	71.26	3600.48
69	9061	100.00	3600.08	*9302*	83.75	3600.16
70	11297	100.00	3600.10	*8736*	89.44	3600.58
71	10731	99.98	3600.07	*9494*	56.91	3600.32
72	12357	100.00	3600.21	*9371*	65.87	3600.04
73	NFS	-	3600.32	*14286*	100.00	3600.06
74	NFS	-	3600.62	*7066*	100.00	3600.25
75	NFS	-	3600.46	*23840*	100.00	3600.05
76	NFS	-	3600.79	*18992*	100.00	3600.16
77	NFS	-	3600.25	*17476*	100.00	3600.11
78	NFS	-	3600.29	*19803*	100.00	3600.40
79	NFS	-	3600.22	*20995*	*0.00*	1696.85
80	NFS	-	3600.36	*26792*	98.64	3600.10
81	NFS	-	3600.50	*20110*	100.00	3600.21
82	NFS	-	3600.09	*25258*	100.00	3600.39
83	NFS	-	3600.65	*38339*	100.00	3600.04
84	NFS	-	3600.20	*32024*	100.00	3600.84
85	NFS	-	3600.43	*14468*	100.00	3600.25
86	NFS	-	3600.57	*18332*	100.00	3600.05
87	NFS	-	3600.92	*35634*	*0.00*	2942.85
88	NFS	-	3600.73	*27719*	99.01	3600.06
89	NFS	-	3600.91	*40865*	100.00	3600.50
90	NFS	-	3600.80	*38737*	100.00	3600.16
91	NFS	-	3600.73	*63913*	99.38	3600.24
92	NFS	-	3600.41	*51450*	99.89	3600.40
93	NFS	-	3600.48	*48924*	100.00	3600.48
94	NFS	-	3600.94	*55396*	100.00	3600.55
95	NFS	-	3600.58	*62653*	99.18	3600.28
96	NFS	-	3600.24	*59560*	98.28	3600.25

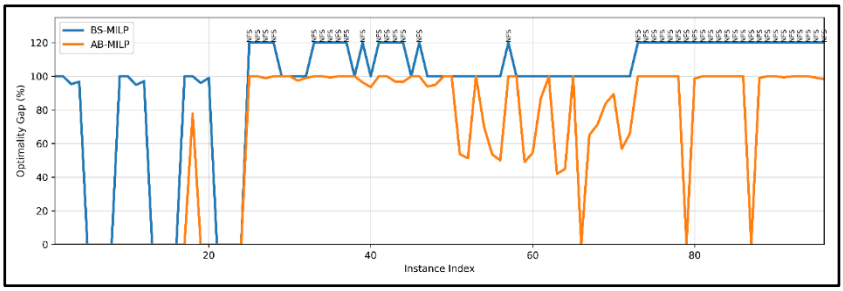


Figure 1. Comparison of optimality gaps for the models.

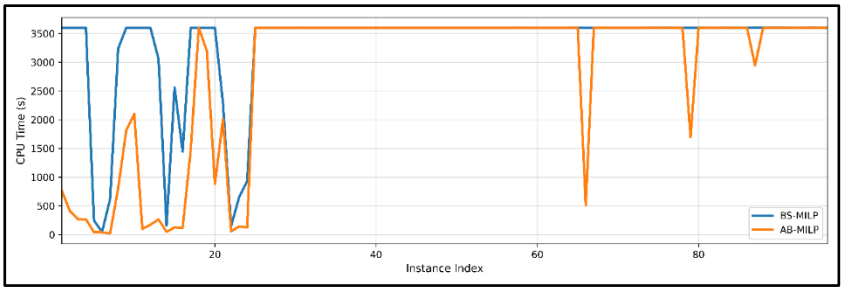


Figure 2. CPU time comparison between the models.

3.3. Performance of the IG-MILP algorithm

The final experimental stage evaluates the performance of the proposed IG-MILP approach. Each instance is solved with a time limit of 1800 seconds, and all experiments are repeated three times. Due to the computational cost of the math–heuristic evaluation and the constructive nature of the algorithm, IG-MILP is applied only to a subset of the benchmark instances, specifically those with 2 machines and 10 jobs, 2 machines and 20 jobs, and 3 machines and 15 jobs. Since IG-MILP builds solutions progressively through local search, a complete job permutation must be generated before the mathematical evaluation can be applied. For larger instances, the algorithm often fails to complete this construction within the time limit, preventing the invocation of the AB-MILP evaluation on the full job set. As a result, IG-MILP cannot be applied to larger problem

sizes within the given computational budget. Tables 10–12 report the results for the selected instances. For the smallest cases with 2 machines and 10 jobs (Table 10), IG-MILP consistently reaches the optimal solutions obtained by the AB-MILP model. The standard deviation is almost zero, indicating highly stable performance across runs. Moreover, average solution times remain below one second, demonstrating that IG-MILP is efficient for small-scale problems.

Table 10. Instances with 2 machines and 10 jobs.

Ins.	AB-MILP	IG-MILP			
		Min.	Std. dev.	Avg. Fitness Eval.	Avg. Time
1	*398*	*398*	0.00	8037.00	0.222
2	*367*	*367*	0.00	9673.33	0.184
3	*2580*	*2580*	0.00	13498.00	0.132
4	*2316*	*2316*	0.00	10597.67	0.168
5	*1300*	*1300*	0.00	33471.67	0.052
6	*1702*	*1702*	0.00	36652.67	0.048
7	*2233*	*2233*	0.00	30343.33	0.058
8	*2989*	*2989*	0.00	24428.67	0.072
9	*1485*	*1485*	0.00	9825.00	0.181
10	*1905*	*1905*	7.50	11159.67	0.159
11	*2632*	*2632*	0.00	14954.33	0.119
12	*2370*	*2370*	0.00	13479.33	0.132
13	*920*	*920*	0.00	13777.33	0.129
14	*1055*	*1055*	0.00	33378.33	0.053
15	*4080*	*4080*	0.00	27985.33	0.063
16	*3040*	*3040*	0.00	20903.33	0.087
17	*1959*	*1959*	0.00	12081.33	0.151
18	*1912*	*1912*	0.00	13930.33	0.127
19	*6655*	*6655*	0.00	14647.67	0.121
20	*5310*	*5310*	0.00	14623.67	0.121
21	*2942*	*2942*	0.00	27657.67	0.064
22	*4528*	*4528*	0.00	35996.67	0.049
23	*6550*	*6550*	0.00	33737.33	0.052
24	*7889*	*7889*	0.00	33114.67	0.053

For instances with 2 machines and 20 jobs (Table 11), IG-MILP continues to deliver competitive solutions, although both the standard deviation and computational time increase noticeably. While solution quality remains close to the optimal

values, average runtimes rise to several seconds for some instances, reflecting the higher computational effort required by the mathematical evaluation.

Table 11. Instances with 2 machines and 20 jobs.

Ins.	AB-MILP	IG-MILP			
		Min.	Std. dev.	Avg. Fitness Eval.	Avg. Time
25	*1751*	2240	1405.40	658.33	2.868
26	*1303*	2010	1321.45	715.33	2.535
27	*10618*	13833	1329.79	884.00	2.054
28	*8063*	10744	3007.89	804.66	2.245
29	4419	*3261*	273.96	9826.66	0.181
30	2972	*2606*	235.22	8351.00	0.214
31	14102	*12508*	254.53	5317.00	0.337
32	12782	*11390*	386.81	6462.66	0.277
33	*2337*	2406	998.47	967.00	1.879
34	*2484*	3750	1373.11	401.00	4.606
35	12591	*9930*	2372.96	1155.33	1.612
36	*10614*	15187	1302.40	1118.66	1.610
37	3110	*2493*	759.25	7725.00	0.231
38	5433	*4857*	493.14	9611.66	0.185
39	18326	*17369*	1087.83	8730.66	0.204
40	14486	*14312*	409.25	8603.33	0.207
41	*9128*	10465	660.71	1706.66	1.054
42	8981	*7265*	535.99	1792.00	1.007
43	26263	26883	646.03	2215.33	0.815
44	*16554*	18821	579.76	2050.00	0.876
45	14354	*11802*	128.03	7863.33	0.227
46	8257	*6404*	241.66	6958.00	0.260
47	19253	*18926*	559.94	5833.00	0.306
48	20467	*19963*	295.57	3668.66	3.897

A similar trend is observed for instances with 3 machines and 15 jobs (Table 12). Although IG-MILP continues to produce high-quality solutions, computational effort increases and variability across runs becomes more pronounced. In particular, the average fitness evaluation time often exceeds several seconds, which limits the number of iterations that can be completed within the available time budget. Additional insight into the computational behavior of IG-MILP is provided by the last two columns of Tables 10–12. The column Avg. Fitness Eval. reports

the average number of mathematical evaluations performed within the 1800-second limit, while Avg. Time indicates the average CPU time required for a single evaluation. For the smallest instances (2 machines and 10 jobs), Avg. Fitness Eval. typically reaches several thousand per run and Avg. Time remains very low, showing that the mathematical model can be solved frequently and efficiently, thereby enabling many improvement steps.

Table 12. Instances with 3 machines and 15 jobs.

Ins.	AB-MILP	IG-MILP			
		Min.	Std. dev.	Avg. Fitness Eval.	Avg. Time
49	*1636*	1930	118.23	1650.33	0.575
50	*1336*	1429	28.64	1394.66	0.459
51	*2667*	2771	6.23	1929.33	0.782
52	*3369*	3398	66.46	1817.00	0.816
53	*1419*	1454	10.84	5057.00	2.622
54	*2094*	2251	31.32	4765.00	2.668
55	*7643*	*7643*	60.81	7787.33	4.624
56	*6780*	6852	42.42	4583.66	2.724
57	*1886*	2003	108.28	2390.00	0.841
58	*1165*	1353	37.72	1805.33	0.560
59	3839	*3801*	227.27	2511.00	0.992
60	6377	*6205*	539.03	2021.33	0.902
61	*2393*	2456	36.57	7409.33	4.430
62	*3201*	3261	40.10	7467.66	4.622
63	4769	*4708*	150.70	5467.33	3.202
64	8526	*8490*	187.14	6744.33	3.932
65	*4203*	4464	20.27	3391.66	1.667
66	6397	*5924*	145.06	3057.00	1.457
67	*9054*	*9054*	91.14	3557.66	1.636
68	7155	*7086*	21.48	2879.33	1.485
69	9302	*8918*	138.65	7288.00	4.318
70	8736	*8486*	223.91	6204.66	3.706
71	*9494*	*9465*	92.86	7503.66	4.778
72	9371	*9170*	52.66	5329.66	3.036

For larger instances, the last two columns reveal a clear shift: the number of fitness evaluations drops sharply, while the average time per evaluation increases to several seconds. Consequently, the 1800-second budget is consumed by a small

number of costly mathematical evaluations, which restricts the extent of the search. This limited exploration leads to higher variability across runs and, in some cases, slightly weaker solution quality. This behavior can be explained by the structural properties of the IG-MILP formulation. Compared to the original AB-MILP model, the math–heuristic evaluation significantly reduces model size by removing all binary server-ordering variables $Z_{i,j}$, since the server sequence is fixed by the permutation derived from the IG solution. As a result, all server-ordering and transitivity constraints are eliminated. In the AB-MILP model, these constraints, particularly the transitivity constraints, constitute the dominant component and grow on the order of J^3. Replacing them with a simple chain of $J-1$ precedence constraints (Constraint set 29) yields a substantial reduction in complexity. Furthermore, the rank-based restriction (Constraint set 28) fixes many sequencing variables $X_{i,j,m}$ to zero, further shrinking the feasible space without introducing additional constraints. These reductions allow IG-MILP to preserve the essential machine-level scheduling structure while removing the most computationally expensive combinatorial elements. This explains its strong performance on small instances, where the model can be solved frequently within the time limit, as observed in Tables 10–12. However, despite these improvements, the remaining model complexity still limits scalability. For larger instances, the combination of repeated MILP evaluations and the constructive nature of the algorithm becomes prohibitively expensive, preventing the generation of complete solutions within the available time. These findings indicate that IG-MILP is well suited for small and medium-sized problems, where it can deliver near-optimal or optimal solutions with reasonable computational effort. Its applicability to larger instances is limited, and it should therefore be viewed primarily as a high-accuracy evaluation mechanism rather than a scalable solution approach for large-scale scheduling problems.

4. CONCLUSION

This study addresses an unrelated parallel machine scheduling problem with sequence-dependent setup times, machine eligibility constraints, and a common setup server, aiming to minimize total weighted tardiness. The problem reflects manufacturing environments with heterogeneous machines and a shared setup resource, where strong interactions between machine operations and server availability significantly increase scheduling complexity.

To model this structure, an arc-based mixed-integer linear programming formulation (AB-MILP) is proposed. By explicitly embedding a global server sequence into the model, AB-MILP consistently coordinates machine-level sequencing and setup operations while preserving feasibility. The formulation produces high-quality solutions for small and medium-sized instances and serves as a reliable benchmark for heuristic approaches.

For larger instances, an Iterated Greedy–based algorithm is developed. The second approach, IG-MILP, represents solutions as job sequences and evaluates them by fixing the server order within the AB-MILP model. This hybrid strategy integrates heuristic exploration with exact optimization, restricting the search to high-quality regions defined by the metaheuristic solution structure. Computational results demonstrate that combining problem-specific mathematical modeling with tailored metaheuristics yields high-quality schedules under common-server constraints, while preserving scalability for large instances.

REFERENCES

Afzalirad, M., & Rezaeian, J. (2016). Resource-constrained unrelated parallel machine scheduling problem with sequence dependent setup times, precedence constraints and machine eligibility restrictions. Computers & Industrial Engineering, 98, 40–52. https://doi.org/10.1016/j.cie.2016.05.020.

Afzalirad, M., & Shafipour, M. (2018). Design of an efficient genetic algorithm for resource-constrained unrelated parallel machine scheduling problem with machine eligibility restrictions. Journal of Intelligent Manufacturing, 29(2), 423–437. https://doi.org/10.1007/s10845-015-1117-6.

Al-Harkan, I. M., & Qamhan, A. A. (2019). Optimize unrelated parallel machines scheduling problems with multiple limited additional resources, sequence-dependent setup times and release date constraints. IEEE Access, 7, 171533–171547. https://doi.org/10.1109/ACCESS.2019.2955975.

Al-Harkan, I. M., Qamhan, A. A., Badwelan, A., Alsamhan, A., & Hidri, L. (2021). Modified harmony search algorithm for resource-constrained parallel machine scheduling problem with release dates and sequence-dependent setup times. Processes, 9(4), 654. https://doi.org/10.3390/pr9040654.

Allahverdi, A. (2015). The third comprehensive survey on scheduling problems with setup times/costs. European Journal of Operational Research, 246(2), 345–378. https://doi.org/10.1016/j.ejor.2015.04.004.

Allahverdi, A., Gupta, J. N. D., & Aldowaisan, T. (1999). A review of scheduling research involving setup

considerations. Omega, 27(2), 219–239. https://doi.org/10.1016/S0305-0483(98)00042-5.

Allahverdi, A., Ng, C. T., Cheng, T. C. E., & Kovalyov, M. Y. (2008). A survey of scheduling problems with setup times or costs. European Journal of Operational Research, 187(3), 985–1032. https://doi.org/10.1016/j.ejor.2006.06.060.

Bektur, G., & Saraç, T. (2019). A mathematical model and heuristic algorithms for an unrelated parallel machine scheduling problem with sequence-dependent setup times, machine eligibility restrictions and a common server. Computers & Operations Research, 103, 46–63. https://doi.org/10.1016/j.cor.2018.10.010.

Bitar, A., Dauzère-Pérès, S., & Yugma, C. (2021). Unrelated parallel machine scheduling with new criteria: Complexity and models. Computers & Operations Research, 132, Article 105291. https://doi.org/10.1016/j.cor.2021.105291.

Bitar, A., Dauzère-Pérès, S., Yugma, C., & Roussel, R. (2016). A memetic algorithm to solve an unrelated parallel machine scheduling problem with auxiliary resources in semiconductor manufacturing. Journal of Scheduling, 19(4), 367–376. https://doi.org/10.1007/s10951-014-0397-6.

Chen, J. F. (2005). Unrelated parallel machine scheduling with secondary resource constraints. The International Journal of Advanced Manufacturing Technology, 26(3), 285–292. https://doi.org/10.1007/s00170-003-1622-1.

Chen, J. F. (2006). Minimization of maximum tardiness on unrelated parallel machines with process restrictions and setups. The International Journal of Advanced

Manufacturing Technology, 29(5), 557–563. https://doi.org/10.1007/BF02729109.

Chen, J. F., & Wu, T. H. (2006). Total tardiness minimization on unrelated parallel machine scheduling with auxiliary equipment constraints. Omega, 34(1), 81–89. https://doi.org/10.1016/j.omega.2004.07.023.

Cheng, T. C. E., Kravchenko, S. A., & Lin, B. M. T. (2017). Preemptive parallel-machine scheduling with a common server to minimize makespan. Naval Research Logistics, 64(5), 388–398. https://doi.org/10.1002/nav.21762.

Fanjul-Peyro, L. (2020). Models and an exact method for the unrelated parallel machine scheduling problem with setups and resources. Expert Systems with Applications: X, 5, Article 100022. https://doi.org/10.1016/j.eswax.2020.100022.

Fanjul-Peyro, L., Perea, F., & Ruiz, R. (2017). Models and matheuristics for the unrelated parallel machine scheduling problem with additional resources. European Journal of Operational Research, 260(2), 482–493. https://doi.org/10.1016/j.ejor.2017.01.002.

Fleszar, K., & Hindi, K. S. (2018). Algorithms for the unrelated parallel machine scheduling problem with a resource constraint. European Journal of Operational Research, 271(3), 839–848. https://doi.org/10.1016/j.ejor.2018.05.056.

García-de-Alba, H., Nucamendi-Guillén, S., Avalos-Rosales, O., & Ángel-Bello, F. (2025). The unrelated parallel machine scheduling problem with sequence- and machine-dependent setup times and a shared resource without overlap. *EURO Journal on Computational Optimization*, 13, Article 100117.

https://doi.org/10.1016/j.ejco.2025.100117.

Hamzadayı, A., & Arvas, M.A. (2026). Arc-based formulation and GRASP-enhanced iterated greedy algorithm for identical parallel machine scheduling with a common server. Swarm and Evolutionary Computation, 100, 102250. https://doi.org/10.1016/j.swevo.2025.102250.

Hadhbi, Y., Deroussi, L., Grangeon, N., & Norre, S. (2023). Improved formulations and Branch-and-Cut algorithm for the unrelated parallel machines scheduling problem with a common server and job-sequence dependent setup times. In Proceedings of the 9th International Conference on Control, Decision and Information Technologies (CoDIT 2023) (pp. 1506–1511). IEEE. https://doi.org/10.1109/CODIT58514.2023.10284149.

Hasani, K., Kravchenko, S. A., & Werner, F. (2014). Block models for scheduling jobs on two parallel machines with a single server. Computers & Operations Research, 41, 94–97. https://doi.org/10.1016/j.cor.2013.08.015.

Kerkhove, L. P., & Vanhoucke. M. (2014). Scheduling of unrelated parallel machines with limited server availability on multiple production locations: A case study in knitted fabrics. International Journal of Production Research, 52(9), 2630–2653. https://doi.org/10.1080/00207543.2013.865855.

Lei, D., & He, S. (2022). An adaptive artificial bee colony for unrelated parallel machine scheduling with additional resource and maintenance. Expert Systems with Applications, 205, Article 117577. https://doi.org/10.1016/j.eswa.2022.117577.

Li, M., Xiong, H., & Lei, D. (2022). An artificial bee colony with adaptive competition for the unrelated parallel machine

scheduling problem with additional resources and maintenance. Symmetry, 14(7), Article 1380. https://doi.org/10.3390/sym14071380.

Lin, S.-W., & Ying, K.-C. (2014). ABC-based manufacturing scheduling for unrelated parallel machines with machine-dependent and job sequence-dependent setup times. Computers & Operations Research, 51, 172–181. https://doi.org/10.1016/j.cor.2014.05.013.

Lopez-Esteve, A., Perea, F., & Yepes-Borrero, J. C. (2023). GRASP algorithms for the unrelated parallel machines scheduling problem with additional resources during processing and setups. International Journal of Production Research, 61(17), 6013–6029. https://doi.org/10.1080/00207543.2022.2121869.

Lozano, A. J., & Medaglia, A. L. (2014). Scheduling of parallel machines with sequence-dependent batches and product incompatibilities in an automotive glass facility. Journal of Scheduling, 17(6), 521–540. https://doi.org/10.1007/s10951-012-0308-7.

Manupati, V. K., Rajyalakshmi, G., Chan, F. T. S., & Thakkar, J. J. (2017). A hybrid multi-objective evolutionary algorithm approach for handling sequence- and machine-dependent setup times in unrelated parallel machine scheduling problem. Sādhanā, 42(3), 391–403. https://doi.org/10.1007/s12046-017-0611-2.

Özpeynirci, S., Gökgür, B., & Hnich, B. (2016). Parallel machine scheduling with tool loading. Applied Mathematical Modelling, 40(9), 5660–5671. https://doi.org/10.1016/j.apm.2016.01.006.

Pinedo, M. (2012). Scheduling: Theory, algorithms and systems. Springer, New York. A.B.D.

Potts, C. N., & Strusevich, V. A. (2009). Fifty years of scheduling: A survey of milestones. Journal of the Operational Research Society, 60(1), 41–68. https://doi.org/10.1057/jors.2009.2.

Qamhan, A. A., & Alharkan, I. M. (2019). Note on a two-stage adaptive fruit fly optimization algorithm for unrelated parallel machine scheduling problem with additional resource constraints. Expert Systems with Applications, 128, 81–83. https://doi.org/10.1016/j.eswa.2019.03.033.

Rabadi, G., Moraga, R. J., & Al-Salem, A. (2006). Heuristics for the unrelated parallel machine scheduling problem with setup times. Journal of Intelligent Manufacturing, 17(1), 85–97. https://doi.org/10.1007/s10845-005-5514-0.

Raboudi, H., Alpan, G., Mangione, F., Tissot, G., & Noel, F. (2022). Scheduling unrelated parallel machines with a common server and sequence dependent setup times. IFAC PapersOnLine, 55(10), 2179–2184. https://doi.org/10.1016/j.ifacol.2022.10.031.

Şaştım, Ö., & Hasgül, S. (2024). A mathematical model for the unrelated parallel machine scheduling problem with common server and process resource constraints. Journal of the Faculty of Engineering and Architecture of Gazi University, 39(1), 607–619. https://doi.org/10.17341/gazimmfd.1099034.

Su, B., Xie, N., & Yang, Y. (2021). Hybrid genetic algorithm based on bin packing strategy for the unrelated parallel workgroup scheduling problem. Journal of Intelligent Manufacturing, 32, 957–969. https://doi.org/10.1007/s10845-020-01597-8.

Torabi, S. A., Sahebjamnia, N., Mansouri, S. A., & Bajestani, M. A. (2013). A particle swarm optimization for a fuzzy

multi-objective unrelated parallel machines scheduling problem. Applied Soft Computing, 13(12), 4750–4762. https://doi.org/10.1016/j.asoc.2013.07.029.

Villa, F., Vallada, E., & Fanjul-Peyro, L. (2018). Heuristic algorithms for the unrelated parallel machine scheduling problem with one scarce additional resource. Expert Systems with Applications, 93, 28–38. https://doi.org/10.1016/j.eswa.2017.09.054.

Yepes-Borrero, J. C., Villa, F., Perea, F., & Caballero-Villalobos, J. P. (2020). GRASP algorithm for the unrelated parallel machine scheduling problem with setup times and additional resources. Expert Systems with Applications, 141, Article 112959. https://doi.org/10.1016/j.eswa.2019.112959.

Yunusoglu, P., & Yildiz, Ş. A. (2022). Constraint programming approach for multi-resource-constrained unrelated parallel machine scheduling problem with sequence-dependent setup times. International Journal of Production Research, 60(7), 2212–2229. https://doi.org/10.1080/00207543.2021.1885068.

Zheng, X. L., & Wang, L. (2016). A two-stage adaptive fruit fly optimization algorithm for unrelated parallel machine scheduling problem with additional resource constraints. *Expert Systems with Applications*, 65, 28–39. https://doi.org/10.1016/j.eswa.2016.08.039.

Zhang, L., Deng, Q., Lin, R., Gong, G., & Han, W. (2021). A combinatorial evolutionary algorithm for unrelated parallel machine scheduling problem with sequence- and machine-dependent setup times, limited worker resources and learning effect. Expert Systems with Applications, 175, Article 114843. https://doi.org/10.1016/j.eswa.2021.114843.

Zhang, Z., Zheng, L., Hou, F., & Li, N. (2011). Semiconductor final test scheduling with Sarsa (λ,k) algorithm. European Journal of Operational Research, 215(2), 446–458. https://doi.org/10.1016/j.ejor.2011.05.052.

Zhu, W., & Tianyu, L. (2019). A novel multi-objective scheduling method for energy based unrelated parallel machines with auxiliary resource constraints. IEEE Access, 7, 168688–168699. https://doi.org/10.1109/ACCESS.2019.2954601.

Zhu, Y., He, S., & Lei, D. (2024). Dynamical artificial bee colony for energy-efficient unrelated parallel machine scheduling with additional resources and maintenance. Computers, Materials & Continua, 81(1), 843–866. https://doi.org/10.32604/cmc.2024.054473.

SPHERICAL FUZZY SETS IN THEIR FIRST DECADE: A BIBLIOMETRIC AND THEMATIC ANALYSIS FROM OPEN CROSSREF DATA

Doğan ŞENGÜL[1]

1. INTRODUCTION

Since their near-simultaneous emergence in 2018 and 2019, spherical fuzzy sets have become an active area of fuzzy decision making. The concept developed through two closely related routes. Kutlu Gündoğdu and Kahraman (2019b) defined a spherical fuzzy set by membership, non-membership and hesitancy degrees whose squared sum is bounded by one and extended TOPSIS to the new environment, while Ashraf, Abdullah, Mahmood, Ghani and Mahmood (2019) developed a closely related formulation through a quadratic generalisation of picture fuzzy sets, with Mahmood, Ullah, Khan and Jan (2019) adding the T-spherical generalisation during the same period. Both routes build on the fuzzy set framework introduced by Zadeh (1965); together they contributed to the emergence of a research stream that now includes methods, theory and a wide range of applications.

The field-level structure of this stream remains less systematically documented than its individual methods and applications. One dedicated early survey, by Özceylan, Ozkan, Kabak and Dağdeviren (2021), reviewed 43 papers at an early stage of the field. Neighbouring fuzzy-set families have received

[1] PhD in Industrial Engineering. Assistant Professor, Department of Software Engineering, Faculty of Engineering and Natural Sciences, Istanbul Sabahattin Zaim University, Istanbul, Türkiye, ORCID: 0000-0002-2285-3907.

more quantitative mapping attention than spherical fuzzy research: Lin, Chen and Chen (2021) mapped the Pythagorean fuzzy literature and Shukla et al. (2020) mapped type-2 fuzzy sets and systems. Beyond the early survey, no dedicated open-metadata bibliometric map of the title-visible spherical fuzzy corpus was identified for the present chapter; this chapter offers such a map using open evidence: the records, counts and citation figures reported here come from the public Crossref registry (Hendricks, Tkaczyk, Lin, & Feeney, 2020), so that readers can re-run the harvest and check the reported numbers without relying on subscription databases.

This chapter addresses four questions. How fast has the field grown and through which publication channels? Which works carry most of its citation weight and what does their composition reveal about the two founding routes? Which methods, variants and application domains structure the literature thematically? Which issues merit attention in the coming years? The analysis follows the general guidance for bibliometric studies given by Donthu, Kumar, Mukherjee, Pandey and Lim (2021), adapted to an open data source, while reporting its own protocol in full so that the study is reproducible end to end.

2. DATA AND METHOD

The corpus is defined as Crossref records whose titles contain the phrase spherical fuzzy, with a hyphenated variant admitted by the screening pattern. Crossref was chosen as the data source because it is the registry where publishers themselves deposit metadata, it is community governed, it imposes no access barrier and it offers a public application interface (Hendricks et al., 2020). The harvest was run on 12 June 2026 through the works endpoint with a title term query, deterministic offset windows covering the first 9,992 relevance ranks and a client-side

screen that kept only records whose registered title matches the phrase. Saturation was monitored throughout: matches arrive in relevance bands and the final 1,376 consecutively scanned ranks contributed no new match, suggesting that the harvest had reached the practical limit of the queryable population. Twenty-eight ranks inside that empty tail zone could not be retrieved before the interface quota was reached; the surrounding band structure suggests that any remaining loss is likely to be very small.

