Journal of Construction Engineering, Management & Innovation
2025 8(4):438-476
DOI 10.31462/jcemi.2025.04438476

RESEARCH ARTICLE

Construction labor productivity estimation through machine learning: Performance comparison of regression algorithms

Ahmet Esat Keser[1], Şirin Şevval Uçak[1], Murat Kuruoğlu[1], Onur Behzat Tokdemir[1]

[1] Istanbul Technical University, Faculty of Civil Engineering, Civil Engineering Department, İstanbul, Türkiye

Article History

Received 19 August 2025
Revised 17 September 2025
Accepted 07 October 2025

Keywords

Constuction
Labor productivity
Machine learning
Regression algorithms

Abstract

Labor productivity is a critical factor that directly affects project performance in terms of time, cost, and quality. However, accurately predicting this productivity is challenging due to variability in construction site conditions, individual characteristics of the workforce, and environmental factors. In this study, a total of 18 different regression algorithms, including ensemble methods, tree-based structures, linear regression models, support vector regressions, and artificial neural networks, were systematically evaluated to predict productivity. In the modeling process, a multi-source dataset from large-scale energy infrastructure projects was utilized. The data includes attributes such as environmental conditions, team experience, task complexity, and daily production output for six different activity types. All models are trained within a robust framework, featuring a 70% training and 30% testing separation, 10-fold cross-validation, and hyperparameter optimization using GridSearchCV. Performance evaluation is based on metrics such as R^2, RMSE, and MAE. The results of the comparative analysis revealed that the ensemble models, particularly CatBoost, XGBoost, and Bagging, outperformed the others in almost all activity types, achieving higher generalization accuracy and lower prediction error. Beyond model accuracy, attribute importance analyses also provided estimates of the determinants of productivity. Variables such as task complexity, temperature, and team experience were among the prominent factors. These findings demonstrate that data-driven models can also be applied to identify variables that influence productivity, thereby supporting planning processes. The study contributes to the literature by providing a comprehensive model comparison framework and offering practical implications for productivity management in construction projects.

1. Introduction

1.1. Labor Productivity: Definition and challenges

Construction projects usually demand significant time and effort for completion, the utilization of multiple resources, and substantial capital investment [1]. However, most projects are completed later than planned, with higher costs, and often fail to meet quality standards or customer expectations. Although companies cannot completely eliminate these challenges, they aim to improve project performance by enhancing productivity [2]. Because when labor productivity is low, both cost and schedule performance are

Correspondence Onur Behzat Tokdemir otokdemir@itu.edu.tr

adversely affected. If the project is accelerated, the duration may be shortened but costs increase; if costs are reduced, schedule performance tends to decline. Therefore, the fundamental condition for achieving satisfactory results in both dimensions is high labor productivity [3].

Consequently, the ability to accurately estimate labor productivity and develop strategies to improve it has become a critical necessity. In this context, the labor-intensive nature of the construction industry, in which labor costs typically represent between 30% and 50% of total project expenditures, highlights the practical importance of improving labor productivity to enhance both project competitiveness and long-term sustainability [4]. While construction project success has traditionally been evaluated within the triangle of time, cost, and quality, in recent years, factors such as occupational health and safety, environmental sustainability, and customer satisfaction have also been incorporated into this evaluation framework. These indicators are not independent of one another; they often interact [5]. Therefore, labor productivity plays a critical role not only in production output but also in all performance indicators that determine project success. Increasing productivity can enhance both the quality of work and production volume.

1.2. Need for accurate forecasting in project management

Although labor productivity is universally regarded as a key determinant of project success, ongoing challenges remain in precisely defining, reliably quantifying, and effectively optimizing it [6, 7]. This challenge is primarily because productivity is not a directly observable metric; rather, it is derived from a set of interrelated performance variables, each influenced by contextual, temporal, and behavioral factors on-site. In this regard, recent conceptual frameworks define labor productivity as the ratio of output to labor input, encompassing both physical and value-based measures [7, 8]. However, in practice, capturing both physical outputs (e.g., m^3 of concrete poured) and intangible value-based outputs (e.g., quality or client satisfaction) within a unified model remains a major challenge in construction productivity research.

Timely forecasts enable early adjustments in workforce allocation, equipment usage, and scheduling, thereby reducing delays and cost overruns. However, numerous factors affecting productivity make this estimation process quite complex [9]. As construction projects often operate on tight budgets and strict timelines, labor productivity plays a vital role in achieving cost-efficiency and client satisfaction [10]. Therefore, accurate productivity estimation is essential for successful project management throughout the planning and construction processes.

It is widely recognized that human performance naturally varies based on several factors, making it unrealistic to expect uniform output across a workforce. The purpose of labor productivity forecasting is not to assume consistency but to anticipate these variations and utilize the results into project planning. By forecasting productivity, managers can estimate expected performance ranges under different site conditions. This approach helps identifying potential delays or opportunities for improvements, enabling timely schedule adjustments and robust risk management strategies.

1.3. Data-driven approaches in construction

In response to this challenge, data-driven decision-making approaches have become an essential element in engineering and construction projects [11-13]. Such approaches are efficient in high-variability environments, where rule-based systems struggle to adapt to dynamic inputs and complex variable interactions. The complex and dynamic nature of projects requires systems with advanced analytical capabilities. Indeed, comparative analyses show that expert opinions often fail to align with the actual parameters observed in the field. This finding highlights the need for future labor productivity research to support expert judgments with field data [14]. Therefore, machine learning models stand out as powerful tools that improve performance in project management processes through predictive analysis [15]. In particular, ensemble-based models have shown superior ability to generalize across varied construction contexts while maintaining robustness against outliers and missing data [16]. The growing interest in these models, particularly in construction management, has led to studies focusing on predicting labor productivity [17].

1.4. Research objectives and contributions

The primary objective of this study is to evaluate the potential of machine learning algorithms in predicting labor productivity in construction projects. Unlike prior studies that focus on isolated tasks or limited algorithms, this research employs a comprehensive approach by benchmarking 18 regression models across six distinct activity types using multi-source, real-world data. The study compares the performance of different algorithms not only to determine the most accurate model but also to analyze the strengths and limitations of each, thereby guiding model selection for different project contexts and data structures.

In doing so, the study makes several contributions to the literature. First, it extends beyond the single-country focus that dominates most prior research by integrating project data from Senegal, Uzbekistan, and Iraq, thereby enhancing the external validity and generalizability of findings. Second, it advances the scope of analysis by covering multiple activity types—including formwork, rebar, concrete, steel structure installation, prefabrication, and alignment/welding—which allows for the evaluation of task-specific variations in algorithm performance. Third, it provides one of the most extensive comparative assessments to date by systematically testing 18 regression algorithms with rigorous procedures such as GridSearchCV-based hyperparameter tuning, 10-fold cross-validation, and multiple error metrics. Finally, it demonstrates the advantages of ensemble methods and complements them with cross-model feature importance analyses, offering both theoretical insights and practical guidance for workforce planning, resource allocation, and risk management.

