Journal of Optimization in Industrial Engineering

Vol.18, Issue 1, Winter & Spring 2025, 337-350

Research Article

A data-driven approach to enhance resilience and sustainability in agricultural waste management network facility location

Seyyed Ali Mousavikhorram [1], Mohamad Mahdi Nasiri [2,*], Matineh Ziari [3]

1. School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
2. School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
3. School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran

https://doi.org/10.71720/joie.2025.1200908

Received: 02 Narch 2025
Revised: 07 October 2025
Accepted: 15 October 2025

Keywords:
Waste Management;
Sustainability;
Resilience;
Data-driven Models;
Agricultural industry.

Abstract

This study focuses on one of the crucial issues in today's world named the agricultural waste management by considering the sustainability and resilience features. In this regard, this research proposes a hybrid data-driven decision framework by combining the multiple-attribute decision-making and machine learning methods. In this way, first of all, the major indicators of the considered problem are specified and their weights are calculated using a recently developed method called the stochastic Best-Worst Method (BWM). In the next stage, the potential locations for establishing the collection centers and recycling centers are evaluated using a machine learning approach. Overall, this research has contributed to the literature by addressing the sustainable-resilient agricultural waste management problem using a data-driven model. The results show that land cost, impact on ecological landscape, pollution prevention and control, and capacity expansion capability are the most significant indicators for the research problem. Also, the achieved outputs confirm the effectiveness and robustness of the developed data-driven decision framework. The developed model is based on a hybrid method combining Data Envelopment Analysis (DEA) and the Gradient Boosting algorithm, achieving 98% accuracy in evaluating potential locations.

Citation:
Mousavikhorram, S.A., Nasiri, M.M., & Ziari, M.. (2025). A data-driven approach to enhance resilience and sustainability in agricultural waste management network facility location. *Journal of Optimization in Industrial Engineering*, 18(1), 337-350. https://doi.org/10.71720/JOIE.2025.1200908

* **Corresponding Author:**
Mohamad Mahdi Nasiri
School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
E-Mail: **mmnasiri@ut.ac.ir**

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (https://creativecommons.org/ licenses/by/4.0/).

1. Introduction

Waste management in the agricultural industry is crucial for promoting sustainability and protecting the environment (Rezaei Zeynali and Tajally, 2024; Waqas et al., 2023). As agricultural production increases to meet the demands of a growing population, the volume of waste generated, such as crop residues, packaging materials, and livestock manure, also rises significantly. Effective waste management practices help minimize pollution, conserve natural resources, and reduce greenhouse gas emissions. By recycling and repurposing agricultural waste through methods like composting and anaerobic digestion, farmers can not only enhance soil health and fertility but also generate renewable energy and valuable by-products (Phiri et al., 2024; Tavakoli et al., 2023). Furthermore, proper waste management contributes to economic efficiency by lowering disposal costs and creating new revenue streams, ultimately fostering a more resilient and sustainable agricultural sector.

Sustainability refers to the ability to meet present needs without compromising the ability of future generations to meet their own needs, encompassing environmental, economic, and social dimensions (Vaezi et al., 2024). In the context of agricultural waste management, sustainability is vital as it encourages practices that minimize waste generation, enhance resource efficiency, and promote ecological balance. Effective management of agricultural waste not only reduces pollution and greenhouse gas emissions but also conserves natural resources by recycling organic materials back into the ecosystem (Koul et al., 2022; Tajally et al., 2025). By adopting sustainable waste management practices, farmers can improve soil health through composting, reduce reliance on chemical fertilizers, and create renewable energy through biogas production. Ultimately, prioritizing sustainability in agricultural waste management leads to a healthier environment, increased productivity, and long-term viability of agricultural systems, benefiting both current and future generations. Moreover, one of the important concepts that has attracted the attention of managers after the COVID-19 pandemic is resilience which aims to enhance the ability of a system to deal with disruptions (Javan-Molaei et al., 2024; Namdar et al., 2021). The resilience concept in agricultural waste management refers to the ability of agricultural systems to withstand, adapt to, and recover from disturbances, such as environmental changes, market fluctuations, or extreme weather events. Additionally, resilience fosters innovation and adaptability, enabling farmers to respond to challenges and seize opportunities for resource recovery and circular economy initiatives (Sawicka, 2020; Shekhawat et al., 2019).

According to the crucial role of the explained points, the current article attempts to evaluate the major indicators to select the best locations for establishing the collection centers and recycling centers for the waste management network based on the sustainability and resilience features for the agricultural sector. To this end, first of all, the major indicators are identified according the experts and literature. Then, the importance (weights) of the determined indicators is measured using the stochastic BWM. In the next step, using a machine learning method, the potential points to establish the collection and recycling centers are ranked based on the considered indicators. The main objectives of this work are as follows: (i) incorporating the sustainability and resilience dimensions into the agricultural waste management problem and (ii) developing an effective machine learning-based decision framework for the evaluation of the potential locations to establish the collection and recycling centers. Moreover, the current article contributes to the existing literature by developing a data-driven model to investigate the resilient and sustainable agricultural waste management problem under uncertainty.

In the rest of this work, Section 2 presents the literature review. Section 3 presents the case study and methodology. Section 4 presents the numerical results. Finally, Section 5 provides the conclusions.

2. Literature Review

Various studies have focused on the issue of selecting optimal locations for specific facilities, often employing intuitive approaches and targeting specific domains such as energy and public stations. For example, Ozdemir and Sahin, (2018) developed a framework for selecting the optimal location for a solar power plant using multi-criteria decision-making methods. Their goal was to evaluate three potential locations based on various qualitative and quantitative criteria. In this study, daily and monthly solar radiation were identified as the most important evaluation criteria and weighted using the AHP method. The potential locations were then assessed based on these criteria. Shayani Mehr et al., (2022) proposed a comprehensive framework for selecting solar panel technologies. They evaluated potential sites using a combined BWM-MULTIMOOSRAL approach. In this study, 20 evaluation criteria were identified and categorized into electrical, technical, economic, climatic, and mechanical sustainability. Similarly, Ghodusinejad et al., (2022) focused on selecting the optimal location and size for solar EV charging stations. These stations were designed to charge electric vehicles, and the study employed a combined GIS-AHP method to evaluate the locations. An optimization model was then proposed to solve the multi-objective problem. The results of this study indicated that the proposed model could identify suitable locations for EV charging stations that meet the demand for electric vehicle charging within a specified access distance. Alkan and Kahraman, (2022) introduced an advanced method for selecting hospital locations during pandemics, which was developed based on the TOPSIS method. They utilized Circular Intuitionistic Fuzzy Sets (CIFS) in combination with the TOPSIS approach to solve this problem. Key criteria considered in this study included the population volume in target areas, ease of access to transportation routes, and availability of suitable vacant land. Using the CIF-TOPSIS technique, their proposed method successfully identified optimal locations while considering various demand scenarios.