Screening returned 1,031 unique records carrying 18,090 citations in the registry at harvest time. Three records predate the 2018 concept and concern spherical fuzzy c-means clustering, a separate earlier use of the same phrase in image segmentation; they are retained as a documented case of terminological overlap and excluded from thematic claims about the decision making field. Three limitations follow directly from the design and are stated openly. First, Crossref citation counts reflect the references deposited by publishers and therefore run lower than subscription indexes; they are used here for internal comparison, not as absolute impact claims. Second, author affiliations are deposited too unevenly for reliable geographic analysis (Hendricks et al., 2020), so country and institution rankings, a standard component of subscription-based bibliometrics, are not reported here. Third, the title screen captures only works that announce the concept in the title; studies that use spherical fuzzy tools under other titles fall outside the corpus, so every count reported here is a lower bound on the full literature and the corpus is best read as the title-visible Crossref literature of the field. Titles, venues, years, types and citation counts are sufficiently available to support the present analysis. Thematic classification assigns each title to method families, variant families and application domains through a transparent keyword rule set (the rules, the corpus file

and the reproducibility materials are available from the author upon reasonable request).

Table 1. The corpus at a glance

Indicator	Value
Records with spherical fuzzy in the title	1,031
Window covered	2014 to 12 June 2026
Total Crossref citations to the corpus	18,090
Corpus h-index	63
Works with at least 100 citations	28
Uncited works	266 (25.8 per cent)
Share of corpus published since the start of 2024	49.4 per cent
Journal articles	827 (80.2 per cent)
Book chapters	109 (10.6 per cent)
Conference papers and preprints	83
Other Crossref record types	12

Source: Author's computations from the Crossref corpus, harvested 12 June 2026.

3. GROWTH AND PUBLICATION STRUCTURE

Figure 1 shows the annual output. After the founding papers, recorded by the registry between 2018 and 2019, output more than quadrupled between 2019 and 2022 and kept climbing to 208 works in 2024. The 199 works of 2025 and the 102 works registered in the first five and a half months of 2026 indicate a continued high output rather than a clear decline: nearly half of the title-visible corpus appeared after the start of 2024. This indicates a rapid growth pattern among fuzzy set extensions; the Pythagorean fuzzy corpus mapped by Lin, Chen and Chen (2021) comprised 354 publications across its first eight years, a total the title-visible spherical fuzzy corpus passed in its sixth year.

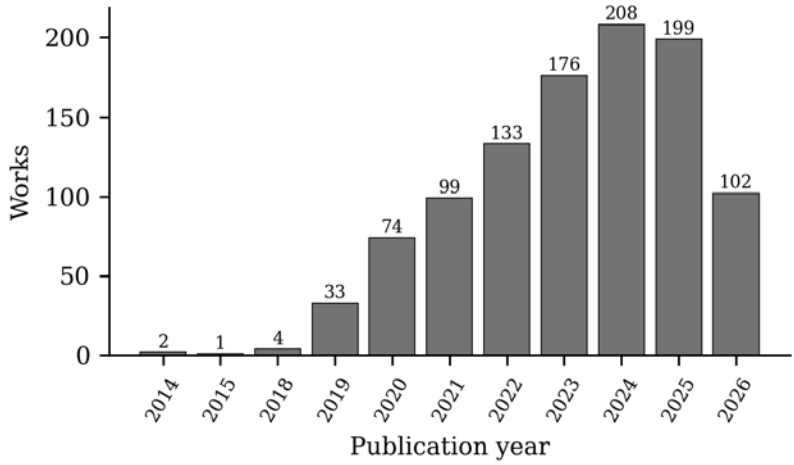


Figure 1. Annual output of spherical fuzzy research, 2014 to 12 June 2026 (author's computations from the Crossref corpus; 2026 is a partial year).

Table 2 sets out the exact counts behind Figure 1. The cumulative column shows the corpus crossing one hundred works in 2020, five hundred during 2023 and one thousand in the spring of 2026. The share column makes the recency of the field concrete: the three most recent registry years alone supply nearly half of the corpus recorded to date.

Table 2. Annual and cumulative output of the corpus

Year	Works	Cumulative	Share (%)
2014	2	2	0.2
2015	1	3	0.1
2018	4	7	0.4
2019	33	40	3.2
2020	74	114	7.2
2021	99	213	9.6
2022	133	346	12.9
2023	176	522	17.1
2024	208	730	20.2
2025	199	929	19.3
2026	102	1,031	9.9

Source: Author's computations from the Crossref corpus; 2026 covers 1 January to 12 June; the two clustering records of 2014 and the one of 2015 are included in the totals; no records carry the years 2016 or 2017.

The publication-channel structure is also informative. Approximately four fifths of the corpus consists of journal articles, while book chapters form the second-largest publication channel ahead of conference papers, with 109 items, many of them associated with Springer volumes on early spherical fuzzy research, above all the edited collection of Kahraman and Kutlu Gündoğdu (2021). Table 3 lists the ten most frequent publication venues. The Journal of Intelligent & Fuzzy Systems, the publication venue of two early journal papers on spherical fuzzy sets, leads with 51 items; broad open-access outlets such as IEEE Access and Scientific Reports account for part of the recent volume; Engineering Applications of Artificial Intelligence and Applied Soft Computing represent a substantial applied component of the field; the Lecture Notes in Networks and Systems series is associated with the INFUS conference community. The venue list shows a field publishing simultaneously in methodological journals, applied engineering journals and proceedings series, which suggests that the field continues to extend from its methodological core into applied areas.

Table 3. The ten most frequent publication venues in the corpus

Venue	Works
Journal of Intelligent & Fuzzy Systems	51
IEEE Access	40
Engineering Applications of Artificial Intelligence	38
Lecture Notes in Networks and Systems	38
Applied Soft Computing	30
Scientific Reports	30
Studies in Fuzziness and Soft Computing	26
Symmetry	23
Expert Systems with Applications	19
Soft Computing	18

Source: Author’s computations from the Crossref corpus.

4. HIGHLY CITED WORKS

Table 4 lists the ten most cited works in the corpus. Their composition is informative. The highest-cited record in the corpus is the medical diagnosis paper of Mahmood, Ullah, Khan and Jan (2019), which introduced the T-spherical generalisation; the early TOPSIS paper of Kutlu Gündoğdu and Kahraman (2019b) follows. Both early routes are therefore represented among the highest-cited records; several other highly cited works are also connected with these two research lines. Several works by Kutlu Gündoğdu, Kahraman and collaborators include methodological extensions that are associated with the development of spherical fuzzy decision making as a recognisable methodological framework: the analytic hierarchy process extension (Kutlu Gündoğdu & Kahraman, 2020), the interval-valued TOPSIS (Kutlu Gündoğdu & Kahraman, 2019a) and, through collaboration, the complex spherical VIKOR of Akram, Kahraman and Zahid (2021). Several works by Ashraf, Mahmood, Ullah and collaborators include operator-based and general decision-making studies, with the multi-attribute framework of Ashraf et al. (2019) and the Dombi operators of Ashraf, Abdullah and Mahmood (2020), next to the early T-spherical similarity measures of Ullah, Mahmood and Jan (2018). Application-oriented work is also represented in the top list, through the manufacturing system selection of Mathew, Chakrabortty and Ryan (2020) and the parsimonious hierarchy process of Moslem (2024), the only post-2021 work in the list, suggesting growing interest in more parsimonious methodological formulations.

Table 4. Highly cited works of the corpus, ranked by Crossref citation count

Rank	Work	Focus	Citations
1	Mahmood, Ullah, Khan and Jan (2019)	Spherical and T-spherical sets, medical diagnosis	811
2	Kutlu Gündoğdu and Kahraman (2019b)	Founding set definition and TOPSIS	687
3	Kutlu Gündoğdu and Kahraman (2020)	Analytic hierarchy process extension	396
4	Ashraf et al. (2019)	Multi-attribute decision framework	304
5	Kutlu Gündoğdu and Kahraman (2019a)	Interval-valued TOPSIS	280
6	Mathew, Chakrabortty and Ryan (2020)	AHP-TOPSIS, manufacturing selection	258
7	Akram, Kahraman and Zahid (2021)	Complex spherical VIKOR	232
8	Ashraf, Abdullah and Mahmood (2020)	Dombi aggregation operators	174
9	Ullah, Mahmood and Jan (2018)	T-spherical similarity measures	172
10	Moslem (2024)	Parsimonious hierarchy process	171

Source: Author's computations from the Crossref corpus; years follow the reference list, where online and print years differ the print year is used.

Citation concentration is higher among the leading records and lower across the wider corpus. The ten highly cited works account for 19.3 per cent of all citations in the corpus, the corpus h-index stands at 63 and 28 works have passed one hundred registry citations. At the lower-citation end, approximately one quarter of the corpus had no Crossref citations at the harvest date, which may reflect the recent publication date of many records and the growth of thematically related operator papers.

Table 5 quantifies the skew. The 28 works with at least one hundred registry citations make up 2.7 per cent of the corpus yet hold 31.8 per cent of all citations, while the 660 works cited fewer than ten times, 64.0 per cent of the corpus, hold 8.2 per cent between them. The band of ten to forty nine citations carries the

largest single share of citation weight, a pattern consistent with a recent literature whose knowledge base is still consolidating around its early reference points.

Table 5. Distribution of registry citations across the corpus

Citation band	Works	Share of works (%)	Citations	Share of citations (%)
0	266	25.8	0	0.0
1-9	394	38.2	1,485	8.2
10-49	280	27.2	6,447	35.6
50-99	63	6.1	4,411	24.4
100 and above	28	2.7	5,747	31.8

Source: Author's computations from the Crossref corpus, harvested 12 June 2026.

5. THEMATIC STRUCTURE

Three thematic layers emerge from title classification. The first layer is methodological. Aggregation operator papers form the largest group with 195 works, exceeding the count of each named decision-method family in the title-level classification, indicating that the production of new operators, from Dombi and Einstein families to Heronian and Hamy means, is highly visible in the title-level corpus. Among decision methods the hierarchy process leads with 103 works ahead of TOPSIS with 62, a pattern that may reflect the early arrival of a dedicated spherical fuzzy AHP (Kutlu Gündoğdu & Kahraman, 2020). Information measures, covering distance, similarity, entropy and correlation, account for 94 works. Objective and subjective weighting tools of the CRITIC, SWARA, best-worst and MEREC families appear in 58, while the remaining method families, from WASPAS and DEMATEL to CoCoSo, EDAS, CODAS and MARCOS, each hold between 8 and 24 works. Table 6 reports the full distribution.

Table 6. Method families signalled in corpus titles

Method family	Works	Method family	Works
Aggregation operators	195	VIKOR	18
Analytic hierarchy process	103	CoCoSo	16
Distance, similarity, entropy, correlation	94	CODAS	14
TOPSIS	62	Clustering and c-means	14
CRITIC, SWARA, BWM, MEREC weighting	58	MARCOS	13
FMEA and risk assessment	33	MOORA and MULTIMOORA	8
WASPAS	24	Quality function deployment	7
DEMATEL	23	EDAS	19

Source: Author's computations from the Crossref corpus; a work may signal several families.

The methodological mix changes over time. Table 7 splits the corpus into the founding period 2018 to 2022 and the recent period 2023 onwards and reports each family's share of period output. Three patterns are visible. Objective weighting tools of the CRITIC, SWARA, best-worst and MEREC families more than doubled their share, from 2.9 to 7.0 per cent. The relative share of AHP and TOPSIS was lower in the recent period, the hierarchy process falling from 12.5 to 8.8 per cent and TOPSIS from 7.6 to 5.3 per cent, while newer method families increased their relative shares: CoCoSo more than tripled its share from 0.6 to 2.0 per cent, DEMATEL more than doubled from 1.2 to 2.8 and WASPAS rose from 1.5 to 2.8. Operator-development papers maintained a share near one fifth of output in both periods, which suggests that operator-oriented research remains a recurring feature of the title-visible corpus rather than only a founding-period pattern.

Table 7. Method family shares in the founding and recent periods

Method family	2018-2022	Share (%)	2023-2026	Share (%)
Aggregation operators	62	18.1	133	19.4
Analytic hierarchy process	43	12.5	60	8.8
Distance, similarity, entropy, correlation	38	11.1	56	8.2
TOPSIS	26	7.6	36	5.3
CRITIC, SWARA, BWM, MEREC weighting	10	2.9	48	7.0
FMEA and risk assessment	8	2.3	25	3.6
WASPAS	5	1.5	19	2.8
DEMATEL	4	1.2	19	2.8
EDAS	6	1.7	13	1.9
VIKOR	9	2.6	9	1.3
CoCoSo	2	0.6	14	2.0
CODAS	3	0.9	11	1.6
MARCOS	4	1.2	9	1.3

Source: Author's computations from the Crossref corpus; period bases are 343 works for 2018-2022 and 685 works for 2023-2026; a work may signal several families; the three pre-2018 clustering records are excluded.

The second layer is structural. Figure 2 shows the diversification of the basic model. T-spherical generalisations appear in 236 titles, nearly a quarter of the corpus, with the T-spherical / q-power generalisation of Mahmood, Ullah, Khan and Jan (2019) appearing as the most widely represented extension in the corpus. Interval-valued forms follow with 89 works, rough set hybrids with 66, complex spherical models with 44, linguistic and 2-tuple forms with 44 and soft set hybrids with 32, while newer parametric generalisations are less frequently represented. The distribution of variants resembles earlier developments in intuitionistic and Pythagorean fuzzy-set research: the base model provides a basis for subsequent extensions and recent work increasingly emphasises hybrid forms.

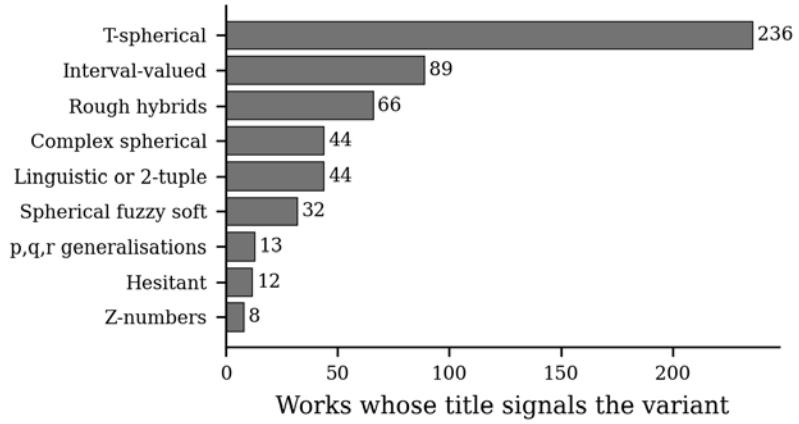


Figure 2. Variant and hybrid families signalled in corpus titles (author's computations from the Crossref corpus).

The third layer is applied. Environment and waste management lead with 193 title matches, followed by technology and digital transformation with 93, energy with 92, transport and logistics with 82 and health with 74. Manufacturing, finance, location and supplier problems fall within established operations research application areas, while education, pattern recognition and construction appear less frequently. Table 8 reports the counts. The domain profile is consistent with sustainability-oriented decision science: many applications are concentrated in areas where conflicting criteria, uncertain or linguistic expert judgement and public-sector or societal considerations are present, which is consistent with application settings discussed in the early papers.

Table 8. Application domains signalled in corpus titles

Domain	Works	Domain	Works
Waste, environment, sustainability	193	Finance and economics	55
Technology and digital	93	Site and location selection	43
Energy	92	Supplier and vendor selection	28
Transport and logistics	82	Education	26
Health and medical	74	Pattern recognition	17
Manufacturing and industry	65	Construction and infrastructure	12

Source: Author's computations from the Crossref corpus; a work may signal several domains.

6. A RESEARCH AGENDA FOR THE DECADE AHEAD

Five observations from the data lead to the following research agenda. First, the data suggest that the title-level corpus gives more visible emphasis to operator-related studies than to comparative validation and benchmark-based consolidation studies: The corpus contains 195 operator papers, while the title-visible set of comparative studies on how operator choice affects decision results is smaller. The corpus also includes many recently published records without Crossref citations at the harvest date. Future work could benefit from continued attention to benchmark studies, reporting standards and consolidation work, alongside additional operator-related developments. Second, the literature continues to use more than one notation and naming convention, while some records do not explicitly state the adopted convention; in the data this appears as parallel terminology for closely related mathematical objects. Explicit convention statements would make the literature easier to integrate. Third, variant-oriented research gives considerable attention to generalisation, while the complementary questions of parsimony, elicitation cost and interpretability are less visible in

the title-level evidence; the citation count of the parsimonious hierarchy process of Moslem (2024) may indicate interest in that direction. Fourth, the channel structure shows a developing field with a substantial presence in book chapters and proceedings; archival consolidation of conference contributions into journal syntheses could strengthen the evidence base. Fifth, open metadata can now support studies of this kind, but geographic patterns cannot be analysed reliably because affiliation deposition remains uneven (Hendricks et al., 2020); future field-level analyses would benefit from richer affiliation metadata deposited by publishers.

7. CONCLUSION

Within a decade of its near-simultaneous emergence, the title-visible Crossref corpus of spherical fuzzy research has passed one thousand works, has recently grown at roughly two hundred works per year, publishes four fifths of its output in journals while sustaining a relatively strong book-chapter channel, shows a citation structure associated with early set-definition papers and later methodological papers and includes both operator-focused methodological studies and sustainability-oriented applications. The analysis is designed to be reproducible: a single public registry, a reported harvest protocol, transparent classification rules and available data and reproducibility materials. These limitations follow from the use of open registry data and should be considered when interpreting the results. The early literature provided a methodological basis; the evidence reported here suggests that future work would benefit from clearer reporting standards, further consolidation and open, reproducible scholarship.

DECLARATIONS

Funding. This research received no external funding.

Conflicts of interest. The author declares no conflict of interest.

Data availability. The corpus of 1,031 records was harvested from the public Crossref interface on 12 June 2026. The corpus file, classification rules and reproducibility materials are available from the author upon reasonable request.

REFERENCES

Akram, M., Kahraman, C., & Zahid, K. (2021). Group decision-making based on complex spherical fuzzy VIKOR approach. Knowledge-Based Systems, 216, 106793. https://doi.org/10.1016/j.knosys.2021.106793

Ashraf, S., Abdullah, S., & Mahmood, T. (2020). Spherical fuzzy Dombi aggregation operators and their application in group decision making problems. Journal of Ambient Intelligence and Humanized Computing, 11(7), 2731-2749. https://doi.org/10.1007/s12652-019-01333-y

Ashraf, S., Abdullah, S., Mahmood, T., Ghani, F., & Mahmood, T. (2019). Spherical fuzzy sets and their applications in multi-attribute decision making problems. Journal of Intelligent & Fuzzy Systems, 36(3), 2829-2844. https://doi.org/10.3233/JIFS-172009

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. Quantitative Science Studies, 1(1), 414-427. https://doi.org/10.1162/qss_a_00022

Kahraman, C., & Kutlu Gündoğdu, F. (Eds.). (2021). Decision making with spherical fuzzy sets: Theory and applications. Springer. https://doi.org/10.1007/978-3-030-45461-6

Kutlu Gündoğdu, F., & Kahraman, C. (2019a). A novel fuzzy TOPSIS method using emerging interval-valued spherical fuzzy sets. Engineering Applications of Artificial

Intelligence, 85, 307-323. https://doi.org/10.1016/j.engappai.2019.06.003

Kutlu Gündoğdu, F., & Kahraman, C. (2019b). Spherical fuzzy sets and spherical fuzzy TOPSIS method. Journal of Intelligent & Fuzzy Systems, 36(1), 337-352. https://doi.org/10.3233/JIFS-181401

Kutlu Gündoğdu, F., & Kahraman, C. (2020). A novel spherical fuzzy analytic hierarchy process and its renewable energy application. Soft Computing, 24(6), 4607-4621. https://doi.org/10.1007/s00500-019-04222-w

Lin, M., Chen, Y., & Chen, R. (2021). Bibliometric analysis on Pythagorean fuzzy sets during 2013-2020. International Journal of Intelligent Computing and Cybernetics, 14(2), 104-121. https://doi.org/10.1108/IJICC-06-2020-0067

Mahmood, T., Ullah, K., Khan, Q., & Jan, N. (2019). An approach toward decision-making and medical diagnosis problems using the concept of spherical fuzzy sets. Neural Computing and Applications, 31(11), 7041-7053. https://doi.org/10.1007/s00521-018-3521-2

Mathew, M., Chakrabortty, R. K., & Ryan, M. J. (2020). A novel approach integrating AHP and TOPSIS under spherical fuzzy sets for advanced manufacturing system selection. Engineering Applications of Artificial Intelligence, 96, 103988. https://doi.org/10.1016/j.engappai.2020.103988

Moslem, S. (2024). A novel parsimonious spherical fuzzy analytic hierarchy process for sustainable urban transport solutions. Engineering Applications of Artificial Intelligence, 128, 107447. https://doi.org/10.1016/j.engappai.2023.107447

Özceylan, E., Ozkan, B., Kabak, M., & Dağdeviren, M. (2021). A survey on spherical fuzzy sets and clustering the

literature. In C. Kahraman, S. Cevik Onar, B. Oztaysi, I. U. Sari, S. Cebi, & A. C. Tolga (Eds.), Intelligent and fuzzy techniques: Smart and innovative solutions (pp. 87-97). Springer. https://doi.org/10.1007/978-3-030-51156-2_12

Shukla, A. K., Banshal, S. K., Seth, T., Basu, A., John, R., & Muhuri, P. K. (2020). A bibliometric overview of the field of type-2 fuzzy sets and systems. IEEE Computational Intelligence Magazine, 15(1), 89-98. https://doi.org/10.1109/MCI.2019.2954669

Ullah, K., Mahmood, T., & Jan, N. (2018). Similarity measures for T-spherical fuzzy sets with applications in pattern recognition. Symmetry, 10(6), 193. https://doi.org/10.3390/sym10060193

Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

EVALUATION OF SUSTAINABLE AGRICULTURE GOALS WITH THE SWARA-TOPSIS-ARAS APPROACHES

Melek IŞIK[1]

1. INTRODUCTION

Sustainable agriculture includes agricultural systems and practices that do not harm the agricultural ecosystem and allow nature to renew. The aim of sustainable agriculture, or in other words sustainable agricultural practices, is to produce safe food without consuming natural resources and without harming the natural environment. Another aim is to increase the economic level and quality of life of producers by ensuring rural development. Since the release of the Brundtland Report in 1987, sustainable agriculture has emerged as an important global concern (Tait and Morris, 2000). By highlighting areas of overlap and concern between evolving conceptions of sustainable agriculture, Velten et al. (2015) sought to expand understandings of sustainable agriculture. Janker et al.'s (2018) attempt to explain the distinctions between the political and scientific discourses'interpretations of sustainable agriculture was brought to light. A narrative systematic review, bibliometric analysis, and idea network analysis were used by Chami et al. (2020) to evaluate the potential of sustainable agriculture to strengthen agroecosystem resilience under climate change conditions.

[1] Assoc. Prof. Dr., Department of Industrial Engineering, Faculty of Engineering, Çukurova University, ORCID: 0000-0001-6078-7026.

The use of MCDM techniques in sustainable agriculture is growing. A hybrid mul-ti-criteria evaluation method was introduced by Radulescu et al. (2010) and used to assess the South Muntenia Region of Romania's performance from a sustainable agricultural perspective. The method merged the Analytical Hierarchy Process (AHP) and the TOPSIS. In order to analyze agricultural sustainability and examine the advantages and dis-advantages of the MCDM approach Elimination, Talukder et al. (2017) used it to a case study in coastal Bangladesh. In order to evaluate and compare the sustainability of various agricultural systems, Talukder et al. (2018) suggested that indicators for six important aspects of agricultural sustainability be used inside an MCDM structure. In their study, the agricultural systems of Bangladesh's southwest coast were assessed and rated according to their degree of agricultural sustainability. Rad et al. (2024) used the DEMATEL (Decision Making Trial and Evaluation Laboratory) approach to ascertain the impact of the criteria in the relation map. Iran was chosen as a case study because of the enormous difficulty this nation faces in supplying a sustainable electricity source for its agricultural needs. A study of the literature on the use of MCDM approaches to agricultural sustainability assessment was the goal of Cicciù et al. (2022). Kumar and Pant (2023) examined the models, data sources, and overall precision obtained utilizing the various sustainable agricultural performance criteria when applying AHP to a variety of farm-related challenges. The goal of Tirth et al. (2021) was to apply MCDM to create a model for sustainable agriculture practices in the Kingdom of Saudi Arabia (KSA). To determine which crop would be best, three MCDM approaches were used, and the outcomes were validated and compared. A framework for assessing and prioritizing various options for meeting agricultural water demand systems in the Poldasht basin, which has five dams in the West Azerbaijan Province, in line with

sustainable development standards was presented by Ardestani et al. (2020) by using TOPSIS method. Marks et al. (1995) studied that the aggregate of MCDM and fuzzy good judgment supplied a promising theoretical framework for the assessment of opportunity farming systems. Qureshi et al. (2018) targeted at the crop choice sample in Indian surroundings that taken into consideration complete standards associated with sustainable farming practices used fuzzy MCDM. Namiotko et al. (2022) evaluated agri-environmental scenario of decided on European Union (EU) nations the use of the MCDM techniques to become aware of the capability techniques for development of agricultural sports and environmental scenario. Deepa et al. (2019) proposed Modified Integrated Weighting (MIW) technique and Complex Proportional Assessment (COPRAS) technique to rank the given groundnut crop dataset. Junior et al. (2022) identified the literature through the lens of the sustainability triple bottom line and employed the Fuzzy MCDM method to determine which dimensions exert greater influence and are more prone to being affected by external factors. Cao and Solangi (2023) used MCDM strategies to research the demanding situations and possibilities dealing with sustainable agriculture in China`s economy. Poursaeed et al. (2010) carried out in Ilam Province of Iran to decide agricultural sustainability standards and surest partnership fashions for agricultural sustainability.

In this study, firstly, the strategies used in sustainable agriculture such as ecology-based strategy, knowledge & science, co-operation, economics-based strategy, adaptive management, subsidiarity, holistic & complex systems thinking were weighted with the SWARA method. Then, it was decided with the TOPSIS and ARAS methods that economic, environmental, social and overarching goals should be focused on in order to implement these strategies.

2. MATERIALS AND METHODS

In this study, firstly, the importance levels of sustainable agriculture strategies are determined with the SWARA MCDM method, and secondly, the goals of the problem are ranked with the TOPSIS and ARAS methods. The decision maker is chosen by three people who are experts in sustainable agriculture and the brainstorming technique scores are taken.

The SWARA method is a tool that allows expert opinions to be easily integrated into the process and allows simple relative comparisons to be used. The criteria are ranked from important to unimportant according to their importance, and the importance weights are taken into account by the score of each decision maker (Yurdoğlu and Kundakçı, 2017). The steps are as follows (Adalı and Ayşegül, 2017):

S. 1: It is supposed that the problem has n criteria (Cj, j=1,2,…,n) and k decision makers (DM_k, k=1,2,…,K).

S. 2: Each DM evaluates the standards primarily based totally on their very own revel in and knowledge. C1 represents the best criterion and Cn represents the lowest criterion.

S. 3: DM's provide a rating of 1.00 to the maximum essential criterion. The scores given to other criteria are compared with each other by taking the most important criterion as reference. All scores are given between 0 and 1 and in multiples of 5. The average of the comparative weights (s_j) for each criterion is determined by taking DM's.

S. 4: A coefficient (k_j) is calculated as shown in Eq (1).

$$k_j = \begin{cases} 1, & j = 1 \\ s_j + 1, & j > 1 \end{cases} \quad (1)$$

S. 5: The weight (w_j) value for each criterion is calculated using Eq (2). The value of the most important criterion is 1.

$$w_j = \begin{cases} 1 & , j = 1 \\ \frac{w_{j-1}}{k_j} & , j > 1 \end{cases} \quad (2)$$

S. 6: The final weights (q_j) of each criterion are calculated by dividing the criterion weights found in the previous step by the total of the criterion weights using Eq (3).

$$q_j = \frac{w_j}{\sum w_j} \quad (3)$$

TOPSIS was developed by Hwang and Yoon in 1981. Below are the steps of the method.

S. 1: The decision matrix is composed.

S. 2: The standard decision matrix is calculated the usage of the factors of the decision matrix and the usage of the subsequent Eq (4). For example, to calculate the yij element of matrix Y, the bij element of matrix B is received through dividing through the square root of the sum of squares of the only column factors of the matrix.

$$y_{ij} = \frac{b_{ij}}{\sqrt{\sum_{k=1}^{r} b_{kj}^2}} \quad (4)$$

S. 3: The weight values for w_i are determined by using SWARA method. Then, the weighted standard decision matrix consists with the aid of using multiplying the factors in every column of the usual decision matrix with the aid of using the corresponding wi value.

S. 4: In order to compose an ideal set (H*) of solutions, the most important of the assessment elements within the weighted standard decision matrix, are selected (Eq 5).

$$H^* = \left\{ (\max_i x_{ij} \,|\, j \in J), (\min_i x_{ij} \,|\, j \in J' \right\} \quad (5)$$

H set of negative ideal (H-) solutions is composed via way of means of deciding on the smallest of the assessment elements within side the weighted standard decision matrix (Eq 6).

$$H^{-} = \left\{ (\min_{i} x_{ij} \,\middle|\, j \in J), (\max_{i} x_{ij} \,\middle|\, j \in J' \right\} \tag{6}$$

Benefit (maximization) in both formulas, J' however, it shows the value of loss (minimization).