By presenting these contributions upfront, the study clarifies its novelty and practical relevance early in the paper, while later sections provide detailed methodological, empirical, and comparative evidence.

2. Labor Productivity and Affecting Factors

Labor productivity is directly related to project outcomes, including time, cost, and quality, and is influenced by numerous factors [18]. The types of operations in construction projects (e.g., on-site or off-site applications) and the construction methods employed are among the key factors that shape labor productivity in different ways. Therefore, project managers have the opportunity to enhance labor productivity by selecting the most appropriate method in alignment with the available technology, materials, and equipment [19]. In this context, a comprehensive and systematic examination of the complex interactions among these factors is essential for developing effective strategies to improve productivity [20].

2.1. Individual factors

At the individual scale, factors such as educational background, professional experience, skillset, age, motivation, and personal outlook are among the key determinants that shape labor performance and ultimately influence construction productivity. These attributes, which develop uniquely in each worker, have been consistently highlighted in the literature as critical to workforce efficiency [18, 21]. Regarding age, evidence suggests that cognitive abilities such as problem-solving, learning speed, and mental flexibility gradually decline throughout adulthood, with meta-analytic findings indicating particularly noticeable reductions after the age of 50 [22]. Complementary research also shows a more continuous, linear decrease in processing speed from early adulthood, with sharper declines in executive functions and mental flexibility observed after the age of 70 [23]. Conversely, it has been reported that older employees can exhibit higher performance in jobs that require experience and verbal skills [24].

Furthermore, a strong relationship has been demonstrated between employees' professional competence and productivity. It has been emphasized that experienced employees complete tasks faster and with fewer errors [25-27].

2.2. Environmental factors

It is frequently stated in the literature that morale and motivation are determinants of employee commitment and project outcomes, and that there is even a bidirectional relationship between them [28-30]. Construction activities are dependent on environmental conditions. Physical space constraints and logistical barriers can lead to delays and increased costs during the project process [27]. Evidence from the wall construction case study demonstrates that while increasing the number of workers initially enhances labor productivity, this effect reverses beyond a certain threshold due to workspace congestion and movement restrictions [31].

Additionally, weather conditions such as temperature, precipitation, and wind have been reported to affect labor productivity. Harsh weather conditions not only jeopardize worker safety but also reduce productivity by creating difficulties in material handling or production [18, 32, 33]. Sevim and Kuruoğlu [34] showed that the impact of weather conditions on productivity can vary depending on the type of work performed. This finding suggests that not only weather conditions but also the nature of the work should be considered when assessing environmental factors.

2.3. Managerial factors

Each construction project has its unique dynamics, and achieving the required project performance requires a balance among project dimensions such as time, cost, and quality. The involvement of numerous factors and stakeholders throughout the project lifecycle makes the fulfillment and control of key objectives more challenging [35]. Effective management and supervision are crucial components of workforce productivity. They play a critical role in predicting potential problems, adapting to time transfers, and maintaining quality [36, 37]. Managerial factors show that adopting a comprehensive approach to scheduling and planning enhances labor productivity, thereby reducing the need for rework and revisions [38].

On the other hand, issues such as poor planning, payment delays, frequent change orders, and the necessity of rework negatively affect workforce motivation and continuity, leading to project delays and productivity losses [25, 39, 40]. In addition, delays in material delivery and resource shortages disrupt production processes and further increase rework rates, thereby exacerbating productivity decline [40].

2.4. Task-specific factors

Construction projects differ in terms of scope, contract type, specifications, applied methods, and environmental conditions. This diversity increases the complexity of projects and makes each project inherently subject to uncertainty and risk. As a result, variations and deviations often occur in construction budgets and schedules [41]. In addition, task-specific factors, such as job complexity, level of standardization, and specialized skill requirements, can also influence labor productivity. Cheng et al. [32] and Karataş & Budak [16] have shown that complex jobs slow down the production process, while repetitive and standardized tasks increase productivity. These findings emphasize the importance of considering task characteristics in productivity analyses.

Table 1 compiles the key factors influencing labor productivity in the construction industry within a four-dimensional classification framework: Individual, Task-Specific, Managerial, and Environmental. Commonly cited elements across various studies—such as supervision quality, experience level, rework, weather conditions, and material supply—are systematically grouped, and each factor is associated with its corresponding references in the literature. In doing so, the table reveals which variables have been most frequently addressed in prior research. It aims to reduce conceptual ambiguity in the literature and provide a structured basis for comparative analysis.

Table 1. Frequently reported factors affecting labor productivity in the literature

Category	Factor	References
Environmental	Congestion	[27, 31, 33]
	Weather	[16, 28, 32, 33, 34, 36, 37, 43, 44, 45, 46, 47, 48, 49, 50]
	Working Conditions	[33, 37, 44, 45, 49, 51]
	Safety	[26, 46, 49, 51, 52]
Individual	Experience/Skill/Age	[16, 21, 24, 25, 26, 27, 28, 29, 32, 36, 37, 42, 45, 46, 47, 48, 49, 50, 51, 52, 53]
	Labor Morale	[21, 28, 29, 30]
	Compatibility of Workers	[21, 26, 29, 33, 43, 50]
Managerial	Supervision	[26, 36, 37, 44, 45, 46, 47, 48, 50, 51, 53]
	Overtime	[26, 36, 52]
	Planning	[36, 44, 46, 51, 53]
	Logistics/Material Availability	[26, 27, 28, 29, 36, 47, 49, 52, 53]
	Management/Coordination	[36, 44, 46, 47, 51, 53]
	Payment	[29, 36, 45, 53]
	Rework	[25, 26, 27, 28, 37, 44, 46, 47, 51, 52, 53]
Task-Specific	Task Type	[16, 32, 33, 48, 53]
	Structural Element Characteristics	[32, 33, 37, 42, 48, 49]
	Gang Size	[16, 32, 48, 50]

The selected features in this study were derived from the four most commonly cited categories in the literature: individual, environmental, managerial, and task-specific factors. Each variable was chosen based on its measurability in the field, data reliability, and modeling relevance. Among the hundreds of factors identified in the literature, the attributes were selected by balancing data availability with analytical impact.

3. Traditional and Machine Learning Methods in Construction Labor Productivity Estimation

Traditional methods such as regression analysis and expert surveys have long been used to study labor productivity, but in recent years machine learning has been increasingly applied to better capture complex relationships and improve prediction accuracy. Logistic regression, SVM, and RF have shown effectiveness in supervised learning [42, 43], while workforce profiling with data science methods has also been explored [54]. Daily work report data—covering factors like weather, number of managers, and day of the week—have been modeled with logistic regression, kNN, decision trees, and SVM [50]. Other studies compared kNN, ANN, RF, and ANFIS using attributes such as task complexity and crew compatibility [33], while Gradient Boosting, RF, SVR, and kNN were tested with variables including temperature, floor height, work method, and crew size [16]. SVM models optimized with feature selection and metaheuristic search further reduced error rates [32]. ANN-based models also achieved high accuracy using factors such as experience, health, weather, and material availability [36, 37, 46, 49] and were extended to generate prediction intervals that account for uncertainty [48]. More recent approaches include XGBoost-based probabilistic time series models for short- and long-term forecasting [55] and Bayesian structural models incorporating external factors such as inflation, unemployment, and management practices [56], both of which provide uncertainty ranges to support decision-making.