Additionally, Kaya et al., (2022) addressed the issue of selecting locations for shared electric vehicle stations from a sustainability perspective. In this study, a multi-criteria approach based on GIS data was proposed. In the first step, decision-making indicators were weighted using the Fuzzy Analytic Hierarchy Process (FAHP), and the required data were then extracted from GIS maps. In the final stage, the potential options were evaluated using the ELECTRE method. This approach was fully implemented in Istanbul, and the results demonstrated its high efficiency in identifying optimal locations for electric vehicle charging stations. In another study, Zhang and Wei, (2023) proposed a hybrid method for selecting locations for electric vehicle charging stations. In this study, four evaluation criteria were considered, including the distance to high-traffic centers, the distance to main roads, the distance to the city center, and the distance to the outskirts. These criteria were analyzed using Spherical Fuzzy Sets (SFSs), and then weighted and prioritized using a hybrid decision-making approach (Combined Compromise Solution-CoCoSo). Sensitivity analysis and comparisons with existing methods revealed that the SF-CPT-CoCoSo model offers high accuracy and efficiency and can effectively be used for identifying optimal locations. Razeghi et al., (2023) presented a multi-criteria approach for selecting suitable locations for solar energy in Iran. The main evaluation criteria in this study included installation costs, distance to urban centers, and environmental impacts. These criteria were weighted using the AHP method, and locations were evaluated using the VIKOR method. In this study, a wide range of data, including climatic conditions, distance to sensitive points, and annual traffic volumes, were examined. By utilizing this data and analyzing their values over different time intervals, machine learning algorithms can be employed to evaluate locations and develop a model capable of analyzing various sites at any given time. The data-driven hybrid model discussed in this article can serve as a foundation for diverse case studies.

This topic has attracted attention over the years. Krishankumar and Ecer, (2024) developed a multi-criteria framework for selecting locations for electric vehicle charging stations. In this study, criteria such as serviceability, environmental impacts, land cost, and traffic density were identified as the most important factors in selecting optimal locations. They introduced a multi-stage hierarchical method that provides high accuracy and efficiency. Similarly, Liang et al., (2024) evaluated and selected an internal terminal location using the fuzzy multi-stakeholder best-worst method. In this study, market volume potential was identified as the most critical evaluation criterion. Their method utilized the opinions of various groups to identify optimal points. However, the methods employed in these two studies rely on expert intuitive and comparative approaches, making it impossible to develop a model that can independently evaluate a location at any moment. Li et al., (2024) proposed a three-stage framework for selecting the optimal locations for large-scale solar farms. They used a combined DBSCAN clustering method and cost-benefit analysis to identify efficient and cost-effective land parcels. Their findings suggest that the developed model is highly suitable for use by organizations and governments. In another study, Alavi et al., (2024) developed a model for selecting the optimal location for hybrid wind systems. They employed the ELECTRE method, and their findings indicate that the Asadabad Nahbandan region is the best location for these systems, with the startup cost of this hybrid system estimated to be nearly half of the actual expense. Additionally, Di Grazia and Tina, (2024) proposed a framework for selecting optimal sites for floating photovoltaic systems. Their approach was based on a combined GIS-AHP-TOPSIS method. In this study, four main criteria were chosen for site evaluation, with cost and regional energy output identified as the most important indicators.

Additionally, Lee and Jeong, (2025) investigated the facility location problem for senior centers in the context of super-aging societies. They developed an optimization model using a genetic algorithm to minimize travel distances for older adults while considering mobility limitations and the existing distribution of senior centers. Using open data from Seoul, including floating population and geographic information, the study demonstrated that adding 15 new centers reduced average travel distances by 24%, thereby enhancing accessibility and quality of life for older adults. Legault and Frejinger, (2025) proposed a model-free simulation-based approach for solving choice-based competitive facility location problems. By reformulating the stochastic optimization problem into a deterministic maximum covering model and applying submodular cuts through a partial Benders decomposition, their method effectively balanced computational efficiency and accuracy, outperforming existing heuristics in large-scale instances. Lyu et al., (2025) integrated Material Flow Analysis (MFA) with Geographic Information Systems (GIS) to estimate construction and demolition (C&D) waste generation and assess suitable disposal facility locations. Using real-estate and building data at a fine spatial scale in Beijing, the study revealed that only 3% of land is suitable for new facilities, yet these locations can significantly improve source separation and sustainable waste management. Li et al., (2025) developed a two-stage robust optimization model for emergency service facility location–allocation under demand uncertainty, incorporating sustainability considerations. The model minimized both preparedness and response costs while accounting for deprivation and environmental impacts, and was validated using the COVID-19 case in Wuhan. The results underscored the value of robust planning in enhancing resilience and sustainability in emergency service networks. Table 1 provides a summary of the literature review conducted on the evaluation and selection of optimal facility locations.

As shown in the literature, in spite of publishing a considerable number of papers in the field of the waste management problem in the agriculture industry, there are still some significant gaps. For instance, the application of the data-driven methods to evaluate the potential locations for establishing the facilities of the waste management network has been rarely addressed by previous works. Moreover, the simultaneous consideration of the sustainability and resilience in assessing the potential locations for opening collection centers and recycling centers for the agricultural waste management problem. Hence, to bridge these gaps, the current article has developed an efficient hybrid data-driven decision framework to select the best location for the agricultural waste management network facilities based on resilience and sustainability dimensions. Indeed, this is the first application of the combination weighted DEA, stochastic BWM, and Gradient Boosting in the waste management problem

Table 1
Literature review summary

Study	Evaluation criteria		Methodology	Case study
	Sustainability	Resilience		
Ozdemir and Sahin, (2018)	×		AHP	Solar Power Plant
Shayani Mehr et al., (2022)			BWM-MULTIMOOSRAL	Solar Panel Technology
Ghodusinejad et al., (2022)	×		GIS-AHP	Solar Vehicle Charging Station
Alkan and Kahraman, (2022)	×		CIF-TOPSIS	Hospital Location Selection
Kaya et al., (2022)	×		GIS – FAHP – ELECTRE	Electric Vehicle Charging Station
Zhang and Wei, (2023)		×	SF-CPT-CoCoSo	Electric Vehicle Charging Station
Razeghi et al., (2023)		×	AHP – VIKOR	Suitable Locations for Solar Energy
Krishankumar and Ecer, (2024)			AHP – GIS	Electric Vehicle Charging Station
Liang et al., (2024)			FBWM	Indoor Terminal Location
Li et al., (2024)	×		DBSCAN	Large-Scale Solar Farm Location
Alavi et al., (2024)	×		ELECTRE	Optimal Location for Wind Hybrid Systems
Di Grazia and Tina, (2024)	×		GIS-AHP-TOPSIS	Optimal Location for Floating Photovoltaic Systems
Lee and Jeong (2025)	×		Genetic Algorithm	Senior centers, Seoul (South Korea)
Legault and Frejinger (2025)		×	Simulation-based model-free approach	Competitive facility location (benchmark instances)
Lyu et al. (2025)	×		Material Flow Analysis + GIS	C&D waste facilities, Beijing (China)
Li et al. (2025)	×	×	Two-stage Robust Optimization	Emergency service facilities, Wuhan (COVID-19)
This study	×	×	**SBWM – WDEA – ML**	**Optimal Location for Pistachio Waste Recycling Points**

3. Case study and Methodology

In this section, the case study is first explained, and based on it, the evaluation criteria for selecting optimal locations for waste management facility siting are described. Additionally, the algorithms and methods used in this study for data analysis are examined, and based on them, the hybrid approach employed in this study is precisely explained with a flowchart.

3.1. Case study and indicators

Damghan County, one of the most important hubs for pistachio production in Iran, has been selected as the case study for this research due to the vastness and high productivity of its pistachio orchards. According to available data, Damghan has 16,900 hectares of pistachio orchards, of which approximately 15,000 hectares are productive and bear fruit. In recent years, pistachio production in this region has significantly increased, with a record yield of 32 tons of pistachios per hectare in one of its orchards in 2024. Furthermore, it is predicted that over 45,000 tons of pistachios will be harvested in Damghan this year, representing a 25% increase compared to the previous year. Given the high volume of pistachio production and the subsequent generation of considerable pistachio green hull waste, the efficient management of this waste has become a necessity. This research aims to identify optimal locations for establishing facilities to manage pistachio green hull waste in Damghan County.