S. 5: The deviation values for the decision points are called the ideal distinction (T*) in Eq (7) and the negative ideal distinction (T-) in Eq (8) calculated measure.

$$T_i^{*} = \sqrt{\sum_{j=1}^{c} (x_{ij} - x_j^{*})^2} \tag{7}$$

$$T_i^{-} = \sqrt{\sum_{j=1}^{c} (x_{ij} - x_j^{-})^2} \tag{8}$$

S. 6: Ideal and negative ideal discrimination measures are used to determine the relative proximity (U*) of each choice point to an ideal solution in Eq (9).

$$U_i^{*} = \frac{T_i^{-}}{T_i^{-} + T_i^{*}} \tag{9}$$

ARAS method was developed by Zavadskas and Turskis (2010) presented as an approach to solving MCDM problems. ARAS method consists of four steps (Zavadskas and Turskis, 2010);

S.1: On the decision matrix x, x_{ij}; the performance value of alternative i of criterion j, x_{0j}; it represents the optimum value

of criterion j. If the optimum value of criterion j is not known, the optimal value is found using the formula in Eq (10).

$$x_{ij} = \max x_{ij}$$

$$x_{0j} = \min x_{0j} \tag{10}$$

S. 2: The normalization matrix is calculated of all criteria in Eq (11).

$$\bar{x}_{ij} = \frac{x_{ij}}{\sum_{i=0}^{m} x_{ij}} \tag{11}$$

If it is preferred that the performance value of the criterion be low, it is calculated in the Eq (12).

$$x^{*}_{ij} = \frac{1}{x_{ij}}, \quad \bar{x}_{ij} = \frac{x^{*}_{ij}}{\sum_{i=0}^{m} x^{*}_{ij}} \tag{12}$$

S. 3: According to the Eq (13), the weight of criterion j is w_j; represents the weighted normalized value x_{ij}.

$$\hat{x}_{ij} = \bar{x}_{ij} * w_{ij} \tag{13}$$

S. 4: The S_i value in the Eq (14) is the optimality function value of alternative i. S_0 optimal function value gives the K_i benefit degree and is calculated with the Eq (15).

$$S_i = \sum_{j=1}^{n} \hat{x}_{ij} \tag{14}$$

$$K_i = \frac{S_i}{S_0} \tag{15}$$

With the calculated K_i values, these values are ranked from largest to smallest and an evaluation of the decision alternatives is investigated.

3. RESULTS

The results of the MCDM analysis indicated a diverse set of sustainable agriculture strategies that can be organized into seven categories. The strategies are ecology-based strategy, knowledge & science, co-operation, economics-based strategy, adaptive management, subsidiarity, holistic & complex systems thinking (Velten et al., 2018). The strategies are weighted with the SWARA method. Then, it is used TOPSIS and ARAS methods that economic, environmental, social and overarching goals (Figure 1).

Figure 1. Goals and Strategies of Sustainable Agriculture (Velten et al., 2018)

Table 1 demonsrates the ranking of the three decision makers from important to unimportant strategies and their scores in multiples of 5. They identified the most important strategy as ecology-based strategy and the least important strategy as holistic & complex systems thinking.

Table 1. Scores of the decision-makers

Strategies	DM1	DM2	DM3
Ecology-based Strategy (S1)	1,00	1,00	1,00
Knowledge & Science (S2)	0,85	0,80	0,75
Co-operation (S3)	0,60	0,70	0,50
Economics-based Strategy (S4)	0,55	0,50	0,40
Adaptive Management (S5)	0,40	0,30	0,35
Subsidiarity (S6)	0,20	0,25	0,30
Holistic & Complex Systems Thinking (S7)	0,10	0,20	0,25

The S_j value consists of the average of the scores given by each decision maker about that strategy. The k_j, w_j, q_j value are analyzed according to Eq 1, Eq 2, Eq 3, respectively. As can be seen in Table 2, S7 has the least weight with a value of 0.04 after S1 with the largest weight value of 0.39.

Table 2. Weighted of strategies

Strategies	S_j	k_j	w_j	q_j
S1	1,00	1,00	1,00	0,39
S2	0,80	1,80	0,56	0,22
S3	0,60	1,60	0,35	0,14
S4	0,48	1,48	0,23	0,09
S5	0,35	1,35	0,17	0,07
S6	0,25	1,25	0,14	0,05
S7	0,18	1,18	0,12	0,04

The scores given between 1 and 9 regarding which strategies the goals in sustainable agriculture are more related to by the decision makers are given in Table 3. In addition, the weights found by the SWARA method have been added to the Table 3.

Table 3. Decision matrix for TOPSIS

Weight	0,39	0,22	0,14	0,09	0,07	0,05	0,04
Strategies/Goals	S1	S2	S3	S4	S5	S6	S7
Economic	7	8	5	9	6	8	7
Environmental	9	4	2	3	2	7	6
Social	5	7	6	7	8	5	9
Overarching	3	6	4	5	3	4	5

The standard decision matrix, the elements of the decision matrix is calculated in the Table 4 using Eq (4). The calculated values are used in the weighted standard decision matrix.

Table 4. Standard Decision Matrix for TOPSIS

Strategies/Goals	S1	S2	S3	S4	S5	S6	S7
Economic	0,55	0,62	0,56	0,70	0,56	0,64	0,51
Environmental	0,70	0,31	0,22	0,23	0,19	0,56	0,43
Social	0,39	0,54	0,67	0,55	0,75	0,40	0,65
Overarching	0,23	0,47	0,44	0,39	0,28	0,32	0,36

The weighted standard decision matrix is formed by multiplying each value of the Table 4 with weights from the SWARA method (Table 5). Then, according to the lowest and highest values in each column, the positive ideal (Eq 5) and negative ideal (Eq 6) numbers are found.

Table 5. Weighted Standard Decision Matrix for TOPSIS

Weight	0,39	0,22	0,14	0,09	0,07	0,05	0,04
Strategies/Goals	S1	S2	S3	S4	S5	S6	S7
Economic	0,21	0,14	0,08	0,06	0,04	0,03	0,02
Environmental	0,27	0,07	0,03	0,02	0,01	0,03	0,02
Social	0,15	0,12	0,09	0,05	0,05	0,02	0,03
Overarching	0,09	0,10	0,06	0,04	0,02	0,02	0,01
Ideal+	0,27	0,14	0,09	0,06	0,05	0,03	0,03
Ideal-	0,09	0,07	0,03	0,02	0,01	0,02	0,01

The positive and negative ideal values found with the help of Eqs 7 and 8 are given in Table 6. The ideal solution values are also reached with Eq 9.

Table 6. Ranking of Goals for TOPSIS

Goals	T+	T-	T++T-	T-/ T++T-	Rank
Economic	0,06	0,16	0,22	0,72	1
Environmental	0,11	0,18	0,29	0,63	2
Social	0,12	0,12	0,24	0,49	3
Overarching	0,19	0,05	0,24	0,21	4

Obtained the result of the calculations for TOPSIS method, it is determined that Economic>

Environmental>Social>Overarching is for sustainable agriculture from the most important goal to the least important.

Seven strategies are scored for four goals by using ARAS method in Table 7. The optimal row shows data for the highest values of the strategies.

Table 7. Decision Matrix for ARAS

Weight	0,39	0,22	0,14	0,09	0,07	0,05	0,04
Strategies/Goals	S1	S2	S3	S4	S5	S6	S7
Optimal	9	8	6	9	8	8	9
Economic	7	8	5	9	6	8	7
Environmental	9	4	2	3	2	7	6
Social	5	7	6	7	8	5	9
Overarching	3	6	4	5	3	4	5

In this step, normalization is performed to ensure ease of operation by reducing the magnitudes of the alternatives to lower levels in order to help them be compared with each other. Normalized values are obtained with Eq (11) in Table 8.

Table 8. Normalized Decision Matrix for ARAS

Weight	0,39	0,22	0,14	0,09	0,07	0,05	0,04
Strategies/Goals	S1	S2	S3	S4	S5	S6	S7
Optimal	0,27	0,24	0,26	0,27	0,30	0,25	0,25
Economic	0,21	0,24	0,22	0,27	0,22	0,25	0,19
Environmental	0,27	0,12	0,09	0,09	0,07	0,22	0,17
Social	0,15	0,21	0,26	0,21	0,30	0,16	0,25
Overarching	0,09	0,18	0,17	0,15	0,11	0,13	0,14

In the model weights are determined by the SWARA method, they are calculated with Eq 13 shown in Table 9.

Table 9. Weighted Normalized Decision Matrix for ARAS

Strategies/Goals	S1	S2	S3	S4	S5	S6	S7
Optimal	0,11	0,05	0,04	0,02	0,02	0,01	0,01
Economic	0,08	0,05	0,03	0,02	0,02	0,01	0,01
Environmental	0,11	0,03	0,01	0,01	0,01	0,01	0,01
Social	0,06	0,05	0,04	0,02	0,02	0,01	0,01
Overarching	0,04	0,04	0,02	0,01	0,01	0,01	0,01

After the weighted normalized decision matrix, the optimality function value is calculated. At this stage, S and K values are found using Eq 14 and 15 and the alternatives are listed in Table 10.

Table 10. Optimality Function Value for ARAS

	S	K	Ranking
Optimal	0,26		
Economic	0,23	0,86	1
Environmental	0,18	0,67	3
Social	0,20	0,76	2
Overarching	0,13	0,50	4

Obtained the result of the calculations for ARAS method, it is determined that is Economic > Social > Environmental > Overarching for sustainable agriculture from the most important goal to the least important.

4. CONCLUSION

In recent years, the concept of sustainability and its relationship with agriculture have been on the agenda. In this study, the strategies used in sustainable agriculture such as ecology-based strategy, knowledge & science, co-operation, economics-based strategy, adaptive management, subsidiarity, holistic & complex systems thinking are weighted with the SWARA method. Then, it is ranked the TOPSIS and ARAS methods that economic, environmental, social and overarching goals. Economic> Environmental>Social>Overarching is ranked for sustainable agriculture for TOPSIS.

Similar results are obtained in the ARAS method, only the social and environmental goals changed. It has become clear that the most important economic goals are the ones that need to be focused on, and the most important strategy is the ecology-based strategy. This shows that it is determined that it is important to protect nature and to realize this at a low cost. In

the continuation of this study, selection can be carried out with new criteria and alternatives and different MCDM methods.

KAYNAKÇA

Adalı, E. A., & Ayşegül, T. (2017). Bir Tedarikçi Seçim Problemi için Swara Ve Waspas Yöntemlerine Dayanan Karar Verme Yaklaşımı. International Review of Economics and Management, 5(4), 56-77.

Cao, J., & Solangi, Y. A. (2023). Analyzing and prioritizing the barriers and solutions of sustainable agriculture for promoting sustainable development goals in China. Sustainability, 15(10), 8317.

Cicciù, B., Schramm, F., & Schramm, V. B. (2022). Multi-criteria decision making/aid methods for assessing agricultural sustainability: A literature review. Environmental Science & Policy, 138, 85-96.

Deepa, N., Ganesan, K., Srinivasan, K., & Chang, C. (2019). Realizing sustainable development via modified integrated weighting MCDM model for ranking agrarian dataset. Sustainability 11 (21), 6060.

Ebad Ardestani, M., Sharifi Teshnizi, E., Babakhani, P., Mahdad, M., & Golian, M. (2020). An optimal management approach for agricultural water supply in accordance with sustainable development criteria using MCDM (TOPSIS)(Case study of Poldasht catchment in West Azerbaijan Province-Iran). Journal of Applied Water Engineering and Research, 8(2), 88-107.

El Chami, D., Daccache, A., & El Moujabber, M. (2020). How can sustainable agriculture increase climate resilience? A systematic review. Sustainability, 12(8), 3119.

Hwang, C.-L., & Yoon, K. (1981). Methods for multiple attribute decision making. Multiple attribute decision making: methods and applications a state-of-the-art survey, 58-191.

Janker, J., Mann, S., & Rist, S. (2018). What is sustainable agriculture? Critical analysis of the international political discourse. Sustainability, 10(12), 4707.

Junior, M. B., Pinheiro, E., Sokulski, C. C., Huarachi, D. A. R., & de Francisco, A. C. (2022). How to Identify Barriers to the Adoption of Sustainable Agriculture? A Study Based on a Multi-Criteria Model. Sustainability, 14(20), 13277.

Kumar, A., & Pant, S. (2023). Analytical hierarchy process for sustainable agriculture: An overview. MethodsX, 10, 101954.

Marks, L., Dunn, E., Keller, J., & Godsey, L. (1995). Multiple criteria decision making (MCDM) using fuzzy logic: an innovative approach to sustainable agriculture. Paper presented at the Proceedings of 3rd International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society.

Namiotko, V., Galnaityte, A., Krisciukaitiene, I., & Balezentis, T. (2022). Assessment of agri-environmental situation in selected EU countries: a multi-criteria decision-making approach for sustainable agricultural development. Environmental Science and Pollution Research, 29(17), 25556-25567.

Poursaeed, A., Mirdamadi, M., Malekmohammadi, I., & Hosseini, J. F. (2010). The partnership models of agricultural sustainable development based on Multiple Criteria Decision Making (MCDM) in Iran. African Journal of Agricultural Research, 5(23), 3185-3190.

Qureshi, M. R. N., Singh, R. K., & Hasan, M. A. (2018). Decision support model to select crop pattern for

sustainable agricultural practices using fuzzy MCDM. Environment, development and sustainability, 20, 641-659.

Rad, M. A. V., Fard, H. F., Khazanedari, K., Toopshekan, A., Ourang, S., Khanali, M., . . . Kasaeian, A. (2024). A global framework for maximizing sustainable development indexes in agri-photovoltaic-based renewable systems: Integrating DEMATEL, ANP, and MCDM methods. Applied Energy, 360, 122715.

Radulescu, C. Z., Rahoveanu, A. T., & Radulescu, M. (2010). A hybrid multi-criteria method for performance evaluation of romanian South Muntenia Region in context of sustainable agriculture. Paper presented at the Proceedings of the international conference on applied computer science.

Tait, J., & Morris, D. (2000). Sustainable development of agricultural systems: competing objectives and critical limits. Futures, 32(3-4), 247-260.

Talukder, B., Blay-Palmer, A., Hipel, K. W., & VanLoon, G. W. (2017). Elimination method of multi-criteria decision analysis (mcda): A simple methodological approach for assessing agricultural sustainability. Sustainability, 9(2), 287.

Talukder, B., Hipel, K. W., & vanLoon, G. W. (2018). Using multi-criteria decision analysis for assessing sustainability of agricultural systems. Sustainable Development, 26(6), 781-799. https://doi.org/10.1002/sd.1848

Tirth, V., Singh, R. K., Islam, S., & Algahtani, A. (2021). Optimal selection of agricultural crops for sustainable

farming in nutrient-rich Wadi areas of Saudi Arabia using MCDM. Journal of Engineering Research, 9(4B).

Velten, S., Leventon, J., Jager, N., & Newig, J. (2015). What is sustainable agriculture? A systematic review. Sustainability, 7(6), 7833-7865.

Yurdoğlu, H., & Kundakcı, N. (2017). SWARA ve WASPAS yöntemleri ile sunucu seçimi. Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 20(38), 253-270.

Zavadskas, E. K., & Turskis, Z. (2010). A new additive ratio assessment (ARAS) method in multicriteria decision-making. Technological and Economic Development of Economy, 16(2), 159-172.

A LITERATURE REVIEW ON DATA-ORIENTED APPROACHES TO ENERGY CONSUMPTION FORECASTING[1]

Vildan ARSLANTÜRK[2]

Betül TURANOĞLU ŞİRİN[3]

1. INTRODUCTION

In recent years, a significant share of energy consumption has originated from buildings. Buildings are considered one of the primary contributors to global energy consumption and greenhouse gas emissions. This situation highlights the need for energy-efficient building designs to ensure environmental sustainability. Energy waste in buildings leads not only to economic losses but also to climate change, air pollution, and the excessive use of natural resources. With population growth and rapid urbanization, building-related energy consumption continues to increase annually, making effective energy management planning even more essential (Olu-Ajayi, Alaka, Sulaimon, Sunmola, & Ajayi, 2022). Accurately forecasting energy consumption in buildings is a key tool for enhancing the effectiveness of energy efficiency policies. However, energy consumption is influenced by numerous factors. Variables such as weather conditions, building characteristics, occupant

---

[1] This paper is derived from the master's thesis completed by Vildan ARSLANTÜRK at the Department of Industrial Engineering, Institute of Science, Atatürk University.

[2] Master's student, Atatürk University, Faculty of Engineering, Department of Industrial Engineering, ORCID: 0009-0005-7349-6099.

[3] Associate Professor, Atatürk University, Faculty of Engineering, Department of Industrial Engineering, ORCID: 0000-0002-7910-6312.

behavior, and seasonal variations result in a complex and nonlinear energy demand structure. Therefore, energy forecasting studies require sophisticated analytical approaches and often necessitate methods beyond traditional forecasting techniques (Olu-Ajayi et al., 2022).

Accurate energy demand forecasting is of strategic importance for maintaining the balance between energy supply and demand, planning production facilities, and ensuring energy security at the national level. Overestimating demand may lead to resource waste and unnecessary investments, whereas underestimating demand can result in energy shortages and supply disruptions (Hwang, Suh, & Otto, 2020). Therefore, the use of reliable forecasting models in energy management constitutes a cornerstone of both economic and environmental sustainability.

Energy represents a substantial expenditure item for both households and national economies. The General Circular on Energy Conservation Measures No. 2024/7, published in the Official Gazette No. 32549 on May 17, 2024, introduced a series of measures aimed at reducing energy consumption and raising awareness of energy efficiency in public institutions in Turkey. In this context, accurately identifying energy needs and developing data-driven plans constitute the first and most critical step toward the successful implementation of energy conservation policies.

2. LITERATURE REVIEW

Numerous studies have been conducted in the literature on electricity consumption forecasting. Some of these studies are summarized below:

Hamzaçebi and Kutay (2004) investigated the applicability of artificial neural network (ANN) techniques for forecasting electricity consumption. In their study, annual electricity consumption and population data for Turkey covering the period 1970–2002 were used to forecast electricity consumption up to 2010. The forecasting results were compared with those obtained from regression analysis and Box–Jenkins models, and the findings indicated that ANN techniques provide an effective tool for electricity consumption forecasting.

Bilgili (2009) used Turkey's electricity consumption data for the period 1990–2007 and conducted a five-year forecasting study extending to 2012 under two scenarios: low and high consumption. Linear regression, nonlinear regression, and ANN methods were employed, and the results were compared with each other, official projections from the Ministry of Energy and Natural Resources, and findings reported in previous studies. Among the developed models, the ANN model achieved the highest forecasting accuracy.

Oğcu et al. (2012) utilized monthly electricity consumption data for Turkey spanning January 1970 to December 2011 and employed support vector regression (SVR) and ANN models to forecast electricity consumption for the period 2010–2011. The study proposed a seasonal SVR model, which had received limited attention in previous research, and demonstrated that the seasonal SVR model outperformed the ANN model in terms of forecasting performance.

Cao et al. (2014) developed an energy demand forecasting model by combining support vector regression (SVR) and quantum-behaved particle swarm optimization (QPSO) techniques using eight economic input variables. The results demonstrated that the proposed model outperformed several

widely used forecasting models in predicting energy demand in China.

Pérez-García and Moral-Carcedo (2016) proposed an alternative analytical framework for electricity demand forecasting based on a simple growth-rate decomposition scheme that incorporates multiple determinants of electricity demand. The authors argued that this framework can serve as a foundation for developing long-term forecasting models capable of generating electricity demand projections while accounting for the expected evolution of relevant factors. Spain was used as a case study to illustrate the proposed methodology and to obtain electricity demand projections through 2030.

Başoğlu and Bulut (2017) developed a hybrid forecasting system that integrates artificial neural networks and expert systems while considering Turkey's market and seasonal conditions. Using this hybrid system, named EPSIM-NM, electricity demand forecasts were generated and compared with actual consumption values. The results revealed a very high level of forecasting accuracy. The authors also suggested that the system could be further enhanced to produce weekly, monthly, and annual demand forecasts.

He and Lin (2018) aimed to identify the most appropriate model for forecasting China's energy structure and demand and developed an autoregressive distributed lag mixed-data sampling (ADL-MIDAS) model to support the analysis of future energy scenarios and carbon emissions in China and other developing countries. To construct the model, mixed-frequency data, including quarterly GDP, quarterly value-added, and annual energy consumption data from various industrial sectors, were utilized. Different combinations of weighting functions and forecasting methods were evaluated, and the model with the best forecasting performance was selected. The results indicated that

the forecasting errors of the selected model ranged from 0.02% to 2%, demonstrating the effectiveness of the ADL-MIDAS approach in forecasting China's energy demand.

De Oliveira and Oliveira (2018) applied a combination of bootstrap aggregating (bagging) and forecasting methods to the electric power sector to improve the accuracy of electricity demand estimation. An out-of-sample comparative analysis of monthly electricity consumption data from different countries was conducted using time-series techniques. The results demonstrated that the proposed methodologies significantly enhanced the accuracy of final consumer demand forecasts in both developed and developing countries.

Li and Jones (2019) proposed a point-process model based on extreme value theory to estimate the maximum demand of an electrical substation. The model incorporated the number of customers, average demand, and installed photovoltaic capacity as explanatory variables within the trend function. The authors noted that forecasting energy consumption and maximum demand separately in utility systems often results in inconsistent trends and suboptimal policy decisions. To address this issue, the proposed approach provides both realistic and flexible maximum demand forecasts while ensuring consistency between energy consumption and peak demand forecasting outcomes.

Uzlu and Dede (2020) developed an artificial neural network (ANN) model trained using the Jaya Algorithm (JA) based on Turkey's electricity consumption data covering the period 1980–2014. The model was employed to forecast electricity consumption through 2023 under two different scenarios. The forecasting results were compared with projections published by TEİAŞ and findings reported in previous studies. The results indicated that the proposed model offers an

effective and advantageous approach for electricity consumption forecasting.

Kazemzadeh et al. (2020) emphasized the importance of load forecasting for power system operation planning and capacity expansion. In their study, a long-term hybrid forecasting model integrating time-series analysis and data mining techniques was proposed to capture the nonlinear and complex patterns of energy demand and annual peak load data. First, the parameters of the support vector regression (SVR) model and the size of the input dataset were optimized using Particle Swarm Optimization (PSO), resulting in the development of an SVR-based forecasting algorithm. Subsequently, a hybrid model combining artificial neural networks (ANN), autoregressive integrated moving average (ARIMA), and the proposed SVR approach was developed to reduce forecasting errors in long-term annual peak load and total electricity demand predictions. The proposed hybrid framework dynamically prioritizes forecasting methods according to their prediction errors. The model was further applied to forecast total energy demand and annual peak load in the Iranian National Electric Power System.

Zhang et al. (2020) proposed a novel approach for forecasting half-hourly building energy consumption (BEC). The proposed method decomposes energy consumption data into cyclic and stochastic components to fully exploit the characteristics of the available data and improve forecasting accuracy. A hybrid model was subsequently developed to model these two components separately. The cyclic component was identified through spectral analysis, whereas the stochastic component was approximated using a novel ensemble model that combines a Deep Belief Network (DBN) and an Extreme Learning Machine (ELM), referred to as DEEM. To validate the proposed DEEM + CF model, energy consumption experiments were conducted on two operational buildings. The performance

of the model was compared with those of DBN, DBN + CF, ELM, ELM + CF, Support Vector Regression (SVR), and SVR + CF models. The results demonstrated that incorporating the cyclic component improved forecasting accuracy by approximately 20%. Furthermore, according to the Mean Absolute Error (MAE) criterion, the DEEM + CF model achieved at least 3%, 6%, and 10% higher forecasting accuracy than the DBN + CF, ELM + CF, and SVR + CF models, respectively, making it the best-performing model among those evaluated.

Liu et al. (2020) developed a Support Vector Machine (SVM)-based model to forecast the energy consumption of public buildings using 11 input variables, including temporal factors, climatic conditions, and historical energy consumption data. The study focused on the period during which air-conditioning systems were actively used in Wuhan. Data from June and July were used for model training, August data were used for testing, and September data were utilized to identify abnormal air-conditioning energy consumption. The results revealed abnormal energy consumption patterns during a four-day period in September. Based on a causal analysis of the findings, the authors proposed several policy recommendations.

Fan et al. (2020) developed a novel hybrid model, termed EMD-SVR-PSO-AR-GARCH, for electricity consumption forecasting. The model integrates Empirical Mode Decomposition (EMD), Support Vector Regression (SVR), Particle Swarm Optimization (PSO), thermal reaction dynamics theory, and autoregressive generalized autoregressive conditional heteroskedasticity (AR-GARCH) techniques. By combining these approaches, the study introduced a new perspective on electricity consumption and the economic behavior associated with energy use. The proposed model was applied to electricity consumption data from the New South Wales (NSW) electricity market in Australia. In addition, complex electricity consumption

patterns and consumer economic behaviors were analyzed using Nash equilibrium and Porter's Five Forces framework to support the sustainable development of the electricity sector.

Jamil (2020) forecasted Pakistan's hydroelectric energy consumption through 2030 using 53 years of historical data and the autoregressive integrated moving average (ARIMA) model. The reliability of the proposed model was assessed by comparing the forecasting results with actual observations, which demonstrated a good fit and minimal deviation. The forecasting results were further compared with the Pakistani government's hydroelectric power generation plans. In addition, a sensitivity analysis was conducted to examine the relationship between hydroelectric energy consumption, annual population growth, and gross domestic product (GDP) growth. The findings indicated that the proposed approach could contribute significantly to the effective planning and management of Pakistan's water and energy resources.

Wang et al. (2020) proposed a novel Long Short-Term Memory (LSTM)-based model for forecasting periodic energy consumption. Initially, latent features were extracted from real industrial data using autocorrelation analysis, and the time variable was refined to ensure complete periodicity. Subsequently, an LSTM network was developed to model and forecast sequential energy consumption data. Experimental results obtained from a cooling system dataset demonstrated that the proposed model outperformed several conventional forecasting approaches, including Backpropagation Neural Networks (BPNN), Autoregressive Moving Average (ARMA), and Autoregressive Fractionally Integrated Moving Average (ARFIMA) models. Analysis of the test data revealed that the Root Mean Square Error (RMSE) of the LSTM model was 19.7%, 54.85%, and 64.59% lower than those of the BPNN, ARMA, and ARFIMA models, respectively.

Somu et al. (2021) introduced the kCNN-LSTM deep learning model for accurate building energy consumption forecasting using energy consumption data recorded at predefined intervals. The proposed model was evaluated using real-time energy consumption data collected from a building at the Indian Institute of Technology (IIT) Mumbai, India. The results demonstrated the effectiveness and practical applicability of the model. Its performance was compared with those of state-of-the-art energy demand forecasting approaches and their k-means-based variants using widely accepted forecasting performance metrics. The findings indicated that the model's ability to capture spatio-temporal dependencies within energy consumption data significantly improved forecasting accuracy, highlighting its suitability for building energy consumption forecasting applications.

Hu et al. (2021) proposed a first-order univariate Grey Model, GM(1,1), for energy demand forecasting without relying on conventional statistical assumptions. To enhance the forecasting accuracy of the GM(1,1) model, upper and lower prediction bounds were generated for individual observations using nonlinear interval regression analysis based on neural networks. These bounds were subsequently incorporated into a nonlinear interval grey forecasting framework, in which residual variables were modeled separately. Empirical results obtained from real-world energy demand datasets demonstrated that the proposed models outperformed existing grey forecasting models and exhibited strong applicability for energy demand forecasting.

Dash et al. (2021) proposed a Recurrent Neural Network–Gradient Boosting Regression Tree (RNN-GBRT) model for electricity demand forecasting. The model analyzes consumers' daily energy consumption patterns to improve forecasting accuracy and support demand management efforts. Its performance was compared with those of several conventional

forecasting models, demonstrating its effectiveness. In addition, the study addressed the monitoring of electricity consumption data to detect irregularities such as power theft and meter tampering, which can lead to substantial financial losses for utility providers. By identifying abnormal and sudden changes in meter readings, the proposed approach enhances data consistency monitoring and facilitates the distinction between electricity theft and meter manipulation. This issue is particularly significant in developing and underdeveloped economies, such as India. The experimental results confirmed the effectiveness and reliability of the proposed model.

Zhou et al. (2023) proposed an integrated forecasting framework based on Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to predict both electricity generation and electricity consumption. The primary objective was to mitigate the effects of supply–demand uncertainty on energy system operations. Based on the forecasting results, the authors further developed an interactive energy supply–demand model capable of dynamically matching electricity generation and consumption. Experimental evaluations verified the effectiveness of the proposed framework. The results demonstrated that accurate forecasting of electricity generation and consumption can significantly reduce supply–demand uncertainty and contribute to more transparent, efficient, stable, and sustainable energy system operations. The study also incorporated the commercial preferences of electricity producers and consumers, aiming to reduce electricity costs while improving overall economic benefits within the energy market.