Despite these advances, important gaps remain. Traditional statistical methods continue to dominate exploratory studies, survey-based approaches are still widely used for managerial and organizational aspects, and there is little standardization across ML-based research. Moreover, contextual limitations such as single-country focus, small datasets, or narrow task-specific scopes constrain the generalizability of findings.

Table 2 presents a comparative summary of the most representative works across traditional and ML-based approaches, helping to situate the present study within this evolving landscape. This condensed overview highlights the methodological diversity, key findings, and remaining limitations in the literature. The extended version of Table 2, including all reviewed studies, is provided in Appendix A for reference.

Table 2. Comparative summary of studies on construction labor productivity: factors, methods, and key findings

Ref.	Activity Types	ML Models Used	Key Findings / Notes
58	Concrete Pouring, Formwork, Finishing	Neural Networks (ANN)	ANN-based models were developed for three activity types. Most influential factor was completed quantity. ANN outperformed regression in some tasks. Weekly data were collected from 8 construction sites.
49	Marble Finishing (Floors)	BPNN	A simple ANN model predicted marble flooring productivity with 90.9% accuracy. Main factors: labor age, experience, and assist workers. $R^2 = 80.19\%$, MAPE = 9.1%. A rare case study from Iraq.
64	Residential Masonry	Linear Regression	35 factors evaluated in masonry work. Top three: material shortage (RII = 0.780), tool/equipment shortage (0.760), and poor site conditions (0.751). Regression yielded $R^2 = 0.94$.
72	Formwork	EPR, GRNN, Linear Regression	EPR was applied to formwork productivity using 221 observations. Top predictors: temperature and crew size. EPR outperformed GRNN and regression ($R^2 = 52.69\%$). Managerial aspects were excluded.
43	Masonry Tasks	KNN, DNN, Logistic Regression, SVM, ResNet18	KNN (K=10) achieved 97.7% accuracy with feature selection. Top variables: temperature, task difficulty, experience, team size, personality fit. Deep models (e.g. ResNet18) supported fine-grained classification.
16	Formwork Installation	ET + GBM + LightGBM + CatBoost + MLP	A stacking meta-ensemble model achieved the best results ($R^2 = 0.7967$, RMSE = 0.0658). Top factors: floor height and weather. Ensemble methods outperformed single models in formwork tasks.
75	Reinforcement (Rebar)	Multivariate Linear Regression	VSM-based redesign guided by ML regression improved reinforcement productivity by 13.7%. Sub-processes like cutting and bending were modeled independently. Lean principles enhanced cycle efficiency.
84	Multiple trades (e.g., concreting, masonry, HVAC)	XGBoost, compared with ANN, RF, GBDT, DT	SHAP-based inefficiency analysis across 10 trades. Key factors: direct work loss, skill gaps, rework, high turnover. Strategy proposals grouped under 5 themes: labor, safety, training, communication, and site logistics.
85	Bricklaying, Rebar Work, Carpentry	Random Forest + 8 others incl. XGBoost, CatBoost	SHAP-based fatigue prediction model highlighted top predictors: work duration, shift (AM/PM), WBGT, and worker age. Model aids safety and task optimization.
86	Rebar-Fixing Works	RF, XGBoost, KNN, SVR	RF achieved best performance ($R^2=0.874$, RMSE≈22.3). SHAP analysis showed temperature (negative) and large-diameter rebars (positive) as key drivers. Demonstrated value of explainable ML for construction productivity forecasting.

4. Research Objectives and Questions

The primary objective of this study is to systematically compare and evaluate the predictive performance of various machine learning algorithms, taking into account the numerous factors that influence labor productivity across different task types.

Hyperparameter optimization is performed using the GridSearchCV method to maximize the performance of each algorithm. This method systematically tries all combinations within the specified parameter ranges, aiming to determine the configuration that provides the best performance. Predefined hyperparameter values for each model are provided to this function in the form of a dictionary, and GridSearchCV evaluates all possible combinations of these values. The system tests each hyperparameter combination using cross-validation to achieve robust overall performance [87]. Cross-validation is a widely used method for evaluating and comparing the performance of machine learning algorithms. In its most well-known application, k-fold cross-validation, the dataset is divided into k equal parts; at each iteration, one part is reserved for validation, while the remaining k−1 parts are used for training the model. This process is repeated k times, ensuring that each data point is included in the validation process at least once. The generalizability of the model is evaluated by averaging the performance metrics obtained from each iteration [88].

In this study, the dataset was first split into two parts: 70% for training and 30% for testing. Model training was performed solely on the training data, and 10-fold cross-validation was applied to this training set. This enables reliable model accuracy measurements and mitigates the risk of overfitting. This approach ensures that each model is internally optimized and fairly compared with the others.

Another objective of the study was to analyze the impact of features on the learning process of the models used. To this end, feature importance levels derived from tree-based models were examined, and the relative contributions of factors affecting labor productivity were assessed. This aimed to provide decision support for project managers and planners to interpret field data more meaningfully.

This study seeks to answer the following research questions:

- To what extent can labor productivity be predicted using machine learning using data from different activity types in construction projects?
- Among different machine learning algorithms, which methods have the highest accuracy and generalization success in predicting labor productivity?
- Among environmental, managerial, task-specific, and individual attribute groups, which factors are more influential on labor productivity?

5. Methodology

In this study, machine learning models developed to predict labor productivity were tested on a multi-source, field-based dataset. The dataset was created by integrating data from various sources related to large-scale power plant projects in four different countries. This provides a more realistic forecasting environment that incorporates both environmental and operational factors. The goal is to develop a machine learning model that can predict labor productivity (in Unit Person-Hours) daily across different activity types.

5.1. Data sources

The dataset used in the modeling process was compiled from the following four primary sources:

- Daily Construction Site Reports: These contain daily production quantities, task types, date, project name, and country information for each activity difficulty. This data directly reflects field performance and forms the basis for productivity analyses.
- SAP Database [89]: The SAP enterprise system used by the company provides weekly updates on each construction activity. It includes critical project-related data such as labor hours, overtime, resource allocation, material usage, and progress tracking. The accuracy of data entries was verified using SAP timesheet logs.
- Oracle Primavera P6 EPPM [90]: Oracle Primavera P6 Enterprise Project Portfolio

Management (EPPM) was used to extract project schedules, timelines, and planned workloads. This data helped validate actual versus planned progress and ensured consistency in productivity modeling.