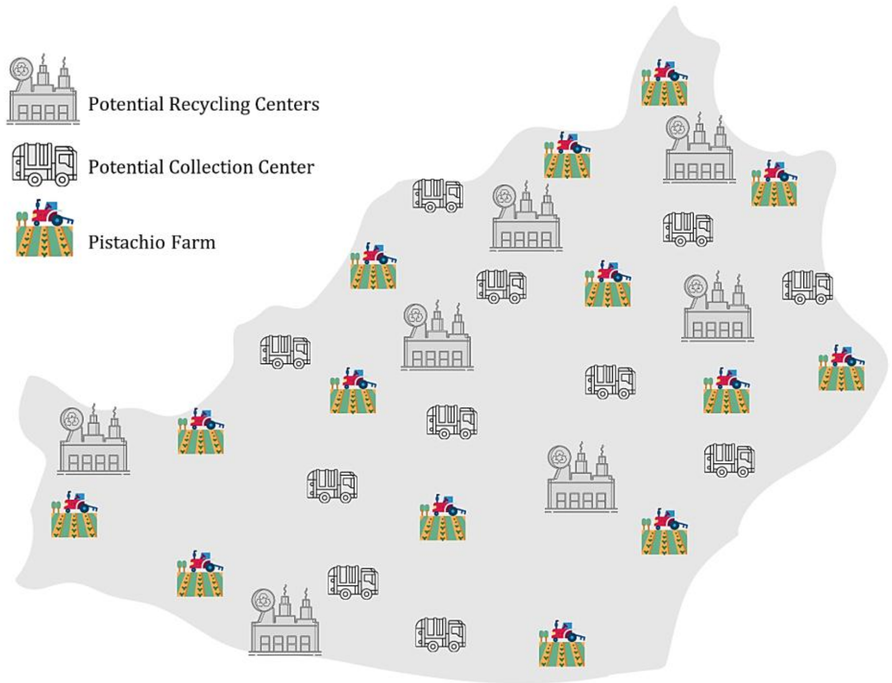

Fig. 1. Case study map

The region's vast agricultural lands, proximity to waste production centers, and existing capacities make it an ideal candidate for this study.

The selection of optimal locations for establishing waste management facilities in such a region is of critical importance. These facilities can not only help reduce environmental pollution caused by waste accumulation but also convert this waste into valuable products, such as compost and animal feed, thereby creating added economic value for farmers and the region. Consequently, this research can provide scientific and practical solutions for optimizing waste management in one of Iran's most prominent agricultural hubs. In summary, the structure of potential points for establishing waste collection and recycling facilities is illustrated in Figure 1.

Finally, based on the literature and experts, the considered indicators have been listed in in Table 2.

Table 2
The considered indicators

Criteria	Reference
Land Cost (C1)	Experts +(Kumar et al., 2020; Sagnak et al., 2021)
Tax (C2)	Experts +(Ayvaz et al., 2015; Kumar et al., 2020)
Work Safety (C3)	Experts +(Ayvaz et al., 2015; Kumar et al., 2020)
Society Benefits (C4)	Experts +(Kheybari et al., 2019; Kumar et al., 2020)
Impact on ecological landscape (C5)	Experts +(Guo and Zhao, 2015; He et al., 2017)
Pollution Prevention and Control (C6)	Experts +(Kumar and Dixit, 2019; Rostami et al., 2023; Sagnak et al., 2021)
Capacity expansion capability (C7)	Experts +(Agrebi and Abed, 2021; Nong, 2022)
Smart infrastructure (C8)	Experts +(Nessari et al., 2024; Xu et al., 2015)
Natural conditions (C9)	Experts +(Chou et al., 2008; He et al., 2017)

3.2. Stochastic BWM

The BWM is one of the most popular approaches in the decision-making research area. In general, this method has several advantages compared to similar methods such as reducing the computational burden and increasing the reliability of the outputs. Recently, a new variant of this method has been proposed by Nayeri et al. (2023), called the stochastic BWM, to deal with the stochastic uncertain environment. The steps of this method are as follows:

(i) Determining the least desirable (worst) and most desirable (best) indicators.

(ii) Forming the comparison vectors based on 1-9 numbers. In this regard, a_{Bjs} represents the Best-to-Other comparison under scenario s. Moreover, a_{jWs} denotes the Other-to-Worst comparison under scenario s.

(iii) Solving Model (1) to achieve the weighs. In this model, P_s shows the probability of occurrences of scenario s. $\xi_s$is the consistency ration under scenario s, ws_{js} represents the weight of indicator j under scenario s, and w_j is the weight of indicator j.

$$
\begin{aligned}
& Min \sum_s P_s.\xi_s && \\
& \left|ws_{Bs} - a_{Bjs}.ws_{js}\right| \le \xi_s && \forall j,s \\
& \left|ws_{js} - a_{jWs}.ws_{Ws}\right| \le \xi_s && \forall j,s \\
& \sum_j ws_{js} = 1 && \forall s \\
& w_j = \sum_s P_s.ws_{js} && \forall j \\
& ws_{js}, w_j \ge 0 && \forall j,s
\end{aligned} \tag{1}
$$

3.3. Weighted DEA

The primary objective of this study is to develop a data-driven model for evaluating potential locations for the establishment of agricultural waste collection and recycling facilities. In this regard, the existing dataset is unlabeled, and for the development of a machine learning model, the dataset must first be labeled. To achieve this, the Weighted Data Envelopment Analysis (WDEA) method has been utilized. In this method, the weights of the data are applied using the Stochastic BWM method as inputs, making the evaluation scenario-based. Moreover, specific weights are assigned to each input and output variable. By executing this stage, the dataset becomes labeled. The steps of the WDEA method are summarized as follows (Kim et al., 2023).

Step 1: Identifying variables

In this step, the input and output variables used for evaluating the efficiency of Decision-Making Units (DMUs) are identified. The variables are divided into two categories: inputs (resource consumption) and outputs (results achieved). Additionally, pre-determined scenario-based weights for each variable are introduced to the model.

Step 2: Constructing the dataset for evaluating locations

Data related to the values of input and output variables for each DMU is collected. This data includes the values corresponding to the location evaluation indicators. The final dataset contains the values of all variables for each DMU.

Step 3: Solving the WDEA model based on BCC

The Weighted DEA model based on BCC is executed using the specified input and output data and weights. The BCC model is chosen due to its variable returns to scale, ensuring that the differences in scale among units are considered in efficiency evaluation. For each DMU, the efficiency is calculated by solving a linear programming problem. The efficiency score in this evaluation ranges between 0 and 1.

Step 4: Assigning efficiency labels to locations

Based on the calculated efficiency scores, each location is labeled according to the following criteria:

- Efficient (Selected): Efficiency score between 0.9 and 1.
- Reserved: Efficiency score between 0.7 and 0.9.
- Inefficient (Rejected): Efficiency score less than 0.7.

Based on these steps, the dataset becomes labeled.