Mikayilov et al. (2023) examined industrial electricity consumption in Saudi Arabia to forecast regional-level electricity demand. Structural time-series modeling was applied using annual data covering the period 1990–2019. The results revealed that electricity demand levels, as well as long-run price and

income elasticities, differ across regions. Analysis of the underlying energy demand trend indicated efficiency-related improvements in industrial electricity consumption in all regions. Based on the baseline scenario, electricity demand for 2030 was projected to reach 11.6 TWh, 63.5 TWh, and a total of 82.5 TWh across the studied regions. These findings provide useful insights for regional energy planning and policy development.

Li et al. (2024) proposed a hybrid model for electricity demand forecasting. First, noise in the electricity demand series was removed using Kalman filtering. The demand series was then decomposed using Empirical Mode Decomposition (EMD), and forecasting was performed using a Genetic Algorithm (GA)-optimized Support Vector Machine (SVM). The performance of the proposed model was evaluated using real electricity demand data from Australia. The results indicated that the model achieved the highest forecasting accuracy, with a Mean Absolute Percentage Error (MAPE) of 0.25%. Compared with conventional SVM, EMD-SVM, and Kalman Filter-SVM models, forecasting accuracy improved by 74%, 73%, and 51%, respectively. Furthermore, the proposed approach outperformed advanced deep learning models, including LSTM, Transformer, and LSTM-AdaBoost, by 63%, 47%, and 36%, respectively. These findings confirm the effectiveness and superior forecasting capability of the proposed hybrid framework.

Aldarraji et al. (2024) utilized data obtained from the Iraqi Ministry of Electricity for the period 2019–2021 to forecast Iraq's electricity demand and supply. The study evaluated the performance of several advanced forecasting techniques, including Linear Regression (LR), Random Forest (RF), XGBoost, Long Short-Term Memory (LSTM), Temporal Convolutional Networks (TCN), and Multilayer Perceptron (MLP) models. Historical data from the previous 24, 48, 72, and 168 hours were used across four forecasting horizons. The results

indicated that the LR model achieved the best performance in demand forecasting, whereas XGBoost produced the most accurate supply forecasts. The authors emphasized the importance of accurate energy forecasting for enhancing Iraq's energy security, resource allocation, and policy development. In addition, the study provided valuable insights into the dynamics of Iraq's electricity supply and demand, supporting effective energy planning, management, and sustainable development.

Nikseresht and Amindavar (2024) proposed a novel adaptive hybrid framework for energy consumption forecasting based on statistical modeling techniques. To demonstrate the effectiveness of the proposed approach, the authors employed widely used energy demand forecasting methods as benchmark models. The Autoregressive Fractionally Integrated Moving Average (ARFIMA) model was adopted as the core forecasting framework. A dynamic statistical structural-break detection procedure was developed to identify change points and their corresponding time indices within the time series. Subsequently, the adaptive ARFIMA model was applied to the segmented time series, and the resulting forecasts were combined to generate the final prediction. The performance of the proposed model was evaluated using the Mean Absolute Percentage Error (MAPE) metric. For four real-world case studies, the proposed adaptive ARFIMA model achieved MAPE values of 1.72%, 1.35%, 0.22%, and 4.40%, respectively. The findings demonstrated that the proposed model outperformed existing forecasting approaches and provided highly accurate energy demand forecasts.

Erten and İnanç (2024) investigated load forecasting models within Demand Side Management (DSM)-based Building Energy Management Systems (BEMS) using a commercial building in Ankara, Turkey, as a case study. Deep Learning (DL) models were employed to enhance energy efficiency and provide

inputs for rule-based control mechanisms. The primary contribution of the study was the development of the ANFIS-IC algorithm, which aims to maximize demand-side participation in commercial buildings. By integrating Adaptive Neuro-Fuzzy Inference System (ANFIS) controllers with LSTM-based load forecasting, the proposed ANFIS-IC algorithm achieved a 33.14% reduction in energy consumption and a 39.22% reduction in energy costs. These results exceeded those obtained using rule-based controllers alone, which reduced energy consumption and energy costs by 25.34% and 34.03%, respectively.

Cen and Lim (2024) proposed a multi-task learning framework that integrates Patch Embedding, Temporal Convolutional Networks (TCN), and Time Series Transformers (TST), referred to as PatchTCN-TST. The proposed framework employs a channel-independent strategy for forecasting indoor environmental conditions and multiple electrical loads at the floor level. The PatchTCN-TST model was evaluated using data collected from an office building in Bangkok, Thailand, and was used to generate forecasts with one-, two-, and three-step prediction horizons. Empirical results demonstrated that the proposed model outperformed widely used forecasting approaches, including LSTM, GRU, TCN, Transformer, Informer, and Autoformer models. Compared with the best-performing baseline model, PatchTCN-TST achieved improvements of 34%, 23%, 12%, and 36.4% in MAE, MSE, RMSE, and aSMAPE metrics, respectively. These results confirm the superior forecasting accuracy of the proposed framework across all forecasting scenarios.

Li et al. (2024) proposed a novel neural network-based optimization framework for energy consumption forecasting. Initially, consumer energy demand was estimated using Deep Belief Networks (DBNs). Subsequently, the structure and performance of the DBN model were enhanced through the

Gorilla Troops Optimization (GTO) algorithm. The resulting hybrid model, referred to as DBN-GTO, was then evaluated using a standard case study involving short-, medium-, and long-term energy demand forecasting scenarios. The forecasting results were compared with those of several previously published methods. The findings demonstrated that the proposed model achieved competitive forecasting performance and can be considered an effective approach for energy consumption forecasting.

3. CONCLUSION

Energy consumption forecasting remains a critical research area for the efficient management of energy resources, cost reduction, and the development of sustainable energy policies. The studies reviewed show that machine learning and deep learning methods can provide high accuracy levels in energy consumption forecasting compared to traditional statistical approaches. In particular, ANN, SVM, RF, LSTM, GRU, and hybrid models have been found to be widely used in energy consumption forecasting.

The literature has determined that model performance depends significantly not only on the algorithm used but also on data quality, data preprocessing processes, feature selection, and forecast horizon. Furthermore, it is noteworthy that there has been a significant increase in the use of deep learning-based approaches in recent years, and successful results have been obtained in the analysis of large volumes of time series data.

However, it is difficult to make direct comparisons between studies due to the use of different datasets and performance metrics. Future research focusing on explainable artificial intelligence, hybrid models, ensemble learning methods, and real-time prediction systems can make significant

contributions to the field of energy consumption forecasting. The widespread adoption of machine learning applications, particularly in public institutions to improve energy management, will offer significant economic and environmental benefits. In this context, it is assessed that machine learning techniques are effective and reliable tools in the field of energy consumption forecasting, and that applications in this area will become even more widespread in the future with the development of data analytics and artificial intelligence technologies.

REFERENCES

Aldarraji, M., Vega-Márquez, B., Pontes, B., Mahmood, B., & Riquelme, J. C. (2024). Addressing energy challenges in Iraq: Forecasting power supply and demand using artificial intelligence models. *Heliyon, 10,* e25821.

Başoğlu, B., & Bulut, M. (2017). Kısa dönem elektrik talep tahminleri için yapay sinir ağları ve uzman sistemler tabanlı hibrit sistem geliştirilmesi. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 32(2),* 575–583.

Bilgili, M. (2009). Estimation of net electricity consumption of Turkey. *Isı Bilimi ve Tekniği Dergisi, 29(2),* 89–98.

Cao, Z., Yuan, P., & Ma, Y. B. (2014). Energy demand forecasting based on economy-related factors in China. *Energy Sources, Part B: Economics, Planning, and Policy, 9(2),* 214–219.

Cen, S., & Lim, C. G. (2024). Multi-task learning of the PatchTCN-TST model for short-term multi-load energy forecasting considering indoor environments in a smart building. *IEEE Access, 12,* 19553–19568.

Dash, S. K., Roccotelli, M., Khansama, R. R., Fanti, M. P., & Mangini, A. M. (2021). Long-term household electricity demand forecasting based on RNN-GBRT and a novel energy theft detection method. *Applied Sciences, 11(18),* 8612.

De Oliveira, E. M., & Oliveira, F. L. C. (2018). Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing. *Energy, 144,* 776–788.

Erten, M. Y., & İnanç, N. (2024). Forecasting electricity consumption in commercial buildings with deep learning

models to facilitate demand response programs. *Electric Power Components and Systems, 52(9),* 1636–1651.

Fan, G. F., Wei, X., Li, Y. T., & Hong, W. C. (2020). Forecasting electricity consumption using a novel hybrid model. *Sustainable Cities and Society, 61,* 102320.

Hamzaçebi, C., & Kutay, F. (2004). YSA ile Türkiye elektrik tüketiminin 2010'a kadar tahmini. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 19(3),* 227–233.

He, Y., & Lin, B. (2018). Forecasting China's total energy demand and structure using ADL-MIDAS model. *Energy, 151,* 420–429.

Hu, Y. C., Chiu, Y. J., Yu, C. Y., & Tsai, J. F. (2021). Integrating nonlinear interval regression with grey prediction for energy demand forecasting. *Applied Artificial Intelligence, 35(15),* 1490–1507.

Hwang, J., Suh, D., & Otto, M. O. (2020). Forecasting electricity consumption in commercial buildings using machine learning approach. *Energies, 13(22),* 5885.

Jamil, R. (2020). Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030. *Renewable Energy, 154,* 1–10.

Kazemzadeh, M. R., Amjadian, A., & Amraee, T. (2020). A hybrid data mining-driven algorithm for long-term electric peak load and energy demand forecasting. *Energy, 204,* 117948.

Li, Q., Zhou, K., Peng, B., & Mashhadi, A. (2024). Optimal deep belief networks for energy demand forecasting using a developed version of the gorilla troops optimization method. *Journal of Electrical Engineering & Technology, 19(1),* 177–191.

Li, X., Jiang, M., Cai, D., Song, W., & Sun, Y. (2024). A hybrid forecasting model for electricity demand in sustainable power systems based on support vector machine. *Energies, 17(17),* 4377.

Li, Y., & Jones, B. (2019). The use of extreme value theory for forecasting long-term substation maximum electricity demand. *IEEE Transactions on Power Systems, 35(1),* 128–139.

Liu, Y., Chen, H., Zhang, L., Wu, X., & Wang, X. J. (2020). Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China. *Journal of Cleaner Production, 272,* 122542.

Mikayilov, J. I., Alyamani, R., Darandary, A., Javid, M., & Hasanov, F. J. (2023). Modeling and forecasting industrial electricity demand for Saudi Arabia: Uncovering regional characteristics. *The Electricity Journal, 36(8),* 107331.

Nikseresht, A., & Amindavar, H. (2024). Energy demand forecasting using adaptive ARFIMA based on a novel dynamic structural break detection framework. Applied *Energy, 353,* 122069.

Oğcu, G., Demirel, O. F., & Zaim, S. (2012). Forecasting electricity consumption with neural networks and support vector regression. *Procedia - Social and Behavioral Sciences, 58,* 1576–1585.

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Residential building energy consumption prediction using deep learning. *Journal of Building Engineering, 45,* 103406.

Pérez-García, J., & Moral-Carcedo, J. (2016). Analysis and long-term forecasting of electricity demand through a

decomposition model: A case study for Spain. *Energy, 97,* 127–143.

Somu, N., M. R., G. R., & Ramamritham, K. (2021). A deep learning framework for building energy consumption forecast. *Renewable and Sustainable Energy Reviews, 137,* 110591.

Uzlu, E., & Dede, T. (2020). Estimating electric energy consumption in Turkey using artificial neural networks optimized with Jaya algorithm. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C, 8(3),* 511–528.

Wang, J. Q., Du, Y., & Wang, J. (2020). LSTM based long-term energy consumption prediction with periodicity. *Energy, 197,* 117197.

Zhang, G., Tian, C., Li, C., Zhang, J. J., & Zuo, W. (2020). Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature. *Energy, 201,* 117531.

Zhou, K., Chu, Y., & Hu, R. (2023). Energy supply-demand interaction model integrating uncertainty forecasting and peer-to-peer energy trading. *Energy, 285,* 129436.

AĞAÇ TABANLI TOPLULUK MODELLERİNDE AÇIKLANABİLİRLİK: ÖZNİTELİK ÖNEMİ, SHAP VE LIME'IN KARŞILAŞTIRILMASI

Tuba IRMAK[1]

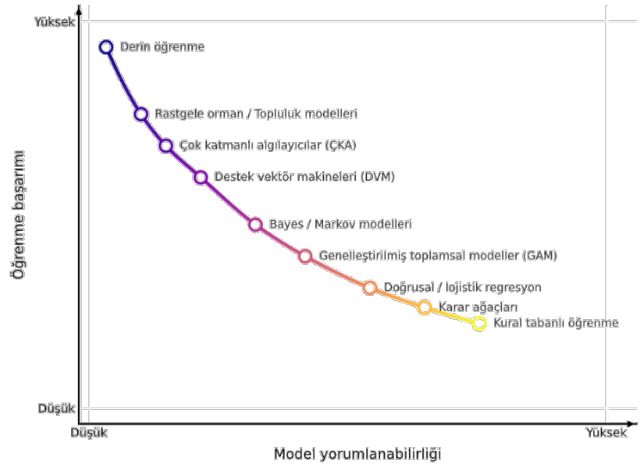
1. GİRİŞ

Yapay zekâ (YZ), son yıllardaki yükselişi sayesinde, yalnızca bir teknolojik yenilik ve akademik bir ilgi alanı olmaktan çıkıp pek çok sektörün karar sürecinin doğrudan bir parçası haline gelmiştir. YZ, köklü bir geçmişe sahip olmakla birlikte, veri işleme kapasitesindeki artış ve öğrenme algoritmalarındaki ilerlemeler sayesinde günümüzde karmaşık problemleri insan performansına yaklaşan hatta bazı durumlarda onu aşan yüksek başarımlarla çözebilir hâle gelmiştir (Arrieta et al., 2020). Bu gelişmeler, makine öğreniminin ilerlemesine olanak sağlamış ve YZ'nin ürün ve film önerileri gibi basit karar süreçlerinden; hastanelerde hastalık teşhisi (Ghaffar Nia et al., 2023), bankalarda yatırım ve kredi yönetimi (Sadok et al., 2022), sigortacılıkta risk değerlendirmesi (Leidman, 2025), araçlarda kaza önleme (Nascimento et al., 2019) ve işyerlerinde iş akışı süreçlerinin yönetimi gibi karmaşık ve kritik uygulamalara kadar geniş bir yelpazede kullanılmasına olanak tanımıştır (Ali et al., 2023).

Farklı alanlardaki (sağlık, ekonomi vb.) yüksek tahmin doğrulukları genellikle daha karmaşık model mimarilerinin geliştirilmesiyle mümkün olmuştur (Şekil 1). Bu karmaşık mimarilere, derin öğrenme ve topluluk öğrenme modellerini örnek olarak verebiliriz (Linardatos et al., 2020). Derin öğrenme

[1] Dr. Öğr. Üyesi, Atatürk Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, ORCID: 0000-0003-4749-5226.

modelleri, çok sayıda katmandan oluşan yapay sinir ağları aracılığıyla veriden temsiller öğrenen ve yüksek boyutlu verilerde üstün başarım sağlayan modellerdir. Topluluk öğrenme modelleri ise tek bir model sonucuna güvenmek yerine çok sayıda temel öğreniciyi birleştirerek tahmin üretir. Her iki yaklaşımın da yüksek başarım sergilemesi doğrudan iç yapılarındaki karmaşıklıktan kaynaklanır. Bir derin sinir ağından üretilen tahmin, milyonlarca parametrenin doğrusal olmayan etkileşimiyle; bir topluluk öğrenme modelinde ise yüzlerce ağacın bileşik kararıyla oluşur. Bu karmaşıklık düzeyinin, çok büyük miktarda verinin kullanımıyla birleşmesi sistemlerin tahmin performansını artırırken, iç işleyişlerinin ve karar mekanizmalarının açıklanabilirliğini azaltmaktadır. Bu durum, verilen kararların ardındaki mantığın anlaşılmasını zorlaştırmakta ve elde edilen sonuçların yorumlanmasını güçleştirmektedir.



Şekil 1. Öğrenme başarımı ile model yorumlanabilirliğine göre modeller (Ortigossa et al., 2024)

Makine öğrenmesi algoritmaları açıklanabilirlik düzeylerine göre farklı kategorilerde değerlendirilebilmektedir. Doğrusal regresyon ve karar ağacı gibi bazı yöntemler, karar verme süreçlerini kullanıcıya açık bir şekilde sunabildikleri için

beyaz kutu veya şeffaf kutu modeller olarak adlandırılmaktadır. Bu modellerde tahminlerin hangi değişkenlerden ve hangi kurallardan etkilenerek üretildiği kolaylıkla izlenebilmektedir. Buna karşılık derin öğrenme ve birçok topluluk öğrenme yöntemi, yüksek tahmin başarısı sunmalarına rağmen karar mekanizmalarını doğrudan açıklayamayan kara kutu (black-box) modeller arasında yer almaktadır (Minh et al., 2022).

Karar süreçlerinin anlaşılabilir olmaması, özellikle sağlık, finans, hukuk ve otonom sürüş sistemleri gibi kritik uygulama alanlarında önemli bir sorun oluşturmaktadır. Çünkü bu alanlarda verilen kararlar bireylerin yaşamını, güvenliğini ve ekonomik durumunu doğrudan etkileyebilmektedir. Bu nedenle günümüzde yalnızca yüksek doğruluk oranlarına sahip modeller geliştirmek yeterli görülmemekte; aynı zamanda bu modellerin hangi gerekçelerle belirli sonuçlara ulaştığının açıklanabilmesi de önemli bir gereklilik olarak kabul edilmektedir. Bu ihtiyaç, yapay zekâ sistemlerinin daha şeffaf, güvenilir ve hesap verebilir hale getirilmesine yönelik çalışmaların hız kazanmasına katkı sağlamıştır.

Açıklanabilirlik yalnızca kullanıcıya yönelik bir gereksinim değildir; modeli geliştiren araştırmacı için de işlevsel bir araçtır. Bir modelin neden belirli kararlar ürettiğinin görünür kılınması hem gizli kusurların saptanıp giderilmesini hem de sistemin çıktısına duyulan güvenin sağlam bir temele oturmasını sağlar. Bu katkı üç başlıkta somutlaşır (Şekil 2). Birincisi adalettir: açıklamalar, modelin veri kümesindeki önyargıları öğrenip öğrenmediğini görünür kılarak bu yanlılıkların düzeltilmesine olanak tanır. İkincisi sağlamlıktır; bir tahminin hangi etkenlere dayandığının izlenebilmesi, başarımı bozabilecek gürültü ya da kararsız davranışların erkenden fark edilmesini mümkün kılar. Üçüncüsü ise modelin gerçekten anlamlı özniteliklere dayanıp dayanmadığının doğrulanmasıdır; yüksek başarımın, ilgisiz ya da tesadüfi örüntüler yerine konuyla ilgili

bilgiden kaynaklandığından emin olmayı sağlar. Açıklanabilir yapay zekâ (Explainable Artificial Intelligence – XAI), tam da bu gereksinimleri karşılamak üzere, model başarımını korurken yorumlanabilirliği artırmayı amaçlayan yöntemler bütünü olarak ortaya çıkmıştır (Escalante et al., 2018).

Şekil 2. Açıklanabilirliğin sağladığı kazanımlar

XAI alanının son yıllarda hızla genişlemesiyle birlikte, çok sayıda kapsamlı derleme çalışması yapılmıştır (Saranya & Subhashini, 2023; Karim et al., 2023; Roshinta & Gábor, 2024; Salih et al., 2025). Bu çalışmalar; açıklama yöntemlerini model-içkin ve model-sonrası, yerel ve küresel, modele özgü ve modelden bağımsız gibi temel ayrımlar üzerinden ele almakta ve LIME ile SHAP'i model-sonrası açıklamanın en yaygın iki temsilcisi olarak öne çıkarmaktadır.

Bu çalışmada, ağaç tabanlı makine öğrenmesi modellerinin açıklanabilirliği; öznitelik önem yöntemleri, SHAP ve LIME ekseninde derli toplu bir biçimde ele alınmaktadır. Amaç, bu yöntemleri yalnızca ayrı ayrı tanıtmak değil; aynı model ve veri üzerinde karşılaştırarak hangi durumlarda hemfikir oldukları, hangi durumlarda ayrıştıkları ve bu ayrışmanın nedenleri üzerine bütüncül bir bakış sunmaktır. Bu doğrultuda saflık tabanlı önem, permütasyon önemi, SHAP ve LIME yöntemleri hem kavramsal temelleriyle hem de güçlü ve zayıf yönleriyle incelenmekte; ardından açık erişimli bir veri kümesi üzerinde gerçekleştirilen illüstratif bir uygulamayla somutlaştırılmaktadır. Böylece çalışmanın, açıklanabilirlik

yöntemlerine yönelik kavramsal bir çerçeve sunmanın yanı sıra, bu yöntemlerin pratikte nasıl yorumlanması gerektiğine dair uygulamalı bir rehber niteliği taşıması hedeflenmektedir.

Çalışmanın geri kalanı şu şekilde düzenlenmiştir. İkinci bölümde, yorumlanabilirlik ve açıklanabilirlik kavramları tanımlanmakta ve açıklama yöntemlerinin sınıflandırma eksenleri sunulmaktadır. Üçüncü bölüm, ağaç tabanlı modellerin içsel önem ölçütlerini ve bunların bilinen yanlılıklarını ele almakta; permütasyon tabanlı alternatifi tartışmaktadır. Dördüncü ve beşinci bölümler sırasıyla LIME ve SHAP yöntemleri ayrıntılı olarak incelemektedir. Altıncı bölümde, California Housing veri kümesi üzerinde üç ağaç tabanlı model (rastgele orman, XGBoost ve CatBoost) kullanılarak dört açıklama yöntemi karşılaştırmalı olarak uygulanmakta ve sonuçları yorumlanmaktadır. Yedinci ve son bölümde ise temel çıkarımlar özetlenmekte ve gelecekteki araştırma yönelimleri tartışılmaktadır.

2. KAVRAMSAL ÇERÇEVE

Bu bölümde, çalışmanın kavramsal temeli oluşturulmaktadır. İlk olarak yorumlanabilirlik ile açıklanabilirlik arasındaki ayrım ele alınmakta; ardından açıklama yöntemlerinin kapsam, zamanlama ve modele bağımlılık eksenlerinde sınıflandırılması sunulmaktadır.

2.1. Yorumlanabilirlik ve Açıklanabilirlik

Yorumlanabilirlik (interpretability) ve açıklanabilirlik (explainability) terimleri sıklıkla birbirinin yerine kullanılsa da ince bir ayrım taşımaktadırlar. Yorumlanabilirlik, genellikle, modelin yapısına ve parametrelerine bakarak kararlarının insanlarca doğrudan anlaşılabilmesi anlamına gelir; örneğin sığ karar ağaçları veya basit lineer modeller, içsel olarak yorumlanabilir kabul edilir. Buna karşılık açıklanabilirlik,

doğrudan anlaşılır olmayan (kara kutu) bir modelin kararlarını sonradan açıklamaya çalışan harici yöntemleri ifade eder. Bu bölümde ele alınan yöntemlerin tamamı bu ikinci anlama, yani model-sonrası açıklamaya karşılık gelir (Marcinkevičs & Vogt, 2023).

2.2. Açıklanabilirlik Yöntemlerinin Sınıflandırılması

Açıklama yöntemleri üç temel eksende konumlandırılabilir. Birinci eksen, açıklamanın kapsamıdır: küresel (global) açıklamalar modelin bütüncül davranışını, hangi özniteliklerin genel olarak ne kadar etkili olduğunu betimlerken; yerel (local) açıklamalar tek bir gözlem için yapılan tahmini gerekçelendirir. İkinci eksen, açıklamanın zamanlamasıdır: model-içkin (intrinsic) yaklaşımlarda yorumlanabilirlik modelin kendi yapısından gelirken, model-sonrası (post-hoc) yaklaşımlarda açıklama eğitilmiş bir model üzerine sonradan inşa edilir. Üçüncü eksen, yöntemin modele bağımlılığıdır: modelden bağımsız (model-agnostic) yöntemler modelin iç yapısına bakmaksızın yalnızca girdi–çıktı ilişkisini kullanırken, modele özgü (model-specific) yöntemler belirli bir mimarinin yapısından yararlanır (Mersha et al., 2024; Ortigossa et al., 2024).

Bu bölümde incelenen dört yöntem bu eksenlerde farklı konumlara yerleşir. Saflık tabanlı önem modele özgü ve küreseldir; permütasyon önemi modelden bağımsız ve küreseldir; LIME modelden bağımsız ve yereldir; SHAP ise hem yerel açıklamalar üretip hem de bunları küresel görünüme toplayabilen, ağaç modellerinde modele özgü (TreeSHAP) verimli bir biçime sahip olan esnek bir çerçevedir. Tablo 1 bu sınıflandırma eksenlerini özetlemektedir.

Tablo 1. Açıklama yöntemlerinin sınıflandırma eksenleri

Eksen	Uç 1	Uç 2
Kapsam	Küresel (global)	Yerel (local)
Zamanlama	Model-içkin (intrinsic)	Model-sonrası (post-hoc)
Bağımlılık	Modele özgü	Modelden bağımsız (agnostic)

3. AĞAÇ TABANLI MODELLERDE İÇSEL ÖNEM ÖLÇÜTLERİ

Bu bölümde, ağaç tabanlı modellerin doğrudan kendi yapısından türettiği içsel önem ölçütleri — saflık tabanlı önem ve permütasyon tabanlı önem — ele alınmaktadır.

3.1. Saflık Tabanlı Önem (Gini / Gain)

Karar ağaçları, her düğümde veriyi bir öznitelik üzerinden ikiye bölerek hedef değişkene göre daha saf (homojen) alt kümeler oluşturur. Saflık tabanlı önem, bir özniteliğin ağaç boyunca gerçekleştirdiği bölünmelerin sağladığı toplam saflık azalımının — sınıflandırmada Gini ya da entropi azalımı, regresyonda varyans azalımı — bir ölçüsüdür. Rastgele ormanlar ve gradyan artırmalı ağaçlar gibi ağaç tabanlı modellerde, küresel öznitelik önemi çoğunlukla düğüm bölünmelerinin sağladığı toplam "gain" (saflık azalımı) üzerinden tanımlanır; her bölünmede kayıp (Gini, entropi, varyans) düşüşü o özelliğe atanır ve toplulukta tüm ağaçlar boyunca toplanıp ortalanır (Lundberg et al., 2019).

Saflık tabanlı önemin yaygın kullanımına karşın, bu bölümün vurgulamak istediği kritik bir kısıtı vardır. Yöntem, yüksek kardinaliteli ve sürekli özniteliklere sistematik biçimde meyleder; çünkü çok sayıda olası bölünme noktası sunan bu öznitelikler, tesadüfen de olsa saflığı azaltacak bir eşik bulma şansına daha fazla sahiptir. Ayrıca önem, eğitim verisi üzerinden hesaplandığından modelin aşırı öğrenmesini de yansıtabilir; eğitimde önemli görünen bir öznitelik, görülmemiş veride aynı katkıyı sağlamayabilir. Korelasyonlu öznitelikler söz konusu olduğunda da önem, ilişkili öznitelikler arasında öngörülemez biçimde paylaşılır. Bu üç etken birlikte, saflık tabanlı önem sıralamalarını yanıltıcı kılabilir (Lundberg et al., 2019).

3.2. Permütasyon Tabanlı Önem

Permütasyon önemi bulurken temel fikir şu şekildedir: eğitilmiş bir modelin başarımı önce bozulmamış veride ölçülür; ardından tek bir özniteliğin değerleri rastgele karıştırılarak o öznitelik ile hedef arasındaki ilişki koparılır ve başarımdaki düşüş kaydedilir. Düşüş ne kadar büyükse, öznitelik o kadar önemlidir. Ölçüm tercihen eğitimde kullanılmamış bir doğrulama kümesinde yapıldığından, yöntem aşırı öğrenmeye karşı saflık tabanlı önemden daha dayanıklıdır. Bununla birlikte permütasyon önemi de korelasyonlu özniteliklerde zayıflar: bir özniteliğin karıştırılması, gerçekçi olmayan girdi bileşimleri yaratarak modeli eğitim dağılımının dışına itebilir ve önemi olduğundan farklı gösterebilir (Covert et al., 2021; Lundberg et al., 2019).

