The Journal of Transport and Land Use jtlu.org
Vol. 19 No. 1 [2026] pp. 381-412
https://doi.org/10.5198/jtlu.2026.2836

JTLU

Learning-based multi-criteria decision-making model for sawmill location problems

Mahid Ahmed, School of Computing Science & Computer Engineering, University of Southern Mississippi, mahid.ahmed@usm.edu
Ali K. Dogru, School of Management, University of Southern Mississippi, ali.dogru@usm.edu
Chaoyang Zhang, School of Computing Science & Computer Engineering, University of Southern Mississippi, chaoyang.zhang@usm.edu
Chao Meng, School of Marketing, University of Southern Mississippi, chao.meng@usm.edu

Abstract: Strategically locating a sawmill is vital for enhancing the efficiency, profitability, and sustainability of timber supply chains. Our study proposes a Learning-Based Multi-Criteria Decision-Making (LB-MCDM) framework that integrates machine learning (ML) with GIS-based spatial location analysis via MCDM. The proposed framework provides a data-driven, unbiased, and replicable approach to assessing site suitability. We demonstrate the utility of the proposed model through a case study in Mississippi (MS). We apply five ML algorithms (Random Forest Classifier, Support Vector Classifier, XGBoost Classifier, Logistic Regression, and K-Nearest Neighbors Classifier) to identify the most suitable sawmill locations in MS. Among these models, the Random Forest Classifier achieves the highest performance. We use the SHAP (SHapley Additive exPlanations) technique to determine the relative importance of each criterion, revealing the Supply-Demand Ratio, a composite feature that reflects local market competition dynamics, as the most influential factor, followed by Road, Rail Line and Urban Area Distance. The validation of suitability maps generated by our LB-MCDM model suggests that 10%-11% of the MS landscape is highly suitable for sawmill location.

Keywords: Multi-criteria decision-making, Machine learning, GIS-based spatial location analysis, Facility location

1 Introduction

Sawmill location decisions have profound implications not only for the sawmill itself, but also for the broader timber supply chain and the local communities they serve. Sawmills play a vital role in rural economies through direct job creation and by supporting family-owned small and medium-sized enterprises (SMEs), such as logging, trucking, and timber management companies, in regions where forestry-related industries are major employers (Lamsal et al., 2025). Sawmills process timber into lumber, boards, beams, and other wood products that support essential industries such as construction, housing, and furniture, which affect everyday life (Dogru et al., 2024).

Locating a sawmill requires substantial long-term capital investment, making relocation difficult once the facility is operational. Between 1999 and 2011, the number of active sawmills across southern U.S. states declined by nearly 30% to over 70% of their initial sawmill base (Brandeis & Hodges, 2018), due to a combination of factors including evolving market competition, changing spatial distribution and procurement patterns, and location decisions that failed to adapt to these shifting industry dynamics.

Copyright 2026 Mahid Ahmed, Ali K. Dogru, Chaoyang Zhang & Chao Meng
ISSN: 1938-7849 | Licensed under the Creative Commons Attribution – Noncommercial License 4.0
Received November 29, 2025; Revised September 20, 2026; Accepted September 28, 2026; Available online November 2, 2026

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

These trends highlight the strategic importance of more nuanced, data-driven sawmill location decisions that account for spatial factors and evolving market dynamics, while underscoring the need for systematic decision-support methodologies. Motivated by these challenges, this study focuses on the sawmill location problem. However, the proposed solution methodology is broadly applicable to a wide range of industrial facility location challenges across sectors such as logistics (Tirkolaee et al., 2023; Zhang et al., 2021), renewable energy (Ahmad et al., 2022), hospitality and tourism (Chen et al., 2022), manufacturing (Liu et al., 2021), and healthcare facility location (Ahmadi-Javid et al., 2017).

The sawmill location selection problem presents unique complexities. For example, sawmills must be located near both forest resources and market destinations (Dogru et al., 2024), have convenient access to roads and railways for efficient transportation (Palander et al., 2021; Vaughan et al., 2022), be positioned close to a skilled timber labor force, such as log truck drivers, logging crews, and crane operators (Koirala et al., 2017), and be situated in areas where high precipitation does not impede timber sourcing (Geisler et al., 2016). In addition, terrain slope plays an important role in industrial site selection because steep topography can increase construction costs, reduce transportation accessibility, and negatively affect logging and hauling operations (Susiati et al., 2022). Proximity to urban areas is also an important consideration, as it improves access to utilities, communication infrastructure, healthcare services, banking facilities, and labor markets necessary for sustained sawmill operations (Adhikari et al., 2023). A well-chosen sawmill site can reduce logistical inefficiencies, improve the use of available resources, lessen environmental impact, and contribute to the sustainable development of surrounding communities (Khatri et al., 2025). In regions where the economy depends heavily on timber production, thoughtful site selection can support local economies while maintaining ecological balance (Jimoh & Ogungbemi, 2024). Given the wide range of and often competing factors, sawmill site selection can be considered as a multi-criteria decision-making (MCDM) problem that requires careful joint analysis of geospatial, topographical, socio-economic, transportation, weather, operational, and market factors.

The facility location literature encompasses a variety of methodological approaches, including exact and heuristic optimization techniques (Boloori Arabani & Farahani, 2012), MCDM models (Witkowski et al., 2018), GIS-based spatial analysis (Church & Murray, 2008), and other frameworks such as the Analytic Hierarchy Process (AHP) (Colak et al., 2020) and Fuzzy Logic (Adhikari et al., 2023). While GIS and MCDM methods enable the integration of diverse criteria, they often rely on subjective factor weighting based on expert opinions or researchers' judgment, which can introduce bias into site selection decisions. Conversely, exact and heuristic optimization models tend to prioritize proximity measures (i.e., transportation distance or cost), while overlooking other influential factors such as labor and resource availability, market competition, and weather. As a result, these models may produce biased or limited outputs that are partially applicable to real-world settings. This highlights an important research gap: how can we objectively determine the relative weights of multiple relevant criteria while integrating large volumes of spatial and non-spatial data from diverse sources to efficiently assess and rank the suitability of candidate locations?

To address this research gap in MCDM, we propose a Learning-Based Multi-Criteria Decision-Making (LB-MCDM) framework that integrates Machine Learning (ML) models with GIS-based spatial analysis and MCDM to objectively incorporate a wide range of factors and predict the suitability of candidate sawmill locations by adaptively tuning the weights of these contributing factors. Unlike traditional MCDM approaches guided by expert input, our method is primarily driven by data and computation from the outset. While expert input is considered during the factor selection, the proposed model

does not depend on subjective weighting; instead, relative factor importance is derived directly from the ML process. Utilizing real-world raster and tabular datasets, we applied the proposed LB-MCDM framework to a case study of strategic location planning for sawmills in Mississippi, one of the top timber-producing states in the U.S. (Dogru et al., 2024). We trained five classification models, namely Random Forest Classifier (RF), Support Vector Classifier (SVC), XGBoost Classifier (XGB), Logistic Regression (LR), and K-Nearest Neighbors Classifier (KNN), using a comprehensive dataset of more than 11,000 random candidate locations, each containing values for ten relevant features, including Road Distance, Rail Line Distance, Urban Area Distance, Unemployment Rate, Terrain Slope, Market Revenue, Supply-Demand Ratio, National Land Cover and Precipitation and their estimated suitability scores (target) extracted from ArcGIS Pro. We validate the results of the proposed LB-MCDM framework using direct comparisons with past location decisions.

This research effort provides the following methodological and practical contributions.

- From a methodological point of view, we integrate ML with GIS-based spatial analysis and MCDM to minimize the reliance on subjective expert judgment and automate the industrial site selection assessment process. We further improve the transparency and interpretability of the proposed LB-MCDM framework via SHAP analysis, which reveals the relative influence of each factor in determining site suitability. We also introduce a novel county-level composite feature named the Supply-Demand Ratio (SDR), designed to capture the trade-offs between local supply and demand when a new site is established, reflecting how existing market competition dynamics will be affected. Our findings show that SDR consistently emerges as a highly influential factor, regardless of the ML technique applied.
- From a practical standpoint, our model dynamically generates a suitability map instead of a static list of candidate locations. The suitability map is continuously updated as new facilities open or existing ones close. Consequently, for any given list of candidate site selections (latitudes and longitudes), the model can automatically provide a rank-ordered list with corresponding suitability scores almost instantly. We applied the proposed LB-MCDM framework to a case study in MS using real datasets and validated the results against the actual spatial distribution of existing sawmills. The results demonstrate the robustness and applicability of the proposed methodology in real-world scenarios.
- Beyond serving as a practical decision support tool, the proposed framework is intended to complement, not replace, established facility location optimization models by serving as an upstream screening stage. Given that facility location problems are NP-hard, meaning that no exact algorithm is known to solve them efficiently, and the computational complexity grows exponentially with the number of candidate sites (Kariv & Hakimi, 1979; Krarup & Pruzan, 1983), processing large candidate sets to identify a smaller, high-potential subset via the proposed LB-MCDM framework can significantly reduce problem size, making the proximity-based facility location models more tractable and practically relevant. Within this sequential workflow, LB-MCDM performs suitability-based screening over a broad spatial domain, while a location-allocation model subsequently identifies the optimal site from the resulting shortlist. The proposed method can also be utilized in other application domains, where its recommendations can be

validated against historical siting decisions made using conventional MCDM or raster-based spatial analysis.

The remainder of this paper is structured as follows. Section 2 reviews recent and relevant literature. Section 3 introduces the LB-MCDM framework and provides an overview of key steps. Section 4 demonstrates the practical applicability of the proposed methodology in a case study aimed at evaluating suitable sawmill locations in Mississippi. Section 5 discusses the managerial insights derived from the case study. Finally, Section 6 summarizes the key findings, discusses the limitations of the study, and suggests potential directions for future research.

2 Related literature

The literature relevant to our work lies at the intersection of four research streams: MCDM for site selection, GIS-based spatial analysis, ML-assisted site selection, and plant location problems. Since the literature is quite rich, we will focus on the most recent and relevant papers. Although an exhaustive review is beyond our scope, we refer interested readers to surveys on the details of facility location problems (Celik Turkoglu & Erol Genevois, 2020), multi-criteria location models (Farahani et al., 2010), and GIS-based approaches (Kuhaneswaran et al., 2025).

2.1 Multi-criteria decision making for facility location problem

MCDM has long been used to evaluate facility locations, particularly when multiple conflicting criteria must be considered. These techniques have proven effective across a range of applications, including urban infrastructure planning (Banaei-Kashani et al., 2014), logistic center location (Witkowski et al., 2018), and locating Photovoltaic (PV) systems in the renewable energy sector (de Luis-Ruiz et al., 2024). For instance, Banaei-Kashani et al. (2014) introduced the Multi-Criteria Optimal Location (MCOL) problem and developed the Maximal Reverse Skyline Query (MaxRSKY) method to identify optimal site selections in multi-dimensional space. Their approach expanded traditional proximity-based location decisions by integrating spatial diversity and preference modeling. Witkowski et al. (2018) conducted a comparative MCDM analysis for logistics center placement in Poland, incorporating economic zones, market access, and transport connectivity. In the renewable energy sector, de Luis-Ruiz et al. (2024) applied a weighted MCDM model using ten spatial factors to rank land suitability for PV installations, showing how GIS integration enhances spatial decision-making. Despite their practical relevance, these MCDM methods often involve subjective factor weight assignments, which may introduce bias and constrain the generalizability of site assessment decisions.

Fuzzy logic extends traditional MCDM approaches by accounting for uncertainty in criteria evaluation. For example, Adhikari et al. (2023) used fuzzy logic and location-allocation modeling to determine optimal sites for hardwood cross-laminated timber (CLT) plants in Tennessee based on proximity to sawmills, transportation networks, and timber supply. AHP is another widely adopted MCDM extension. Examples of the use of AHP for solar PV site design and landfill selection can be found in Colak et al. (2020), Shriki et al. (2023), and Islam et al. (2018). Similar to traditional MCDM approaches, these structured and intuitive extensions also rely on expert-defined weights and continue to face challenges related to generalizability. Our study addresses this gap by introducing an objective, data-driven, generalizable, and replicable MCDM methodology that integrates ML and GIS-based spatial location analysis to improve industrial facility location decisions.

2.2 GIS-based spatial location analysis

Using common factor analysis and geostatistical regression, Aguilar (2007) analyzed the spatial distribution of the softwood lumber industry in the U.S. South, identifying timber availability, supplier proximity, and vertical integration as key drivers, with spatial dependencies and clustering influenced by production efficiencies and local resources. While not explicitly focused on site selection, these findings underscore the importance of spatial patterns and resource-based advantages in sawmill placement. Diaz-Balteiro and Romero (2008) illustrated the integration of GIS into the MCDM process to support forest management strategies. Richardson (2016) employed GIS analysis to evaluate how sawmill locations affect the forest products industry in northern Colorado. Adhikari et al. (2023) demonstrated the use of GIS in identifying optimal sites for hardwood CLT plants in Tennessee, highlighting three locations with substantial annual production capacity. Beyond forestry, GIS has also been utilized in site selection for wind farms (Bilgili et al., 2024; Razeghi et al., 2023), landfills (Islam et al., 2018), and power plants (Colak et al., 2020). None of these studies, however, incorporate ML to objectively determine factor weights, which play a critical role in site assessment decisions.