3.4. Gradient boosting algorithm

Gradient Boosting is a powerful and flexible method for solving classification and regression problems. This algorithm builds a strong model by sequentially combining a set of weak models (typically decision trees). At each step, the errors of the previous model are reduced to enhance the final model's accuracy (Liu et al., 2024; Nessari et al., 2024). The steps of this algorithm are summarized as follows (Callens et al., 2020):

Step 1: Data Preparation

In this step, the required data, including 9 features (Land Cost, Tax, Capacity Expansion Capability, Smart Infrastructure, Natural Conditions, Work Safety, Society Benefits, Impact on Ecological Landscape, and Pollution Prevention and Control), were collected and preprocessed. The data were obtained from reliable sources, and missing values were either removed or replaced using appropriate methods to ensure data quality. Additionally, the data were normalized or standardized to ensure uniform feature scaling. Finally, the data were divided into two parts: training and testing sets, with a portion of the training data also allocated for validation.

Step 2: Defining the Gradient Boosting Model

The Gradient Boosting algorithm was used for classifying the points. The model was designed based on shallow decision trees and optimized sequentially to reduce the residual errors from the previous stages. The main parameters of the model included the number of trees (n_estimators), the learning rate (learning_rate), the maximum depth of the trees (max_depth), and the percentage of samples used in each step (subsample). The model was implemented using the XGBoost library, which provides advanced tools for executing Gradient Boosting algorithms. Initial parameter settings were based on experience and adjusted for optimization.

Step 3: Model Training
The Gradient Boosting model was trained using the training data. At each step, the model learned from the residual errors of the previous step, improving incrementally. Validation data were used to monitor the model's performance throughout the training process, preventing overfitting. By the end of this stage, the model was sufficiently trained, and metrics such as validation error and prediction accuracy were evaluated based on the validation data.

Step 4: Evaluation and Classification of Results
After training, the model was evaluated on the test data. The predictions were compared to the actual labels, and metrics such as accuracy and F1-Score were calculated to measure the model's performance. Based on the predicted probabilities, final labels were assigned to the points. The labels included three categories: Selected, Rejected, and Reserved. The advantages of this method include its interpretability and its suitability for implementation on the target dataset.

3.5. Hybrid mythology

This study consists of several stages. In the first stage, the evaluation criteria for potential locations for facility siting are identified and weighted using the stochastic BWM method. Then, the dataset is labeled using the Weighted DEA method, enabling the development of an algorithm for evaluating new locations based on the labeled dataset. After creating the labeled dataset, the Gradient Boosting algorithm is employed to assess each new potential location in accordance with the rules identified from the data. Overall, the main steps of the research are illustrated in Figure 2.

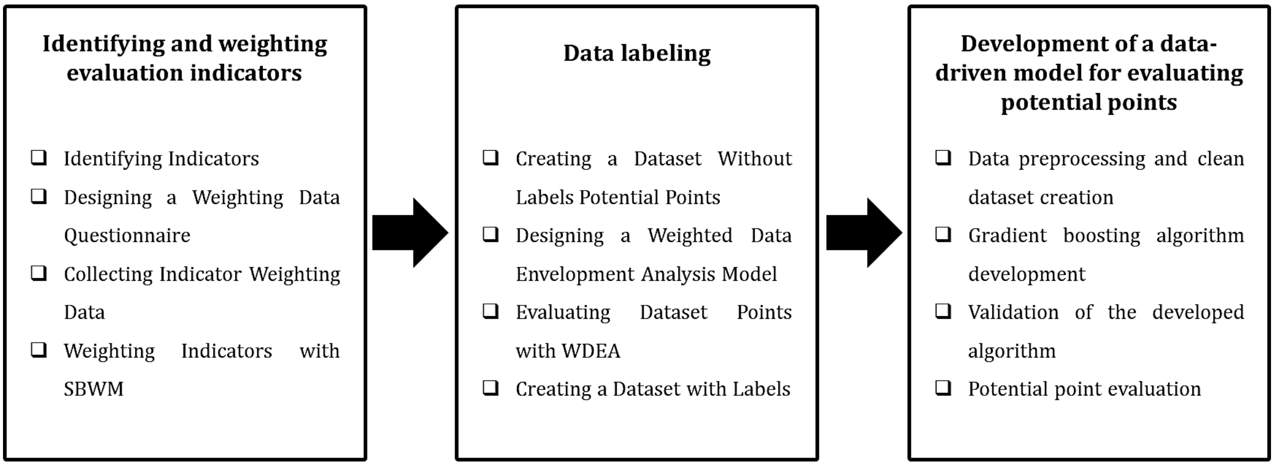

Fig. 2. Hybrid methodology flowchart

4. Numerical Results

4.1. The weights of indicators

In this section, we calculate the indicators' weights using the stochastic BWM approach. For this purpose, we have defined three scenarios as pessimistic, most likely, and pessimistic scenarios so that their occurrence probabilities are respectively equal to 0.25, 0.5, and 0.25. Also, to forming the pairwise comparison matrix, the relevant questionnaires were dispatched among three groups of experts and the average of reached vectors for best and worst indicators has been shown in Tables 3 and 4. Also, Table 5 presents the obtained weights. According to the achieved results, Land cost, Impact on ecological landscape, Pollution Prevention and Control, and Capacity expansion capability have been specified as the best indicators

Table 3
Comparison vector between the best indicator and other ones

Expert		C1			C2			C3			C4			C5		
		s1	s2	s3	s1	s2	s3	s1	s2	s3	s1	s2	s3	s1	s2	s3
1	**C1**	1	1	1	2	2	2	2	3	2	2	3	2	2	1	2
2		1	1	1	2	2	2	2	2	2	2	3	2	2	2	2
3		1	1	1	2	2	2	2	2	2	2	2	2	2	1	2
Average		**1.00**	**1.00**	**1.00**	**2.00**	**2.00**	**2.00**	**2.00**	**2.33**	**2.00**	**2.00**	**2.67**	**2.00**	**2.00**	**1.33**	**2.00**
Expert		**C6**			**C7**			**C8**			**C9**					
		s1	s2	s3	s1	s2	s3	s1	s2	s3	s1	s2	s3			
1	**C1**	2	1	2	2	2	2	3	3	3	2	2	2			
2		2	2	2	2	1	2	3	3	3	2	2	2			
3		2	1	2	2	1	2	3	3	3	2	2	2			
Average		**2.00**	**1.33**	**2.00**	**2.00**	**1.33**	**2.00**	**3.00**	**3.00**	**3.00**	**2.00**	**2.00**	**2.00**			

Table 4
Comparison vector between the worst indicator and other ones

Expert / Criteria	1	2	3	Average
		C8		
C1	3	3	3	**3.00**
	3	3	3	**3.00**
	3	3	3	**3.00**
C2	2	2	2	**2.00**
	2	2	2	**2.00**
	2	2	2	**2.00**
C3	2	3	2	**2.33**
	3	3	3	**3.00**
	2	3	2	**2.33**
C4	2	2	2	**2.00**
	2	2	2	**2.00**
	2	2	2	**2.00**
C5	2	3	3	**2.67**
	2	3	2	**2.33**
	3	3	3	**3.00**
C6	3	3	3	**3.00**
	2	2	2	**2.00**
	2	3	2	**2.33**
C7	2	3	2	**2.33**
	2	2	2	**2.00**
	3	3	3	**3.00**
C8	1	1	1	**1.00**
	1	1	1	**1.00**
	1	1	1	**1.00**
C9	2	2	1	**1.67**
	3	3	3	**3.00**
	2	2	2	**2.00**

Table 5
Indicators' weights

Indicator	Weight
Land Cost (C1)	0.1776451
Tax (C2)	0.1080945
Work Safety (C3)	0.1005574
Society Benefits (C4)	0.09474053
Impact on ecological landscape (C5)	0.1263352
Pollution Prevention and Control (C6)	0.1190829
Capacity expansion capability (C7)	0.1190829
Smart infrastructure (C8)	0.04636708
Natural conditions (C9)	0.1080945

4.2. Assessing the potential locations

In this section, the proposed hybrid data-driven model for evaluating and selecting potential locations is developed. In the first step, the dataset is labeled using the WDEA method. It is important to note that the dataset used for developing the data-driven model is initially unlabeled. The efficiency labels for the dataset are first calculated using the WDEA method, and the labeled dataset is then fed into the Gradient Boosting algorithm.