4. LIME YÖNTEMİ

LIME (Local Interpretable Model-agnostic Explanations – Yerel Yorumlanabilir Modelden Bağımsız Açıklamalar), tek bir tahmini yerel olarak açıklamayı amaçlara. Yöntem, açıklanacak gözlemin çevresinde girdiyi küçük biçimlerde değiştirerek (perturbasyon) yeni örnekler üretir, bu örneklerin kara kutu modeldeki tahminlerini toplar ve açıklanan gözleme yakınlığa göre ağırlıklandırır. Ardından bu yerel komşulukta, kara kutu modelin davranışını taklit eden basit ve yorumlanabilir bir vekil model — tipik olarak ağırlıklı bir doğrusal regresyon — uydurulur. Bu vekil modelin katsayıları, ilgili gözlem için özniteliklerin yerel katkıları olarak okunur. LIME'ın temel varsayımı, karmaşık bir karar yüzeyinin yeterince küçük bir komşulukta doğrusal olarak yaklaşıklanabileceğidir (Ribeiro et al., 2016).

LIME'ın en güçlü yanı, modelden tümüyle bağımsız olması ve sezgisel, yerel bir açıklama sunmasıdır; herhangi bir model türüne aynı mantıkla uygulanabilir. Buna karşın yöntem

birkaç açıdan eleştirilir. Bunların başında kararsızlık gelir: perturbasyon rastgele olduğundan, aynı gözlem için farklı çalıştırmalar farklı açıklamalar üretebilir. Ayrıca yerel komşuluğun nasıl tanımlanacağı (çekirdek genişliği) büyük ölçüde keyfîdir ve sonucu belirgin biçimde etkiler. Son olarak, yerel doğrusallık varsayımı oldukça eğri karar yüzeylerinde geçerliliğini yitirebilir. Bu nedenle LIME açıklamaları, tek başına kesin gerekçeler olarak değil, yerel davranışa dair göstergeler olarak yorumlanmalıdır.

5. SHAP YÖNTEMİ

SHAP (SHapley Additive exPlanations), açıklamayı işbirlikçi oyun teorisinin sağlam bir zeminine oturtur. Burada bir tahmin, oyuncuların (özniteliklerin) ortak ürettiği bir kazanç olarak görülür; soru, bu kazancın öznitelikler arasında adil biçimde nasıl paylaştırılacağıdır. Shapley değeri, bir özniteliğin olası tüm öznitelik koalisyonlarına eklenmesinin sağladığı marjinal katkının ortalaması olarak tanımlanır ve dört temel adalet aksiyomunu sağlayan tek paylaştırmadır: eksiksizlik (katkıların toplamı tahmin ile taban değer arasındaki farka eşittir), simetri (aynı katkıyı yapan öznitelikler eşit değer alır), kukla (hiçbir katkısı olmayan öznitelik sıfır değer alır) ve toplanabilirlik. Bu aksiyomatik temel, SHAP'i diğer birçok açıklama yönteminden teorik olarak ayırır. Ancak Shapley değerlerinin doğrudan hesabı, öznitelik sayısıyla üstel biçimde büyür ve genel durumda pratik değildir (Lundberg & Lee, 2017).

SHAP, bu teorik çerçeveyi eklemeli öznitelik atfı (additive feature attribution) biçiminde somutlaştırır: her tahmin, bir taban değer (modelin ortalama çıktısı) ile her özniteliğin o gözleme yaptığı katkıların toplamı olarak ifade edilir. Bu yapı, açıklamanın hem yerel düzeyde tek bir tahmini tam olarak ayrıştırmasını hem de bu yerel açıklamaların biriktirilmesiyle

küresel bir görünüme ulaşılmasını sağlar. Örneğin tüm gözlemler üzerinde bir özniteliğin SHAP değerlerinin mutlak ortalaması, o özniteliğin küresel önemini verir. Böylece SHAP, yerel ve küresel açıklamayı tek ve tutarlı bir çerçevede birleştirir.

5.1. TreeSHAP: Ağaç Modelleri için Verimli Hesaplama

SHAP'in ağaç tabanlı modeller için özel önemi, TreeSHAP algoritmasından kaynaklanır. Shapley değerlerinin genel hesabı üstel karmaşıklıkta olmasına karşın, TreeSHAP ağaç yapısının özelliklerinden yararlanarak bu değerleri polinom zamanda ve yaklaşıklama yapmadan, tam olarak hesaplar. Bu verimlilik, yöntemin rastgele orman, XGBoost ve CatBoost gibi yüzlerce ağaçtan oluşan topluluk modellerine pratikte uygulanabilmesini sağlar. Dolayısıyla TreeSHAP hem teorik sağlamlığı hem de hesaplama uygulanabilirliği bakımından ağaç tabanlı modeller için ayrıcalıklı bir konuma sahiptir; bu bölümün ağaç tabanlı modellere odaklanma gerekçelerinden biri de budur (Lundberg et al., 2019; Lundberg et al., 2020).

6. ÖRNEK BİR UYGULAMA: YÖNTEMLERİN KARŞILAŞTIRMALI GÖSTERİMİ

Bu bölümde, önceki bölümlerde kavramsal olarak tanıtılan dört açıklama yöntemi, California konut veri kümesi üzerinde eğitilen üç ağaç tabanlı topluluk modeline uygulanarak karşılaştırmalı biçimde değerlendirilmektedir. Önce veri seti ve deneysel kurulum tanıtılmakta; ardından içsel ve permütasyon önemi, SHAP ve LIME sonuçları sırayla sunularak yöntemlerin nerede örtüştüğü ve nerede ayrıştığı tartışılmaktadır.

6.1. Veri Seti ve Deneysel Kurulum

Bu bölümde, önceki bölümlerde kavramsal olarak ele alınan dört açıklama yöntemi açık erişimli bir veri kümesi

üzerinde karşılaştırmalı olarak uygulanmaktadır. Amaç en yüksek tahmin başarımına ulaşmak değil; aynı veri ve model üzerinde yöntemlerin ürettiği açıklamaların ne ölçüde örtüştüğünü, nerede ayrıştığını ve bu ayrışmanın nedenlerini somut biçimde göstermektir.

Uygulamadaki tüm deneyler Google Colab bulut ortamında, Python programlama dili kullanılarak yapılmıştır. Uygulamada California konut veri kümesi kullanılmıştır (Pace & Barry, 1997). 1990 ABD nüfus sayımından türetilen bu küme, blok grubu düzeyinde 20.640 gözlem ve sekiz öznitelik içerir; hedef değişken medyan konut değeridir (100.000 ABD doları biriminde). Öznitelikler medyan gelir (MedInc), konut yaşı (HouseAge), ortalama oda ve yatak odası sayısı (AveRooms, AveBedrms), nüfus (Population), ortalama hane doluluğu (AveOccup) ve coğrafi konumdan (Latitude, Longitude) oluşur. Veri kümesi %80–%20 oranında eğitim (16.512) ve test (4.128) olarak ayrılmış, tüm rastgelelik kaynakları sabit bir tohum değeriyle (seed = 42) denetlenmiştir. Ayrıca ortalama oda sayısı ile yatak odası sayısı arasındaki yüksek korelasyon (0,85), yöntemlerin korelasyonlu öznitelikler karşısındaki davranışını gözlemlemek için elverişli bir zemin sunar.

Topluluk (ensemble) modelleri, tek bir öğreniciye güvenmek yerine çok sayıda temel modeli birleştirerek daha kararlı ve başarılı tahminler üretir. Ağaç tabanlı topluluklar, bu yaklaşımın temel öğrenicisi olarak karar ağaçlarını kullanır ve iki ana strateji etrafında toplanır. Birincisi torbalama (bagging) yaklaşımıdır; burada çok sayıda ağaç birbirinden bağımsız olarak eğitilir ve tahminleri ortalanır; rastgele orman (Random Forest) bu stratejinin en yaygın temsilcisidir. İkincisi artırma (boosting) yaklaşımıdır; burada ağaçlar ardışık olarak, her biri bir öncekinin hatalarını düzeltecek biçimde eklenir; XGBoost, LightGBM ve CatBoost bu ailenin öne çıkan üyeleridir. Bu modeller, tablosal veride sıklıkla en yüksek başarımı sağladıkları için pratikte fiilî

standart hâline gelmiştir. Bu çalışmada da hem torbalama hem de artırma kanadını temsil edecek biçimde üç model kullanılmıştır: rastgele orman, XGBoost ve CatBoost.

Bu üç model, ortak bir kurulumla eğitilmiştir. Ağaç tabanlı modeller bölünme temelli çalıştığından ve özniteliklerin ölçeğine duyarsız olduğundan, ölçekleme ya da normalleştirme gibi bir ön işleme uygulanmamıştır; bu, yöntemin doğasından kaynaklanan bilinçli bir tercihtir. Modellerin test kümesi üzerindeki başarımları Tablo 2'de sunulmaktadır. Üç model de 0,81–0,84 aralığında belirleme katsayısına (R^2) ulaşarak, açıklama yöntemlerini anlamlı biçimde karşılaştırmak için yeterli ve birbirine yakın bir başarım sergilemiştir. İzleyen ayrıntılı SHAP ve LIME çözümlemelerinde, temsil gücü yüksek bir model olarak CatBoost birincil model seçilmiştir. Bu aşamada tek bir modelin seçilmesi anlatıyı sadeleştirmek içindir; üç modelin başarımı birbirine yakın olduğundan, bu seçim başarım sıralamasına değil, gösterim uygunluğuna dayanmaktadır. CatBoost, TreeSHAP ile verimli çalışması ve dengeli davranışı nedeniyle temsilî birincil model olarak belirlenmiştir.

Tablo 2. Modellerin test kümesi üzerindeki başarımı

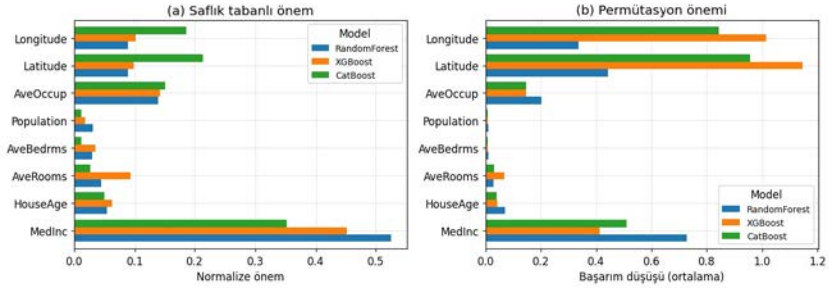
Model	MAE	RMSE	R^2
Rastgele orman	0,326	0,502	0,807
XGBoost	0,306	0,463	0,836
CatBoost	0,322	0,480	0,824

6.2. İçsel ve Permütasyon Önem Sonuçları

Şekil 3, saflık tabanlı önem ile permütasyon önemini üç model için yan yana göstermektedir. Her iki yöntem de medyan geliri (MedInc) açık ara en önemli öznitelik olarak işaret etmekte; bu, konut değerinin gelirle güçlü ilişkisini doğrulamaktadır. Ancak iki yöntem arasında dikkat çekici bir ayrışma, coğrafi özniteliklerde (Latitude ve Longitude) ortaya çıkmaktadır. Bu öznitelikler saflık tabanlı önemde orta sıralarda yer alırken, permütasyon öneminde neredeyse medyan gelir kadar belirleyici

hâle gelir. Ayrışmanın nedeni öğreticidir: coğrafi koordinatlar ağaç içinde görece az sayıda bölünmeye yol açtığından saflık temelli ölçüt onları düşük önemli gösterir; oysa bu öznitelikler karıştırıldığında modelin başarımı belirgin biçimde düşer; bu da onların tahmin için kritik olduğunu ortaya koyar. Dolayısıyla saflık tabanlı önem tek başına yanıltıcı olabilmektedir.

Korelasyonlu oda ve yatak odası sayısı öznitelikleri ise her iki yöntemde de düşük önem almakta; ancak saflık tabanlı önemde XGBoost'un bu özniteliklerden birine diğer modellerden daha fazla ağırlık vermesi, korelasyonlu öznitelikler arasındaki önemin öngörülemez biçimde paylaşıldığını göstermektedir.

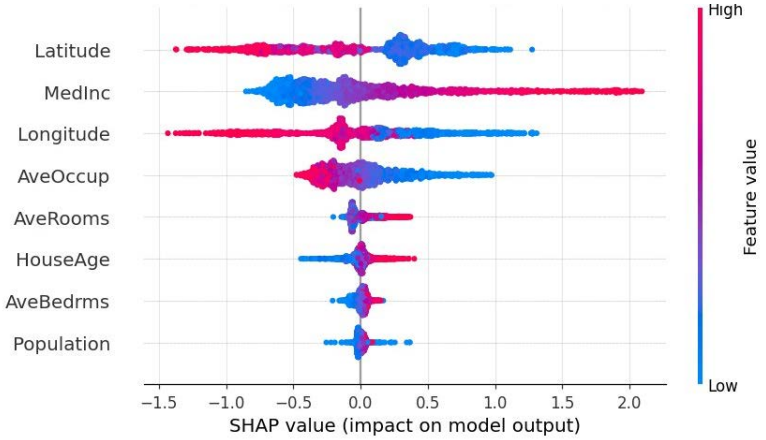


Şekil 3. Saflık tabanlı ve permütasyon öneminin model ailelerine göre karşılaştırması

6.3. SHAP Analizi

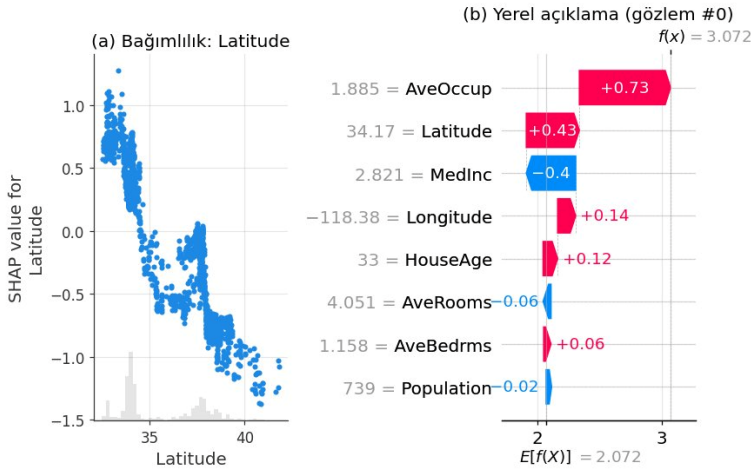
SHAP çözümlemesi, küresel önemin ötesine geçerek her özniteliğin etkisinin yönünü ve büyüklüğünü de görünür kılar. Şekil 4'teki özet grafiği, CatBoost modeli için en etkili özniteliklerin Latitude, MedInc ve Longitude olduğunu göstermektedir. Bu sıralamanın permütasyon önemiyle örtüşmesi, buna karşılık saflık tabanlı önemden ayrışması dikkat çekicidir; başarım temelli iki yaklaşım (permütasyon ve SHAP) benzer bir tablo çizerken, modelin iç yapısına dayanan saflık temelli ölçüt farklı bir sıralama önermektedir.

Etkinin yönü açısından MedInc sezgisel ve tek yönlü bir örüntü sergiler: yüksek gelir değerleri tahmini konut değerini güçlü biçimde artırır. Buna karşılık coğrafi öznitelikler tek yönlü değildir; yüksek ve düşük değerleri hem pozitif hem negatif katkılara karşılık gelebilir. Bu durum, küresel önemin yüksek olmasına rağmen ilişkinin monoton olmadığını gösterir.



Şekil 4. CatBoost modeli için SHAP özet grafiği (beeswarm)

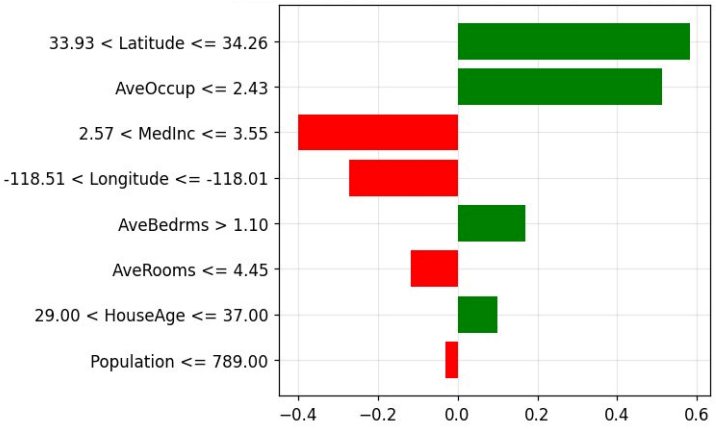
Şekil 5a, bu doğrusal olmayan ilişkiyi enlem (Latitude) özniteliği için açar: yaklaşık 33–35 derece aralığında enlemin katkısı pozitifken, 37 derecenin ötesinde negatife döner. Bu eşik benzeri davranış, bölgenin coğrafi fiyat örüntüsünü yansıtır ve bağımlılık grafiğinin neden gerekli olduğunu gösterir. Şekil 5b ise tek bir gözlem için yerel açıklamayı sunar: tahmin, modelin ortalama çıktısı olan taban değerden (yaklaşık 2,07) başlayarak özniteliklerin eklemeli katkılarıyla nihai değere (yaklaşık 3,07) ulaşır. Bu gözlem için en büyük pozitif katkıyı ortalama hane doluluğu ve enlem sağlarken, medyan gelir negatif yönde katkı yapmaktadır. Bu, SHAP'ın eklemeli yapısının tek bir kararı nasıl şeffaf biçimde ayrıştırdığını somutlaştırır.



Şekil 5. SHAP bağımlılık grafiği (Latitude) ve tek bir gözlem için şelale (waterfall) açıklaması

6.4. LIME Analizi

Şekil 6, aynı gözlem için LIME tarafından üretilen yerel açıklamayı göstermektedir. İki yöntem, baskın öznitelikler ve katkı yönleri konusunda büyük ölçüde uyumludur: hem SHAP hem LIME, enlem ve ortalama hane doluluğunu pozitif, medyan geliri ise negatif yönde etkili bulur. Bununla birlikte yöntemler sıralama ve büyüklükte ayrışır; örneğin LIME enlemi en güçlü katkı olarak öne çıkarırken SHAP ortalama hane doluluğunu birinci sıraya koyar. Bu ayrışma beklenen bir sonuçtur: LIME kararı yerel bir doğrusal vekil modelle yaklaşıklarken, SHAP oyun teorisine dayalı eklemeli bir atıf hesaplar; aynı gözlem için iki yöntemin tıpatıp aynı açıklamayı vermesi beklenmez. Dolayısıyla LIME ve SHAP, birbirinin yerine geçen değil, birbirini tamamlayan araçlar olarak görülmelidir.



Şekil 6. Aynı gözlem için LIME yerel açıklaması (CatBoost)

6.5. Yöntemlerin Karşılaştırılması

Yöntemlerin karşılaştırmalı bir özeti Tablo 3'te sunulmaktadır. Modeller arası tutarlılık, hem saflık tabanlı önem (ortalama Spearman sıra korelasyonu 0,90) hem de SHAP (0,92) için yüksek ve birbirine yakın bulunmuştur. Bu, güçlü sinyalli ve iyi ayrışmış bir veri kümesinde farklı model ailelerinin benzer öznitelik sıralamaları öğrendiğini; dolayısıyla her iki yöntemin de modelden modele büyük ölçüde kararlı kaldığını gösterir.

Tablo 3. Yorum yöntemlerinin karşılaştırılması

Yöntem	Kapsam	Model bağımlılığı	Modeller arası tutarlılık (Spearman)
Saflık tabanlı önem	Küresel	Ağaca özgü	0,90
Permütasyon önemi	Küresel	Model-agnostik	—
SHAP (TreeSHAP)	Küresel + yerel	Ağaca özgü	0,92
LIME	Yerel	Model-agnostik	—

Bu uygulamadan üç temel çıkarım elde edilir. Birincisi, en belirgin ayrışma yöntemler arasında — özellikle saflık tabanlı önem ile başarım temelli yöntemler (permütasyon ve SHAP)

arasında — coğrafi öznitelikler örneğinde ortaya çıkmıştır; bu da tek bir önem ölçütüne güvenmenin riskini göstermektedir. İkincisi, SHAP ve LIME aynı gözlemde baskın öznitelikler ve yönler konusunda uzlaşsa da sıralama ve büyüklükte ayrışabilmektedir. Üçüncüsü, bu veri kümesinde yöntem seçimi, model ailesi seçiminden daha belirleyici olmuştur; çünkü modeller arası tutarlılık yüksekken yöntemler arası farklar daha barizdir.

7. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, ağaç tabanlı topluluk öğrenme modellerinin açıklanabilirliği saflık tabanlı önem, permütasyon önemi, SHAP ve LIME yöntemleri kullanılarak incelenmiştir. Elde edilen sonuçlar, farklı açıklanabilirlik yöntemlerinin aynı model üzerinde farklı değerlendirmeler üretebildiğini göstermiştir. Özellikle performans temelli yöntemlerin bazı önemli öznitelikleri daha doğru şekilde öne çıkardığı görülmüştür. Ayrıca SHAP ve LIME benzer yorumlar sunsa da katkı değerlerinin büyüklüğü ve sıralaması değişebilmektedir. Bu nedenle açıklanabilirlik çalışmalarında tek bir yönteme bağlı kalmak yerine, birden fazla yöntemin birlikte kullanılması daha güvenilir sonuçlar sağlayacaktır. Bununla birlikte, elde edilen açıklamaların değişkenler arasındaki ilişkileri ortaya koyduğu, ancak doğrudan nedensel bir ilişkiyi kanıtlamadığı unutulmamalıdır.

Gelecek çalışmalarda, açıklanabilir yapay zekâ yöntemlerinin daha güvenilir ve tutarlı hale getirilmesine odaklanılabilir. Özellikle korelasyon temelli açıklamalar yerine nedensel ilişkileri ortaya koyan yaklaşımların geliştirilmesi, açıklamaların doğruluğunu artıracaktır. Ayrıca SHAP ve LIME gibi yöntemlerin kararlılığının değerlendirilmesi ve açıklama kalitesini ölçen standart kriterlerin oluşturulması önemli bir

araştırma alanıdır. Bunun yanı sıra, ağaç tabanlı modellerin yanında derin öğrenme modelleri için geliştirilen açıklanabilirlik yöntemlerinin karşılaştırmalı olarak incelenmesi literatüre katkı sağlayacaktır.

KAYNAKÇA

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., ... & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Information fusion, 99, 101805.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.

Covert, I., Lundberg, S., & Lee, S. I. (2021). Explaining by removing: A unified framework for model explanation. Journal of Machine Learning Research, 22(209), 1-90.

Deliloğlu, R. A. S. (2025). Açıklanabilir yapay zekâ tasarımı ve sürdürülebilirlik, Doktora Tezi, İstatistik anabilim dalı.

Escalante, H. J., Escalera, S., Guyon, I., Baró, X., Güçlütürk, Y., Güçlü, U., ... & van Lier, R. (Eds.). (2018). Explainable and interpretable models in computer vision and machine learning. Cham: Springer International Publishing.

Ghaffar Nia, N., Kaplanoglu, E., & Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease diagnosis and prediction. Discover Artificial Intelligence, 3(1), 5.

Karim, M. R., Shajalal, M., Graß, A., Döhmen, T., Chala, S. A., Boden, A., ... & Decker, S. (2023, October). Interpreting black-box machine learning models for high dimensional datasets. In 2023 IEEE 10th international conference on data science and advanced analytics (DSAA) (pp. 1-10). IEEE.

Leidman, B. (2025). The impact of artificial intelligence and predictive analytics on insurance risk assessment in the

digital age. Periodicals of Engineering and Natural Sciences, 13(2), 375-388.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. Entropy, 23(1), 18.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2019). Explainable AI for trees: From local explanations to global understanding. arXiv preprint arXiv:1905.04610.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1), 56-67.

Marcinkevičs, R., & Vogt, J. E. (2023). Interpretable and explainable machine learning: A methods-centric overview with concrete examples. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(3), e1493.

Mersha, M., Lam, K., Wood, J., Alshami, A. K., & Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. Neurocomputing, 599, 128111.

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. Artificial Intelligence Review, 55(5), 3503-3568.

Nascimento, A. M., Vismari, L. F., Molina, C. B. S. T., Cugnasca, P. S., Camargo, J. B., de Almeida, J. R., ... & Hata, A. Y. (2019). A systematic literature review about the impact of

artificial intelligence on autonomous vehicle safety. IEEE Transactions on Intelligent Transportation Systems, 21(12), 4928-4946.

Ortigossa, E. S., Gonçalves, T., & Nonato, L. G. (2024). Explainable artificial intelligence (xai)—from theory to methods and applications. IEEE Access, 12, 80799-80846.

Pace, R. K., & Barry, R. (1997). Sparse spatial autoregressions. Statistics & Probability Letters, 33(3), 291–297.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

Roshinta, T. A., & Gábor, S. (2024, December). A comparative study of lime and shap for enhancing trustworthiness and efficiency in explainable ai systems. In 2024 IEEE International Conference on Computing (ICOCO) (pp. 134-139). IEEE.

Sadok, H., Sakka, F., & El Maknouzi, M. E. H. (2022). Artificial intelligence and bank credit analysis: A review. Cogent Economics & Finance, 10(1), 2023262.

Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. Advanced Intelligent Systems, 7(1), 2400304.

Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. Decision analytics journal, 7, 100230.

SÜRDÜRÜLEBİLİR TEDARİKÇİ SEÇİMİNDE ÇOK KRİTERLİ KARAR VERME TABANLI DEĞERLENDİRME: GÜNCEL EĞİLİMLERİN ANALİZİ

Özge ALBAYRAK ÜNAL[1]

1. GİRİŞ

Küreselleşme, artan rekabet baskısı, çevresel sorunlar ve değişen toplumsal beklentiler, modern iş dünyasında işletmelerin yalnızca ekonomik performansa odaklanan geleneksel yaklaşımlarını yetersiz hale getirmiş ve daha sürdürülebilir yönetim anlayışlarını benimsemelerini zorunlu kılmıştır. Doğal kaynakların tükenme riski ve paydaşların sürdürülebilirlik konusundaki artan duyarlılığı, işletmelerin faaliyetlerini ekonomik, çevresel ve sosyal boyutları birlikte ele alan bütüncül bir bakış açısıyla şekillendirmesine yol açmıştır. Bu yaklaşımın operasyonel düzeydeki en önemli yansımalarından biri ise tedarik zinciri süreçlerinde ortaya çıkmaktadır.

Tedarik zinciri yönetimi, geçmişte daha çok maliyet, kalite ve teslimat performansı gibi ekonomik hedeflere odaklanırken, çevresel ve sosyal boyutlar çoğu zaman ikincil planda kalmıştır. Günümüzde ise söz konusu baskılar işletmeleri tedarik zinciri süreçlerine çevresel ve sosyal kriterleri de entegre etmeye yöneltmektedir (Erdebilli & Sıcakyüz, 2024). Özellikle tedarikçiler, ürün ve hizmetlerin yaşam döngüsü boyunca ortaya çıkan çevresel ve sosyal etkilerin önemli bir bölümünden sorumlu olduklarından, sürdürülebilir tedarik zinciri yönetiminin temel

[1] Dr. Öğr. Üyesi, Atatürk Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği ORCID: 0000-0001-7798-8799

unsurlarından biri haline gelmiştir. Bu doğrultuda sürdürülebilirlik, işletmeler için yalnızca bir sorumluluk alanı değil, aynı zamanda önemli bir rekabet avantajı kaynağı olarak değerlendirilmektedir.

Bu çok boyutlu yapısı nedeniyle sürdürülebilir tedarikçi seçimi, birbiriyle çoğu zaman çelişen ekonomik, çevresel ve sosyal kriterlerin eş zamanlı olarak değerlendirilmesini gerektiren karmaşık bir karar problemi haline gelmektedir. Söz konusu karmaşıklık karşısında çok kriterli karar verme (ÇKKV) yöntemleri, farklı nitelikteki kriterleri sistematik bir yapı içinde bir araya getirerek tedarikçi alternatiflerinin objektif biçimde değerlendirilmesine imkân tanıması nedeniyle literatürde yaygın biçimde benimsenmiştir (Govindan, Rajendran, Sarkis, & Murugesan, 2015). Bu bağlamda mevcut çalışma, sürdürülebilir tedarikçi seçimi problemlerinin çözümünde kullanılan ÇKKV metodolojilerinin bütünsel bir haritasını çıkarmak ve literatürün gelişim seyrini yapısal bir perspektifle analiz etmeyi amaçlamaktadır. Bu doğrultuda, ilgili literatürde kullanılan başlıca yöntemler, dikkate alınan kriterler ve değerlendirme yaklaşımları sistematik biçimde incelenerek, alanın mevcut eğilimleri ve eksik kalan yönleri ortaya konmaya çalışılmaktadır. Elde edilen bulgular ışığında, tedarikçi değerlendirme sürecinde dikkate alınan ekonomik, çevresel ve sosyal kriterler bütüncül biçimde sentezlenerek hem akademik literatüre hem de uygulamaya yönelik çıkarımlar sunulması hedeflenmektedir.