2.3 Machine learning assisted site selection

Recent advances in ML offer data-driven alternatives to traditional MCDM techniques, improving the scalability and accuracy of site selection models. For example, Verma and Sood (2024) utilized ML models to optimize distributed energy generation placement by analyzing load profiles and minimizing losses in power grids. In the renewable sector, Bilgili et al. (2024) used a suite of ML models such as XGB and LightGBM to evaluate wind farm sites in Türkiye based on twelve environmental, economic, and social features. Their model achieved high accuracy (XGB: 96.07%) and is aligned well with existing wind turbine locations, underscoring ML's potential for spatial planning. Similarly, Liu et al. (2024) applied ML to the placement of oil wells in the petroleum industry, reporting an increase in cumulative oil production compared to traditional models. In the hospitality sector, Yang et al. (2015) combined ML with GIS to assess optimal hotel locations in Beijing, offering dynamic and web-based visualizations to support decision-making. Al-Ruzouq et al. (2022) proposed an AHP-based framework integrated with ML algorithms to optimize the location of waste-to-energy (WTE) facilities in UAE, considering social, environmental, and economic factors. This approach achieved up to 94.6% accuracy and identified 16.6% of the area as highly suitable. Yunusoglu et al. (2024) presented an ML-based two-stage approach for biomass-to-bioenergy supply chain network design to locate undesirable facilities in Türkiye. These studies, however, primarily focus on other sectors like energy or urban development, with relatively limited application in timber-related industries. Our work addresses this gap by focusing specifically on the sawmill site selection problem, an economically significant yet underexplored area.

2.4 Plant location models

The plant location problem (a.k.a., the facility location problem) involves selecting the optimal locations from a set of candidate sites, which is NP-hard (Tadić et al., 2023). Most plant location models in the timber industry rely on heuristics for large problem instances or exact optimization techniques for computationally tractable problem instances (Weintraub & Epstein, 2002). For instance, Vanzetti et al. (2017) presented a MILP model to compare individual plant locations versus industrial clusters within the forest supply chain, aiming to optimize production and logistics decisions in the Argentine forestry sector. Similarly, Duarte et al. (2014) proposed a MILP model for

locating biofuel plants in Colombia, specifically applying the model to a second-generation bioethanol plant using coffee cut stems. Some studies focus on locating satellite wet yards, a storage area where the harvested timber is sorted and temporarily stored before being transported to sawmills. Aguiar et al. (2020) evaluated the allocation of wood storage yards in an Amazonian sustainable forest by comparing exact optimization methods with metaheuristic approaches. Similarly, Chan et al. (2009) employed MILP to determine optimal locations for satellite yards in Canadian forestry. Some studies incorporate MCDM and plant location. Adhikari et al. (2023), for instance, used fuzzy MCDM and location-allocation modeling to determine suitable locations for hardwood CLT manufacturing, considering the proximity to sawmills, timber supply, and transportation networks. Our work advances plant location modeling by integrating GIS-based raster analysis and ML techniques to systematically evaluate and rank a large number of candidate sites, thereby reducing problem size and enhancing computational tractability before applying exact optimization methods.

3 Methodology: LB-MCDM framework

This section outlines the key components of the proposed LB-MCDM framework, which integrates GIS-based geospatial analysis with ML to address limitations in traditional MCDM approaches. Consider a decision maker who seeks to select a subset of promising facility locations denoted by V^s from the set of N candidate locations $V^c = \{v_1, v_2, \dots, v_N\}$ such that $V^s \subset V^c$. The decision maker evaluates these candidate locations using K criteria, each criterion $i \in K$ represented by a raster layer. A raster layer $\boldsymbol{R}$ is a two-dimensional matrix with m rows and n columns, where each cell r_{xy} contains a value of the spatial feature at location (x, y). Given a vector of raster layers $\vec{X} = (\boldsymbol{R}_1, \boldsymbol{R}_2, \dots, \boldsymbol{R}_K)$ and corresponding vector of weights $\vec{w} = (w_1, w_2, \dots, w_K)$, the weighted sum layer $\widehat{\boldsymbol{R}}(\vec{X}, \vec{w})$ is computed as:

$$\widehat{\boldsymbol{R}}(\vec{X}, \vec{w}) = \sum_{i=1}^{K} w_i \boldsymbol{R}_i \tag{1}$$

In practice, these weights are often determined subjectively, based on expert opinion survey results, empirical findings in the literature, or researchers' own assessment. To reduce subjectivity and potential bias, we propose employing ML techniques. The proposed LB-MCDM framework consists of four key phases: 1) data collection and preprocessing, 2) initial suitability mapping, 3) feature weight tuning and suitability map reconstruction, and 4) validation and rank-ordering of candidate locations. The process flow is summarized in Figure 1. In the following subsections, we elaborate on the key phases.

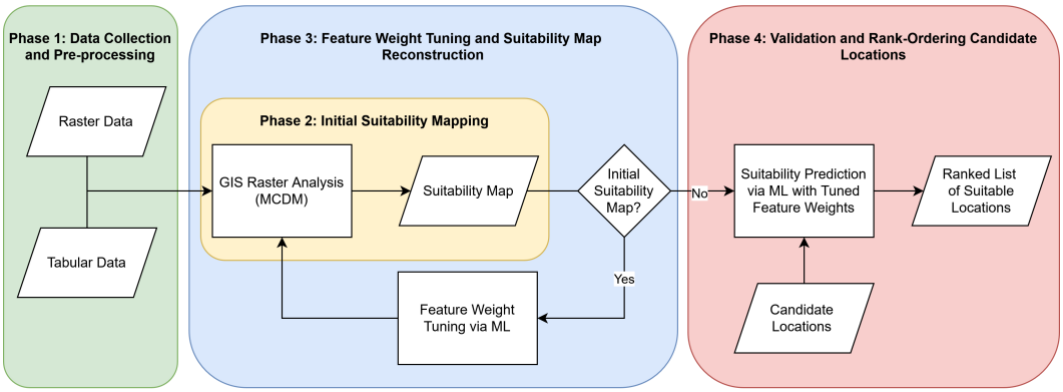


Figure 1. LB-MCDM process flow

3.1 Data collection and preprocessing

In this phase, decision makers gather both raster and tabular data relevant to the facility location problem under investigation. In the sawmill location problem domain, raster data, such as land cover, terrain slope, proximity to roads and railways, and proximity to urban areas, represent spatially continuous variables derived from geospatial layers. In contrast, tabular data, including labor statistics, timber supply and demand, local market revenue, and precipitation, provide structured and attribute-based information. Once the relevant data is collected, tabular data is converted into a raster format to enable spatial processing within a GIS environment. This ensures consistency in data format and facilitates seamless integration during subsequent analysis. Data preprocessing is an essential step because raw data is often inconsistent or incomplete, which can negatively impact the ML model's performance. After collecting the data, it is essential to clean the datasets, handle missing values, perform necessary transformations, and scale the features.

3.2 Initial suitability mapping

With the raster preprocessed, a weighted sum overlay is performed in the GIS system to generate an initial suitability map. At this stage, equal weights are assigned to each of the K features to avoid premature bias ($w_i = \frac{1}{K}, \forall i \in K$). The weighted sum layer $\widehat{\boldsymbol{R}}(\vec{X}, \vec{w})$ provides location-specific suitability scores visually represented through a color gradient. The decision maker then samples a large number of random points from the initial suitability map, each with associated predictor values (features) and a target to be used in training ML classifiers.

3.3 Feature weight tuning and suitability map reconstruction

The ML phase follows a structured process that includes data partitioning (training and testing), feature selection, model development, hyperparameter tuning, and model evaluation. Since multiple ML classifiers can be used in the proposed framework, let $\overrightarrow{y_j} = (y_1, y_2, \dots, y_N) \in \{0,1,2,3\}^N$ denote the multi-class target variable (i.e., 0: not suitable, 1: somewhat suitable, 2: suitable, 3: highly suitable) for the ML classifier $j \in J$. Each classifier j predicts $\overrightarrow{y_j}$ based on the vector of sample candidate location input features $\vec{x}$ derived from $\vec{X}$ and the initial vector of equal weights $\vec{w}$ such that $\overrightarrow{y_j} = f_j(\vec{x}, \vec{w})$ and yields $\overrightarrow{\alpha_j} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_N) \in [0,1]^N$, the resulting vector of the likelihoods of candidate locations in V^c being suitable. Then, the initial feature weight vector $\vec{w}$ is

updated using the SHAP values computed for feature i by ML classifier j such that $w_{ij} = \varphi_{ij}, \forall i \in K$ and $\forall j \in J$, resulting in the updated feature weight vector denoted by $\vec{w}'$. These tuned weights are re-integrated into the GIS system to reconstruct the suitability map $\widehat{\boldsymbol{R}}(\vec{X}, \vec{w})$, reflecting the updated factor of importance. The decision makers then resample from this refined map to generate a new dataset, which is used to retrain the ML classifiers. Then employing standard model evaluation techniques, the decision maker identifies the best-performing ML classifier out of J classifiers denoted by $\vec{y}^* = f^*(\vec{x}, \vec{w}')$ with likelihood scores $\vec{\alpha}^* = (\hat{\alpha}_1^*, \hat{\alpha}_2^*, \dots, \hat{\alpha}_N^*) \in [0,1]^N$.

3.4 Validation and rank-ordering candidate locations

After retraining, the best-performing ML classifier's final predictions are validated using existing facility locations and expert opinions. Once validated, the classifier can assess and rank candidate sites, enabling data-driven prioritization based on predicted suitability. The candidate locations are first sorted in descending order of their estimated suitability scores: $V^c = \{v_{(1)}, v_{(2)}, \dots, v_{(N)}\}$ such that $\hat{\alpha}_{(1)}^* \geq \hat{\alpha}_{(2)}^* \geq \cdots \geq \hat{\alpha}_{(N)}^*$. Then, the decision maker chooses M best suitable for N candidate locations: $V^s = \{v_{(m)} \in V^c | m \leq M\}$.

4 Case study: Sawmill location in Mississippi via LB-MCDM

To demonstrate the applicability and effectiveness of the proposed LB-MCDM framework, we conducted a case study focusing on the state of Mississippi (MS). The state of MS is one of the leading timber-producing states in the U.S., with forests covering 63% of its land equivaling to about 19 million acres of forestland (Mississippi Department of Agriculture and Commerce, 2025). Timber ranks as the state's second-largest commodity, supporting approximately 85,000 jobs and supplying over 140 sawmills (Dogru et al., 2024). These factors make MS a strong candidate for studying the sawmill location assessment problem.

4.1 Data collection and preprocessing

We identified ten key criteria influencing sawmill site suitability, guided by insights from prior literature, consultations with policymakers, industrial relevance, and the available data. Table 1 summarizes the data sources, measurements, and intended purposes for each feature. We use these datasets to construct the suitability layer in ArcGIS software and train ML classifiers. Each criterion (feature) is briefly described below to contextualize its role in LB-MCDM framework:

National Land Cover Data (NLCD): Sourced from the U.S. Census Bureau, the NLCD layer classifies land into usability categories. Highly suitable areas include barren land, open space, and low-intensity development. Suitable areas consist of forest land and grassland, while somewhat suitable areas encompass developed zones situated on shrubland and cultivated land with medium to high development intensity. Areas classified as not suitable include water bodies, wetlands, and national forests. The dataset allows us to distinguish candidate facility locations based on their relative suitability and potential for sustainable development (Adhikari et al., 2023; Susiati et al., 2022).

Transportation Data: The U.S. Geological Survey's National Transportation Dataset enables us to measure proximity to major roads and rail lines. Proximity to roads and rail lines was calculated using the direct Euclidean distance from each candidate location to the nearest road and railway line extracted from the dataset (de Luis-Ruiz et al., 2024; Khavarian-Garmsir & Rezaei, 2015). Access to roads and rail lines reduces transportation

costs and facilitates the efficient movement of raw materials, finished products, and workers.

Slope Data: Open topography data from the U.S. Geological Survey allows the assessment of terrain slope, classifying potential sites from flat to very steep slopes (Susiati et al., 2022). This raster data helps identify land that supports easy construction and transportation routes. Avoiding steep areas reduces development costs and improves transportation efficiency.

Urban Data: Proximity to urban centers enhances operational feasibility by ensuring seamless access to essential infrastructure, including power, water, telecommunications, banking, and healthcare. Prior literature provides precedent for these proximity boundaries: Adhikari et al. (2023) identified sites within 8 km of an urban area as suitable for industrial timber manufacturing, whereas Migose et al. (2018) categorized regions within 15 km as urban, 20–50 km as mid-rural, and beyond 50 km as extremely rural. Synthesizing these benchmarks, we selected <10 km as a balanced threshold for highly suitable sites, while stratifying suitable (10–20 km), somewhat suitable (20–50 km), and not suitable (>50 km).