To develop the WDEA model, input and output variables must be selected from the location evaluation indicators, which are summarized as follows:

Input Variables:

- Land Cost (C1)
- Tax (C2)
- Capacity Expansion Capability (C7)
- Smart Infrastructure (C8)
- Natural Conditions (C9)

Output Variables:

- Work Safety (C3)
- Society Benefits (C4)
- Impact on Ecological Landscape (C5)
- Pollution Prevention and Control (C6)

In the WDEA model, the weight of each input and output variable is provided as an input parameter to the model. This approach ensures that the evaluation of locations is based on the importance of each specific criterion. The dataset used consists of a total of 300 records, with an average of 10 records representing each specific time interval. The WDEA model is executed 30 times, and the efficiency labels for the 300 records are determined. Sample values for one evaluation using the WDEA model are presented in Table 6.

Table 6.
Sample dataset value for running WDEA

DMU	(C1)	(C2)	(C3)	(C4)	(C5)	(C6)	(C7)	(C8)	(C9)
1	309506.9	1	3	1	9	41	5	1	2
2	318206.9	0	4	7	10	91	4	6	4
3	325617.4	1	9	9	6	51	1	4	7
4	331895.1	0	2	1	8	33	1	1	3
5	337718.1	0	8	6	5	71	2	9	5
6	342400.5	0	5	4	1	54	1	6	9
7	382726.6	0	8	10	5	59	4	2	4
8	383164.6	1	4	3	2	50	3	8	4
9	390198.8	0	4	10	3	46	4	6	6
10	391669.1	1	2	8	6	28	3	8	5

By executing the WDEA model, the efficiency of each option has been determined, as shown in Table 7, which presents the efficiency values along with their respective labels. As explained in Section 3.3, options with an efficiency score greater than 0.9 are labeled as Selected, those with a score between 0.7 and 0.9 are labeled as Reserved, and those with a score below 0.7 are labeled as Rejected.

Table 7
The efficiency value and label are sample data.

DMU	Efficiency	Label
1	0.378	Rejected
2	0.505	Rejected
3	0.936	Selected
4	0.172	Rejected
5	0.460	Rejected
6	0.332	Rejected
7	0.780	Reserved
8	0.704	Reserved
9	0.840	Reserved
10	0.976	Selected

Based on the evaluation sample provided using the WDEA method, the entire dataset has been labeled. One of the initial steps in developing data-driven models is examining the correlation coefficient. The correlation coefficient is a numerical measure that indicates the strength and direction of a linear relationship between two variables. This value lies within the range [-1, 1]. A positive correlation coefficient (>0) indicates that as one variable increases, the other variable also increases, while a negative correlation coefficient (<0) indicates an inverse relationship. A value close to 1 or -1 represents a strong correlation, while a value near zero indicates no linear relationship between the variables. For instance, a coefficient of 0.8 shows a strong positive relationship, whereas 0.2 indicates a weak correlation. Based on the correlation matrix and heatmap presented in Figure 3, the relationships between most features in the dataset are weak to moderate, with correlation values rarely approaching 1 or -1. This suggests that the variables act significantly independently, and there is no severe multicollinearity in this dataset. Therefore, no issues are expected in developing the prediction or classification model, and all features can be utilized in modeling. However, the relative relationships between some variables may contribute to a better interpretation of the model's results.

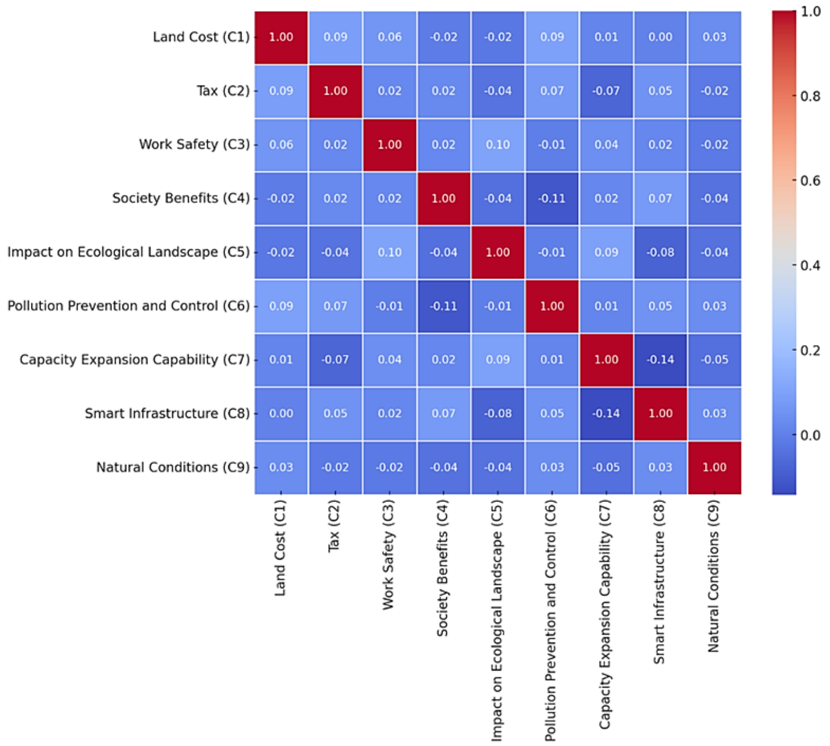

Fig. 3. Heatmap diagram of feature correlation coefficient

The Gradient Boosting algorithm has been used to evaluate and label the points. This algorithm was selected as an appropriate approach for our problem due to its high flexibility and advanced learning capabilities in identifying complex patterns. The primary goal of this algorithm is to predict the labels of points based on input features while minimizing prediction errors. In this study, the Gradient Boosting model was trained using input data such as Land Cost, Capacity Expansion Capability, Smart Infrastructure, and other features. After training, the model successfully labeled the points into three categories: Selected (Efficient), Reserved, and Rejected (Inefficient).

Figure 4 illustrates a part of the output of the Gradient Boosting algorithm. The decision tree shown represents one of the trees in the Gradient Boosting model, which functions as a weak learner. In the Gradient Boosting model, hundreds or even thousands of small trees are combined sequentially to build a strong model. Each decision tree is responsible for learning a portion of the residual errors from the previous trees and ultimately reduces the overall error of the model. The internal nodes of these trees are based on conditions related to the feature values, dividing the data into two branches. Each branch represents a decision made by the model that applies to a specific subset of the data. The leaf nodes provide the final output of the tree, which helps the model correct inaccurate predictions.

Individually, this tree does not perform strongly and only compensates for a small part of the model's errors. However, in combination with other trees, it plays a critical role in improving the model's overall performance. Visualizing decision trees is particularly important for analyzing and interpreting the model. This visualization demonstrates how features contribute to the model's decision-making process and identifies which features have a greater impact. For instance, features that appear more frequently in the internal nodes are likely more important to the model. Such visualizations can aid in better understanding the model's performance, identifying patterns in the data, and making the model's analysis clearer to audiences.