2. ARAŞTIRMA METODOLOJİSİ

Bu çalışma, literatürde sürdürülebilir tedarikçi seçimi sürecinde ÇKKV yöntemlerinin kullanımına yönelik kapsamlı bir analiz sunmayı amaçlamaktadır. Çalışmanın odak noktasını doğru bir biçimde sınırlandırmak, ilgili verileri sistematik olarak analiz etmek ve sürece dahil etmek amacıyla, kapsamlı bir

inceleme gerçekleştirilmiştir. Bu doğrultuda, araştırma sorularının ve anahtar kelimelerin belirlenmesini içeren beş aşamalı bir metodolojik süreç kurgulanmıştır. Bu süreç Şekil 1'de verilmiştir.

Şekil 1.Araştırma metodolojisi

Araştırma metodolojisi beş adımdan oluşmaktadır:

1. Araştırma sorularının belirlenmesi: Araştırmanın amacını ve kapsamını netleştirmek için aşağıdaki araştırma soruları oluşturulmuştur:

1) Sürdürülebilir tedarikçi seçiminde ÇKKV yöntemlerinin kullanımına ilişkin yayın eğilimleri (yıl, ülke, dergi, yöntem vb.) nasıl bir gelişim göstermektedir?

2) Sürdürülebilir tedarikçi seçimi literatüründe kriter kullanım tercihleri zaman içerisinde nasıl bir değişim göstermektedir?

3) Mevcut literatürde kullanılan yöntemler ve kriterler açısından öne çıkan araştırma eğilimleri ve gelecekteki araştırma fırsatları nelerdir?

2. Veri tabanlarının belirlenmesi: Veri kalitesine ve akademik seçiciliğe katkı sağlamak amacıyla, alanında kabul görmüş ve yüksek etki faktörlü dergileri indeksleyen Web of Science (WoS) ve Scopus veri tabanlarından yararlanılmıştır. Bu veri tabanlarının sunduğu gelişmiş arama altyapısı, sürdürülebilir tedarikçi seçimi literatüründeki olgun ve nitelikli çalışmaların sistematik bir yaklaşımla incelenmesine zemin hazırlamıştır.

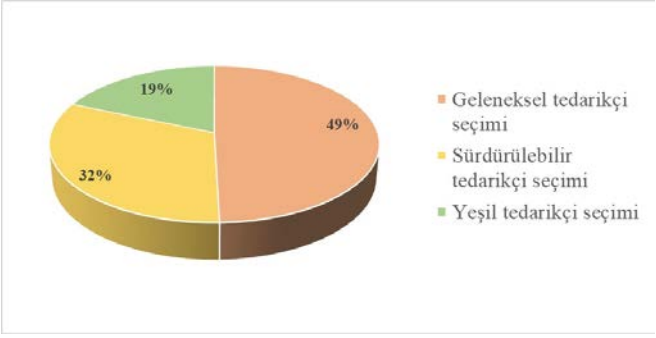
3. Anahtar kelimelerin belirlenmesi: Anahtar kelimeler belirlenirken Web of Science ve Scopus veri tabanlarında başlık, özet ve anahtar kelime alanlarında tarama yapılmıştır. Aynı gruptaki terimler "OR (VEYA)", farklı gruplardaki terimler ise "AND (VE)" operatörleri kullanılarak birleştirilmiştir. Bu kapsamda ("supplier selection" OR "vendor selection" OR "supplier evaluation") AND ("multi-criteria decision making" OR "multi-criteria decision analysis" OR "MCDM") anahtar kelimeleri kullanılarak literatür taranmıştır.

4. Dahil etme/hariç tutma kriterlerinin belirlenmesi: Literatürdeki güncel gelişmeleri yansıtmak ve özgün uygulamalar ile hibrit yaklaşımlar gibi yenilikçi katkıları analiz ederek günümüz karar alma süreçlerine katkı sağlamak amacıyla, araştırma kapsamı 2020–2025 yılları arasında İngilizce olarak yayımlanan makalelerle sınırlandırılmıştır. Araştırmanın temel eksenini sürdürülebilir tedarik zinciri yönetimi, çok kriterli karar verme yöntemleri ve tedarikçi seçim süreçleri oluşturmaktadır. Bu bağlamda, çalışmanın teorik tutarlılığını ve kapsam geçerliliğini korumak adına, doğrudan bu tematik alanlara dahil olmayan, operasyonel içerikten yoksun veya diğer disiplinlerin teorik çerçevesine odaklanan çalışmalar hariç tutma kriterleri

kapsamında değerlendirilerek araştırma evreninin dışında bırakılmıştır.

5. Sonuçların analiz edilmesi: Yapılan ilk taramalar sonucunda Web of Science (WoS) veri tabanından 1680, Scopus veri tabanından ise 1253 çalışma tespit edilmiştir. Belirlenen dahil etme ve hariç tutma kriterleri doğrultusunda uygulanan filtreleme işlemleri neticesinde, WoS bünyesindeki yayın sayısı 220'ye, Scopus bünyesindeki yayın sayısı ise 315'e düşürülmüştür. Elde edilen bu makaleler EndNote yazılımına aktarılarak birleştirilmiş ve tekrarlayan kayıtların ayıklanmasıyla veri seti 286 makaleye indirilmiştir. Son aşamada, araştırmanın kapsam geçerliliğini güvence altına almak adına makalelerin tam metinleri titizlikle incelenmiş ve araştırma odağı ve konusuyla doğrudan örtüşmeyen çalışmalar elenmiştir.

Tarama süreci sonucunda ulaşılan makalelerin tematik dağılımı Şekil 2'de sunulmuştur.



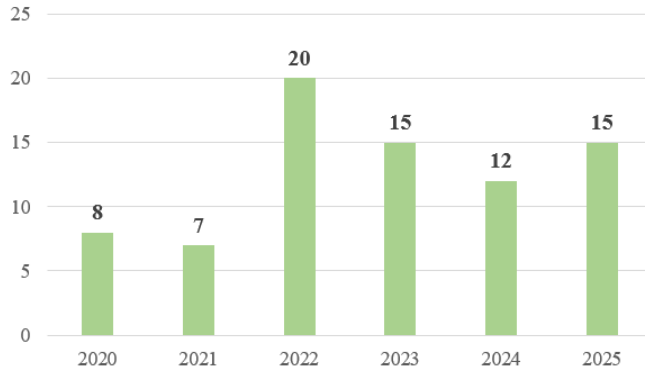
Şekil 2. Tarama sonuçları

İlgili yazın genel olarak geleneksel, yeşil ve sürdürülebilir tedarikçi seçimi başlıkları altında kümelenmiştir. Elde edilen bulgular, sürdürülebilir ve yeşil tedarikçi seçimi çalışmalarının literatürde daha sınırlı bir yer tuttuğunu göstermektedir. Bu araştırmada, hem küresel sürdürülebilirlik standartlarına uyum sağlama gerekliliği hem de literatürdeki mevcut araştırma boşluklarını doldurma motivasyonuyla sürdürülebilir tedarikçi

seçimine odaklanılmıştır. Geleneksel yaklaşımların aksine, ekonomik, çevresel ve sosyal boyutları bir arada değerlendiren bu alan, yenilikçi karar mekanizmalarını incelemek için daha stratejik bir zemin sunmaktadır.

2.1. Yayın Eğilimi

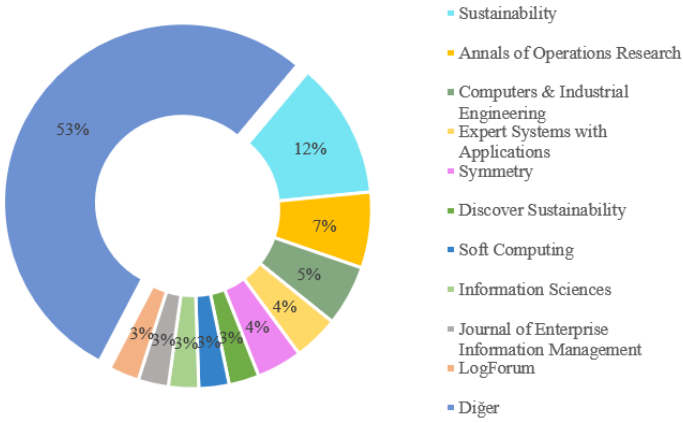
Bu bölümde, 1. araştırma sorusuna yanıt verebilmek amacıyla sürdürülebilir tedarikçi seçimi için yayınların eğilimleri incelenmiştir. Öncelikle 2020–2025 yıllarını kapsayan 77 çalışma incelenmiş ve çalışmaların yayın yılına göre dağılımı Şekil 3'te verilmiştir.



Şekil 3. Makalelerin yıllara göre dağılımı

Şekil 3 incelendiğinde, ele alınan konuya yönelik akademik ilginin zamansal süreçte dalgalı bir seyir izlemekle birlikte dinamik yapısını koruduğu görülmektedir. En yüksek yayın sayısı 2022 yılında 20 çalışma ile gerçekleşmiştir. 2024 yılında çalışma sayısının biraz düşmesi, araştırma faaliyetlerinde geçici bir yavaşlamaya işaret etmektedir. Bununla birlikte, 2025 yılında yayın sayısının tekrar yükselmesi, konuya yönelik ilginin devam ettiğini ve araştırmaların yeniden ivme kazandığını göstermektedir. Son yıllardaki bu genel yoğunluk, ele alınan problemin güncel tedarik zinciri yönetimindeki stratejik önemini ve akademik çevrelerdeki sürdürülebilir çözüm arayışlarının devam ettiğini açıkça ortaya koymaktadır.

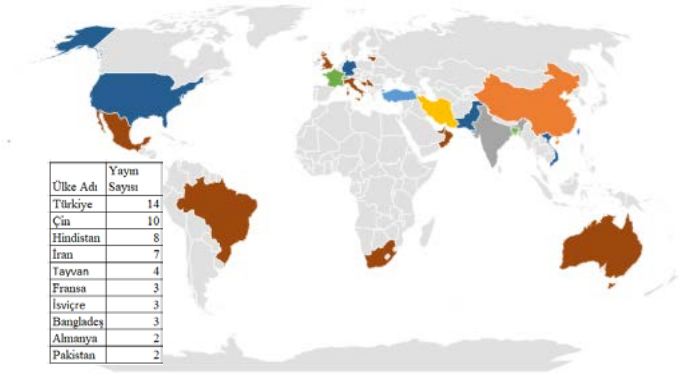
Şekil 4'te ise çalışmaların yayımlandıkları dergilere göre dağılımı gösterilmektedir.



Şekil 4. En çok yayın yapılan dergiler

Bulgular, yayınların %12'sinin Sustainability, %7'sinin Annals of Operations Research ve %5'inin Computers & Industrial Engineering dergilerinde yayımlandığını göstermektedir. Bununla birlikte, yayınların %53'lük kısmının "Diğer" kategorisinde yer alan çeşitli dergilerde yayımlandığı görülmektedir. Bu durum, sürdürülebilir tedarikçi seçimi konusundaki araştırmaların belirli dergilerde yoğunlaşmakla birlikte geniş bir dergi yelpazesine yayıldığını ve sürdürülebilirlik, operasyon yönetimi, bilgi sistemleri ile yumuşak hesaplama gibi farklı akademik topluluklar tarafından paralel olarak ele alındığını, dolayısıyla alanın disiplinlerarası bir araştırma yapısına sahip olduğunu ortaya koymaktadır.

Şekil 5, araştırma kapsamında değerlendirilen ülkelerin dünya üzerindeki coğrafi dağılımını göstermektedir. Yayınları ülkelere göre ayırırken, sorumlu yazarın bağlı olduğu kurumun kökeni dikkate alınmıştır.



Şekil 5. Yayınların coğrafi dağılım analizi

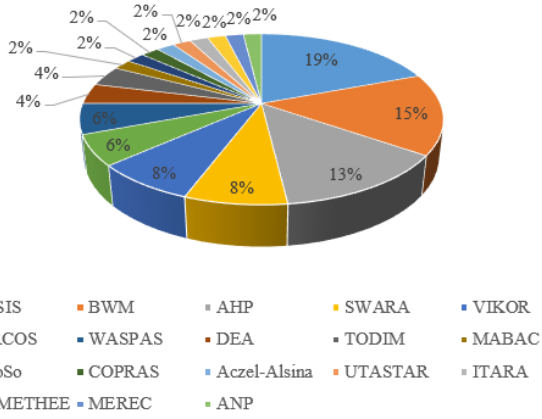
Çalışmaya 24 ülke katkıda bulunmuştur. Coğrafi dağılım incelendiğinde, Türkiye'nin %26'lık bir pay ile literatürde baskın bir üretkenliğe sahip olduğu görülmektedir. Burada örneklemin Kuzey Amerika, Güney Amerika, Avrupa, Asya, Afrika ve Okyanusya kıtalarına yayıldığı görülmektedir. Bu dağılım, çalışmanın yalnızca belirli bir bölgeye odaklanmadığını, farklı ekonomik, kültürel ve coğrafi özelliklere sahip ülkeleri kapsayarak karşılaştırmalı bir analiz imkânı sunduğunu göstermektedir.

Tablo 1'de incelenen çalışmalar arasında en yüksek atıf alan beş makalenin yayınlandıkları dergiler ve uygulandıkları sektörler gösterilmektedir.

Tablo 1. En fazla atıf alan beş yayının analizi

Yazar	Alıntı Sayısı	Yayınlandığı Dergi	Uygulanan Sektör
(Stević, Pamučar, Puška, & Chatterjee, 2020)	1727	Computers & Industrial Engineering	Sağlık
(Kusi-Sarpong et al., 2023)	218	Production Planning & Control	Tekstil
(Chai, Zhou, & Jiang, 2023)	173	Information Sciences	Elektrikli Bisiklet Paylaşım Hizmetleri
(Rahman, Bari, Ali, & Taghipour, 2022)	124	Resources, Conservation & Recycling Advances	Tekstil
(Ortiz-Barrios et al., 2021)	120	Annals of Operations Research	Madencilik

Tablo 1 incelendiğinde, (Stević et al., 2020) çalışmasının literatürde önemli bir referans noktası haline geldiği görülmektedir. Yüksek atıf alan diğer çalışmaların ise tekstil, sağlık, madencilik ve elektrikli bisiklet paylaşım hizmetleri gibi farklı sektörlerde uygulandığı dikkat çekmektedir. Özellikle sağlık sektöründe gerçekleştirilen çalışma, sahip olduğu yüksek atıf sayısıyla öne çıkarken, tekstil sektörünün sürdürülebilirlik ve çevresel etkiler açısından taşıdığı önem nedeniyle literatürde tekrarlayan bir araştırma alanı oluşturduğu görülmektedir. Sonuçlar, sürdürülebilir tedarikçi seçimi konusunun farklı sektörlerde yaygın olarak uygulandığını ve alanın disiplinlerarası bir araştırma yapısına sahip olduğunu göstermektedir.

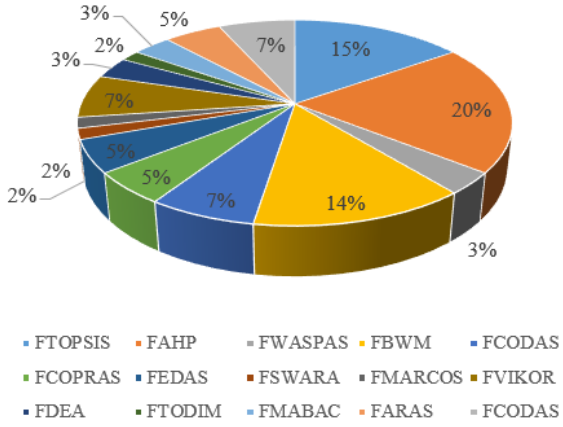


Şekil 6. Sürdürülebilir tedarikçi seçiminde kullanılan yöntemler

Şekil 6'da sunulan oransal dağılım incelendiğinde, sürdürülebilir tedarikçi seçimi literatürünün belirli yöntemler etrafında yoğunlaştığı görülmektedir. TOPSIS (%19), BWM (%15) ve AHP (%13) yöntemleri toplamda yazının yaklaşık yarısını (%47) oluşturarak en çok tercih edilen üç temel metodoloji olarak öne çıkmaktadır. Ancak araştırmacıların aynı zamanda MARCOS, WASPAS, CoCoSo gibi yeni nesil ÇKKV tekniklerini de aktif olarak kullandıkları görülmektedir. Yöntem çeşitliliğinin yüksek olması, alanın metodolojik açıdan olgun

ancak hâlâ yenilikçi arayışlara açık bir karaktere sahip olduğunu göstermektedir.

Şekil 7'de sürdürülebilir tedarikçi seçimi literatüründe kullanılan bulanık çok kriterli karar verme yöntemlerinin dağılımı sunulmaktadır.



Şekil 7. Çalışmada kullanılan bulanık ÇKKV yöntemleri

FAHP yönteminin %20'lik payla en sık tercih edilen bulanık teknik olduğu görülmekte olup, bu durum geleneksel AHP yönteminin belirsizlik ve subjektiflik içeren karar verme ortamlarına uyarlanmış bulanık versiyonunun literatürde geniş kabul gördüğünü göstermektedir. FTOPSIS (%15) ve onu izleyen FBWM (%14) yöntemleri de klasik ÇKKV tekniklerinin bulanık mantıkla genişletilmiş hallerinin yaygın olarak kullanıldığı görülmektedir. Bu durum, bulanık küme teorisinin sürdürülebilir tedarikçi seçimi problemindeki belirsizlik ve öznel yargı unsurlarını modellemede vazgeçilmez bir araç haline geldiğini ve mevcut ÇKKV yöntemlerinin çoğunun bulanık versiyonlarının geliştirilmiş olduğunu ortaya koymaktadır.

Bulanık küme tabanlı çalışmaların alt türlere göre dağılımı incelendiğinde, sürdürülebilir tedarikçi seçimi literatüründe standart bulanık mantığın (Fuzzy Logic) baskın bir konuma sahip

olduğu görülmektedir. Bulanık gruptaki toplam 42 çalışmanın %69'una karşılık gelen 29 çalışma, klasik (Type-1) bulanık küme teorisini kullanmaktadır (Ben Abdallah, El-Amraoui, Delmotte, & Frikha, 2024; Gidiagba, Okwu, & Tartibu, 2025; Orji & Ojadi, 2021; Ortiz-Barrios et al., 2021). Bu bulgu, (Zadeh, 1965)'in öncü çalışmasıyla ortaya çıkan standart bulanık mantığın, sürdürülebilir tedarikçi seçimi probleminde hâlâ en çok tercih edilen yaklaşım olduğunu göstermektedir.

Aralık değerli sezgisel bulanık kümeler (IVIF) 6 çalışmada (%14) kullanılmıştır (Afzali, Afzali, & Pourmohammadi, 2022; Hendiani & Walther, 2025; Salimian, Mousavi, & Antucheviciene, 2022). IVIF yaklaşımı, üyelik ve üyelik-dışı derecelerini aralıklarla ifade ederek belirsizlik ve tereddütleri daha esnek biçimde modellemeyi sağlar (K. Atanassov & Gargov, 1989). Bu nedenle özellikle yüksek düzeyde belirsizlik içeren karar verme problemlerinde tercih edildiği görülmektedir.

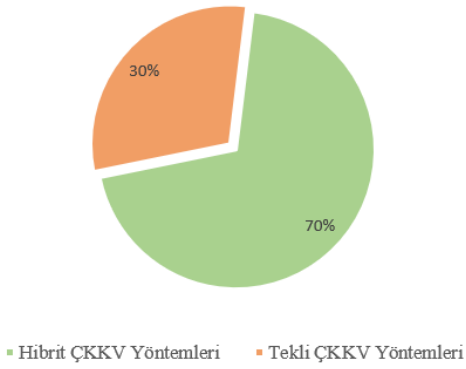
Küresel bulanık kümeler (Spherical Fuzzy Sets) ise 4 çalışmada (%10) kullanılmıştır. Son yıllarda geliştirilen bu yaklaşım, Pisagor bulanık kümeler ve nötrosofik bulanık kümelerin güçlü yönlerini bir araya getiren üç boyutlu bir bulanık küme yaklaşımıdır. Üyelik, üyelik olmama ve tereddüt derecelerini daha geniş bir karar alanında ele alması nedeniyle karar verme problemlerinde giderek daha fazla tercih edilmektedir (Kutlu Gündoğdu & Kahraman, 2019).

Pisagor ve Fermatean bulanık kümelerin kullanıldığı 3 çalışma (%7), literatürde klasik bulanık ve sezgisel bulanık yaklaşımların sınırlılıklarını aşmaya yönelik geliştirilen yöntemler arasında yer almaktadır (Demiralay & Paksoy, 2022; Farid, Bouye, Riaz, & Jamil, 2023). Nötrosofik ve Picture Fuzzy gibi diğer gelişmiş bulanık yaklaşımlar ise yalnızca 2 çalışmada

yer alarak metodolojik çeşitliliğin en güncel örneklerini oluşturmaktadır (Sun, Yu, Li, Yuan, & Zhao, 2025).

Bulanık küme temelli yaklaşımların yanı sıra, son yıllarda belirsizliğin farklı boyutlarını modelleyebilmek amacıyla gri sayılar (Sun et al., 2025; Ulutaş et al., 2022), kaba sayılar (ForouzeshNejad, 2023) ve D sayıları (Gökler & Boran, 2025) gibi alternatif yöntemler de kullanılmaya başlanmıştır. Her ne kadar kullanım oranları bulanık küme temelli yöntemlere kıyasla daha düşük olsa da bu yaklaşımlar eksik ve kesin olmayan bilgilerin temsilindeki yetenekleri sayesinde literatürde giderek daha fazla ilgi görmektedir.

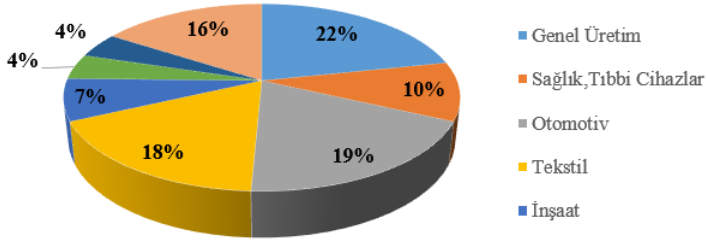
İncelenen çalışmada kullanılan karar verme yaklaşımlarının dağılımı Şekil 8'de gösterilmektedir.



Şekil 8. Sürdürülebilir tedarikçi seçiminde hibrit ve tekli ÇKKV yöntemlerinin kullanımı

Çalışmaların büyük çoğunluğunda (%70) hibrit ÇKKV yöntemlerinin kullanıldığı görülmektedir. Bu sonuç, sürdürülebilir tedarikçi seçimi problemlerinin çok sayıda kriter ve belirsizlik içermesi nedeniyle araştırmacıların farklı yöntemlerin avantajlarını bir araya getiren hibrit yaklaşımları daha fazla tercih ettiğini göstermektedir.

İncelenen çalışmaların sektörlere göre dağılımı Şekil 9'da verilmiştir.



Şekil 9. Çalışmalarının sektörlere göre dağılımı

Literatür taramasından elde edilen çalışmaların %22'sinin genel üretim ve %19'unun otomotiv alanlarında uygulandığı ve çalışmaların odak noktasını oluşturduğu görülmektedir.

2.2. Kriterlerin Değerlendirilmesi

İkinci araştırma sorusu kapsamında incelenen yayınların kriter yapıları analiz edildiğinde, sürdürülebilir tedarikçi seçimi literatüründe ekonomik, çevresel ve sosyal boyutları içeren klasik üçlü temel çerçevenin baskın bir referans modeli olarak kullanıldığı görülmektedir. Tablo 2'de literatürde yaygın olarak kullanılan kriterlerin değerlendirmesi verilmiştir.

Tablo 2. Literatürde kullanılan sürdürülebilir tedarikçi seçim kriterleri

Ana Kriterler	Referanslar	Kriter adı	Tanımı
Ekonomik Kriterler	(Afzali et al., 2022), (Aditi, Kannan, Darbari, & Jha, 2023),	Kalite	Tedarikçinin ürün/hizmetinin standartları ve müşteri beklentilerini karşılama düzeyidir.
	(Giri & Roy, 2024), (Orji & Ojadi, 2021), (Kazancoglu, Ozturkoglu, Mangla, Ozbiltekin-Pala, & Ishizaka, 2023)	Maliyet	Tedarik sürecine ilişkin satın alma fiyatı ve toplam finansal yüktür.

	(Liu, Rani, & Pachori, 2022), (Nguyen et al., 2022)	Teslimat	Siparişlerin zamanında ve eksiksiz teslim edilme performansıdır.
	(Aslani, Rabiee, & Tavana, 2021), (Ulutaş et al., 2022)	Esneklik	Değişen talep ve koşullara hızlı uyum sağlayabilme kapasitesidir.
	(Aditi et al., 2023), (Chai et al., 2023)	Teknik Yetenek	Tedarikçinin mühendislik ve ürün geliştirme kapasitesidir.
	(Liaqait et al., 2021), (Liaqait, Warsi, Agha, Zahid, & Becker, 2022)	Hacim Esnekliği	Üretim miktarını talebe göre artırıp azaltabilme yeteneğidir.
Çevresel Kriterler	(Chang et al., 2023), (Bonab et al., 2023)	Çevre Yönetim Sistemi	Çevresel etkileri planlama ve izlemeye yönelik yönetim uygulamalarıdır.
	(Chai et al., 2023), (Lo, Liaw, Gul, & Lin, 2021)	Kirlilik	Üretim faaliyetlerinin yol açtığı emisyon ve atık düzeyidir.
	(Aslani et al., 2021), (Liu et al., 2022)	Eko-tasarım	Ürünün yaşam döngüsü boyunca çevresel etkisini azaltacak şekilde tasarlanmasıdır.
	(Orji & Ojadi, 2021), (Bonab et al., 2023)	Yeşil ürün	Çevreye verdiği zararı en aza indiren ürünleri ifade eder.
Sosyal Kriterler	(Torğul & Paksoy, 2022), (Hendiani, Torkayesh, Venghaus, & Walther, 2025)	İş sağlığı ve güvenliği	Çalışanların sağlığını ve güvenliğini koruma düzeyidir.
	(Chang et al., 2023), (Giri & Roy, 2024)	Çalışan hakları	Adil çalışma koşulları ve insan haklarına saygı düzeyidir.
	(Eghbali-Zarch, Zabihi, & Masoud, 2023), (Rahman et al., 2022)	Personel eğitimi	Çalışanların mesleki gelişimine yönelik eğitim faaliyetleridir.
	(Badulescu, Hameri, & Cheikhrouhou, 2022), (Lo et al., 2021)	İstihdam olanakları	Tedarikçinin faaliyet gösterdiği bölgede yarattığı iş imkânları ve istihdama katkı düzeyidir.

Çalışmaların yaklaşık %10'luk bir bölümünde ise klasik (ekonomik, çevresel ve sosyal) yapının dayanıklılık (Bonab et al., 2023), risk yönetimi (Giri & Roy, 2024), pandemi sonrası kriz yanıtları (Dang, Nguyen, Nguyen, & Dang, 2022) ve akıllı/Endüstri 4.0 (Demiralay & Paksoy, 2022) boyutlarıyla

genişletildiği görülmektedir. Bu genişleme eğiliminin özellikle 2022 yılından itibaren belirginleşmesi, COVID-19 sonrası dönemde tedarik zinciri kırılganlıklarının literatüre doğrudan yansıdığını göstermektedir. Söz konusu çalışmalar, sürdürülebilirliği artık tek başına yeterli bir değerlendirme çerçevesi olarak görmemekte tedarikçilerin beklenmedik kesintiler karşısındaki uyum ve toparlanma kapasitesini de değerlendirme sürecine dahil etmektedir. Bu durum, literatürde "sürdürülebilir-dayanıklı (sustainable-resilient) tedarikçi seçimi" şeklinde adlandırılabilecek yeni bir hibrit araştırma çizgisinin ortaya çıktığına işaret etmekte ve sürdürülebilirlik ile dayanıklılığın artık birbirinden bağımsız değil, bütünleşik biçimde ele alınması gereken iki boyut olarak konumlandığını ortaya koymaktadır.

3. SONUÇ

Bu çalışmada, sürdürülebilir tedarikçi seçimi alanında çok kriterli karar verme yöntemlerinin kullanımına ilişkin literatür sistematik biçimde incelenmiş, ele alınan çalışmaların yöntemsel eğilimleri, kriter yapıları ve zaman içerisindeki gelişimi ortaya konmuştur. Elde edilen bulgular genel hatlarıyla özetlenerek araştırma soruları doğrultusunda değerlendirilmiştir. Üçüncü araştırma sorusu kapsamında elde edilen bulgular, FAHP, FTOPSIS ve FBWM gibi bulanık ÇKKV yöntemlerinin literatürde yaygın olarak kullanıldığını, standart bulanık mantığın ise belirsizliklerin modellenmesinde baskın yaklaşım olmaya devam ettiğini göstermektedir. Bununla birlikte, IVIF, Küresel Bulanık, Pisagor ve Fermatean bulanık kümeler gibi gelişmiş yaklaşımların kullanımında artış gözlemlenmekte, ayrıca gri sayılar, kaba sayılar ve D sayıları gibi alternatif yöntemler de araştırmacıların ilgisini çekmektedir. Kriterler açısından değerlendirildiğinde, geleneksel ekonomik kriterlerin önemini

koruduğu, ancak çevresel ve sosyal sürdürülebilirlik kriterlerinin giderek daha fazla çalışmaya dahil edildiği görülmektedir. Bu durum, sürdürülebilir tedarikçi değerlendirmesinin çok boyutlu bir yapıya dönüştüğünü ortaya koymaktadır.