Table 1. Summary of datasets

Data	Source	Measurement	Purpose
National Land Cover Data	US Census Bureau	Highly suitable: Developed areas located on barren land with open space and low intensity, Suitable: Forest land and grassland, somewhat suitable: Developed areas located on shrubland and cultivated land with medium and high intensity, not suitable: Water space, and wetlands.	To distinguish candidate locations based on their land suitability and potential for sustainable development.
Transportation Data	USGS National Transportation Dataset	Highly suitable: $0\text{m} < \text{Value} \leq 500\text{m}$, Suitable: $500\text{m} < \text{Value} \leq 1000\text{m}$, Somewhat suitable: $1000\text{m} < \text{Value} \leq 2000\text{m}$, Not suitable: $\text{Value} > 2000\text{m}$. Source: (de Luis-Ruiz et al., 2024)	To measure proximity to major roads and rail lines to improve transportation access and reduce logistical costs.
Slope Data	USGS open topography	Highly suitable: $0 \leq \text{Slope} \leq 7\%$, Suitable: $8 \leq \text{Slope} \leq 13\%$, Somewhat suitable: $14 \leq \text{Slope} \leq 20\%$, Not suitable: $21 \leq \text{Slope} \leq 90\%$. Source: (Susiati et al., 2022)	To determine the slope of potential sites to ensure ease of access and support more efficient transportation and construction.
Urban Data	US Census Bureau	Highly suitable: $\text{Value} < 10{,}000$ m, Suitable: $10{,}000\text{ m} \leq \text{Value} < 20{,}000$ m, Somewhat suitable: $20{,}000\text{ m} \leq \text{Value} < 50{,}000$ m, Not suitable: $\text{Value} \geq 50{,}000$ m Source: (Adhikari et al., 2023; Migose et al., 2018)	To ensure better access to utilities and services such as electricity, gas, water, communication, banking, and healthcare facilities.
Labor Data	U.S. Bureau of Labor Statistics	County-level unemployment rate for labor, whose age is greater than 16.	To assess the availability of labor in the area to support sawmill operations.
Timber Demand Data	Forisk Mill Capacity Database	The data provides detailed information on sawmills' timber processing capacities and demands in tons.	To align site selection with existing timber demand and operational capacity constraints.
Timber Supply Data	MS Dept. of Agriculture & Commerce	Total available timber volume in tons at the county level.	To prioritize locations with sufficient timber availability.
Revenue Data	Nexis Uni	The annual average revenue of wood manufacturing companies in USD.	To estimate the local timber market demand.

Data	Source	Measurement	Purpose
Precipitation Data	National Center for Environmental Information	County-level rainfall in inches in 2024.	To prioritize areas with lower precipitation levels for operational efficiency, significant precipitation can hinder timber harvesting and transportation activities.

Labor Data: Labor force availability is vital for sustaining sawmill operations, with sites near urban areas preferred due to better access to skilled workers. We assess the current labor supply in MS using Local Area Unemployment Statistics (LAUS) as defined by the U.S. Bureau of Labor Statistics. The tabular dataset allows for retrieving the county-level unemployment rate for the labor population in 82 MS counties aged 16 and older. A higher unemployment rate may indicate greater labor availability, making the location more favorable for sawmill operations.

Market Revenue Data: We retrieved revenue data from Nexis Uni for Mississippi companies that depend on lumber and wood products, excluding existing sawmills, in tabular format. To evaluate local market potential, we calculated the total revenue of these companies located within a 75-mile straight-line Euclidean radius of each sawmill, rather than using actual transport-network route distance. The 75-mile rule is a widely used rule of thumb in the timber industry, reflecting the typical procurement or sales radius (Anderson & Germain, 2007). It balances transportation costs and supply chain efficiency, making it a practical benchmark for location decisions. Proximity to markets offers significant operational advantages to sawmills, including increased sales opportunities, faster delivery times, lower outbound transportation costs, and a smaller carbon footprint (Dogru et al., 2024).

Precipitation Data: We obtained county-level precipitation data (in inches) for all 82 Mississippi counties from the year 2024, sourced from the National Centers for Environmental Information. Precipitation refers to any form of water, such as rain, sleet, or snow, that falls from the atmosphere to the ground, typically measured in inches. High precipitation negatively impacts upstream timber supply chains by reducing accessibility for logging, damaging forest roads, and causing safety hazards and transportation delays (Dogru et al., 2024). Additionally, wet logs are prone to quality degradation, increased drying costs, and operational inefficiencies.

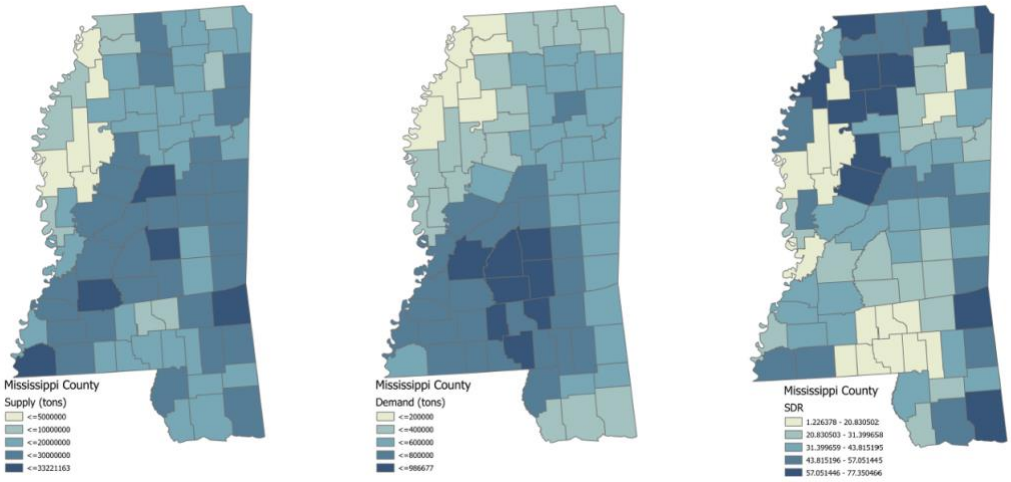


Figure 2. Comparison of timber supply, timber demand, and supply demand ratio in MS

Supply-Demand Ratio (SDR): To capture local market competition dynamics, we propose a county-level composite metric called the Supply-Demand Ratio (SDR), which reflects the trade-offs between timber supply and timber demand following the introduction of a new sawmill in a county. Timber demand data were obtained from the Forisk 2024 Mill Capacity Database, which provides timber processing capacities, timber demands, and geolocations of more than 140 wood-consuming mills in Mississippi. Timber supply data were collected from the Mississippi Department of Agriculture and Commerce and provide county-level available timber volumes in tons across all 82 counties. The SDR combines these two datasets to represent the balance between available timber resources and existing market competition. A higher SDR indicates a more favorable location for a new sawmill, reflecting greater timber availability and lower competition for timber resources in the surrounding area. Table 3 shows that the Timber Supply and Timber Demand datasets were combined to construct the composite SDR feature within the LB-MCDM framework.

Definition 1. Assume there are J counties, indexed by $j \in J$. Let N_j and S_j denote the set of existing sawmills and the available timber supply in tons in county j, respectively. In addition, let D_{ij} be the annual tons of timber demanded by the existing sawmill i from county j. Now, suppose that we open a new sawmill in county j, which will demand $D_{new,j}$ tons of timber annually. Assuming the supply remains constant, the supply-demand ratio for county j, denoted by SDR_j, is calculated using the following expression:

$$SDR_j = \frac{S_j}{\sum_{i \in N_j} D_{ij} + D_{new,j}}, \forall j \in J \tag{2}$$

To calculate D_{ij}, we employ the 75-mile radius rule: each sawmill i sources timber within a 75-mile radius. Since this radius may span multiple counties, we allocate the demand proportionally based on the portion of the radius area that lies within each county so that the total demand by sawmill i can be found $D_i = \sum_{j \in J} D_{ij}, \forall i \in N_j$. Figure 2 compares county-level timber supply, timber demand, and the calculated SDRs. Note that SDR is not a simple, static supply-to-demand ratio for each county; rather, it captures the dynamic reallocation of sawmill-level timber demand using the 75-mile radius rule after hypothetically placing a new sawmill in each county.

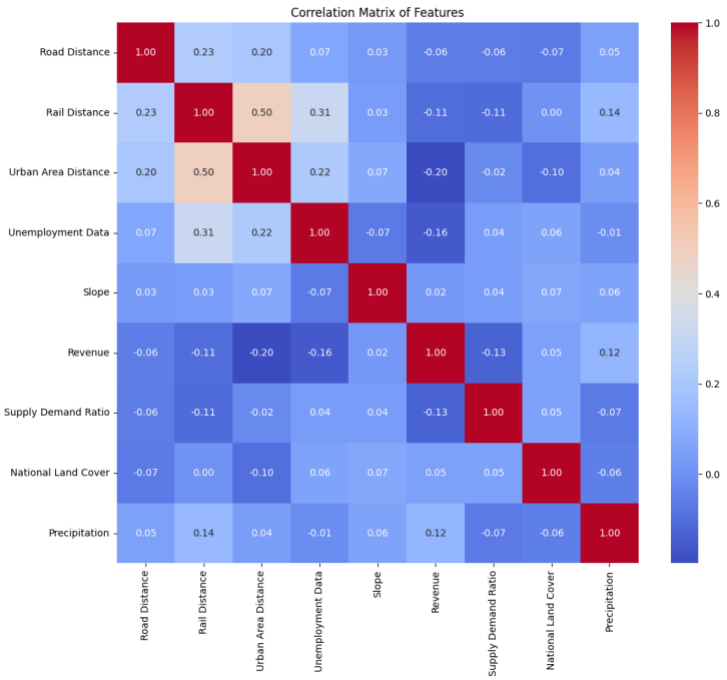


Figure 3. Correlation matrix among the 9 input features

During data preprocessing, we cleaned the datasets by removing records with critical missing values and imputing remaining gaps using the statistical mode to preserve distributional properties. After cleaning, we applied standard scaling to all features, transforming them to have a mean of 0 and a standard deviation of 1 to ensure equal contribution and fair weight computation across variables. Finally, we converted tabular datasets (labor, timber demand, timber supply, revenue, and precipitation datasets) into a raster format to enable spatial processing in the ArcGIS Pro software.

4.2 Initial suitability mapping

Initially, we assigned equal weights to all nine features (i.e., $w_i = \frac{1}{9} = 0.111, \forall i \in K$) and calculated the weighted sum layer $\widehat{\boldsymbol{R}}$, which serves as our initial suitability map. From this map, we initially sampled 10,000 random points (locations), each containing values for nine predictor variables and the corresponding estimated suitability score, which is a non-negative continuous variable. We classified these random locations into four suitability levels: highly suitable, suitable, somewhat suitable, and not suitable, following the classifications adopted in prior studies. To determine thresholds, we calculated the range of suitability scores of these random points and divided the range into four equal intervals, where the highest suitability interval indicates highly suitable, and the lowest interval is not suitable. After observing significant class imbalances among the suitability categories, we drew an additional random sample of 1,467 locations to achieve a more balanced dataset. The final dataset, consisting of 11,467 observations, was reclassified into the same four categories, which mitigated but did not fully eliminate the imbalance issue. To further address this, we decided to employ synthetic minority oversampling techniques during model training. We split the resulting dataset into training (80%) and testing (20%) subsets.

4.3 Feature weight tuning and suitability map reconstruction

At this stage, we have a large dataset to train multiple ML classifiers, consisting of nine features: Road Distance, Rail Line Distance, Urban Area Distance, Unemployment Rate, Terrain Slope, Market Revenue, Supply-Demand Ratio, National Land Cover and Precipitation. Our target variable (Suitability) is a categorical variable with four categories: Highly Suitable, Suitable, Somewhat Suitable, and Not Suitable. In the following subsections, we discuss how we address data imbalances, selected features, evaluated models, and fine-tuned feature weights using the SHAP values. We provide the details regarding the employed ML classifiers, calculation of performance metrics, and SHAP calculations in Sections 7, 8, and 9 of the Appendix, respectively.

4.3.1 Handling imbalances in the dataset

Despite our efforts to incorporate additional random samples, our initial suitability map revealed an unbalanced distribution across the four suitability categories, with 1,306 sites classified as highly suitable, 2,397 as suitable, 6,152 as somewhat suitable, and 1,612 as not suitable. As is widely known, imbalanced datasets can bias ML models toward the majority class, leading to misleading accuracy and poor precision and recall for underrepresented classes. To mitigate this issue, we employed the Synthetic Minority Over-sampling Technique - Edited Nearest Neighbors (SMOTE-ENN), a hybrid of SMOTE and ENN on the training data only to avoid data leakage (all other preprocessing steps were conducted separately on the training and testing datasets to ensure the validity of model evaluation). SMOTE generates synthetic samples for the minority class by interpolating between existing minority class instances and their k-nearest neighbors (Zhu et al., 2017). Following the SMOTE step, ENN is employed to refine synthetic observations located near class boundaries by removing instances whose class labels are inconsistent with the majority of their three nearest neighbors. This helps eliminate noisy data and better preserve the class boundaries. This approach reduces duplication and improves model generalization by enhancing the quality of the synthetic samples while maintaining class separability.