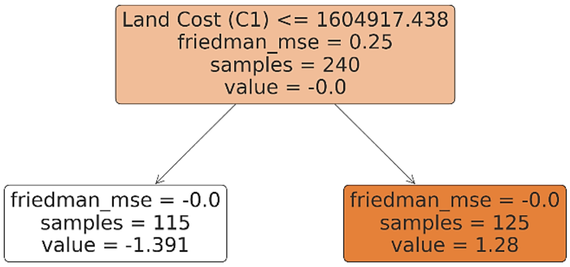

Fig. 4. Part of the gradient boosting algorithm

Based on the developed Gradient Boosting model, the capability to independently evaluate each point exists. This model can estimate the appropriate label for any point using the values of input features (such as land cost, natural conditions, work safety, etc.). In other words, by providing specific values for the defined features, the model automatically and accurately assigns the point's label to one of the three categories: Selected (Efficient), Reserved, or Rejected (Inefficient). This flexible feature of the model allows it to perform quick and accurate evaluations under various conditions and for multiple points, making it a practical tool for operational decision-making. In this context, based on the developed model, the selected points in the present study, which include waste collection and recycling facilities, are illustrated in Figure 5.

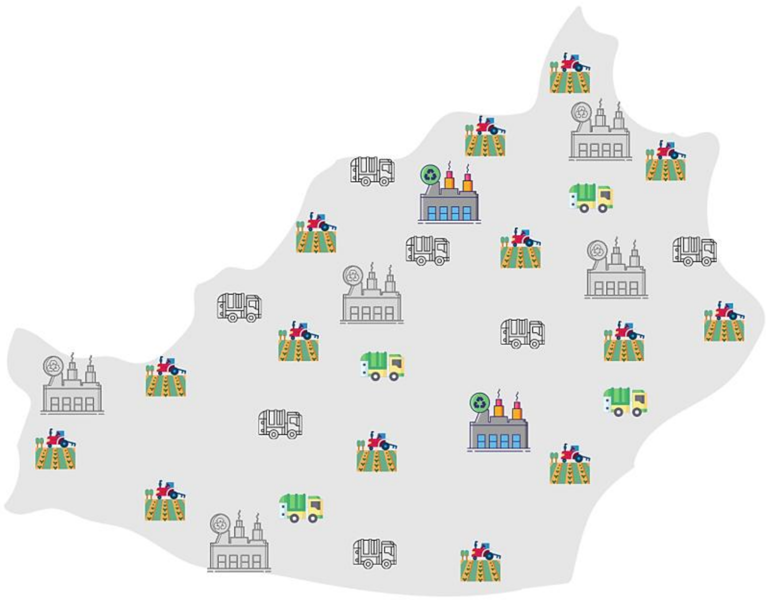

Fig. 5. Points selected using data-driven model

4.3. Effectiveness of the employed stochastic BWM

To examine the performance of the employed stochastic BWM, this section aims to compare its outputs with one of the well-known traditional methods named fuzzy AHP. In this way, Figure 6 depicts the weights achieved by the applied stochastic BWM and the fuzzy AHP.

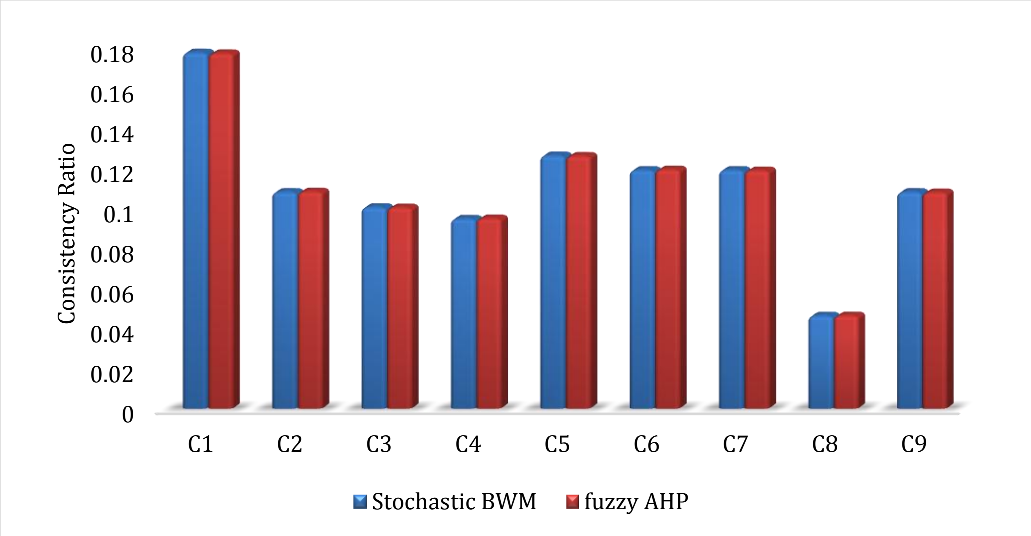

Fig. 6. Comparing the outputs of the stochastic BWM and the fuzzy AHP

The results shows that weights of the majority of indicators are close to each other that confirm the validity of the employed method. However, based on the achieve results, the value of the consistency ratio for the stochastic BWM is equal to 0.03854385 while it is equal to 0.057398 for the fuzzy AHP that confirm the efficiency and reliability of the employed stochastic BWM.

4.4. Effectiveness of the machine learning method

To evaluate the performance and accuracy of the data-driven model, specifically the Gradient Boosting algorithm, two approaches are employed: the Confusion Matrix and evaluation metrics such as Accuracy, Precision, Recall, and F1-Score. The Confusion Matrix is a visual and analytical tool for assessing the performance of classification models. It demonstrates how the model's predictions align with the actual labels. The matrix is divided into four main components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These components provide insights into the number of correct and incorrect predictions. The rows of the matrix represent the true labels, while the columns indicate the predicted labels.

Figure 7 illustrates the Confusion Matrix. This matrix reflects the highly accurate performance of the model. Specifically, out of 31 points with the true label Rejected, the model correctly predicted all cases. Additionally, out of 17 points with the true label Reserved, the model successfully identified 16 cases, misclassifying only 1 point into the Selected category. Furthermore, for the 12 points with the true label Selected, the model correctly predicted all cases without any errors.

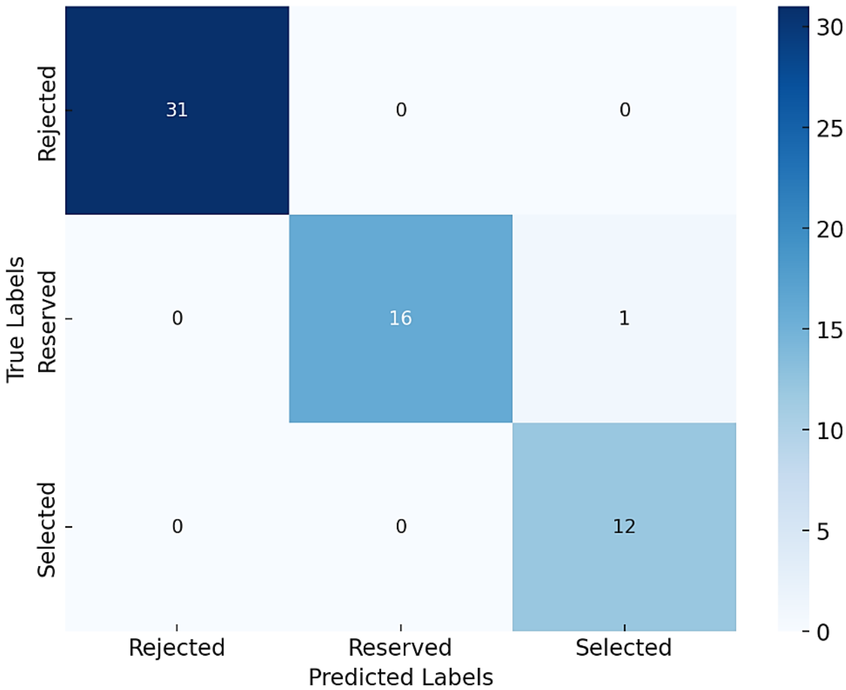

Fig. 7. Algorithm evaluation confusion matrix

These results indicate that the model exhibits a very high level of accuracy in distinguishing and classifying points, with minimal errors. This highlights the model's efficiency and reliability in performing classification tasks. Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms were also executed on the dataset, and their accuracy and performance evaluation results are presented in Table 8. The findings indicate that the performance of the Gradient Boosting algorithm surpasses all other algorithms.