- Meteorological Data: Daily weather conditions such as temperature, humidity, precipitation, and wind speed, based on the geographic location of the project sites.
- Difficulty Scores: Difficulty ratings assigned by field engineers to each activity type from 1 to 10 reflect the complexity of production conditions.

5.2. Datasets

The dataset utilized in this study was compiled from four large-scale energy infrastructure projects carried out in three different countries—Senegal, Uzbekistan, and Iraq. While much of the existing literature on construction labor productivity tends to focus on single-country case studies or broad global overviews, this study provides a distinctive contribution by integrating field-based, real-world data across diverse geographical, environmental, and managerial contexts. The dataset comprises a total of 46,962 records, encompassing six core activity types: formwork installation (formwork), rebar installation (rebar), concrete pouring (concrete), steel structure installation (steel), prefabrication of pipes (prefabrication), and alignment/welding of pipes (alignment) (Table 3). This comprehensive distribution enables a detailed, work-item-specific evaluation of model performance. In addition to quantitative variables such as daily output, person-hours, overtime, and weather conditions, the dataset includes contextual attributes such as crew experience, supervisor-to-worker ratio, and task complexity. All data were aggregated and anonymized from enterprise resource planning (ERP) and scheduling systems, verified through weekly updates across all projects to ensure consistency and reliability. Given that the volume and distribution of data vary across activity types, observed differences in model performance are interpreted as a natural outcome of this structural diversity. This not only allows the study to assess predictive accuracy across different production scenarios but also underscores the critical role of data quantity and quality in the success of machine learning–based modeling approaches.

The dataset consists of 14 attributes in total and covers six different activity types:

As shown in Table 4, one of the key attributes is task complexity. Complexity ratings were obtained through a structured evaluation process involving site engineers with direct project experience. Activities were first categorized into broader work groups (e.g., formwork, reinforcement) and then subdivided into specific activity types, such as “formwork for columns” or “formwork for foundations”. For each activity type, engineers assigned a complexity score between 1 and 10, reflecting the relative level of technical difficulty, resource demand, and degree of coordination required. In some cases, the assigned complexity also accounted for contextual factors such as whether the work was located in a hard-to-reach area, required more detailed workmanship, or formed part of a critical system where precision and reliability were essential.

Table 3. Activity types in the dataset

Activity Type	Unit	Data Records
Formwork Installation	m^2	10,827
Rebar Installation	ton	9,245
Concrete Pouring	m^3	10,349
Steel Structure Installation	ton	1,760
Prefabrication of Pipes	weld diameter inch	6,296
Alignment/Welding of Pipes	weld diameter inch	8,485

Table 4. Selected input features and descriptions

Attribute	Explanation
Characteristics	Job-specific conditions
Complexity	Task complexity (1–10)
Crew Experience	Workforce experience level (1–10)
Resource Name	Activity Type
Unit PH	Person-hour ratio per unit of production
Overtime	Average daily overtime
Temperature	Daily temperature (°C)
Humidity	Relative humidity (%)
Precipitation	Rainfall (mm)
Windspeed	Wind speed (km/h)
Project	Project name
Country	The country where the project is carried out
SV Ratio	Supervisor/worker ratio
Local Personnel Ratio	Local worker/Total worker ratio

Ratings were discussed among engineers within each project to reduce individual bias, and the final scores represent a consensus judgment. This procedure ensured that the dataset captured both the complexity of different activity types and the situational challenges arising from project conditions.

5.3. Preprocessing

For machine learning models to make accurate predictions, raw data must be transformed into a format that is analyzable and consistent. In this study, we conducted a comprehensive preprocessing process on the data, which included the following steps:

- Completion of Missing Data: Missing days were filled in using weekly production data obtained from the SAP system as a reference. For single missing values, the weekly value was directly assigned; for consecutive missing values, weekly production was distributed equally among the remaining weeks.
- Outlier Removal: The target variable "Unit PH" values were analyzed according to the $\mu + 3\sigma$ criterion, and outliers were removed from the dataset.
- Encoding the Categorical Variables: Categorical variables such as Resource Name, Project, and Country were converted to binary columns using One-Hot Encoding and made suitable for the modeling process.
- Feature Scaling: For scale-sensitive algorithms, numerical variables were normalized using the MinMaxScaler or StandardScaler methods.
- Data Split: Across different domains, the 70–30 train–test split has consistently emerged as a reliable benchmark. For instance, Candaş and Tokdemir [91] evaluated both 70–30 and 80–20 ratios in text classification, finding that each was effective depending on the model, while Bichri et al. [92] demonstrated in image classification tasks that allocating more than 70% of the data to training improved accuracy, sensitivity, and specificity. Similarly, a study on soil shear strength prediction confirmed the 70/30 division as the most dependable option [93]. Reflecting this consensus in the literature, the present study adopts a random 70% training and 30% test split, complemented by 10-fold cross-validation within the training set.
- Defining the Target Variable: The target variable, "Unit PH," represents the ratio of total daily person-hours to produce one unit of quantity for an activity type.

The entire process is summarized in Fig. 1, with a flow that includes data collection, preprocessing, modeling, and evaluation steps.

5.4. Model selection and classification

This study evaluates 18 regression algorithms with different structural characteristics to estimate labor productivity. The selected models comprise methods widely used in the machine learning literature and are capable of working with both linear and nonlinear data structures. Each algorithm was evaluated based on criteria such as predictive power and interpretability, and was classified into five main model groups (Fig. 2):

5.4.1. Ensemble learning methods

Ensemble methods aim to increase prediction accuracy by combining multiple weak learners. The core principle is to combine multiple variations of the same method rather than rely on a single model [94]. Mostofi et al. [95], in their study on rework cost, emphasized that the accumulation of project experiences throughout the construction lifecycle enriches the knowledge base of ensemble models, which in turn enhances the accuracy of predictions and strengthens strategic planning.

- CatBoost: This is a gradient boosting decision tree algorithm that can directly process categorical variables [96].