Gelecek çalışmalarda, gelişmiş belirsizlik modelleme yaklaşımlarının yeni nesil ÇKKV yöntemleriyle bütünleştirilmesi ve sürdürülebilirlik performansının daha kapsamlı değerlendirilmesine yönelik hibrit modellerin geliştirilmesi önerilmektedir. Ayrıca yapay zekâ, büyük veri analitiği ve dijital tedarik zinciri uygulamalarının sürdürülebilir tedarikçi seçimi modellerine entegrasyonu, literatüre önemli katkılar sağlayabilecek araştırma alanları olarak değerlendirilmektedir. Bunun yanında, sosyal sürdürülebilirlik kriterlerinin ve dayanıklı tedarik zinciri göstergelerinin daha fazla dikkate alınması, gelecekteki çalışmalar için önemli araştırma fırsatları sunmaktadır.

KAYNAKÇA

Aditi, Kannan, D., Darbari, J. D., & Jha, P. (2023). Sustainable supplier selection model with a trade-off between supplier development and supplier switching. *Annals of operations research, 331*(1), 351-392.

Afzali, M., Afzali, A., & Pourmohammadi, H. (2022). An interval-valued intuitionistic fuzzy-based CODAS for sustainable supplier selection: M. Afzali et al. *Soft Computing, 26*(24), 13527-13541.

Aslani, B., Rabiee, M., & Tavana, M. (2021). An integrated information fusion and grey multi-criteria decision-making framework for sustainable supplier selection. *International Journal of Systems Science: Operations & Logistics, 8*(4), 348-370.

Badulescu, Y., Hameri, A.-P., & Cheikhrouhou, N. (2022). Sustainable partner selection for collaborative networked organisations with risk consideration in the context of COVID-19. *Journal of global operations and strategic sourcing, 15*(2), 197-218.

Ben Abdallah, C., El-Amraoui, A., Delmotte, F., & Frikha, A. (2024). A Hybrid approach for sustainable and resilient farmer selection in food industry: Tunisian case study. *Sustainability, 16*(5), 1889.

Bonab, S. R., Haseli, G., Rajabzadeh, H., Ghoushchi, S. J., Hajiaghaei-Keshteli, M., & Tomášková, H. (2023). Sustainable resilient supplier selection for IoT implementation based on the integrated BWM and TRUST under spherical fuzzy sets. *Decision Making: Applications in Management and Engineering, volume 6, issue: 1*.

Chai, N., Zhou, W., & Jiang, Z. (2023). Sustainable supplier selection using an intuitionistic and interval-valued fuzzy MCDM approach based on cumulative prospect theory. *Information Sciences, 626*, 710-737.

Chang, J.-P., Chen, Z.-S., Wang, X.-J., Martínez, L., Pedrycz, W., & Skibniewski, M. J. (2023). Requirement-driven sustainable supplier selection: Creating an integrated perspective with stakeholders' interests and the wisdom of expert crowds. *Computers & industrial engineering, 175*, 108903.

Dang, T.-T., Nguyen, N.-A.-T., Nguyen, V.-T.-T., & Dang, L.-T.-H. (2022). A two-stage multi-criteria supplier selection model for sustainable automotive supply chain under uncertainty. *Axioms, 11*(5), 228.

Demiralay, E., & Paksoy, T. (2022). Strategy development for supplier selection process with smart and sustainable criteria in fuzzy environment. *Cleaner Logistics and Supply Chain, 5*, 100076.

Eghbali-Zarch, M., Zabihi, S. Z., & Masoud, S. (2023). A novel fuzzy SECA model based on fuzzy standard deviation and correlation coefficients for resilient-sustainable supplier selection. *Expert Systems with Applications, 231*, 120653.

Erdebilli, B., & Sıcakyüz, Ç. (2024). An integrated Q-rung orthopair fuzzy (Q-ROF) for the selection of supply-chain management. *Sustainability, 16*(12), 4901.

Farid, H. M. A., Bouye, M., Riaz, M., & Jamil, N. (2023). Fermatean fuzzy CODAS approach with topology and its application to sustainable supplier selection. *Symmetry, 15*(2), 433.

ForouzeshNejad, A. A. (2023). Leagile and sustainable supplier selection problem in the Industry 4.0 era: a case study of

the medical devices using hybrid multi-criteria decision making tool. *Environmental science and pollution research, 30*(5), 13418-13437.

Gidiagba, J. O., Okwu, M., & Tartibu, L. (2025). Multi-criteria decision support for sustainable supplier evaluation in mining SMEs: A fuzzy logic and TOPSIS approach. *Logistics, 9*(3), 132.

Giri, B. K., & Roy, S. K. (2024). CI-MM-Dombi operator based on interval type-2 spherical fuzzy set and its applications on sustainable supply chain with risk criteria: using CI-TODIM-MARCOS method. *Soft Computing, 28*(17), 10023-10056.

Govindan, K., Rajendran, S., Sarkis, J., & Murugesan, P. (2015). Multi criteria decision making approaches for green supplier evaluation and selection: a literature review. *Journal of cleaner production, 98*, 66-83.

Gökler, S. H., & Boran, S. (2025). A novel resilient and sustainable supplier selection model based on D-AHP and DEMATEL methods. *Journal of engineering research, 13*(1), 57-67.

Hendiani, S., Torkayesh, A. E., Venghaus, S., & Walther, G. (2025). Leading the green charge: A novel type-2 fuzzy VIKOR method applied to eco-conscious freight transport. *Expert Systems with Applications, 285*, 128082.

Hendiani, S., & Walther, G. (2025). Towards sustainable futures: Rethinking supplier selection through interval-valued intuitionistic fuzzy decision-making. *International Journal of Production Economics, 285*, 109620.

K. Atanassov, & Gargov, G. (1989). Interval valued intuitionistic fuzzy sets. *Fuzzy Sets and Systems, 31*(3), 343-349.

Kazancoglu, Y., Ozturkoglu, Y., Mangla, S. K., Ozbiltekin-Pala, M., & Ishizaka, A. (2023). A proposed framework for multi-tier supplier performance in sustainable supply chains. *International Journal of Production Research, 61*(14), 4742-4764.

Kusi-Sarpong, S., Gupta, H., Khan, S. A., Chiappetta Jabbour, C. J., Rehman, S. T., & Kusi-Sarpong, H. (2023). Sustainable supplier selection based on industry 4.0 initiatives within the context of circular economy implementation in supply chain operations. *Production Planning & Control, 34*(10), 999-1019.

Kutlu Gündoğdu, F., & Kahraman, C. (2019). Spherical fuzzy sets and spherical fuzzy TOPSIS method. *Journal of intelligent & fuzzy systems, 36*(1), 337-352.

Liaqait, R. A., Warsi, S. S., Agha, M. H., Zahid, T., & Becker, T. (2022). A multi-criteria decision framework for sustainable supplier selection and order allocation using multi-objective optimization and fuzzy approach. *Engineering Optimization, 54*(6), 928-948.

Liaqait, R. A., Warsi, S. S., Zahid, T., Ghafoor, U., Ahmad, M. S., & Selvaraj, J. (2021). A decision framework for solar PV panels supply chain in context of sustainable supplier selection and order allocation. *Sustainability, 13*(23), 13216.

Liu, C., Rani, P., & Pachori, K. (2022). Sustainable circular supplier selection and evaluation in the manufacturing sector using Pythagorean fuzzy EDAS approach. *Journal of Enterprise Information Management, 35*(4-5), 1040-1066.

Lo, H.-W., Liaw, C.-F., Gul, M., & Lin, K.-Y. (2021). Sustainable supplier evaluation and transportation planning in multi-

level supply chain networks using multi-attribute-and multi-objective decision making. *Computers & industrial engineering, 162*, 107756.

Nguyen, T.-L., Nguyen, P.-H., Pham, H.-A., Nguyen, T.-G., Nguyen, D.-T., Tran, T.-H., . . . Phung, H.-T. (2022). A novel integrating data envelopment analysis and spherical fuzzy MCDM approach for sustainable supplier selection in steel industry. *Mathematics, 10*(11), 1897.

Orji, I. J., & Ojadi, F. (2021). Investigating the COVID-19 pandemic's impact on sustainable supplier selection in the Nigerian manufacturing sector. *Computers & industrial engineering, 160*, 107588.

Ortiz-Barrios, M., Cabarcas-Reyes, J., Ishizaka, A., Barbati, M., Jaramillo-Rueda, N., & de Jesús Carrascal-Zambrano, G. (2021). A hybrid fuzzy multi-criteria decision making model for selecting a sustainable supplier of forklift filters: A case study from the mining industry. *Annals of operations research, 307*(1), 443-481.

Rahman, M. M., Bari, A. M., Ali, S. M., & Taghipour, A. (2022). Sustainable supplier selection in the textile dyeing industry: An integrated multi-criteria decision analytics approach. *Resources, Conservation & Recycling Advances, 15*, 200117.

Salimian, S., Mousavi, S. M., & Antucheviciene, J. (2022). An interval-valued intuitionistic fuzzy model based on extended VIKOR and MARCOS for sustainable supplier selection in organ transplantation networks for healthcare devices. *Sustainability, 14*(7), 3795.

Stević, Ž., Pamučar, D., Puška, A., & Chatterjee, P. (2020). Sustainable supplier selection in healthcare industries using a new MCDM method: Measurement of alternatives

and ranking according to COmpromise solution (MARCOS). *Computers & industrial engineering, 140*, 106231.

Sun, L., Yu, C., Li, J., Yuan, Q., & Zhao, S. (2025). A two-stage decision model for sustainable-resilient supplier selection and order allocation under uncertain environment. *Kybernetes, 54*(8), 4078-4113.

Torğul, B., & Paksoy, T. (2022). Smart and sustainable supplier selection using interval type-2 fuzzy AHP. *Politeknik Dergisi, 26*(4), 1359-1373.

Ulutaş, A., Topal, A., Pamučar, D., Stević, Ž., Karabašević, D., & Popović, G. (2022). A new integrated multi-criteria decision-making model for sustainable supplier selection based on a novel grey WISP and grey BWM methods. *Sustainability, 14*(24), 16921.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control, 8*(3), 338-353.

MACHINE LEARNING-BASED DECISION SUPPORT FOR REMANUFACTURING VIABILITY: A CROSS-MATERIAL COMPARATIVE STUDY OF ECONOMIC AND ENERGETIC OPTIMALITY IN CIRCULAR SUPPLY CHAINS

Ümit YILMAZ[1]

1. INTRODUCTION

The global manufacturing sector is undergoing a structural transformation driven by the convergence of circular economy principles and Industry 4.0-enabled production systems. Remanufacturing, which is the industrial process of restoring end-of-life (EOL) products to an as-new condition, has emerged as one of the most resource-efficient strategies within this transformation, conceptually recognized for its potential to minimize resource input, energy leakage, and waste generation compared to primary manufacturing (Geissdoerfer et al., 2017; Ghisellini et al., 2016). The environmental benefits of remanufacturing have been substantiated by empirical lifecycle assessments and dynamic greenhouse gas emission models, demonstrating substantial reductions in virgin material use and CO_2 emissions relative to new part production within the heavy-duty and automotive machinery sectors (Liu et al., 2014; Shi et al., 2019; Xiao et al., 2018). Beyond environmental merits, the economic rationale is compelling: cost savings relative to new production are frequently reported, and macroeconomic

[1] Instructor Dr., Bursa Technical University, Quality Coordination Office, ORCID: 0000-0003-4268-8598.

assessments project that the European remanufacturing and circular economy market will expand substantially, potentially creating a net benefit of €1.8 trillion by 2030 (MacArthur, 2015).

Despite these advantages, practical adoption remains constrained by the complexity of supply chain design decisions. Manufacturers must choose among competing configurations: a traditional linear scenario without product recovery (SC1), a direct closed-loop configuration in which returned products are processed within a manufacturer–customer dyad (SC2), and an indirect closed-loop configuration mediated by an intermediate distribution centre (SC3). The focal study by Ferraro et al. (2024a), which also generated the dataset (Ferraro et al., 2024b) used in this research, proposed a three-stage methodology based on mathematical modeling to evaluate this choice across economic and energetic dimensions for steel, aluminium, and titanium mechanical components, demonstrating that Remanufacturing Volume Rate and material preserved rate are the most critical variables in scenario selection.

Classical deterministic approaches, such as exact mathematical programming, often struggle with the NP-hard complexity and large search spaces inherent in modern, dynamic remanufacturing networks (Caterino et al., 2022). To address these computational bottlenecks, the literature is increasingly shifting towards data-driven Industry 4.0 paradigms and advanced algorithmic frameworks (Caterino et al., 2022; Kerin & Pham, 2019). Machine learning (ML) offers a complementary paradigm within this transition: once trained on labelled simulation output, classifiers enable rapid scenario prediction for new parameter combinations. However, the application of ML to this specific task has not been explored previously. No study has investigated whether gradient-boosted classifiers can learn the economic and energetic scenario-optimality boundaries from the Ferraro et al. (2024b) data, whether performance differs

systematically across evaluation criteria, or which features drive predictions across material types. This study addresses these gaps directly.

Three specific research questions motivate this work:

- RQ1: Are gradient-boosted classifiers (XGBoost, LightGBM) statistically superior to Random Forest for remanufacturing scenario classification across all material–criterion combinations?
- RQ2: Does ML classification performance differ significantly between economic and energetic evaluation criteria, and what structural feature-space properties explain the observed differences?
- RQ3: Does the SHAP (SHapley Additive exPlanations) feature importance profile vary significantly across steel, aluminium, and titanium?

The principal contributions are: (i) the first ML-based classification pipeline applied to the Ferraro et al. (2024b) remanufacturing simulation dataset; (ii) a statistically rigorous algorithm comparison using Bonferroni-corrected pooled Wilcoxon tests on 30 cross-validation (CV) scores per pair; (iii) a cross-material SHAP analysis demonstrating material-specific feature importance heterogeneity; and (iv) a leakage-free, reproducible methodological template applicable to other multi-class remanufacturing decision support problems. The remainder of this paper is organised as follows. Section 2 reviews the relevant literature. Section 3 describes the dataset, methodology, and experimental protocol. Section 4 presents the results. Section 5 discusses findings. Section 6 concludes.

2. LITERATURE REVIEW

The intersection of remanufacturing decision support, ML for circular supply chains, and model interpretability constitutes the theoretical space within which this study is situated.

Remanufacturing has been extensively studied as a core circular economy strategy to decouple economic value creation from resource depletion (Geissdoerfer et al., 2017; Ghisellini et al., 2016). Geissdoerfer et al. (2017) established a theoretical framework distinguishing the circular economy from prior sustainability constructs, while Kirchherr et al. (2017) provided a comprehensive analysis of circular economy definitions, emphasizing the 4R framework (reduce, reuse, recycle, recover) as the core operational principles. In the supply chain domain, remanufacturing is framed as a product recovery strategy within closed-loop supply chains (CLSCs) (Guide & Van Wassenhove, 2009; Souza, 2013). Souza (2013) synthesised the CLSC literature, identifying channel design, core quality uncertainty, and remanufacturing capacity as the dominant research themes. The impact of channel structures, such as direct versus platform-mediated CLSC configurations, has been analysed from economic and competitive perspectives (Shi et al., 2024). Furthermore, Wan and Xie (2024) demonstrated that government subsidy structures significantly influence investment in remanufacturing technology. Zhang et al. (2023) examined consumer preference and capacity constraints in remanufacturing supply network design. The pivotal work underpinning the dataset used here, Ferraro et al. (2024b), proposed a three-stage methodology based on mathematical modeling comparing three supply chain scenarios across economic and energetic dimensions for steel, aluminium, and titanium, generating 60,000 labelled observations.

The application of ML to circular economy and remanufacturing decision support has grown substantially with Industry 4.0 technologies (Paraschos et al., 2022). Paraschos et al. (2022) proposed ML-integrated design and operation management for resilient circular manufacturing systems. Koulinas et al. (2024) developed an ML framework for explainable knowledge mining and for optimizing quality control in flexible circular manufacturing. Bhattacharya et al. (2024) systematically reviewed Artificial Intelligence (AI) applications in CLSCs, identifying production planning, network design, and disassembly sequence planning as the most extensively explored domains, while highlighting predictive applications as a key area for future research. Kerin et al. (2023) demonstrated digital twin-augmented optimisation for EOL product decision-making. Govindan (2024) identified critical success factors for ML integration in remanufacturing quality inspection. Paschko et al. (2026) applied reinforcement learning for decision-making in disassembly processes. While these works confirm the broad applicability of ML to remanufacturing contexts, none have applied ML-based multi-class classification to the scenario selection problem defined by Ferraro et al. (2024a).

Ensemble tree-based classifiers have emerged as the dominant paradigm for tabular classification in industrial settings (Grinsztajn et al., 2022; Shwartz-Ziv & Armon, 2022). Breiman (2001)'s Random Forest established the foundational ensemble learning architecture through bootstrap aggregation and random feature subsampling. XGBoost, proposed by Chen and Guestrin (2016), extended gradient boosting with second-order Taylor approximations and sparsity-aware split finding. LightGBM, introduced by Ke et al. (2017), improved computational efficiency through Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). Hyperparameter sensitivity is a recognised challenge for all three; to address this,

Akiba et al. (2019) introduced Optuna, featuring a novel define-by-run API that enables efficient Bayesian hyperparameter search using algorithms such as the Tree-structured Parzen Estimator (TPE).

Class imbalance is endemic to remanufacturing datasets where minority-class scenarios are structurally underrepresented. SMOTE (Chawla et al., 2002) addresses this by interpolating within training folds. Joloudari et al. (2023) demonstrated in Applied Sciences that combining SMOTE with deep learning architectures consistently improves minority-class recall across a wide range of imbalanced benchmark datasets. Best practices for robust SMOTE application, particularly the critical need to mitigate partition-induced dataset shift and to ensure rigorous validation, as highlighted by Fernández et al. (2018), are operationalised here through the leakage-free imblearn pipeline used in the present study.

Explainability is a critical requirement for high-stakes industrial ML decisions. Lundberg and Lee (2017) introduced SHAP; the TreeExplainer variant (Lundberg et al., 2020) provides exact values for tree-based models. Applications of SHAP to manufacturing include process quality prediction (Puthanveettil Madathil et al., 2024), explainable reinforcement learning scheduling (Immordino et al., 2026), and a systematic XAI review in manufacturing (Tzionis et al., 2025). Senoner et al. (2022) demonstrated that SHAP-guided process improvement reduced yield loss in semiconductor manufacturing, evidencing the operational actionability of SHAP attributions.

3. MATERIALS AND METHODS

3.1. Dataset

The dataset was generated by Ferraro et al. (2024b) via Monte Carlo simulation to support economic and energetic assessment of remanufacturing scenarios for mechanical components. The repository contains six Excel files corresponding to three materials (steel, aluminium, titanium), each under economic and energetic evaluation, with 10,000 observations per file and 60,000 in total. All datasets share the three-class target variable, Best Scenario (1, 2, 3), encoding SC1 (Linear Supply Chain), SC2 (Direct Closed-Loop), and SC3 (Indirect Closed-Loop).

Economic datasets contain 13 cost-based features (e.g., Cost of Raw Material Manufacturing/Remanufacturing, Cost of Set Up, Cost of Labour, Cost of Effective Production, three Transportation Cost components, Cost of Disposal, Remanufacturing Cost Rate, Remanufacturing Volume Rate). Energetic datasets contain 9 energy-based features (e.g., MJ Primary Material Manufacturing/Remanufacturing, kg initial material, l_rate, MJ Primary Production Manufacturing/Remanufacturing, Remanufacturing Volume Rate, Remanufacturing Finishing Rate, Distance 1). The columns Best Scenario (1,2) and Best Scenario (1,3) were identified as leaky and removed before feature extraction.

All six files contain no missing values or duplicates and have the correct target cardinality ({1, 2, 3}). Class distribution and imbalance ratios are summarised in Table 1. Economic datasets exhibit severe class imbalance (SC1 dominance: 77.8–81.6%; SC3 frequency: 0.6–0.8%), motivating the use of SMOTE and balanced accuracy (BA) as the primary metric. Energetic datasets are more balanced (SC1: 8.4–55.3%).

Table 1. Dataset summary: observations per class, imbalance ratio (IR = majority/minority count), and feature count.

Dataset	SC1 (n)	SC2 (n)	SC3 (n)	IR	Features
Steel – Economic	8,157	1,770	73	~111.7	13
Steel – Energetic	5,527	2,302	2,171	~2.5	9
Aluminium – Economic	8,075	1,851	74	~109.1	13
Aluminium – Energetic	1,854	4,149	3,997	~2.2	9
Titanium – Economic	7,784	2,158	58	~134.2	13
Titanium – Energetic	842	4,652	4,506	~5.5	9

3.2. Experimental Setup

A data preparation pipeline was implemented to address the challenges of class-imbalanced tabular classification (Chawla et al., 2002). Leaky columns were removed before partitioning, and the target was re-encoded to the 0-indexed {0, 1, 2}. Following best practices to mitigate prior-probability shift (Fernández et al., 2018), each dataset was split into training (80%) and held-out test (20%) sets using stratified random sampling (SEED=42). All subsequent pre-processing was applied within a leakage-free imblearn Pipeline to prevent data leakage across folds.

Within each training fold, the pipeline applies: (i) SMOTE oversampling (k_neighbors=5) exclusively to the training fold; (ii) StandardScaler normalisation fitted on the oversampled training fold; and (iii) the classifier. The k_neighbors=5 value follows the original SMOTE recommendation (Chawla et al., 2002). It is consistent with established best practices in comprehensive methodological reviews (Fernández et al., 2018) and diverse benchmark applications (Joloudari et al., 2023). This architecture strictly prevents any validation-fold information from influencing pre-processing.

This pipeline implements a sequential rather than nested CV protocol. Hyperparameter optimisation is performed first on the training partition using a separate five-fold CV; the resulting optimal configuration is then evaluated via a second independent

five-fold CV on the same partition. The held-out test set, participating in neither step, provides an independent final performance estimate.

Three ensemble tree-based classifiers were evaluated: XGBoost (Chen & Guestrin, 2016), LightGBM (Ke et al., 2017), and Random Forest (Breiman, 2001). XGBoost employs second-order gradient boosting with L1/L2 regularisation and sparsity-aware split finding. LightGBM uses leaf-wise growth with GOSS and EFB. Random Forest uses bootstrap aggregation and random feature subsampling (Breiman, 2001). To address class imbalance during model training, the Random Forest implementation was configured with balanced class weighting (class_weight='balanced').

Hyperparameter optimisation was performed for each classifier–dataset combination (18 studies) using Optuna (Akiba et al., 2019) with the TPE sampler (seed=42) and MedianPruner (n_warmup_steps=10), with 80 trials per combination; the objective was the mean BA across five-fold CV on the training set. Search spaces are given in Table 2.

Table 2. Hyperparameter search spaces for Optuna TPE optimisation.

Hyperparameter	XGBoost	LightGBM	Random Forest
n_estimators	{100,150,…,600}	{100,150,…,600}	{100,150,…,600}
max_depth	[3, 9]	[3, 9]	[5, 30]
learning_rate	[0.01, 0.3] (log-uniform)	[0.01, 0.3] (log-uniform)	—
subsample	[0.5, 1.0]	[0.5, 1.0]	—
colsample_bytree	[0.5, 1.0]	[0.5, 1.0]	—
reg_alpha / reg_lambda	[10^{-4}, 10] (log-uniform)	[10^{-4}, 10] (log-uniform)	—
num_leaves	—	[20, 150]	—
min_child_samples	—	[5, 50]	—
min_child_weight	[1, 10]	—	—
gamma	[0.0, 5.0]	—	—
min_samples_split	—	—	[2, 20]
min_samples_leaf	—	—	[1, 10]
max_features	—	—	{sqrt, log2, 0.5}

Performance was evaluated using five-fold stratified CV (StratifiedKFold, n_splits=5, shuffle=True, random_state=42) on the training set, reporting mean ± SD per metric. Final models were trained on the full training set and evaluated once on the held-out test set, using the following metrics: Balanced Accuracy (BA), F1-Macro, MCC, and Cohen's κ. BA is the primary metric.

Statistical significance was assessed in two stages. First, a per-dataset Friedman χ^2 omnibus test was applied to the five-fold BA scores of three classifiers ($\alpha = 0.05$). Second, Bonferroni-corrected Wilcoxon signed-rank tests were applied to pooled BA scores across all six datasets (n = 30: 6 datasets × 5 folds), yielding $\alpha\_Bonf = 0.05/3 = 0.0167$. Pooling was adopted because the per-dataset Wilcoxon test with n=5 has a minimum achievable p-value of 0.0625, which is below the Bonferroni threshold.

Post hoc explainability was performed using the SHAP framework (Lundberg & Lee, 2017), specifically employing the TreeExplainer algorithm (Lundberg et al., 2020), on the final model with the highest BA on the held-out test set for each dataset. SHAP values were computed on the scaled test set for all three classes, producing an array of shape (n_test, n_features, n_classes). Feature importance was visualised as mean absolute SHAP values, aggregated across classes, for global cross-dataset comparison (Lundberg et al., 2020). Furthermore, SHAP dependence plots (Lundberg et al., 2020) illustrate the marginal relationship between standardised Remanufacturing Volume Rate and its SC2 SHAP contribution across materials.

4. RESULTS

4.1. Class Distribution and Correlation Structure

Figure 1 shows that no feature pair exceeded $|\rho| = 0.10$ across all six feature sets, confirming the absence of

multicollinearity and justifying retention of all features without dimensionality reduction.

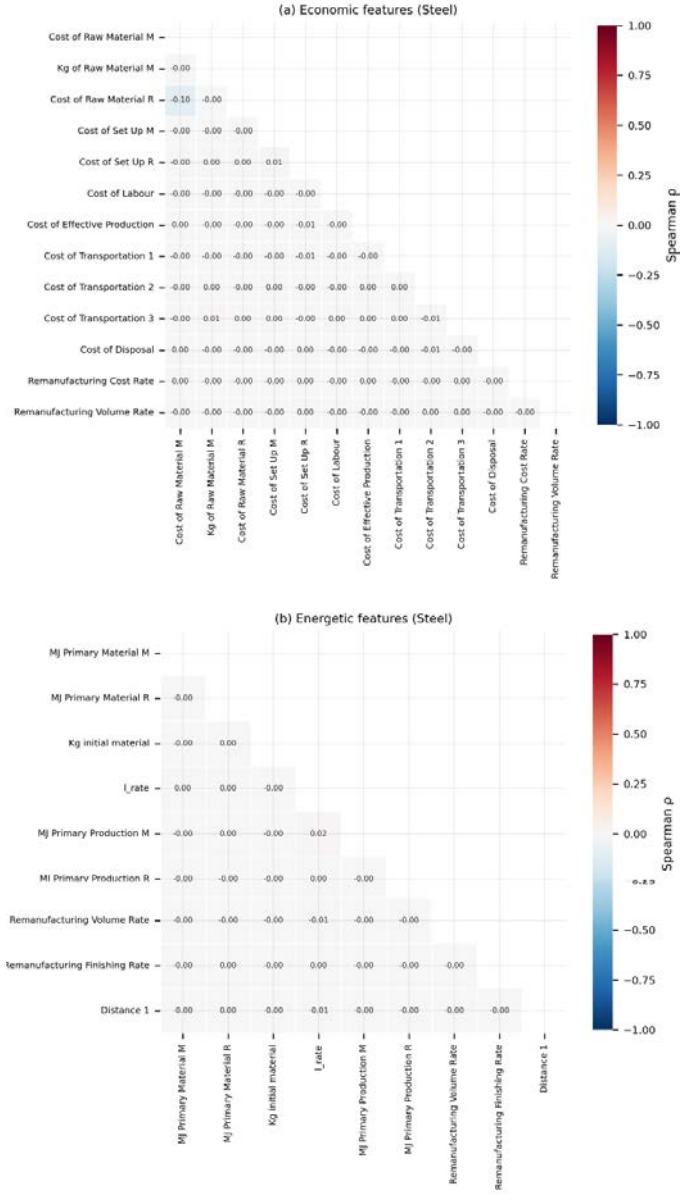


Figure 1. Spearman rank correlation matrices (lower triangle): Steel Economic (a) and Steel Energetic (b). All |ρ| < 0.10.

4.2. Model Performance

Table 3 summarises the best-model five-fold CV and hold-out test performance per dataset. Gradient-boosted classifiers dominated economic datasets, with XGBoost (XGB) and LightGBM (LGBM) achieving CV BA of 0.918–0.928, compared to Random Forest (RF)'s 0.864–0.890. On the energetic datasets, all three models converged to CV BA values of 0.766–0.795, with SD < 0.015. Hold-out test results confirm CV findings; best-model test BA ranged from 0.768 (Aluminium Energetic) to 0.951 (Steel Economic).