Table 8
Accuracy of the implemented algorithm.

Algorithm	Accuracy	Precision	Recall	F1-Score
Gradient Boosting	0.98	0.94	0.94	0.96
SVM	0.86	0.85	0.86	0.86
ANN	0.91	0.92	0.91	0.92

4.5. Managerial insights

This study provides comprehensive managerial insights for optimizing agricultural waste management, particularly within the context of sustainability and resilience in developing agricultural regions such as Iran. The proposed hybrid data-driven framework—integrating the Stochastic Best-Worst Method (SBWM), Weighted Data Envelopment Analysis (WDEA), and Gradient Boosting algorithm—offers a scientifically grounded and practically applicable structure for selecting optimal locations for agricultural waste collection and recycling facilities. By combining multi-criteria decision-making with advanced machine learning, the model bridges the gap between theoretical evaluation and real-world implementation, enabling policymakers and managers to make informed, evidence-based decisions.

A major managerial implication of this research lies in its capability to handle uncertainty in data and decision environments. Through the SBWM component, the model accounts for the variability in experts' opinions and dynamic environmental factors, ensuring that decisions remain robust under different economic or ecological scenarios. This feature is especially relevant in Iran, where agricultural waste generation fluctuates due to seasonal variations, water scarcity, and market instability. The model's robustness thus provides a reliable foundation for long-term infrastructure planning and resource allocation in rural and semi-arid regions.

The integration of Gradient Boosting significantly enhances the predictive and classification power of the model. By accurately categorizing candidate sites into efficient, reserved, and rejected clusters, it allows decision-makers to prioritize high-potential locations. This precision-driven approach helps in optimizing investments, minimizing waste transportation costs, and reducing environmental degradation. Managers can utilize these insights to develop a tiered implementation plan that aligns with both national sustainability strategies and local operational capacities.

The empirical application of this model to Damghan County—a central hub for pistachio cultivation and processing in Iran—further demonstrates its contextual relevance and practical significance. Damghan's agricultural economy produces a large volume of pistachio green hull waste each year, much of which remains underutilized and contributes to soil and water contamination. The implementation of the proposed model in this setting not only identified suitable locations for new waste treatment facilities but also revealed how optimizing facility networks can reduce average waste transportation distances, lower greenhouse gas emissions, and increase recycling efficiency. This practical linkage illustrates how data-driven optimization directly supports sustainable agricultural transformation in Iran's semi-arid ecosystems.

Moreover, the managerial implications extend beyond local operations. By promoting the conversion of agricultural waste into valuable by-products such as compost and animal feed, the framework encourages a circular economy mindset within Iran's agricultural sector. Policymakers can adopt these insights to design incentive systems for farmers and cooperatives, fostering a shift from reactive waste disposal to proactive resource recovery. The model's data-centric structure also provides a scalable template for other Iranian provinces—such as Kerman, Yazd, and Fars—where similar crops and environmental challenges exist, ensuring consistency in national-level sustainability strategies.

From a broader managerial perspective, the integration of machine learning into agricultural waste management decisions represents a pivotal step toward digital and intelligent governance. The proposed model not only enhances operational efficiency and environmental responsibility but also supports the alignment of agricultural planning with Iran's sustainable development goals. Managers and planners can use this framework to build resilient waste management systems that adapt to climate change, improve resource circularity, and contribute to the long-term socio-economic stability of rural communities.

In summary, by explicitly contextualizing the results within the Iranian agricultural sector and demonstrating their tangible benefits in Damghan County, this study strengthens the practical and policy-oriented relevance of its findings. The approach provides a roadmap for decision-makers to integrate data analytics, sustainability principles, and resilience thinking into the strategic design of agricultural waste management systems.

5. Conclusion

Owing to the undeniable importance of the waste management problem in the agricultural sector, the current article has addressed the agricultural waste management problem considering two important aspects namely sustainability and resilience. To this end, this work has developed a machine learning-based model to evaluate the potential points to establish the collection and recycling centers for handling the agricultural waste. In this way, at the outset, the major indicators were specified and their weights were computed employing the stochastic BWM. Then, the potential points for establishing the facilities were assessed using the machine learning approach. Based on the achieved results, land cost, impact on ecological landscape, pollution prevention and control, and capacity expansion capability have been determined the most important ones. Also, the findings indicate that the machine learning model, with an accuracy of 98%, effectively evaluates the points, thereby selecting the optimal location with high flexibility and agility. Moreover, the obtained outputs demonstrated the reliability and effectiveness of the developed data-driven method. Furthermore, future studies can consider other crucial aspects such as agility and digitalization. Also, future works can investigate the research problem under mixed uncertainty (e.g., fuzzy-scenario) to deal with a high level of uncertainty.

Reference

Agrebi, M., Abed, M., 2021. Decision-making from multiple uncertain experts: case of distribution center location selection. Soft Comput. 25, 4525–4544.

Alavi, S.M.S., Maleki, A., Noroozian, A., Khaleghi, A., 2024. Simultaneous optimal site selection and sizing of a grid-independent hybrid wind/hydrogen system using a hybrid optimization method based on ELECTRE: A case study in Iran. Int. J. Hydrogen Energy 55, 970–983.

Alkan, N., Kahraman, C., 2022. Circular intuitionistic fuzzy TOPSIS method: Pandemic hospital location selection. J. Intell. Fuzzy Syst. 42, 295–316.

Ayvaz, B., Bolat, B., Aydın, N., 2015. Stochastic reverse logistics network design for waste of electrical and electronic equipment. Resour. Conserv. Recycl. 104, 391–404.

Callens, A., Morichon, D., Abadie, S., Delpey, M., Liquet, B., 2020. Using Random forest and Gradient boosting trees to improve wave forecast at a specific location. Appl. Ocean Res. 104, 102339.

Chou, S.-Y., Chang, Y.-H., Shen, C.-Y., 2008. A fuzzy simple additive weighting system under group decision-making for facility location selection with objective/subjective attributes. Eur. J. Oper. Res. 189, 132–145.

Di Grazia, S., Tina, G.M., 2024. Optimal site selection for floating photovoltaic systems based on Geographic Information Systems (GIS) and Multi-Criteria Decision Analysis (MCDA): a case study. Int. J. Sustain. Energy 43, 2167999.

Ghodusinejad, M.H., Noorollahi, Y., Zahedi, R., 2022. Optimal site selection and sizing of solar EV charge stations. J. Energy Storage 56, 105904.

Guo, S., Zhao, H., 2015. Optimal site selection of electric vehicle charging station by using fuzzy TOPSIS based on sustainability perspective. Appl. Energy 158, 390–402.