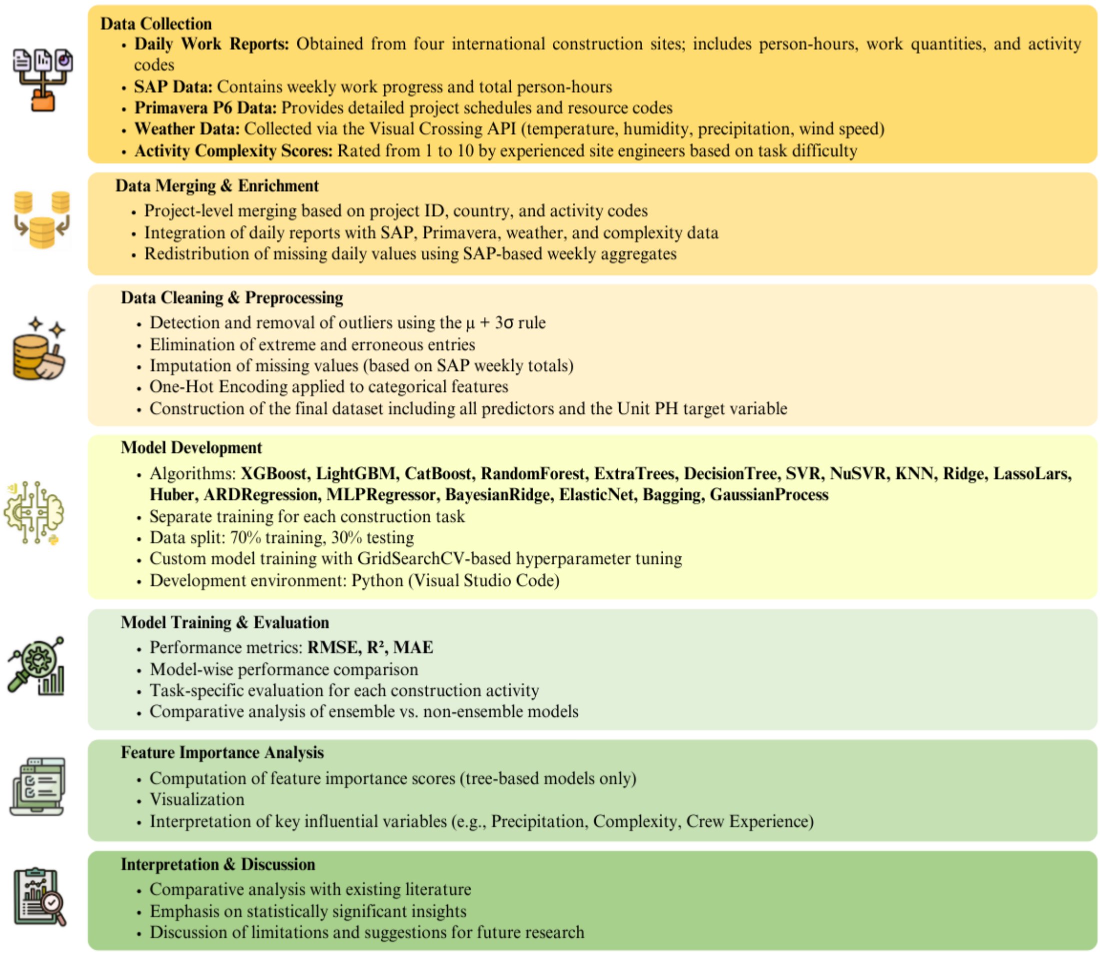

Fig. 1. Methodology

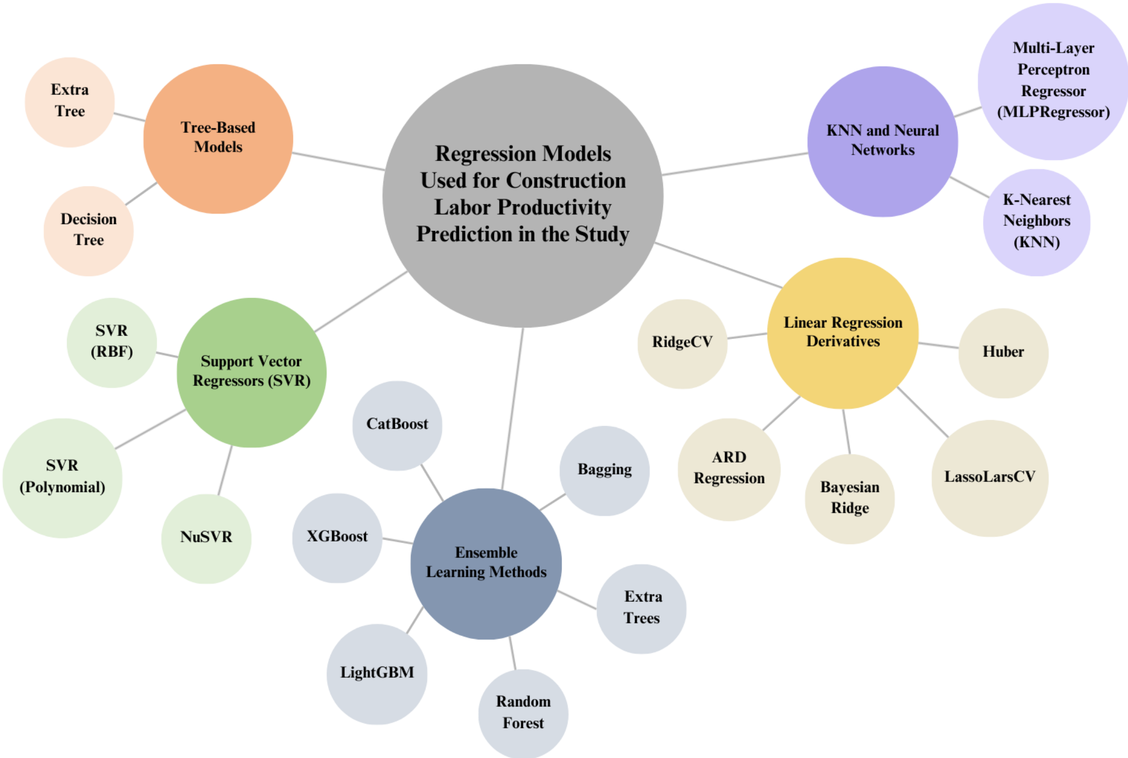

Fig. 2. Machine learning models used

Thanks to its ability to automatically process categorical variables, the CatBoost algorithm stands out as a powerful alternative for researchers, especially in cases where categorical variables are dense and the data structure exhibits heterogeneity [97]. It is robust against overfitting and can achieve high accuracy even with small datasets. However, the complex structure of the model and the ordered boosting approach can complicate hyperparameter tuning and increase computational overhead [98].

• XGBoost: This is a GBDT-based algorithm that provides high speed and accuracy. A weighted quantile sketch algorithm that quickly determines the split point by taking sample weights into account allows meaningful distinctions to be made even in cases with missing or sparse data. It operates with a sparsity-aware split method and cache-aware memory access, enabling more efficient processing of large datasets. Additionally, it can perform out-of-memory processing with out-of-core computation support [99].

• LightGBM: This is a histogram-based GBDT algorithm developed for large, high-dimensional datasets. Thanks to GOSS (Gradient-based One-Side Sampling) and EFB (Exclusive Feature Bundling) techniques, it reduces both training time and memory usage. This allows the model to produce fast and effective results in data-intensive applications [100].

• RandomForest: This method works by combining multiple independent decision trees and is robust against overfitting. It can also be used for feature importance analysis [101]. While RF is valued for its strong predictive accuracy, flexibility, and ease of use, the ensemble of many deep trees is often seen as a 'black box,' since the overall decision-making process is difficult to interpret [102].

• ExtraTrees: This method, with a high level of randomness, is computationally faster and less sensitive to hyperparameter tuning. It can produce successful results even on noisy datasets with a large number of variables [103].

- Bagging: Each base learner is trained on subsets generated by sampling with replacement from the original dataset. The final prediction is the average of these models [104].