Table 2. Variance inflation factor (VIF) and tolerance

Feature	VIF	TOL
Road Distance	1.0762	0.9292
Rail Line Distance	1.4911	0.6706
Urban Area Distance	1.4112	0.7082
Unemployment Rate	1.1471	0.8718
Terrain Slope	1.0271	0.9736
Market Revenue	1.0976	0.9111
Supply Demand Ratio	1.0481	0.9541
National Land Cover	1.0429	0.9589
Precipitation	1.0507	0.9518

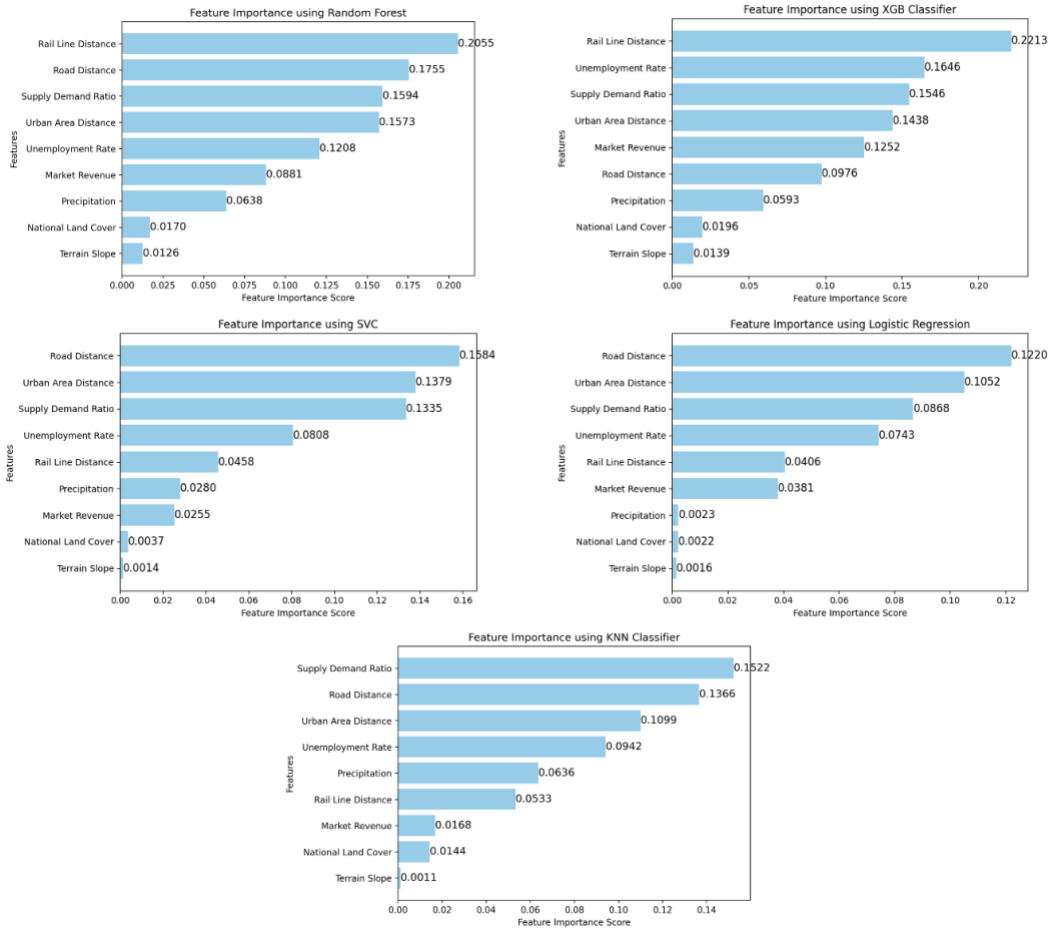


Figure 4. Feature importance scores by five ML classifiers

4.3.2 Feature selection

Feature selection aims to systematically identify and retain the most relevant features that significantly enhance ML model performance by eliminating less important, correlated, or redundant features, thereby improving accuracy, reducing overfitting, and decreasing computational complexity. In this study, we employed a sequential approach combining correlation analysis, multicollinearity testing, feature importance, and SHAP values to identify the most suitable features for our ML model.

We first constructed the correlation matrix of all nine features presented in Figure 3 to identify highly correlated feature pairs ($r > 0.7$ or $r < -0.7$), which could potentially introduce redundancy into our model. As shown in the figure, the urban area distance and rail distance pair have the highest correlation ($r = 0.50$), which is well below the threshold ($r = 0.7$), so we decided to retain all nine features.

Table 3. Evolution of features across the LB-MCDM framework

Phase	Features	Notes	Count
Initial Dataset	Road Distance, Rail Line Distance, Urban Area Distance, Unemployment Rate, Terrain Slope, Market Revenue, Timber Supply, Timber Demand, National Land Cover, Precipitation	Initial set of spatial and non-spatial features	10
Dataset after Feature Engineering	Road Distance, Rail Line Distance, Urban Area Distance, Unemployment Rate, Terrain Slope, Market Revenue, Supply–Demand Ratio, National Land Cover, Precipitation	Supply–Demand Ratio replaces Timber Supply and Timber Demand	9
Final Dataset after Feature Selection	Road Distance, Rail Line Distance, Urban Area Distance, Unemployment Rate, Market Revenue, Supply–Demand Ratio, Precipitation	Terrain Slope and National Land Cover were removed due to their relatively low importance	7

Table 4. Performance comparison. HSC: Highly Suitable Class, SC: Suitable Class, SSC: Somewhat Suitable Class, NSC: Not Suitable Class

Classifier	Accuracy	Recall	Precision	F1 Score	AUC	HSC	SC	SSC	NSC
RF	**0.8648**	**0.8648**	**0.8734**	**0.8663**	**0.9656**	0.9119	0.8017	**0.8530**	**0.9658**
XGB	0.8495	0.8495	0.8588	0.8511	0.9569	0.9080	**0.9596**	0.8424	0.7620
SVC	0.8186	0.8186	0.8286	0.8199	0.9159	**0.9157**	0.7035	0.8172	0.9161
LR	0.5870	0.5870	0.6919	0.5862	0.8597	0.8812	0.5741	0.4460	0.9068
KNN	0.7257	0.7257	0.7600	0.7313	0.8142	0.8008	0.6618	0.6962	0.8727

To further assess multicollinearity and ensure feature independence, we calculated the Variance Inflation Factor (VIF) for each feature. Although many machine learning algorithms can handle correlated inputs, assessing VIF enhances model interpretability and minimizes redundancy that could distort feature importance rankings. VIF quantifies how much the variance of a regression coefficient is inflated due to linear relationships with other variables (Tay, 2017). A VIF close to 1 suggests negligible multicollinearity; values exceeding 10 imply high multicollinearity that could affect model performance. Tolerance indicates how much of the variability in a feature is independent of other features; values closer to 0 suggest stronger dependence. As shown in Table 2, all VIF scores are below 1.50, and all tolerance values exceed 0.67, indicating that there is no multicollinearity problem.

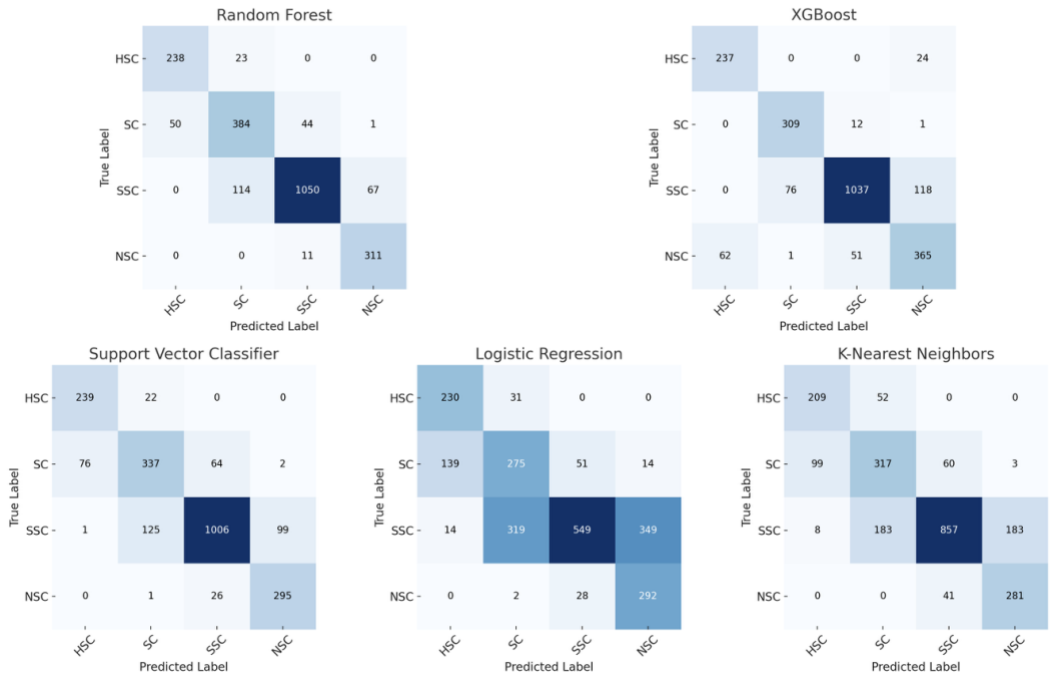


Figure 5. Confusion matrix of five ML classifiers

Next, we analyzed the feature importance scores of five ML models, namely RF, XGB, SVC, LR, and KNN. Figure 4 presents the rankings of feature importance derived from multiple ML models. The results consistently indicate that Terrain Slope and National Land Cover have minimal influence across all models. This finding can be attributed to the relatively flat topography of Mississippi, where slope variations are negligible across the state. Furthermore, Mississippi's extensive forest coverage reduces the variability in land cover types, thereby diminishing the significance of the National Land Cover feature in determining suitable locations for sawmill development. However, it is important to note that in more mountainous geographies with limited timber resources, these two features may play a more critical role. Due to their relatively low importance, we decided to remove Terrain Slope and National Land Cover and continue our analysis with the remaining seven features (SDR, Road Distance, Rail Line Distance, Urban Area Distance, Unemployment Rate, Market Revenue, and Precipitation). Table 3 summarizes the evolution of the datasets and features throughout the LB-MCDM framework.

4.3.3 Model evaluation and classification performance

Using selected seven features, we trained and evaluated five ML classifiers: RF, XGB, SVC, LR, and KNN. We optimized the parameters of each ML model through hyperparameter tuning. Performance metrics are presented in Table 4, with confusion matrices shown in Figure 5 and AUC-ROC curves in Figure 6.

Among all models, the RF classifier demonstrated the highest overall performance, achieving an accuracy of 0.8648, a precision of 0.8734, and the highest AUC score of 0. 9656. RF also achieved the highest per-class accuracy in the Somewhat Suitable Class (0.8530) and the Not Suitable Class (0.9658), while demonstrating strong performance across the remaining classes.

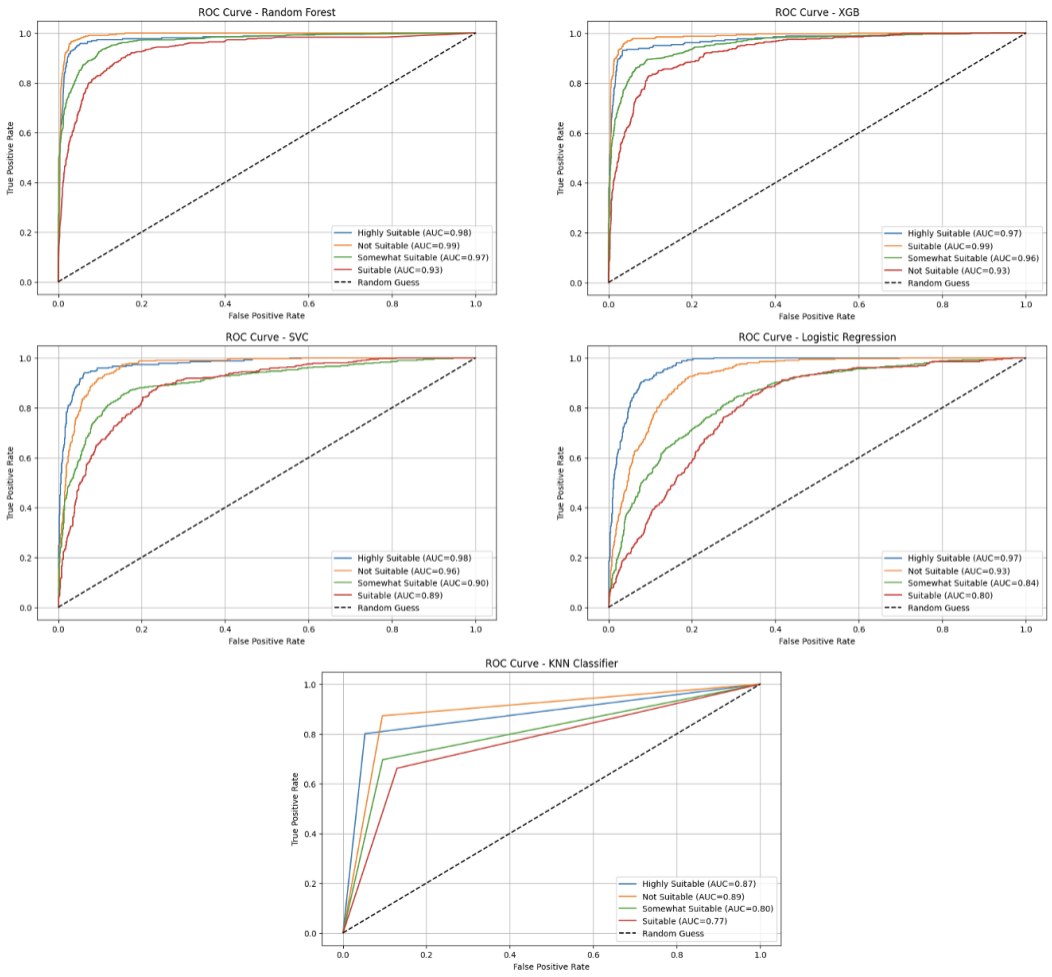


Figure 6. AUCROC curve of five ML classifiers

XGB, SVC, and KNN also delivered a decent performance. XGB slightly outperformed SVC overall, achieving an accuracy of 0.8495 and an AUC of 0.9569. It also produced the best classification accuracy for the Suitable Class (0.9596) and performed competitively in the Highly Suitable Class (0.9080). SVC achieved an accuracy of 0.8186, a precision of 0.8286, and an AUC of 0.9159. It showed strength in classifying extreme categories, with the highest accuracy for the Highly Suitable Class (0.9157). KNN, while lower in overall metrics, reached a respectable accuracy of 0.7257 and achieved the best performance in the Not Suitable Class (0.8727), indicating potential utility in conservative classification contexts. In contrast, LR exhibited the weakest overall performance, with an accuracy of 0.5870, F1 score of 0.5862, and the lowest AUC score of 0.8597. Its poor per-class accuracy, particularly 0.4460 for the Somewhat Suitable Class, limits its effectiveness for this multi-class classification task. The AUC-ROC curves in Figure 6 further support these findings, confirming the superior class separation capabilities of RF, XGB, and SVC, compared to the relatively weaker performance of LR and KNN.