He, Y., Wang, X., Lin, Y., Zhou, F., Zhou, L., 2017. Sustainable decision making for joint distribution center location choice. Transp. Res. Part D Transp. Environ. 55, 202–216.

Javan-Molaei, B., Tavakkoli-Moghaddam, R., Ghanavati-Nejad, M., Asghari-Asl, A., 2024. A data-driven robust decision-making model for configuring a resilient and responsive relief supply chain under mixed uncertainty. Ann. Oper. Res. 1–38.

Kaya, Ö., Alemdar, K.D., Atalay, A., Çodur, M.Y., Tortum, A., 2022. Electric car sharing stations site selection from the perspective of sustainability: A GIS-based multi-criteria decision making approach. Sustain. Energy Technol. Assessments 52, 102026.

Kheybari, S., Kazemi, M., Rezaei, J., 2019. Bioethanol facility location selection using best-worst method. Appl. Energy 242, 612–623.

Kim, N.H., He, F., Kwon, O.C., 2023. Combining common-weights DEA window with the Malmquist index: A case of China's iron and steel industry. Socioecon. Plann. Sci. 87, 101596.

Koul, B., Yakoob, M., Shah, M.P., 2022. Agricultural waste management strategies for environmental sustainability. Environ. Res. 206, 112285.

Krishankumar, R., Ecer, F., 2024. A multi-criteria framework for electric vehicle charging location selection using double hierarchy preferences and unknown weights. Eng. Appl. Artif. Intell. 133, 108251.

Kumar, A., Dixit, G., 2019. A novel hybrid MCDM framework for WEEE recycling partner evaluation on the basis of green competencies. J. Clean. Prod. 241, 118017.

Kumar, A., Wasan, P., Luthra, S., Dixit, G., 2020. Development of a framework for selecting a sustainable location of waste electrical and electronic equipment recycling plant in emerging economies. J. Clean. Prod. 277, 122645.

Lee, E.H., Jeong, J., 2025. Facility location problem for senior centers in an upcoming super-aging society. Sci. Rep. 15, 6317.

Legault, R., Frejinger, E., 2025. A model-free approach for solving choice-based competitive facility location problems using simulation and submodularity. INFORMS J. Comput. 37, 603–622.

Li, H., Yu, D., Zhang, Y., Yuan, Y., 2025. A two-stage robust optimization model for emergency service facilities location-allocation problem under demand uncertainty and sustainable development. Sci. Rep. 15, 2895.

Li, X.-Y., Dong, X.-Y., Chen, S., Ye, Y.-M., 2024. The promising future of developing large-scale PV solar farms in China: A three-stage framework for site selection. Renew. Energy 220, 119638.

Liang, F., Verhoeven, K., Brunelli, M., Rezaei, J., 2024. Inland terminal location selection using the multi-stakeholder best-worst method. Int. J. Logist. Res. Appl. 27, 363–385.

Liu, Z., Jiang, P., De Bock, K.W., Wang, J., Zhang, L., Niu, X., 2024. Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction. Technol. Forecast. Soc. Change 198, 122945.

Lyu, Y., Xiong, N., Liang, Y., Lin, S., Wang, J., 2025. Estimation of construction and demolition waste and assessment of disposal facility location: An integrated Material Flow Analysis and Geographic Information System approach. Resour. Conserv. Recycl. 218, 108226.

Namdar, J., Torabi, S.A., Sahebjamnia, N., Nilkanth Pradhan, N., 2021. Business continuity-inspired resilient supply chain network design. Int. J. Prod. Res. 59, 1331–1367.

Nayeri, S., Sazvar, Z., Heydari, J., 2023. Towards a responsive supply chain based on the industry 5.0 dimensions: A novel decision-making method. Expert Syst. Appl. 213, 119267.

Nessari, S., Ghanavati-Nejad, M., Jolai, F., Bozorgi-Amiri, A., Rajabizadeh, S., 2024. A data-driven decision-making approach for evaluating the projects according to resilience, circular economy and industry 4.0 dimension. Eng. Appl. Artif. Intell. 134, 108608.

Nong, T.N.-M., 2022. A hybrid model for distribution center location selection. Asian J. Shipp. Logist. 38, 40–49.

Ozdemir, S., Sahin, G., 2018. Multi-criteria decision-making in the location selection for a solar PV power plant using AHP. Measurement 129, 218–226.

Phiri, R., Rangappa, S.M., Siengchin, S., 2024. Agro-waste for renewable and sustainable green production: A review. J. Clean. Prod. 434, 139989.

Razeghi, M., Hajinezhad, A., Naseri, A., Noorollahi, Y., Moosavian, S.F., 2023. Multi-criteria decision-making for selecting a solar farm location to supply energy to reverse osmosis devices and produce freshwater using GIS in Iran. Sol. Energy 253, 501–514.

Rezaei Zeynali, F., Tajally, A., 2024. Designing a sustainable waste chain network for banana farm under uncertainty: a case study. J. Optim. Ind. Eng. 37, 175.

Rostami, O., Tavakoli, M., Tajally, A., GhanavatiNejad, M., 2023. A goal programming-based fuzzy best–worst method for the viable supplier selection problem: a case study. Soft Comput. 27, 2827–2852.

Sagnak, M., Berberoglu, Y., Memis, İ., Yazgan, O., 2021. Sustainable collection center location selection in emerging economy for electronic waste with fuzzy Best-Worst and fuzzy TOPSIS. Waste Manag. 127, 37–47.

Sawicka, B., 2020. Resilient agricultural practices, in: Zero Hunger. Springer, pp. 700–712.

Shayani Mehr, P., Hafezalkotob, Ashkan, Fardi, K., Seiti, H., Movahedi Sobhani, F., Hafezalkotob, Arian, 2022. A comprehensive framework for solar panel technology selection: A BWM-MULTIMOOSRAL approach. Energy Sci. Eng. 10, 4595–4625.

Shekhawat, P., Shrivastava, S., Shrivastava, N., 2019. Agricultural waste utilization in sustainable and resilient construction, in: Advances in Waste Management: Select Proceedings of Recycle 2016. Springer, pp. 385–393.

Tajally, A., Vamarzani, M.Z., Ghanavati-Nejad, M., Zeynali, F.R., Abbasian, M., Bahengam, A., 2025. A hybrid machine learning-based decision-making model for viable supplier selection problem considering circular economy dimensions. Environ. Dev. Sustain. 1–33.

Tavakoli, M., Tajally, A., Ghanavati-Nejad, M., Jolai, F., 2023. A Markovian-based fuzzy decision-making approach for the customer-based sustainable-resilient supplier selection problem. Soft Comput. 1–32.

Vaezi, A., Rabbani, E., Yazdian, S.A., 2024. Blockchain-integrated sustainable supplier selection and order allocation: A hybrid BWM-MULTIMOORA and bi-objective programming approach. J. Clean. Prod. 444, 141216. https://doi.org/10.1016/J.JCLEPRO.2024.141216

Waqas, M., Hashim, S., Humphries, U.W., Ahmad, S., Noor, R., Shoaib, M., Naseem, A., Hlaing, P.T., Lin, H.A., 2023. Composting processes for agricultural waste management: a comprehensive review. Processes 11, 731.

Xu, Y., Liu, C.-C., Schneider, K.P., Ton, D.T., 2015. Toward a resilient distribution system, in: 2015 IEEE Power & Energy Society General Meeting. IEEE, pp. 1–5.

Zhang, H., Wei, G., 2023. Location selection of electric vehicles charging stations by using the spherical fuzzy CPT–CoCoSo and D-CRITIC method. Comput. Appl. Math. 42, 60.