Table 3. Best-model performance per dataset: five-fold CV balanced accuracy (mean ± SD) and hold-out test metrics.

Dataset	Best Model	CV BA (mean±SD)	Test BA	Test F1M	Test MCC	Test κ
Steel – Economic	LGBM	0.918 ± 0.031	0.951	0.801	0.867	0.861
Steel – Energetic	XGB	0.775 ± 0.010	0.792	0.791	0.732	0.732
Aluminium – Economic	XGB	0.928 ± 0.042	0.920	0.779	0.836	0.828
Aluminium – Energetic	LGBM	0.768 ± 0.010	0.768	0.749	0.589	0.587
Titanium – Economic	LGBM	0.923 ± 0.043	0.935	0.773	0.865	0.863
Titanium – Energetic	XGB	0.795 ± 0.006	0.792	0.755	0.546	0.544

4.3. Statistical Tests

Table 4 presents the statistical significance results, addressing RQ1. The per-dataset Friedman test detected significant differences for Steel Economic (χ^2=7.6, p=0.022) and Aluminium Economic (χ^2=10.0, p=0.007). No significant Friedman effect was observed for any energetic dataset (all p > 0.07), confirming classifier convergence under energetic criteria. The Bonferroni-corrected pooled Wilcoxon tests (n=30) yielded: XGB vs RF (W=33.0, p=0.000006) and LGBM vs RF (W=36.0, p=0.000009), establishing statistical superiority of both gradient-boosted models over Random Forest. XGB vs LGBM was non-

significant (W=152.0, p=0.100), confirming statistical equivalence of the two boosted models and answering RQ1.

Table 4. Statistical significance results.

Test scope	Test	Statistic	p-value	Significant?
Steel – Economic	Friedman χ^2	χ^2=7.6	0.022	Yes ($p < 0.05$)
Steel – Energetic (per-dataset)	Friedman χ^2	χ^2=1.2	0.549	No
Aluminium – Economic (per-dataset)	Friedman χ^2	χ^2=10.0	0.007	Yes ($p < 0.05$)
Aluminium – Energetic (per-dataset)	Friedman χ^2	χ^2=2.8	0.247	No
Titanium – Economic (per-dataset)	Friedman χ^2	χ^2=4.8	0.091	No
Titanium – Energetic (per-dataset)	Friedman χ^2	χ^2=5.2	0.074	No
XGB vs LGBM (pooled n=30)	Wilcoxon (Bonf.)	W=152.0	0.100	No ($p > 0.0167$)
XGB vs RF (pooled n=30)	Wilcoxon (Bonf.)	W=33.0	0.000006	Yes ($p < 0.0167$)
LGBM vs RF (pooled n=30)	Wilcoxon (Bonf.)	W=36.0	0.000009	Yes ($p < 0.0167$)

4.4. Confusion Matrices and Per-Class Performance

The normalised confusion matrices reveal qualitatively different patterns between economic and energetic datasets, as shown in Figure 2. For economic datasets, SC1 recall is consistently high (0.94–0.97), SC2 recall is near-perfect (0.94–0.97), while SC3 recall, which is the rarest class, ranges from 0.86 (Aluminium Economic) to 0.93 (Steel Economic). For energetic datasets, SC1 recall remains high (0.92–0.95), but SC2 and SC3 recall drop to 0.68–0.73, revealing systematic confusion between these two classes. This SC2–SC3 confusion under energetic criteria is structurally informative: both configurations involve remanufacturing, and their energetic profiles are differentiated primarily by distribution-centre logistics (Distance 1) rather than primary energy metrics, which are similar across both closed-loop configurations. The energetic feature space, therefore, lacks the dimensional contrast needed to separate SC2 from SC3 reliably.

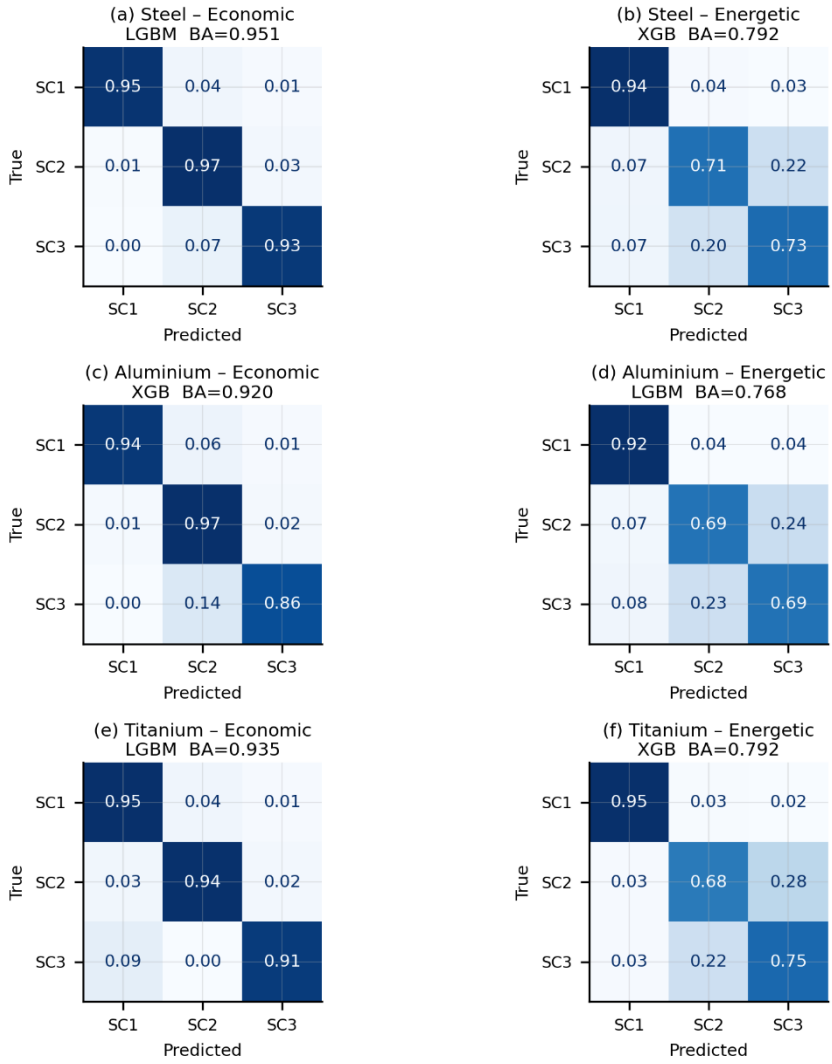


Figure 2. Normalised confusion matrices

4.5. Multi-Metric Radar

Figure 3 presents multi-metric radar charts across BA, F1-Macro, MCC, and Cohen's κ. XGBoost and LightGBM produce nearly overlapping polygons on economic datasets; Random Forest exhibits a visibly smaller area, particularly on MCC and F1-Macro. On energetic datasets, all three polygons converge,

reinforcing the finding that classifier choice has limited practical impact under energetic criteria.

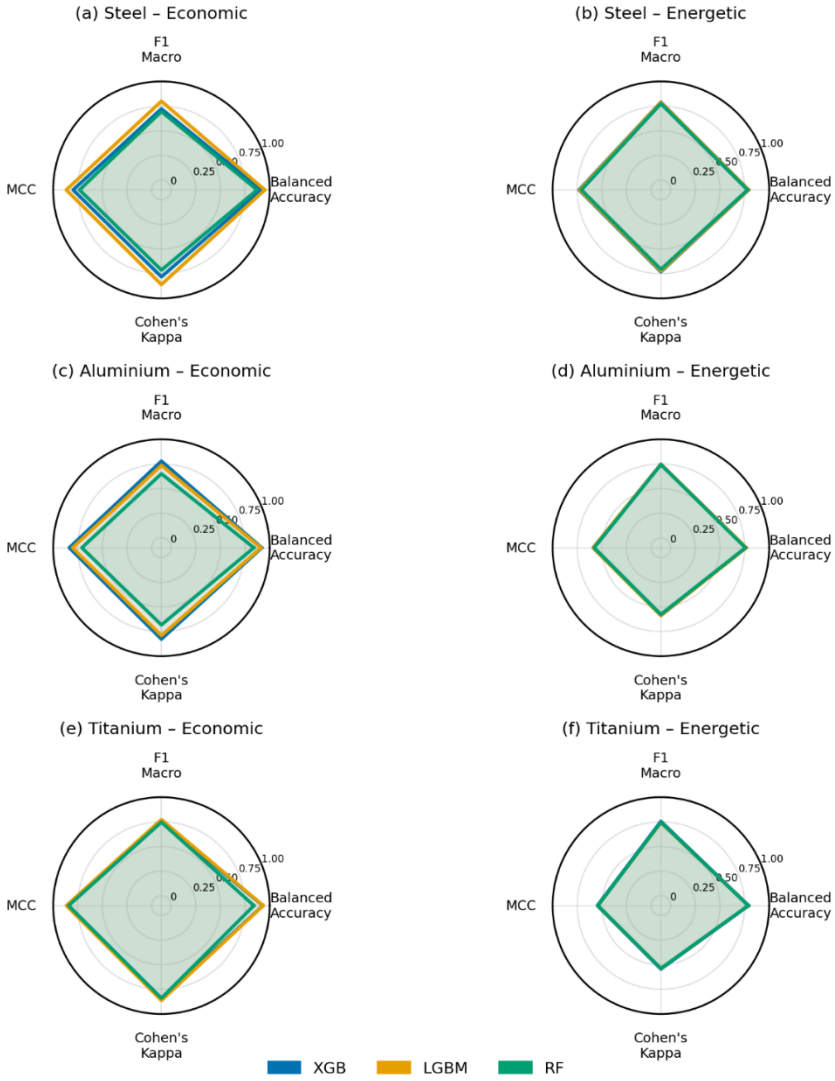


Figure 3. Multi-metric performance radar (hold-out test set).

4.6. SHAP Feature Importance

Figure 4 presents the mean absolute SHAP values per feature for the best-performing model in each dataset, addressing

RQ3. Remanufacturing Volume Rate ranks among the top features across all six datasets and all three materials, confirming the prediction of Ferraro et al. (2024b). Remanufacturing Cost Rate ranks second or third in all economic datasets. Cross-material comparison reveals material-specific heterogeneity in secondary feature importance: Steel and Titanium share Remanufacturing Volume Rate and Cost of Raw Material Manufacturing among their top-5 features. At the same time, Aluminium's profile places greater weight on Cost of Effective Production and transportation costs. For energetic datasets, MJ Primary Material Manufacturing and MJ Primary Production Manufacturing dominate across all three materials. This cross-material heterogeneity answers RQ3 affirmatively.

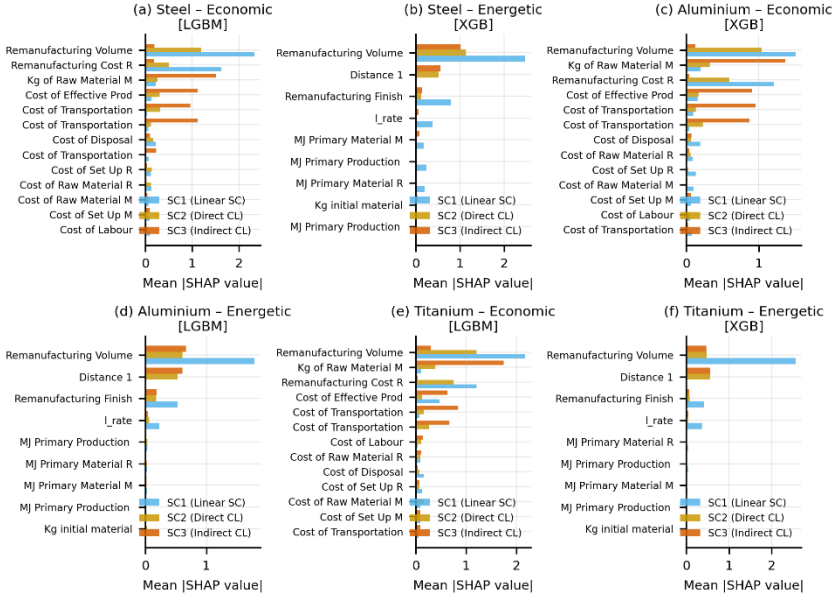


Figure 4. Mean absolute SHAP values per feature (sorted descending) for best-performing model on each dataset.

4.7. SHAP Beeswarm and Dependence

Figure 5 presents the SHAP beeswarm for Steel Economic (LightGBM, SC3 class). High Remanufacturing Volume Rate

values strongly increase SC3 probability; high Remanufacturing Cost Rate values suppress SC3 in favour of SC2. SC3 emerges as economically optimal only when remanufacturing volume is high, and cost rates are high, conditions under which the distribution-centre infrastructure cost is amortised by volume. Figure 6 extends the Volume Rate dependence to SC2 probability across all three materials, showing a positive SHAP–feature relationship for Steel and more dispersed patterns for Aluminium and Titanium, reflecting material-specific moderating effects.

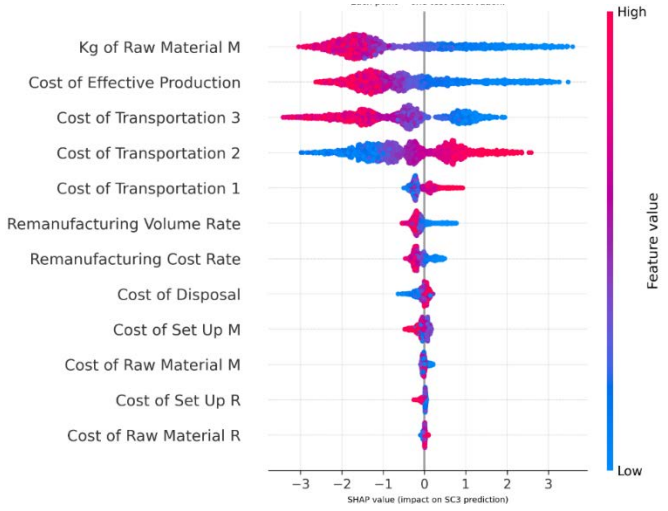


Figure 5. SHAP beeswarm plot for Steel Economic dataset (LightGBM, best model) for SC3 class.

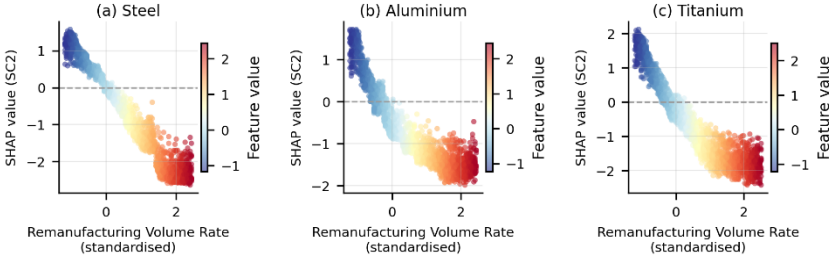


Figure 6. SHAP dependence plots for Remanufacturing Volume Rate (SC2 class): steel (a), aluminium (b), titanium (c).

5. DISCUSSION

The systematic BA gap between economic (BA=0.864–0.951) and energetic (BA=0.766–0.795) evaluation criteria, ranging from 0.07 to 0.19 BA units across all materials and classifiers, answers RQ2 affirmatively. The economic framework incorporates 13 cost components that vary continuously and differentially across configurations, creating wide, discriminative decision boundaries. The energetic framework relies primarily on 9 energy–volume relationships that are highly correlated within material classes, yielding narrower margins separating SC2 from SC3.

The SC2–SC3 confusion identified in the confusion matrix analysis provides the mechanistic explanation. In the original simulation framework, both closed-loop configurations are mathematically identical with respect to energetic criteria, meaning they share the same energetic footprint (Ferraro et al., 2024a). This structural equivalence perfectly explains the classifiers' inability to reliably separate the two classes energetically. Under economic criteria, however, the presence of a distribution centre in SC3 translates into significant holding costs for inventory management, whereas SC2 incurs zero holding costs (Ferraro et al., 2024a). This fundamental structural difference in logistics overhead varies widely across simulated scenarios, providing the richer discriminative signal that enables accurate economic classification.

Supply chain designers using economic criteria can rely on ML classification with high confidence: XGBoost and LightGBM achieve BA $\geq$ 0.90 on all three economic datasets, while even Random Forest exceeds BA = 0.84 in the worst case. When energetic criteria are primary, as required by carbon accounting regulations, ML-based recommendations carry higher uncertainty, suggesting that energetic scenario classification may

require richer feature engineering (e.g., lifecycle energy profiles, material-specific energy recovery rates) or complementary simulation.

The statistical equivalence of XGBoost and LightGBM ($p = 0.100$), combined with their shared superiority over Random Forest (XGB vs RF: $p = 0.000006$; LGBM vs RF: $p = 0.000009$), answers RQ1 directly. The inferior performance of Random Forest on economic datasets, despite class_weight='balanced' and SMOTE, is attributable to its inability to efficiently partition heavily imbalanced distributions: each independent tree must learn the minority-class boundary from scratch, without the iterative residual correction that allows gradient-boosted models to concentrate predictive mass on misclassified minority instances (Chen & Guestrin, 2016; Ke et al., 2017).

From a practitioner standpoint, LightGBM's computational advantage (training approximately 10× faster than XGBoost on large datasets (Ke et al., 2017)) may favour its adoption in high-throughput contexts. Given their statistical indistinguishability, the choice may be made on operational grounds without sacrificing accuracy.

The SHAP analysis confirms and extends Ferraro et al. (2024b): Remanufacturing Volume Rate is the dominant driver across all six datasets. The beeswarm analysis (Figure 5) reveals that the mechanistic condition for SC3 optimality is simultaneously a high volume rate and a high cost rate. This aligns with fundamental CLSC network design principles (Guide & Van Wassenhove, 2009; Souza, 2013), which posit that indirect channels incur additional intermediary logistics overhead but facilitate bulk aggregation. Our SHAP results empirically demonstrate that this theoretical trade-off becomes viable only above a specific volume–cost threshold. The dominance of MJ Primary Material Manufacturing in energetic SHAP profiles

reflects the fundamental structure of the energetic comparison: primary energy for material production dwarfs distribution-related energy differences between SC2 and SC3, explaining both the lower discriminability and the SC2–SC3 confusion. Cross-material heterogeneity in secondary features answers RQ3 affirmatively.

5.1. Limitations and Future Directions

The dataset (Ferraro et al., 2024b) originates from the Monte Carlo simulation framework developed by Ferraro et al. (2024a) rather than empirical field data, limiting direct transferability to specific industrial settings where process variability, quality uncertainty, and demand volatility introduce additional stochasticity. The sequential hyperparameter optimisation and CV protocol may carry a slight optimistic performance bias compared to nested CV. The current framework treats each material–criterion combination independently; a joint multi-task or transfer learning approach could leverage shared structure across materials. The adoption of such advanced ML frameworks aligns with the future research agenda for closed-loop supply chains highlighted by Bhattacharya et al. (2024). The circular economy framing of the present study also opens directions for resilience-oriented scenario assessment (Kennedy & Linnenluecke, 2022). SHAP provides additive attributions but does not explicitly capture non-additive interaction effects. Future work should explore richer energetic feature engineering, extension to additional materials, multi-objective scenario assessment, and federated learning for cross-company benchmarking.

6. CONCLUSIONS

This study presents the first ML-based classification framework for selecting remanufacturing supply chain scenarios,

applied to the multi-material, dual-criterion simulation dataset of Ferraro et al. (2024b), comprising 60,000 observations. The principal conclusions are:

- XGBoost and LightGBM are statistically indistinguishable (Bonferroni-corrected Wilcoxon $p = 0.100$) but both significantly outperform Random Forest (XGB vs RF: $p = 0.000006$; LGBM vs RF: $p = 0.000009$), answering RQ1.
- Economic evaluation criteria yield substantially higher BA (0.864–0.951) than energetic criteria (0.766–0.795) across all materials and classifiers, answering RQ2. The mechanism is the richer cost-feature dimensionality under economic analysis, and the structural difficulty in separating SC2 from SC3 using energy metrics alone.
- SHAP analysis consistently identifies Remanufacturing Volume Rate as the dominant driver across all materials and criteria, corroborating Ferraro et al. (2024b) and adding the mechanistic finding that SC3 requires simultaneously high volume rate and high cost rate to become economically optimal.
- Cross-material SHAP comparison reveals material-specific heterogeneity in secondary feature importance, answering RQ3: steel, aluminium, and titanium feature spaces are not interchangeable.
- The leakage-free pipeline, Bonferroni-corrected testing protocol, and SHAP interpretability framework constitute a reproducible template applicable to multi-class remanufacturing decision support in circular supply chains.

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna: A Next-generation Hyperparameter Optimization Framework* Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA. https://doi.org/10.1145/3292500.3330701

Bhattacharya, S., Govindan, K., Ghosh Dastidar, S., & Sharma, P. (2024). Applications of artificial intelligence in closed-loop supply chains: Systematic literature review and future research agenda. *Transportation Research Part E: Logistics and Transportation Review*, *184*, 103455. https://doi.org/10.1016/j.tre.2024.103455

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Caterino, M., Fera, M., Macchiaroli, R., & Pham, D. T. (2022). Cloud remanufacturing: Remanufacturing enhanced through cloud technologies. *Journal of Manufacturing Systems*, *64*, 133–148. https://doi.org/10.1016/j.jmsy.2022.06.003

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939785

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and

challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, *61*, 863–905. https://doi.org/10.1613/jair.1.11192

Ferraro, S., Baffa, F., Cantini, A., Leoni, L., De Carlo, F., & Campatelli, G. (2024a). Exploring remanufacturing conveniency: An economic and energetic assessment for a closed-loop supply chain of a mechanical component. *Journal of Cleaner Production*, *458*, 142504. https://doi.org/10.1016/j.jclepro.2024.142504

Ferraro, S., Baffa, F., Cantini, A., Leoni, L., De Carlo, F., & Campatelli, G. (2024b). *Supplementary data: Exploring Remanufacturing Conveniency: An Economic and Energetic Assessment for a Closed-Loop Supply Chain of a Mechanical Component* (Zenodo. https://doi.org/10.5281/zenodo.11082914

Geissdoerfer, M., Savaget, P., Bocken, N. M. P., & Hultink, E. J. (2017). The Circular Economy – A new sustainability paradigm? *Journal of Cleaner Production*, *143*, 757–768. https://doi.org/10.1016/j.jclepro.2016.12.048

Ghisellini, P., Cialani, C., & Ulgiati, S. (2016). A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. *Journal of Cleaner Production*, *114*, 11–32. https://doi.org/10.1016/j.jclepro.2015.09.007

Govindan, K. (2024). Unlocking the potential of quality as a core marketing strategy in remanufactured circular products: A machine learning enabled multi-theoretical perspective. *International Journal of Production Economics*, *269*, 109123. https://doi.org/10.1016/j.ijpe.2023.109123

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree-based models still outperform deep learning on*

typical tabular data? Proceedings of the 36th International Conference on Neural Information Processing Systems, New Orleans, LA, USA.

Guide, V. D. R., & Van Wassenhove, L. N. (2009). OR FORUM—The Evolution of Closed-Loop Supply Chain Research. *Operations Research*, *57*(1), 10–18. https://doi.org/10.1287/opre.1080.0628

Immordino, A., Stöckermann, P., Hayen, N., Altenmüller, T., Susto, G. A., Gebser, M., Schekotihin, K., & Seidel, G. (2026). Explainable AI for reinforcement learning based dynamic scheduling solutions in semiconductor manufacturing. *Journal of Intelligent Manufacturing*, *37*(5), 1999–2015. https://doi.org/10.1007/s10845-025-02631-3

Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences*, *13*(6), 4006. https://doi.org/10.3390/app13064006

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). *LightGBM: a highly efficient gradient boosting decision tree* Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA.

Kennedy, S., & Linnenluecke, M. K. (2022). Circular economy and resilience: A research agenda. *Business Strategy and the Environment*, *31*(6), 2754–2765. https://doi.org/10.1002/bse.3004

Kerin, M., Hartono, N., & Pham, D. T. (2023). Optimising remanufacturing decision-making using the bees

algorithm in product digital twins. *Scientific Reports*, *13*(1), 701. https://doi.org/10.1038/s41598-023-27631-2

Kerin, M., & Pham, D. T. (2019). A review of emerging industry 4.0 technologies in remanufacturing. *Journal of Cleaner Production*, *237*, 117805. https://doi.org/10.1016/j.jclepro.2019.117805

Kirchherr, J., Reike, D., & Hekkert, M. (2017). Conceptualizing the circular economy: An analysis of 114 definitions. *Resources, Conservation and Recycling*, *127*, 221–232. https://doi.org/10.1016/j.resconrec.2017.09.005

Koulinas, G. K., Paraschos, P. D., & Koulouriotis, D. E. (2024). A machine learning framework for explainable knowledge mining and production, maintenance, and quality control optimization in flexible circular manufacturing systems. *Flexible Services and Manufacturing Journal*, *36*(3), 737–759. https://doi.org/10.1007/s10696-024-09537-x

Liu, Z., Li, T., Jiang, Q., & Zhang, H. (2014). Life Cycle Assessment–based Comparative Evaluation of Originally Manufactured and Remanufactured Diesel Engines. *Journal of Industrial Ecology*, *18*(4), 567–576. https://doi.org/10.1111/jiec.12137

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions* Proceedings of the 31st

International Conference on Neural Information Processing Systems, Long Beach, California, USA.

MacArthur, E. (2015). Towards a circular economy: business rationale for an accelerated transition. *Greener Manag International*, *20*(3), 22–34.

Paraschos, P. D., Xanthopoulos, A. S., Koulinas, G. K., & Koulouriotis, D. E. (2022). Machine learning integrated design and operation management for resilient circular manufacturing systems. *Computers & Industrial Engineering*, *167*, 107971. https://doi.org/10.1016/j.cie.2022.107971

Paschko, F., Krini, A., Kemke, M., & Knorn, S. (2026). Reinforcement learning approach for decision making in disassembly processes. *Journal of Intelligent Manufacturing*, *37*(5), 1829–1855. https://doi.org/10.1007/s10845-025-02622-4

Puthanveettil Madathil, A., Luo, X., Liu, Q., Walker, C., Madarkar, R., Cai, Y., Liu, Z., Chang, W., & Qin, Y. (2024). Intrinsic and post-hoc XAI approaches for fingerprint identification and response prediction in smart manufacturing processes. *Journal of Intelligent Manufacturing*, *35*(8), 4159–4180. https://doi.org/10.1007/s10845-023-02266-2

Senoner, J., Netland, T., & Feuerriegel, S. (2022). Using Explainable Artificial Intelligence to Improve Process Quality: Evidence from Semiconductor Manufacturing. *Management Science*, *68*(8), 5704–5723. https://doi.org/10.1287/mnsc.2021.4190

Shi, J., Fan, S., Wang, Y., & Cheng, J. (2019). A GHG emissions analysis method for product remanufacturing: A case study on a diesel engine. *Journal of Cleaner Production*,

206, 955–965. https://doi.org/10.1016/j.jclepro.2018.09.200

Shi, Y., Ma, R., & Yang, T. (2024). Remanufacturing and channel strategies in e-commerce closed-loop supply chain. *PLOS ONE*, *19*(5), e0303447. https://doi.org/10.1371/journal.pone.0303447

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, *81*, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Souza, G. C. (2013). Closed-Loop Supply Chains: A Critical Review, and Future Research. *Decision Sciences*, *44*(1), 7–38. https://doi.org/10.1111/j.1540-5915.2012.00394.x

Tzionis, G., Mouratidis, P., Kougka, G., Gialampoukidis, I., Vrochidis, S., Kompatsiaris, I., & Vlachopoulou, M. (2025). A review of explainable AI methods and their application in manufacturing systems. *Discover Applied Sciences*, *8*(1), 52. https://doi.org/10.1007/s42452-025-07908-z

Wan, P., & Xie, Z. (2024). Decision making and benefit analysis of closed-loop remanufacturing supply chain considering government subsidies. *Heliyon*, *10*(19), e38487. https://doi.org/10.1016/j.heliyon.2024.e38487

Xiao, L., Liu, W., Guo, Q., Gao, L., Zhang, G., & Chen, X. (2018). Comparative life cycle assessment of manufactured and remanufactured loading machines in China. *Resources, Conservation and Recycling*, *131*, 225–234. https://doi.org/10.1016/j.resconrec.2017.12.021

Zhang, X., Zhou, G., Cao, J., & Lu, J. (2023). A remanufacturing supply chain network with differentiated new and remanufactured products considering consumer preference, production capacity constraint and

government regulation. *PLOS ONE*, *18*(8), e0289349. https://doi.org/10.1371/journal.pone.0289349

ENDÜSTRİ MÜHENDİSLİĞİ ALANINDA AKADEMİK TARTIŞMALAR

yaz
yayınları

YAZ Yayınları
M.İhtisas OSB Mah. 4A Cad. No:3/3
İscehisar / AFYONKARAHİSAR
Tel : (0 531) 880 92 99
yazyayinlari@gmail.com • www.yazyayinlari.com