Table 5 demonstrates that removing two features from the model did not lead to any significant degradation in performance. In fact, the Random Forest (RF) classifier with seven features achieved comparable, and in several cases slightly superior, results to the nine-feature version. The seven-feature RF model maintained higher accuracy, recall,

precision, and F1 score, while the AUC values (0.9629 for nine features and 0.9656 for seven features) remained nearly identical, indicating consistent discriminative capability.

Table 5. Comparison of model performance with 9 and 7 features using the RF model

Model	Accuracy	Recall	Precision	F1 Score	AUC	HSC	SC	SSC	NSC
RF (9 Features)	0.8404	0.8404	0.8529	0.8430	0.9629	0.8621	0.7912	0.8278	0.9441
RF (7 Features)	0.8648	0.8648	0.8734	0.8663	0.9656	0.9119	0.8017	0.8530	0.9658

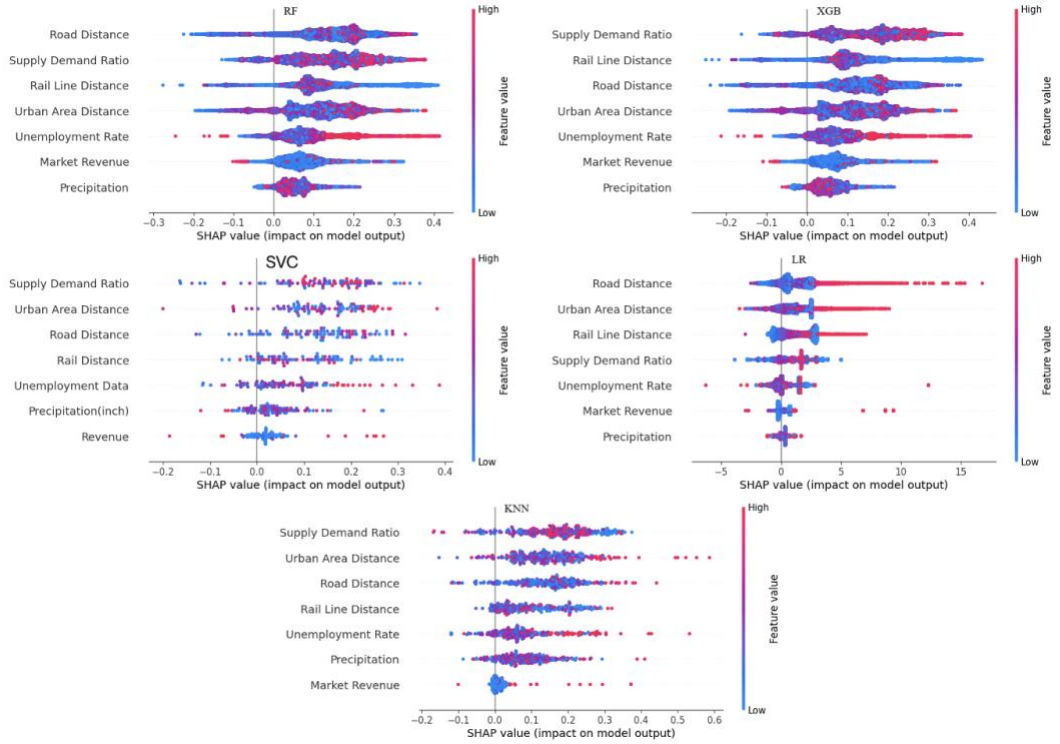


Figure 7. Impact of individual features on the output of five machine learning classifiers (RF, XGB, SVC, LR, and KNN) using SHAP analysis. Each plot visualizes the contribution of features to model predictions. The color scale represents the feature value (red for high and blue for low), while the SHAP values on the x-axis indicate the magnitude and direction of each feature's influence on the model output.

Similarly, per-class accuracy showed only minor variations without any meaningful decline in classification quality. These findings suggest that the reduced seven-feature RF model offers a more efficient and streamlined configuration without compromising accuracy or predictive robustness.

4.3.4 Ensuring explainability via SHAP analysis

Once the models are trained and their performances are evaluated, the next step is to ensure explainability by quantifying the influence of each feature on the model's predictions. While gain-based feature importance metrics can aid feature selection, they often lack consistency and completeness in explaining predictions (Lundberg et al., 2019). SHAP analysis, grounded in cooperative Game Theory and first introduced by

Shapley (Shapley, 1953), addresses this gap by fairly distributing the prediction output among the contributing features. SHAP provides both local and global interpretability through sample-wise feature attributions, typically visualized using Beeswarm and bar plots.



Figure 8. Mean SHAP values of the top features across five machine learning classifiers (RF, XGB, SVC, LR, and KNN). The horizontal bars represent the average absolute SHAP values, quantifying each feature's overall contribution to model predictions.

Figure 7 shows that the SDR has a strong positive impact on model predictions, making it the most influential feature for the XGB, SVC, and KNN models, and the second most important feature for the RF model. This finding underscores the significance of market competition and resource availability in sawmill location decisions. As a composite construct combining timber supply and demand, SDR effectively reduces feature dimensionality while enhancing prediction accuracy, illustrating how meaningful feature engineering can improve model performance. As can be seen in the figure, the Road Distance appears to be the most significant feature in the Logistic Regression model. Features such as Precipitation and Market Revenue consistently exhibit the lowest importance across all models and contribute only marginally and positively to the predictions.

Figure 8 shows the mean SHAP values across the five ML models used. It confirms that SDR, Road Distance, Urban Area Distance, and Rail Line Distance are consistent among the most important predictors across all models. On the other hand, Market

Revenue and Precipitation exhibit relatively lower average SHAP values, highlighting their relatively limited contribution to the sawmill's suitability assessment.

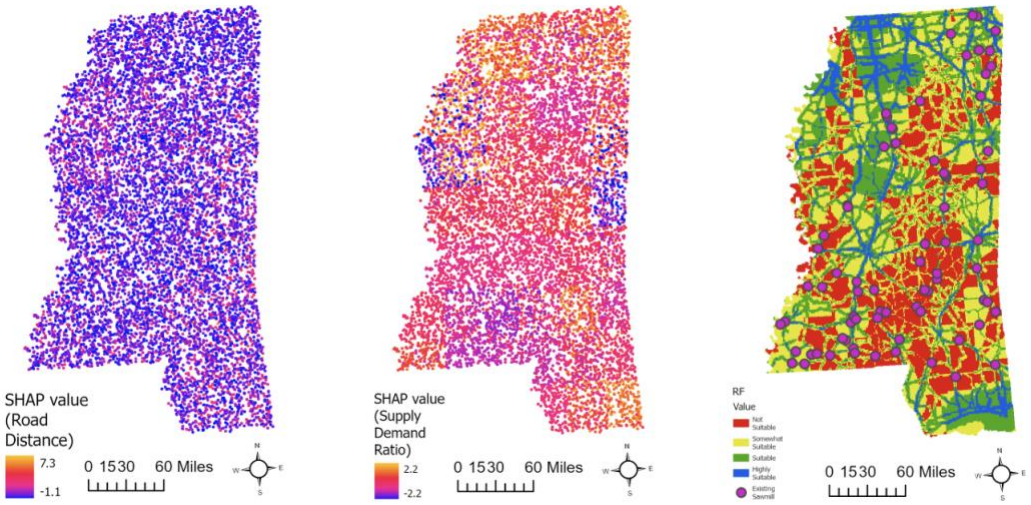


Figure 9. Spatial distribution of local SHAP values for SDR and road distance

Figure 9 further extends the SHAP analysis in Figure 8 by visualizing the spatial distribution of local SHAP values for the two most influential RF model features, namely Road Distance and Supply-Demand Ratio (SDR). For reference, the figure also includes the suitability map generated by the RF model. Locations closer to roads (i.e., low Road Distance values represented by purple dots) generally contribute positively to suitability predictions, whereas more remote locations (yellow and red dots) tend to reduce suitability. In contrast, higher SDR values (yellow and red dots) usually positively influence suitability, suggesting that locations with stronger timber availability relative to demand are preferred by the model to assign high suitability. However, these local SHAP patterns should be interpreted cautiously, as SHAP values are observation-specific and reflect the RF model's learned nonlinear relationships and feature interactions rather than causal effects. Consequently, the influence of a feature may vary across locations and suitability classes depending on surrounding geographic and operational conditions. Therefore, the contribution of individual features should be interpreted within the broader multivariate context of the model.

4.3.5 Suitability map reconstruction with tuned weights

Once the SHAP values are retrieved, they can be used to fine-tune the weights of the raster layers on the ArcGIS Pro software. Then the raster analysis is run once more to generate the corresponding final suitability maps. For each ML model, a separate suitability map was developed by incorporating SHAP-based feature contribution scores. Unlike the initial suitability map, where all features were assigned equal weights, the final suitability maps incorporated feature weights computed from SHAP values, reflecting their relative importance in the model's decision-making process (see Table 6, baseline weight $= \frac{1}{7} = 0.143$). The resulting suitability scores ranged from 0 to 1 and were classified into four categories using the Jenks Natural Breaks method in ArcGIS Pro software. Following classification, the spatial distribution of each suitability class was

calculated for all models to compare performance and spatial agreement across the predictions.

Table 6. Comparison of adjusted weights

Adjusted Weights	Road Dist.	Rail Line Dist.	Urban Dist.	Unemployment Rate	Market Revenue	SDR	Precipitation
Baseline	0.143	0.143	0.143	0.143	0.143	0.143	0.143
RF	0.195	0.182	0.156	0.117	0.091	0.195	0.065
XGB	0.182	0.182	0.156	0.117	0.091	0.208	0.065
SVC	0.189	0.149	0.203	0.122	0.054	0.216	0.068
LR	0.213	0.190	0.204	0.101	0.077	0.176	0.040
KNN	0.192	0.128	0.192	0.128	0.026	0.218	0.115

Figure 10 compares the adjusted weights of the seven relevant features across the five ML models, highlighting SDR and the three proximity features (Road Distance, Rail Line Distance, and Urban Area Distance) as the most significant factors influencing sawmill location suitability.

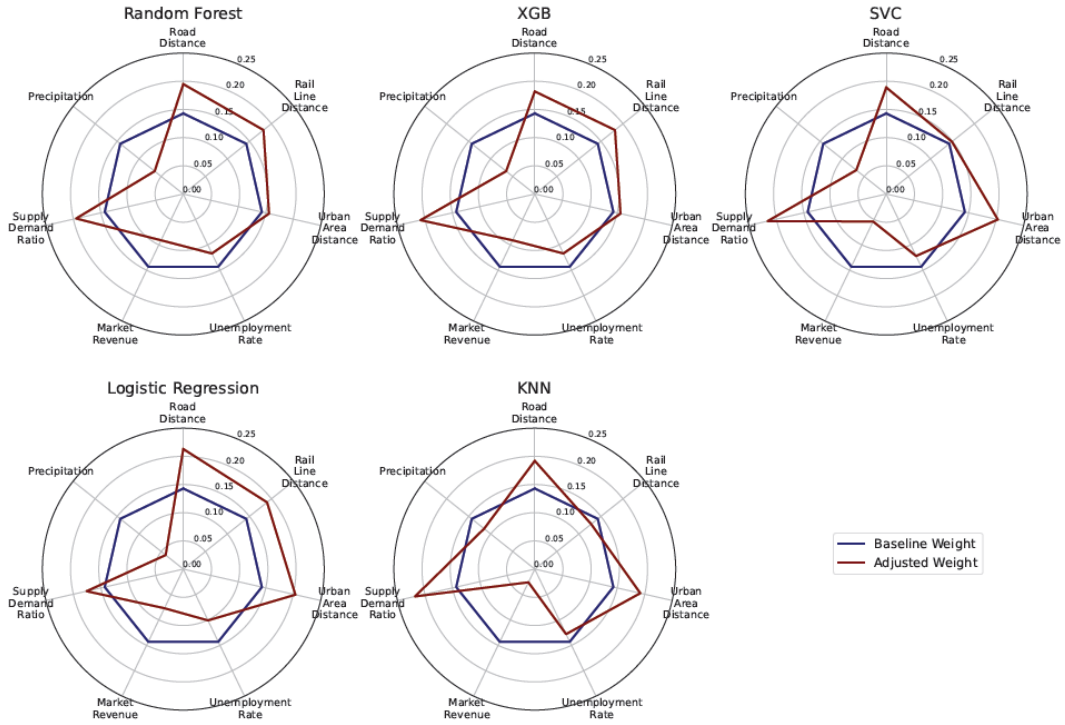


Figure 10. Comparison of adjusted weights by different ML classifiers

4.4 Validation and rank-ordering candidate locations

The site suitability maps generated using SHAP scores from each ML model are shown in Figure 11, while the corresponding areal distribution across suitability classes is

depicted in Figure 12. The results indicate that the majority of locations in Mississippi are classified as Somewhat Suitable, with percentages ranging from 33.6% to 39.3% depending on the model. In contrast, the Highly Suitable category consistently represents the smallest proportion of land area across all models. While the distribution patterns are broadly similar among the models, the KNN model predicts the largest combined percentage of area as either Highly Suitable or Suitable.