5.4.2. Decision tree-based models

Single-decision trees have the advantage of modeling nonlinear relationships with their interpretable and straightforward structures. However, their generalization capabilities are limited compared to ensemble models.

- DecisionTree: Decision trees classify data through branching structures, forming an inverted tree composed of a root node, internal nodes, and leaf nodes. Each path, starting from the root and moving through internal nodes until reaching a leaf, represents a specific decision rule, typically expressed in an "if-then" format. The main steps in building a decision tree include splitting, stopping, and pruning. The process usually begins with growing a large tree, which is then optimized by pruning nodes that provide little or no additional information. The key advantages of decision trees are their ability to simplify complex relationships, their ease of interpretation, and their robustness against outliers [105].
- ExtraTree: ExtraTree is a single-decision tree and an algorithm that performs the learning process by randomly assigning split points. However, because it only contains a single tree, its variance is high and its generalization ability is limited. The results obtained in this study confirm this. The ExtraTree model achieved lower R^2 values and higher RMSE-MAE error levels compared to its ensemble-based counterpart, the ExtraTrees algorithm. This difference demonstrates the capacity of ensemble learning methods to reduce variance and increase overall performance.

5.4.3. Linear regressions

It is preferred for analyzing the interactions between variables in datasets containing linear relationships, revealing the direction and strength of these relationships, and making future predictions [106].

- RidgeCV: This version of the Ridge regression algorithm automatically determines the regularization parameter (alpha) through cross-validation. It was primarily developed to improve the generalization performance of the model in datasets with high dimensions and multiple linear relationships [107].
- LassoLarsCV: This regression method combines Lasso (Least Absolute Shrinkage and Selection Operator) regression with the LARS (Least Angle Regression) algorithm and automatically determines the optimal regularization coefficient (alpha) through cross-validation. This model stands out with its feature selection and model simplification capabilities in high-dimensional datasets containing numerous explanatory variables [108].
- Huber: This is a generalization of the classical linear regression method. It exhibits different behaviors depending on the magnitude of the error term using the Huber loss function. This approach increases the stability and reliability of estimations, especially in datasets with outliers [109].
- Bayesian Ridge: Provides an alternative solution to the ordinary least squares (OLS) method by making probabilistic inferences about the distribution of regression coefficients and the error term [110].
- ARDRegression: Increases interpretability by keeping only significant features in the model, thanks to automatic relevant variable determination (ARD) [111].

5.4.4. Support vector regression

SVR models use an ε-insensitive loss function that only considers deviations exceeding a certain error threshold. This structure strengthens generalization ability by ignoring small errors [112]. Furthermore, thanks to its kernel functions, it can model nonlinear relationships. This flexibility allows it to produce stable and accurate predictions even with complex data structures [113].

- SVR (RBF): Using a Radial Basis Function (RBF) kernel, SVR effectively captures nonlinear relationships by transforming the data into a high-dimensional space.

• SVR (Poly): Using a Polynomial kernel, this method models the relationships between data points to a specified polynomial degree.
• NuSVR: Model complexity is controlled directly by the "nu" parameter instead of the "C" parameter. This parameter simultaneously determines both the ratio of support vectors and the model's error tolerance.

5.4.5. KNN and ANN

• KNN: Each new sample is predicted by averaging the outputs of its nearest neighbors [114]. It performs well on small and low-noise datasets, but noisy data and poorly selected k values can significantly reduce prediction accuracy.
• MLPRegressor: With its multi-layered artificial neural network architecture, this model effectively learns complex patterns and captures nonlinear relationships. [115].

5.5. Evaluation metrics

To evaluate the performance of machine learning models comprehensively and in a multidimensional manner, this study employed multiple statistical metrics. Each metric assesses model performance from a different perspective, enabling a more balanced and reliable comparison. The key evaluation metrics used are summarized below:
• R^2 (Determination Coefficient): R^2 is a measure of fit that indicates how much of the total variance in the dependent variable is explained by the independent variables. It ranges between 0 and 1; the closer it is to 1, the greater the model's explanatory power.

$R^2 = 1$: All variance is explained by the model.

$R^2 = 0$: The model offers no explanatory power.

$R^2 < 0$: The model performs worse than simply predicting the mean.

This metric is one of the fundamental statistical measures for assessing the overall explanatory power of a model [116].

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SSR}}{\mathrm{TSS}} \tag{1}$$

SSR: Sum of Squared Residuals = $\sum(y_i - \hat{y}_i)^2$
TSS: Total Sum of Squares = $\sum(y_i - \bar{y})^2$
• RMSE (Root Mean Squared Error): It is the square root of the average squared difference between the predicted and actual values. Because it is sensitive to arithmetical values, the model has a high ability to reflect high deviations. As the RMSE value decreases, the model's accuracy increases [117].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

y_i: the true value of the observation,
$\hat{y}_i$: estimated value,
n: total number of observations
• MAE (Mean Absolute Error): It represents the average of the absolute values of the differences between the predictions and the actual values. Because it weights all errors equally, it is less sensitive to outliers than RMSE. It is easy to interpret and intuitively reflects the model's mean error. It is a reliable performance indicator, especially in datasets with limited outliers [118]. It is calculated with the following formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

Where the symbols represent the same as defined above.

Evaluating these metrics together reveals not only the average accuracy of the models but also their sensitivity to outliers and their contextual reliability.

6. Model Performance Comparisons

As part of the study, the R^2, RMSE, and MAE performances of 18 different models were analyzed using error metric tables created for each activity type. These analyses reveal not only the overall accuracy of the models but also their sensitivity to various types of errors.
• *Alignment:* The analysis results for the Alignment activity type show that ensemble and gradient boosting-based models stand out for this activity type (Fig. 3). When examining the R^2 metric, the CatBoost model achieved the highest accuracy with a value of 0.73. This was followed by XGBoost (0.69), LightGBM (0.64), and Bagging (0.64).