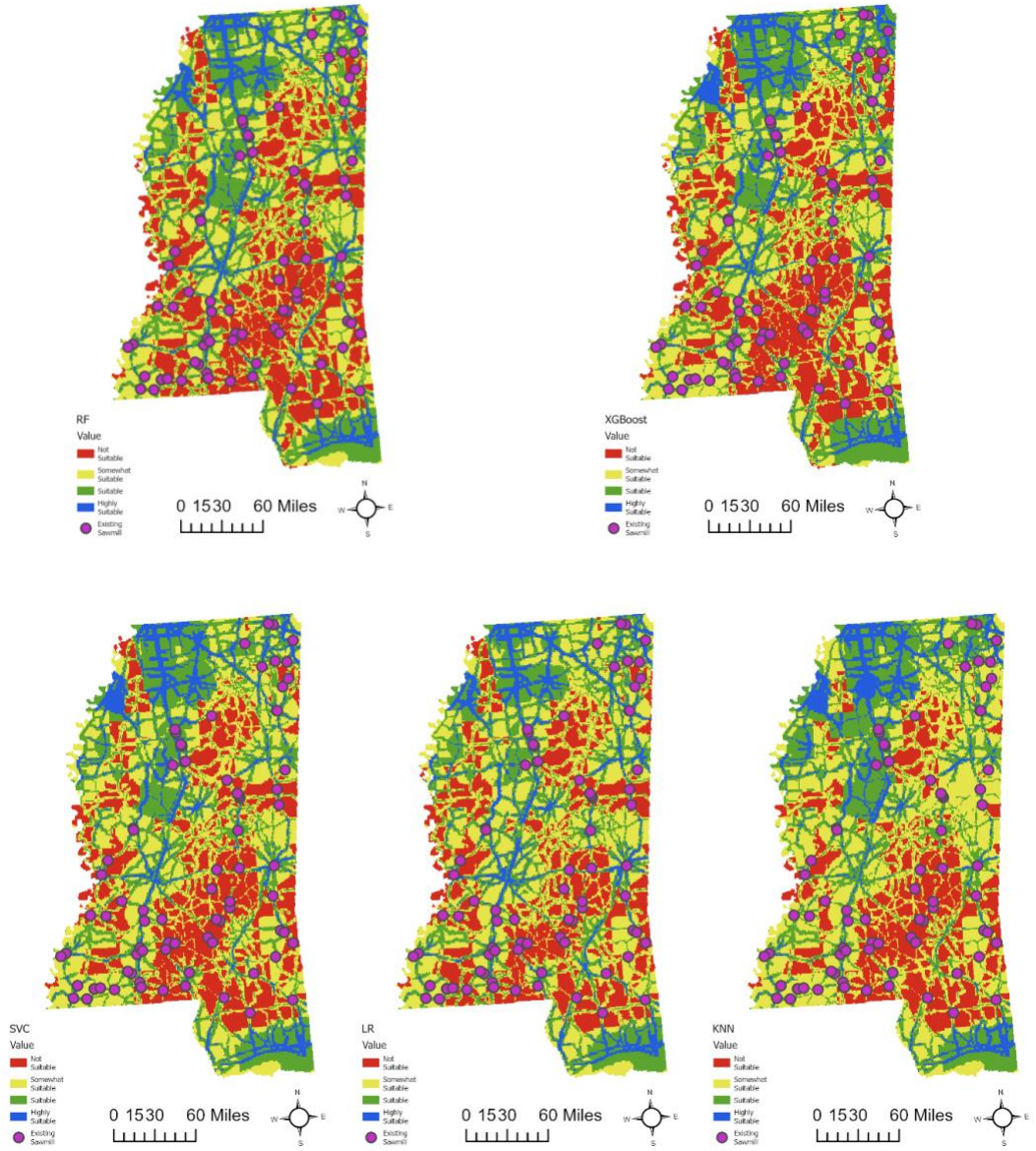


Figure 11. Suitability map of Mississippi using five ML models

To validate the predictive accuracy of the generated suitability maps, we analyzed the spatial distribution of existing sawmills in Mississippi and calculated the percentage of sawmills located within each suitability class. This distribution is shown in Figure 13. As evident from the figure, the majority of existing sawmills are located in areas classified as either Highly Suitable or Suitable. Specifically, 75.9% of sawmills fall within these two classes according to the RF model, while the corresponding values are 69.9% for KNN,

78.4% for LR, 73.5% for XGB, and 77.1% for SVC. Logistic Regression (LR) also reports the lowest percentage of sawmills in the Not Suitable category (3.6%), followed by 7.2% for both RF and KNN, 10.8% for XGB, and 9.6% for SVC. It is important to note that the dynamic nature of sawmill openings and closures continually reshapes market conditions; consequently, locations once considered highly suitable may become less suitable as new sawmills emerge nearby, intensifying market competition.

In addition to classification validation, candidate locations were ranked based on their predicted likelihood scores. Each ML model produced continuous scores for all sites, which were sorted in descending order to create a prioritized list. This allowed for efficient filtering of high-ranking sites (i.e., top 10) for potential development.

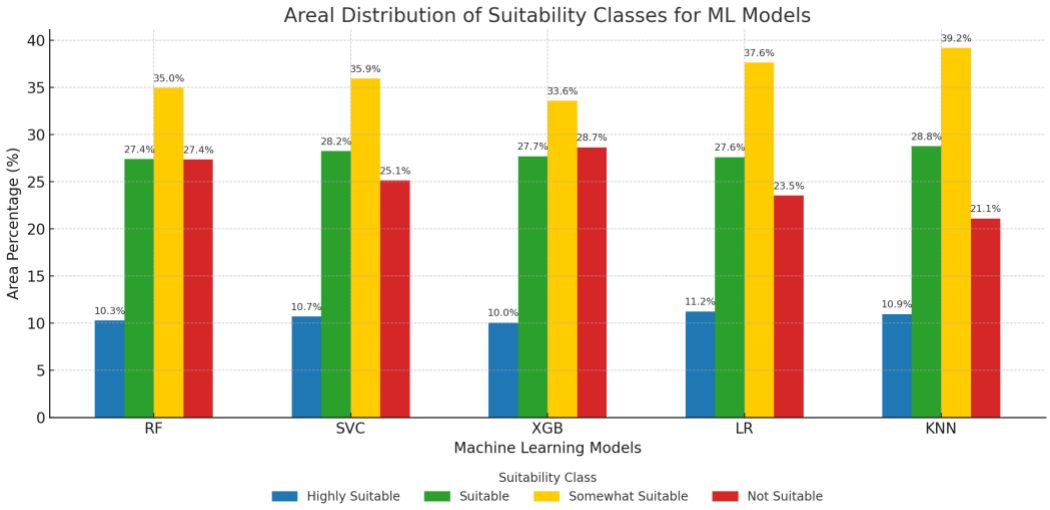


Figure 12. Area distribution of 4 suitability classes using RF, KNN, LR, XGB, and SVC model

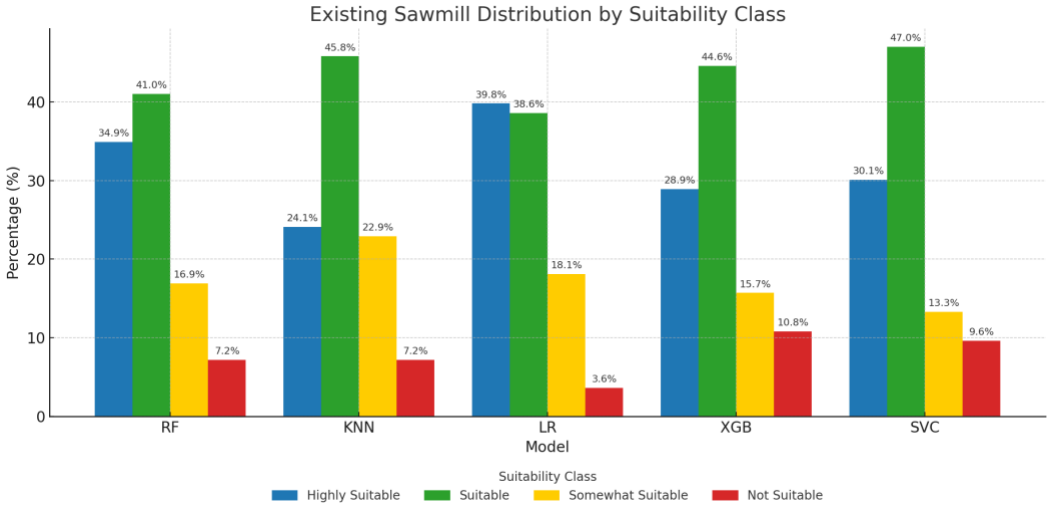


Figure 13. Existing sawmill distribution over the different suitability class

5 Discussion and managerial insights

Our case study offers several insights for decision makers and researchers.

First, geography matters. The minimal relevance of Terrain Slope and National Land Cover on site suitability highlights the context-dependent nature of feature selection. In the case of Mississippi, flat terrain and extensive forest coverage, approximately 62% of the state is forested (Johnson et al., 2023), reducing the discriminating power of these features across candidate facility locations. However, this finding does not necessarily generalize to other regions. In more topographically heterogeneous regions, slope can become a major suitability criterion (Mallick et al., 2022), and greater variation in land cover may similarly increase the predictive importance of that variable. Although direct use of the National Land Cover Database (NLCD) in sawmill site selection studies is limited, land-cover characteristics have been shown to matter in sawmill operations and forest-products facility location (Adhikari et al., 2023; Anderson & Germain, 2009). NLCD data have also been used in facility location analyses in other contexts, including solar development (Hernandez et al., 2015), power-generation infrastructure (Vernon et al., 2023), composting facilities (Harrison et al., 2024), and anaerobic-digestion facilities (Armington et al., 2026). Thus, the low importance observed here should be interpreted as specific to Mississippi and to sawmill site selection, rather than as evidence that terrain slope or national land cover are generally unimportant to site selection criteria.

Second, we find that market and supply-demand dynamics may be more important than proximity measures in Mississippi. The supply-demand ratio (SDR), a composite metric we proposed with the LB-MCDM framework to capture both local timber availability and existing local demand and competition, emerged as the most influential factor, regardless of the classification method used. This finding differs from some of the existing GIS-MCDM studies of the forest-industry site selection problems, in which transportation distance and associated haulage costs are commonly considered among the most influential criteria for site suitability (Ghaffariyan et al., 2013; Woo et al., 2018). We attribute this difference to two factors. First, our framework learns criteria importance from observed spatial patterns rather than relying on expert-defined weights, so the resulting rankings are not anchored to prior expectations about which criteria should dominate. Second, Mississippi's well-developed transportation network (Mississippi Department of Transportation, 2025) may reduce variation in accessibility across candidate locations, allowing timber availability and market competition, as captured by SDR, to play a greater role in determining sawmill suitability. Nevertheless, proximity remains important. Road Distance matched SDR under RF, received the highest weight of any criterion under LR, and remained among the top three under every classifier. Urban Area Distance ranked highest under SVC and second under LR, while Rail Line Distance ranked second under RF and tied first under XGB. Together, these results indicate that accessibility criteria consistently rank among the most influential factors regardless of the classifier used, consistent with prior evidence that transportation connectivity and spatial accessibility are important determinants of industrial and logistics facility location (Adhikari et al., 2023; Jaller et al., 2022; Sakai et al., 2018).

Third, while the relatively low impact of Precipitation, Market Revenue, and Unemployment Rate in this case study suggests that these factors may be of limited importance for sawmill site selection in Mississippi, it does not mean that they are universally unimportant. Socio-economic criteria such as labor availability ranked among the more influential criteria in site selection studies within the wood products industry conducted elsewhere (Singer, 2024), and the relative influence of any criterion depends on how much it varies across the study area. In regions where variability in precipitation significantly affects timber growth and harvesting, or where economic factors such as market size and revenue fluctuate considerably, these variables may play a more critical role.

Fourth, although the majority of existing sawmills were located within Highly Suitable or Suitable zones, approximately 25% were classified within Somewhat Suitable or Not Suitable regions. This does not necessarily imply that these past sawmill location decisions were wrong. The fact that conditions favoring a location shift over time is evident in the industry itself: of the 71 lumber manufacturing facilities recorded for Mississippi, 14 are listed as closed (Mississippi Department of Agriculture and Commerce, 2024). The closure of Hood Industries' Wiggins mill in August 2026, after 56 years of operation (Corder, 2026), illustrates that the conditions supporting continued operation at a given site do not necessarily persist. Some sawmills may have been established decades ago under different market conditions, timber availability, transportation infrastructure, land prices, or policy environments, and at that time, they may indeed have represented favorable locations. Industrial location decisions are made in a changing environment, however, and suitability may change accordingly (De Bok & Van Oort, 2011; Nilsen et al., 2020). As additional sawmills open, market competition changes, timber resources shift or become depleted, transportation patterns evolve, and regional demand structures change (Espinoza et al., 2011; Rose et al., 1983). Therefore, locations that were once highly attractive may become relatively less favorable today (Kang, 2020; Lin et al., 2022). The suitability maps generated by the model reflect current conditions and do not capture factors that may have influenced past location decisions. Moreover, a high suitability score should not be interpreted as a guarantee of business success. A favorable sawmill location under current operating conditions may facilitate business success. However, firms must also make informed and effective operational, financial, and managerial decisions to fully capitalize on this location advantage.

Fifth, transparency builds trust. One of the key strengths of the proposed framework is its enhanced explainability through SHAP analysis, which allows stakeholders to clearly understand the drivers behind site suitability. This transparency fosters stakeholder trust and supports more informed decision-making. Moreover, the model's validation against existing sawmill locations in Mississippi also demonstrates that it works in practice, not just in theory.

Sixth, the proposed LB-MCDM framework provides a practical decision-support tool by screening large sets of potential sites into a refined shortlist based on predicted suitability scores (i.e., top 10 or scores above 0.90). Decision makers can then evaluate these smaller numbers of candidate locations against real-world constraints using various MCDM techniques, such as land availability and regional plans, or feed them into downstream location-allocation models to further optimize spatial coverage, transportation costs, and capacity allocation.

Seventh, the proposed framework can be adapted to other facility location problems beyond the timber industry, where multiple spatial and non-spatial criteria must be considered. The integration of ML, GIS, and MCDM can support decision-making in other sectors requiring location assessments. By reducing subjective biases and providing a systematic approach to evaluating candidate sites, this model serves as a practical decision-support tool for complex location planning challenges.

Last but not least, the proposed framework can provide some relief for traditional facility location optimization models by processing large candidate sets to identify high-potential subsets, reducing problem size considerably and improving computational efficiency for NP-hard problems.

6 Conclusion

Our study contributes to plant location literature in two ways: methodologically and practically. Our methodological contribution is the proposal of a novel methodology that addresses some of the limitations of existing methods. Traditional MCDM approaches rely heavily on subjective factor weighting, whereas conventional optimization and heuristic techniques mainly focus on proximity measures. The proposed LB-MCDM framework, on the other hand, integrates ML, MCDM, and GIS-based spatial analysis to objectively evaluate diverse spatial and non-spatial factors. Our ML models use multiple criteria derived from literature and our conversations with the experts in the field and generate a suitability map that classifies all potential facility locations into four categories as highly suitable, suitable, somewhat suitable, and not suitable. These categories are represented as color gradients on the GIS platform. Unlike traditional MCDM methods that begin with expert surveys to identify, measure, and rank multiple criteria, our approach starts with data and computation, generating unbiased and automated predictions with interpretable outputs to support expert decision-making. As the data changes, the suitability map is automatically updated. This approach is replicable and can be easily adapted to a wide range of industrial site selection problems in other contexts.

In terms of practical contribution, we demonstrate the utility of the proposed model through a case study of the sawmill site selection assessment problem in MS. Our computational results show that the RF model outperforms other algorithms, achieving an accuracy of 86.48% and an AUC score of 0.9656. RF is followed by XGB (84.95%, AUC = 0.9569) and SVC (81.86%, AUC = 0.9159). LR and KNN demonstrated lower predictive performance. These models indicate that around 10–11% of MS is highly suitable for sawmill location, highlighting significant investment potential. The supply demand ratio (SDR) stands out as the most influential factor emphasizing the importance of market dynamics. Proximity to roads, railways, and urban areas also plays a secondary role. We validated the proposed framework through direct comparison with existing and historically established sawmill locations in Mississippi. The results demonstrate that our model predictions align well with real-world data, as 70–80% of existing sawmills are located in areas classified as either highly suitable or suitable. This validation gives us confidence that the LB-MCDM model offers a reliable framework for supporting data-driven industrial site assessments.

This study has several limitations. We focused specifically on sawmill site selection in Mississippi, so some of our findings are inherently context-specific and cannot be generalized directly to other regions. For example, due to Mississippi's flat and forest-rich landscape, terrain slope and land cover had minimal impact. These features would likely play a more crucial role in mountainous regions with larger variations in forest coverage. In addition, the Rail Line Distance criterion measured proximity to the nearest rail infrastructure without distinguishing between active freight lines and passenger-only or inactive lines. Because sawmill operations depend primarily on freight transportation, this criterion may not fully represent usable rail accessibility. Furthermore, future research could examine other algorithms, such as artificial neural networks, to potentially enhance accuracy at the expense of reduced interpretability.

Future research could build on our work in several ways. Comparing multiple states or even countries with diverse geographic and socio-economic conditions would be valuable. Since our case study highlights how context shapes the role of specific features, it would be interesting to see how the importance of these factors (and potentially new ones) shifts across different settings. Future studies could also incorporate freight-specific rail classification data to distinguish active freight lines from passenger-only or inactive lines, providing a more accurate measure of rail accessibility for sawmill operations.

Additionally, incorporating other variables such as environmental regulations, land costs, and alternative transportation modes (i.e., water transportation) could add value. For the methodology, formal expert validation should be conducted. Future location studies can leverage the proposed LB-MCDM framework to efficiently screen large candidate sets and generate a refined shortlist, from which various location-allocation models can subsequently be applied to identify optimal locations. Another promising research direction is to apply the proposed framework across different application domains and benchmark its outputs against historical decisions made using different approaches, such as traditional MCDM or raster-based spatial analysis. Last but not least, facility location optimization studies could use the proposed LB-MCDM framework to process large sets of candidate locations, filtering them down to a smaller subset of high-potential sites that meet multiple criteria. We expect that this will reduce both problem size and computational effort considerably.

Acknowledgements

Funding for this research was made possible by the U.S. Department of Agriculture's (USDA) Agricultural Marketing Service through grant 23FSMIPMS1012. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USDA.

Author contribution

The authors confirm their contribution to the paper as follows: study conception and design: M. Ahmed, A. K. Dogru, C. Zhang, C. Meng; data collection: M. Ahmed, A. K. Dogru, C. Meng; model development and implementation: M. Ahmed; analysis and interpretation of results: M. Ahmed, A. K. Dogru, C. Zhang, C. Meng; visualization: M. Ahmed; supervision: A. K. Dogru, C. Zhang, C. Meng; draft manuscript preparation: M. Ahmed; manuscript review and editing: A. K. Dogru, C. Zhang, C. Meng. All authors reviewed the results and approved the final version of the manuscript.

Appendix

Appendix available as a supplemental file at https://doi.org/10.5198/jtlu.2026.2836.

References

Adhikari, R. K., Poudyal, N. C., Brandeis, C., & Nepal, P. (2023). Identifying optimal locations for hardwood CLT plants in Tennessee: Application of a spatially explicit framework. *Forest Products Journal*, *73*(3), 219–230. https://doi.org/10.13073/FPJ-D-23-00010

Aguiar, M. O., Fernandes da Silva, G., Mauri, G. R., Ferreira da Silva, E., Ribeiro de Mendonça, A., Martins Silva, J. P., Silva, R. F., Santos, J. S., Lavagnoli, G. L., & Figueiredo, E. O. (2020). Metaheuristics applied for storage yards allocation in an Amazonian sustainable forest management area. *Journal of Environmental Management*, *271,* 110926. https://doi.org/10.1016/j.jenvman.2020.110926

Aguilar, F. X. (2007). *Factors influencing the spatial distribution of natural resource-based industries: The softwood lumber industry in the United States South* [Doctoral dissertation, Louisiana State University]. LSU Scholarly Repository. https://doi.org/10.31390/gradschool_dissertations.1645

Ahmad, F., Iqbal, A., Ashraf, I., Marzband, M., & Khan, I. (2022). Optimal location of electric vehicle charging station and its impact on distribution network: A review. *Energy Reports*, *8*, 2314–2333. https://doi.org/10.1016/j.egyr.2022.01.180

Ahmadi-Javid, A., Seyedi, P., & Syam, S. S. (2017). A survey of healthcare facility location. *Computers & Operations Research*, *79*, 223–263. https://doi.org/10.1016/j.cor.2016.05.018

Al-Ruzouq, R., Abdallah, M., Shanableh, A., Alani, S., Obaid, L., & Gibril, M. B. A. (2022). Waste to energy spatial suitability analysis using hybrid multi-criteria machine learning approach. *Environmental Science and Pollution Research*, *29*(2), 2613–2628. https://doi.org/10.1007/s11356-021-15289-0

Anderson, N., & Germain, R. (2007). Variation and trends in sawmill wood procurement in the Northeastern United States. *Forest Products Journal*, *57*(10), 36–44.

Anderson, N., & Germain, R. (2009). Land cover, land use, and mill characteristics as predictors of wood procurement range. *Forest Products Journal*, *59*(11–12), 100–107. https://doi.org/10.13073/0015-7473-59.11.100

Armington, W. R., Shrestha, S., Tomaszewski, B., Babbitt, C. W., & Tyler, A. C. (2026). Integrating regional transportation and ecological factors into anaerobic digestion siting decisions. *NPJ Sustainable Agriculture*, *4*, Article 31. https://doi.org/10.1038/s44264-026-00140-1

Banaei-Kashani, F., Ghaemi, P., & Wilson, J. P. (2014). Maximal reverse skyline query. *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 421–424). https://doi.org/10.1145/2666310.2666435

Bilgili, A., Arda, T., & Kilic, B. (2024). Explainability in wind farm planning: A machine learning framework for automatic site selection of wind farms. *Energy Conversion and Management*, *309*, 118441. https://doi.org/10.1016/j.enconman.2024.118441

Boloori Arabani, A., & Farahani, R. Z. (2012). Facility location dynamics: An overview of classifications and applications. *Computers & Industrial Engineering*, *62*(1), 408–420. https://doi.org/10.1016/j.cie.2011.09.018

Brandeis, C., & Hodges, D. (2018). Sawmill industry in Tennessee: Assessing location pattern changes and their effects on sawlog procurement distribution. *Forest Science, 64(3),* 280–289. https://doi.org/10.5849/FS-2016-124

Celik Turkoglu, D., & Erol Genevois, M. (2020). A comparative survey of service facility location problems. *Annals of Operations Research*, *292*(1), 399–468. https://doi.org/10.1007/s10479-019-03385-x

Chan, T., Cordeau, J.-F., & Laporte, G. (2009). Locating satellite yards in forestry operations. *INFOR: Information Systems and Operational Research*, *47*(3), 223–234. https://doi.org/10.3138/infor.47.3.223

Chen, Y. C., Yao, H. L., Weng, S., & Tai, Y. F. (2022). An analysis of the optimal facility location of tourism industry in Plain Region by utilizing GIS. *SAGE Open*, *12*(2). https://doi.org/10.1177/21582440221095020

Church, R. L., & Murray, A. T. (2008). *Business site selection, location analysis and GIS*. Wiley. https://doi.org/10.1002/9780470432761

Colak, H. E., Memisoglu, T., & Gercek, Y. (2020). Optimal site selection for solar photovoltaic (PV) power plants using GIS and AHP: A case study of Malatya Province, Turkey. *Renewable Energy*, *149*, 565–576. https://doi.org/10.1016/j.renene.2019.12.078

Corder, F. (2026, July 7). Hood Industries closing Wiggins mill after 56 years. *Magnolia Tribune*, Business. https://magnoliatribune.com/2026/07/07/hood-industries-closing-wiggins-mill-after-56-years/

De Bok, M., & Van Oort, F. (2011). Agglomeration economies, accessibility and the spatial choice behavior of relocating firms. *Journal of Transport and Land Use*, *4*(1), 5–24. https://doi.org/10.5198/jtlu.v4i1.144

de Luis-Ruiz, J. M., Salas-Menocal, B. R., Pereda-García, R., Pérez-Álvarez, R., Sedano-Cibrián, J., & Ruiz-Fernández, C. (2024). Optimal location of solar photovoltaic plants using Geographic Information Systems and multi-criteria analysis. *Sustainability*, *16*(7), 2895. https://doi.org/10.3390/su16072895

Diaz-Balteiro, L., & Romero, C. (2008). Making forestry decisions with multiple criteria: A review and an assessment. *Forest Ecology and Management*, *255*(8–9), 3222–3241. https://doi.org/10.1016/j.foreco.2008.01.038

Dogru, A. K., Elmadag, A. B., Gong, K., Travers, J. M., & Meng, C. (2024). Examination of upstream supply chain and logistics issues in the US logging industry. *Transportation Journal*, *63*(2), 74–97. https://doi.org/10.1002/tjo3.12009

Duarte, A. E., Sarache, W. A., & Costa, Y. J. (2014). A facility-location model for biofuel plants: Applications in the Colombian context. *Energy*, *72*, 476–483. https://doi.org/10.1016/j.energy.2014.05.069

Espinoza, O., Buehlmann, U., Bumgardner, M., & Smith, B. (2011). Assessing changes in the U.S. hardwood sawmill industry with a focus on markets and distribution. *BioResources*, *6*(3), 2676–2689. https://doi.org/10.15376/biores.6.3.2676-2689

Farahani, R. Z., SteadieSeifi, M., & Asgari, N. (2010). Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling*, *34*(7), 1689–1709. https://doi.org/10.1016/j.apm.2009.10.005

Geisler, E., Rittenhouse, C. D., & Rissman, A. R. (2016). Logger perceptions of seasonal environmental challenges facing timber operations in the Upper Midwest, USA. *Society & Natural Resources*, *29*(5), 540–555. https://doi.org/10.1080/08941920.2015.1107789

Ghaffariyan, M. R., Acuna, M., & Brown, M. (2013). Analysing the effect of five operational factors on forest residue supply chain costs: A case study in Western Australia. *Biomass and Bioenergy*, *59*, 486–493. https://doi.org/10.1016/j.biombioe.2013.08.029

Harrison, B. P., McNeil, W. H., Dai, T., Campbell, J. E., & Scown, C. D. (2024). Site suitability and air pollution impacts of composting infrastructure for California's organic waste diversion law. *Environmental Science & Technology*, *58*(45), 19913–19924. https://doi.org/10.1021/acs.est.4c06371

Hernandez, R. R., Hoffacker, M. K., Murphy-Mariscal, M. L., Wu, G. C., & Allen, M. F. (2015). Solar energy development impacts on land cover change and protected areas.