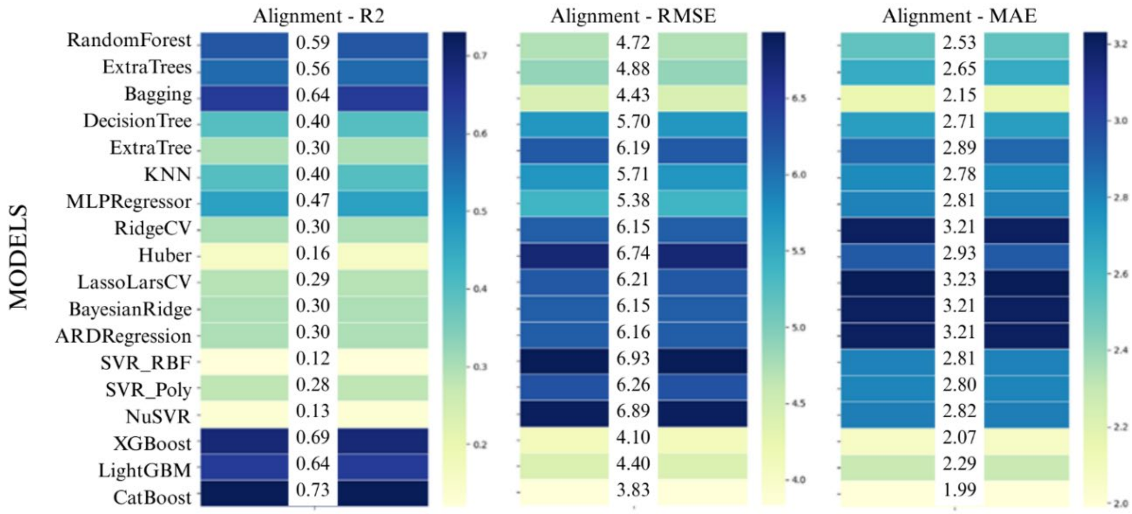

Fig. 3. Comparative heatmaps of model performances for the alignment task

A common feature of these models is their structural reliance on multiple decision trees and ensemble learning. These structures provide advantages in capturing nonlinear relationships. In contrast, models such as SVR_RBF, NuSVR, and Huber have relatively low R^2 values. This indicates that these models cannot adequately represent the patterns in the Alignment data.

A similar pattern was observed in terms of error metrics. The CatBoost model produced not only the most accurate but also the most stable and low-variance predictions, with RMSE (3.83) and MAE (1.99) values. This result demonstrates that the model avoids large errors and generally maintains prediction values close to the target. The XGBoost and LightGBM models also stood out with their low RMSE and MAE values. In contrast, both the error and variance were relatively high in the SVR and linear-based models.

Consequently, for the Alignment activity type, gradient boosting models, particularly CatBoost, yielded the most successful results, with both high accuracy (R^2) and low error (RMSE, MAE) values. This can be explained by CatBoost's ability to compensate for model complexity and its ability to learn relationships in the data effectively. On the other hand, SVR and classical linear regression models were insufficient for this task type, thus demonstrating limited performance in datasets with complex and variable inputs, such as field conditions.

- *Concrete:* Model comparisons for the concrete activities demonstrate that tree-based ensemble and boosting algorithms stand out (Fig. 4). The ExtraTrees models and CatBoost successfully explained a significant portion of the variance in the target variable with R^2 values of 0.68. These models were followed by other ensemble-based algorithms such as Bagging, XGBoost, and LightGBM. A common feature of these models is their ability to capture nonlinear relationships in the data with high precision.

When evaluated in terms of prediction errors, the RMSE and MAE values reveal a similar picture. The CatBoost, ExtraTrees, and Bagging models achieved the lowest error values, producing predictions that were both stable and closer to the target. This result is significant for activity types like concrete, which are sensitive to production quality, environmental conditions, and labor variability. On the other hand, linear regression models and SVR derivatives performed poorly in both accuracy and error metrics. This supports the conclusion that the Concrete data contains multidimensional, nonlinear patterns and is therefore not suitable for simple models.

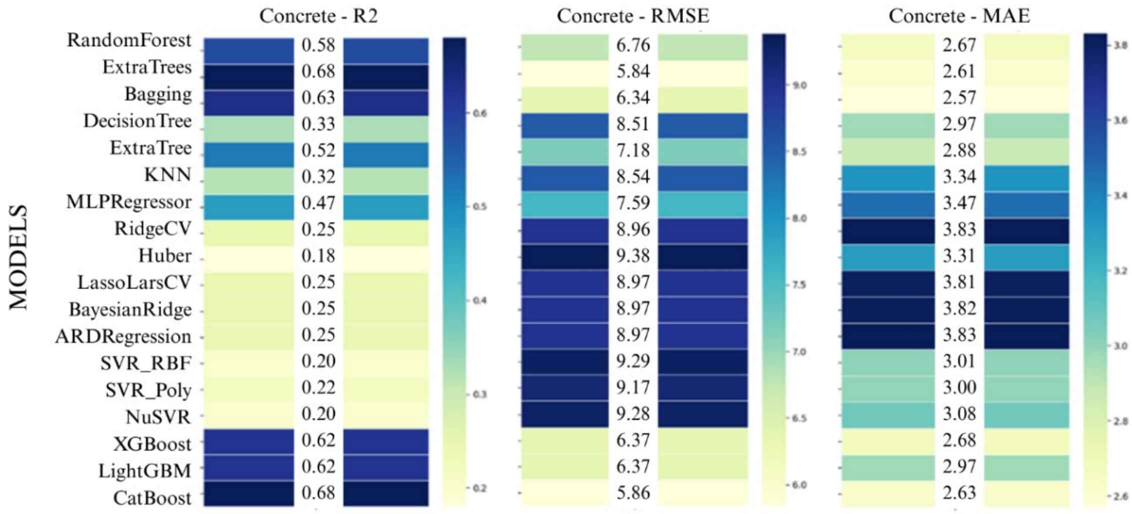

Fig. 4. Comparative heatmaps of model performances for the concrete task

Overall, ensemble methods for the Concrete activity type yielded the most successful modeling results in terms of both accuracy and error levels.

• *Formwork:* For the formwork activities, boosting and ensemble-based models were observed to provide significant superiority (Fig. 5). The XGBoost model achieved the highest accuracy, as indicated by R^2 (0.63), and led in error metrics, with the lowest RMSE (11.89) and MAE (5.04) values. Bagging, CatBoost, and LightGBM models also shared this success closely. A common characteristic of these models is their ability to create a strong predictive capacity by combining multiple weak learners.

On the other hand, linear models (Ridge, LassoLars, ARDRegression) and support vector regression-based methods performed poorly, with high error values and low R^2 scores. This result demonstrates that classical models are unable to adequately capture the complexity of patterns in formwork activities.

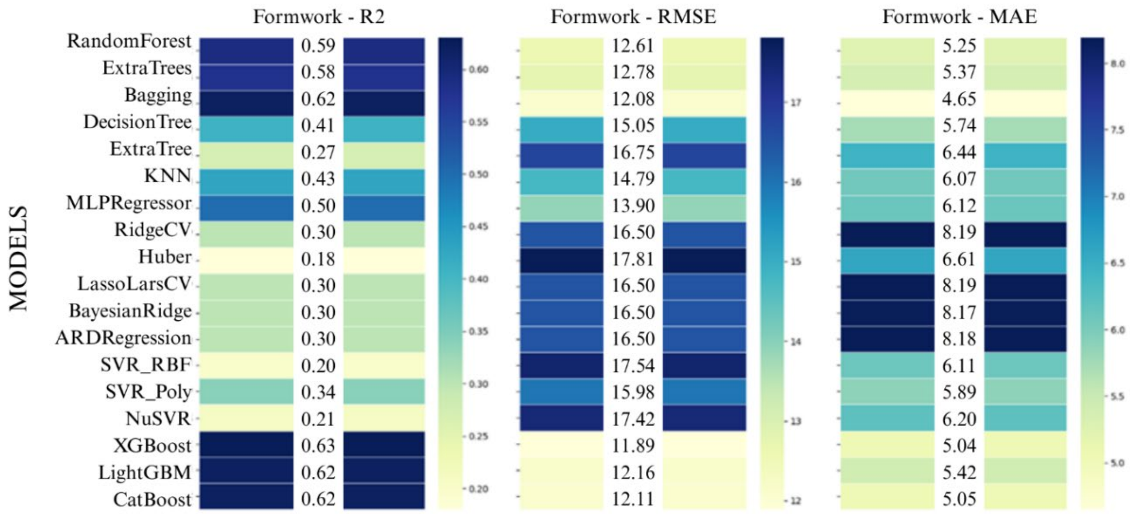

Fig. 5. Comparative heatmaps of model performances for the formwork task

Consequently, the XGBoost, Bagging, and CatBoost models are the top three candidates for formwork predictions, both in terms of accuracy and error.