Proceedings of the National Academy of Sciences of the United States of America, *112*(44), 13579–13584. https://doi.org/10.1073/pnas.1517656112

Islam, A., Ali, S. M., Afzaal, M., Iqbal, S., & Zaidi, S. N. F. (2018). Landfill sites selection through analytical hierarchy process for twin cities of Islamabad and Rawalpindi, Pakistan. *Environmental Earth Sciences*, *77*(3), 72. https://doi.org/10.1007/s12665-018-7239-3

Jaller, M., Qian, X., & Zhang, X. (2022). Distribution facilities in California: A dynamic landscape and equity considerations. *Journal of Transport and Land Use*, *15*(1), 755–778. https://doi.org/10.5198/jtlu.2022.2130

Jimoh, U. U., & Ogungbemi, D. (2024). Socio-economic effects of Orisumbare Sawmilling Industries on the residents of Ikire, Irewole Local Government Area Osun-State, Nigeria. *CSID Journal of Infrastructure Development*, *7*(1). https://doi.org/10.7454/jid.v7.i1.1091

Johnson, D., Lilly, M., Lowery, B., & Bockheim, J. G. (2023). *Land use in Mississippi*. In *The soils of Mississippi* (pp. 143–149). World soils book series. Springer. https://doi.org/10.1007/978-3-031-36235-4_15

Kang, S. (2020). Warehouse location choice: A case study in Los Angeles, CA. *Journal of Transport Geography*, *88*, 102297. https://doi.org/10.1016/j.jtrangeo.2018.08.007

Kariv, O., & Hakimi, S. L. (1979). An algorithmic approach to network location problems. II: The p-medians. *SIAM Journal on Applied Mathematics*, *37*(3), 539–560. https://doi.org/10.1137/0137041

Khatri, P., Loeffler, D., Bergman, R. D., & Nepal, P. (2025). On-site energy consumption and life-cycle environmental impacts of sawmills in California. *Forest Products Journal*, *75*(2), 95–108. https://doi.org/10.13073/FPJ-D-24-00040

Khavarian-Garmsir, A. R., & Rezaei, M. R. (2015). Selection of appropriate locations for industrial areas using GIS-fuzzy methods: A case study of Yazd Township, Iran. *Journal of Settlements and Spatial Planning*, *6*(1), 19–25.

Koirala, A., Kizha, A. R., & De Urioste-Stone, S. M. (2017). Policy recommendation from stakeholders to improve forest products transportation: A qualitative study. *Forests*, *8*(11), 434. https://doi.org/10.3390/f8110434

Krarup, J., & Pruzan, P. M. (1983). The simple plant location problem: Survey and synthesis. *European Journal of Operational Research*, *12*(1), 36–81. https://doi.org/10.1016/0377-2217(83)90181-9

Kuhaneswaran, B., Chamanee, G., & Kumara, B. T. G. S. (2025). A comprehensive review on the integration of geographic information systems and artificial intelligence for landfill site selection: A systematic mapping perspective. *Waste Management & Research: The Journal for a Sustainable Circular Economy*, *43*(2), 137–159. https://doi.org/10.1177/0734242X241237100

Lamsal, B., Poudel, J., & Pokharel, R. (2025). Regional economic and employment impacts of industrial decline in the forest sector: Evidence from sawmill closures in Michigan. *Forest Policy and Economics*, *178*, 103601. https://doi.org/10.1016/J.FORPOL.2025.103601

Lin, Y. H., Tian, Q., & Zhao, Y. (2022). Locating facilities under competition and market expansion: Formulation, optimization, and implications. *Production and Operations Management*, *31*(7), 3021–3042. https://doi.org/10.1111/poms.13737

Liu, A., Zhu, Q., Xu, L., Lu, Q., & Fan, Y. (2021). Sustainable supply chain management for perishable products in emerging markets: An integrated location-inventory-routing model. *Transportation Research Part E: Logistics and Transportation Review*, *150*, 102319. https://doi.org/10.1016/j.tre.2021.102319

Liu, C., Feng, Q., Zhou, W., Zhang, Q., Zhang, K., Dai, Q., & Zhang, W. (2024). Optimization of well location in w reservoir based on machine learning agent model.

In *Proceedings of the International Field Exploration and Development Conference 2023*, (Vol. 5, pp. 1015–1025). Springer Series in Geomechanics and Geoengineering. Springer. https://doi.org/10.1007/978-981-97-0475-0_86

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2019). *Consistent individualized feature attribution for tree ensembles*. http://arxiv.org/abs/1802.03888

Mallick, J., Ibnatiq, A. A., Kahla, N. B., Alqadhi, S., Singh, V. P., Hoa, P. V., Hang, H. T., Hong, N. V., & Le, H. A. (2022). GIS-based decision support system for safe and sustainable building construction site in a mountainous region. *Sustainability*, *14*(2), 888. https://doi.org/10.3390/su14020888

Migose, S. A., Bebe, B. O., de Boer, I. J. M., & Oosting, S. J. (2018). Influence of distance to urban markets on smallholder dairy farming systems in Kenya. *Tropical Animal Health and Production*, *50*(7), 1417–1426. https://doi.org/10.1007/s11250-018-1575-x

Mississippi Department of Agriculture and Commerce. (2024). *Mississippi Forest Product Manufacturers Details*. Mississippi Department of Agriculture and Commerce. https://agnet.mdac.ms.gov/agManage/uploads/1423.pdf

Mississippi Department of Agriculture and Commerce. (2025).Mississippi timber. https://timber.mdac.ms.gov/

Mississippi Department of Transportation. (2025). *Highway and Freight Systems Existing Conditions*. Mississippi Department of Transportation. https://mdot.ms.gov/documents/Planning/Plan/MULTIPLAN%202050/MDOT_T05_Existing_Highways_FINAL.pdf

Nilsen, Ø. L., Díez-Gutiérrez, M., Andersen, S. N., Tørset, T., & Cherchi, E. (2020). Where and why do firms choose to move? Empirical evidence from Norway. *Journal of Transport and Land Use*, *13*(1), 207–225. https://doi.org/10.5198/jtlu.2020.1424

Palander, T., Borz, S. A., & Kärhä, K. (2021). Impacts of road infrastructure on the environmental efficiency of high capacity transportation in harvesting of renewable wood energy. *Energies*, *14*(2), 453. https://doi.org/10.3390/en14020453

Razeghi, M., Hajinezhad, A., Naseri, A., Noorollahi, Y., & Moosavian, S. F. (2023). Multi-criteria decision for selecting a wind farm location to supply energy to reverse osmosis devices and produce freshwater using GIS in Iran. *Energy Strategy Reviews*, *45*, 101018. https://doi.org/10.1016/j.esr.2022.101018

Richardson, E. A. (2016). *Spatial component for the decision support systems of Colorado's forest products industry – industry cluster analysis on sawmills in northern Colorado* [Master's thesis, Colorado State University]. Mountain Scholar. https://doi.org/10.25675/3.019998

Rose, D. W., Montzka, T. B., Bradley, D. P., & Hoganson, H. M. (1983). Assessing locational advantages of markets in terms of delivered cost of wood. *Canadian Journal of Forest Research*, *13*(1), 176–183. https://doi.org/10.1139/x83-024

Sakai, T., Kawamura, K., & Hyodo, T. (2018). The relationship between commodity types, spatial characteristics, and distance optimality of logistics facilities. *Journal of Transport and Land Use*, *11*(1), 575–591. https://doi.org/10.5198/jtlu.2018.1363

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games II* (pp. 307-317). Princeton University Press.

Shriki, N., Rabinovici, R., Yahav, K., & Rubin, O. (2023). Prioritizing suitable locations for national-scale solar PV installations: Israel's site suitability analysis as a case study. *Renewable Energy*, *205*, 105–124. https://doi.org/10.1016/j.renene.2023.01.057

Singer, H. (2024). Location selection for a lumber drying facility via a hybrid Pythagorean fuzzy decision-making approach. *BioResources*, *19*(3), 4120–4134. https://doi.org/10.15376/biores.19.3.4120-4134

Susiati, H., Dede, M., Widiawaty, M. A., Ismail, A., & Udiyani, P. M. (2022). Site suitability-based spatial-weighted multicriteria analysis for nuclear power plants in Indonesia. *Heliyon*, *8*(3), e09088. https://doi.org/10.1016/j.heliyon.2022.e09088

Tadić, S., Krstić, M., Stević, Ž., & Veljović, M. (2023). Locating collection and delivery points using the p-median location problem. *Logistics*, *7*(1), 10. https://doi.org/10.3390/logistics7010010

Tay, R. (2017). Correlation, variance inflation and multicollinearity in regression model. *Journal of the Eastern Asia Society for Transportation Studies*, *12*, 2006–2015. https://doi.org/10.11175/easts.12.2006

Tirkolaee, E. B., Goli, A., & Mardani, A. (2023). A novel two-echelon hierarchical location-allocation-routing optimization for green energy-efficient logistics systems. *Annals of Operations Research*, *324*(1–2), 795–823. https://doi.org/10.1007/s10479-021-04363-y

Vanzetti, N., Corsano, G., & Montagna, J. M. (2017). A comparison between individual factories and industrial clusters location in the forest supply chain. *Forest Policy and Economics*, *83*, 88–98. https://doi.org/10.1016/j.forpol.2017.06.006

Vaughan, D., Edgeley, C., & Han, H. S. (2022). Forest contracting businesses in the US Southwest: Current profile and workforce training needs. *Journal of Forestry*, *120*(2), 186–197. https://doi.org/10.1093/jofore/fvab060

Verma, R., & Sood, Y. R. (2024). Optimal location of distributed generation for loss minimization by application of machine learning. *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India,* (pp. 1-4). https://doi.org/10.1109/SCEECS61402.2024.10482000

Vernon, C. R., Mongird, K., Nelson, K. D., & Rice, J. S. (2023). Harmonized geospatial data to support infrastructure siting feasibility planning for energy system transitions. *Scientific Data*, *10*(1), 786. https://doi.org/10.1038/s41597-023-02694-y

Weintraub, A., & Epstein, R. (2002). The supply chain in the forest industry: Models and linkages. In Geunes, J., Pardalos, P.M., Romeijn, H.E. (Eds.), *Supply chain management: Models, applications, and research directions* (pp. 343–362). Springer. https://doi.org/10.1007/0-306-48172-3_13

Witkowski, K., Mrówczyńska, M., Bazan-Krzywoszańska, A., & Skiba, M. (2018). Methods for determining potential sites for the location of logistics centres on the basis of multicriteria analysis. *LogForum, 14*(3), 279–292. https://doi.org/10.17270/J.LOG.2018.282

Woo, H., Acuna, M., Moroni, M., Taskhiri, M. S., & Turner, P. (2018). Optimizing the location of biomass energy facilities by integrating Multi-Criteria Analysis (MCA) and Geographical Information Systems (GIS). *Forests*, *9*(10), 585. https://doi.org/10.3390/f9100585

Yang, Y., Tang, J., Luo, H., & Law, R. (2015). Hotel location evaluation: A combination of machine learning tools and web GIS. *International Journal of Hospitality Management*, *47*, 14–24. https://doi.org/10.1016/j.ijhm.2015.02.008

Yunusoglu, P., Ozsoydan, F. B., & Bilgen, B. (2024). A machine learning-based two-stage approach for the location of undesirable facilities in the biomass-to-bioenergy supply chain. *Applied Energy*, *362*, 122961. https://doi.org/10.1016/j.apenergy.2024.122961

Zhang, X., Lu, J., & Peng, Y. (2021). Hybrid MCDM model for location of logistics hub: A case in China under the belt and road initiative. *IEEE Access*, *9*, 41227–41245. https://doi.org/10.1109/ACCESS.2021.3065100

Zhu, T., Lin, Y., & Liu, Y. (2017). Synthetic minority oversampling technique for multiclass imbalance problems. *Pattern Recognition*, *72*, 327–340. https://doi.org/10.1016/j.patcog.2017.07.024