- *Prefabrication:* In terms of the R^2 metric, CatBoost achieved the highest score (R^2 = 0.75). CatBoost was followed by XGBoost (0.72), Bagging, LightGBM, and RandomForest (0.71), respectively (Fig. 6). All these models are ensemble or boosting-based structures. They stand out particularly for their ability to handle multidimensional, nonlinear data relationships.

A similar superiority is also observed in terms of error metrics. CatBoost not only delivered the most accurate predictions with RMSE (5.80) and MAE (2.94) values, but also a stable prediction profile. Other boosting models, such as XGBoost (RMSE = 6.16, MAE = 2.98) and Bagging (RMSE = 6.29, MAE = 2.91), also performed well in this regard.

In contrast, linear and kernel-based models, particularly RidgeCV, ARDRegression, SVR variants, and Huber regression, exhibited poor performance, with low R^2 values (generally below 0.30) and high RMSE and MAE scores. The poor performance of these models can largely be attributed to their inability to adequately address the multifaceted, nonlinear, and interactive nature of prefabrication activities. These results suggest that boosting-based models are much more effective in activity types with high process complexity and operational interactions, such as prefabrication.

- *Rebar:* According to the modeling results obtained within the Rebar activity type, tree-based ensemble models (Bagging, Random Forest, Extra Trees, XGBoost, LightGBM, and CatBoost) generally demonstrated consistent and high performance (Fig. 7).

In terms of R^2 scores, the Bagging and Random Forest models achieved the highest accuracy compared to other models, with a performance of 70%, followed by Catboost, XGBoost, and LightGBM, all with a performance of 69%. These five models stand out as the algorithms that best explain the variance of the target variable in the Rebar activity type.

A similar ranking was maintained for the RMSE and MAE metrics. Bagging achieved the lowest error levels with RMSE value of 120.98 and MAE value of 56.21, optimizing both prediction accuracy and error tolerance. The relatively high RMSE and MAE values observed in the Rebar activities are due not only to low model performance but also to the distributional properties of the target variable.

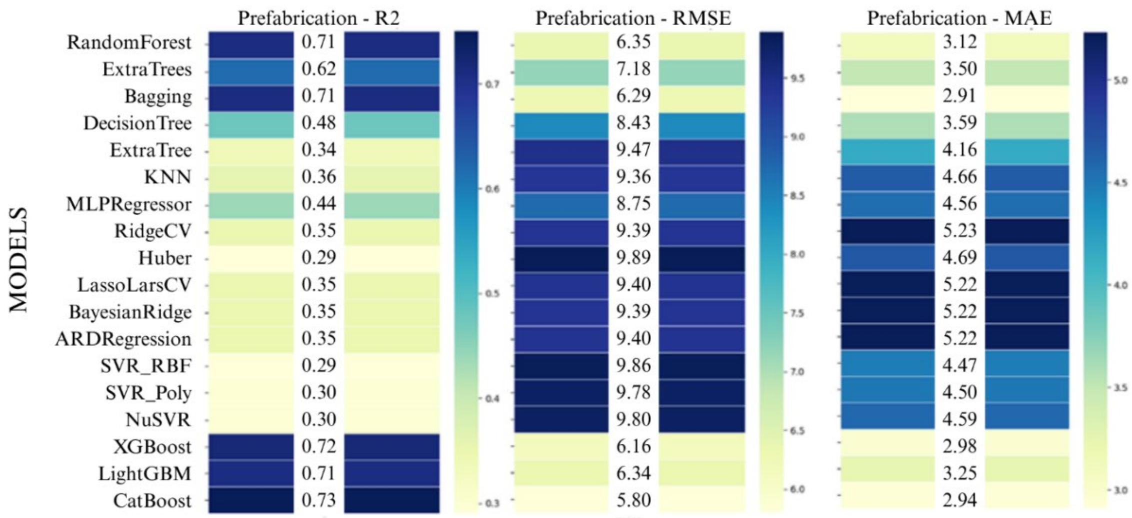

Fig. 6. Comparative heatmaps of model performances for the prefabrication task

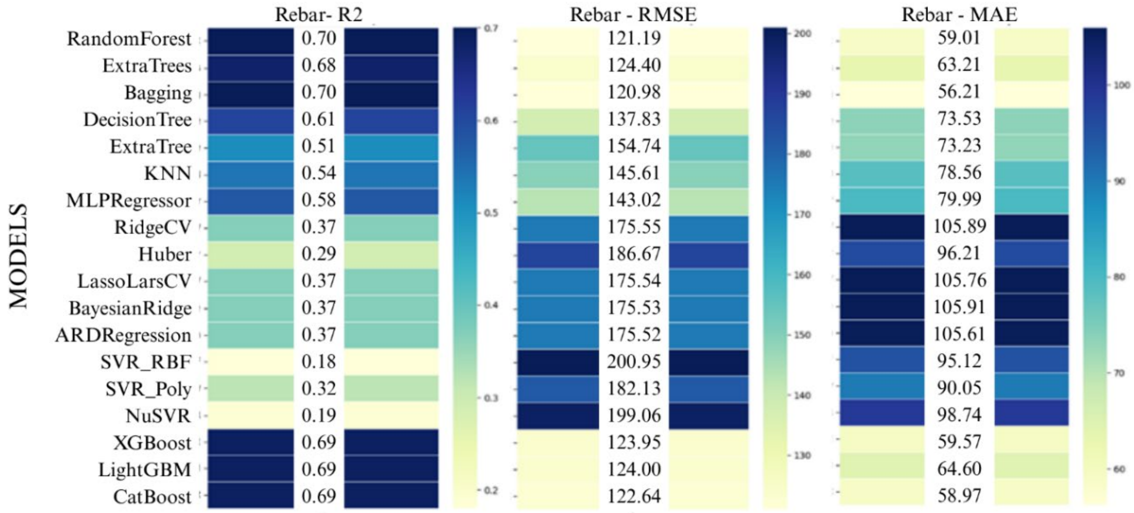

Fig. 7. Comparative heatmaps of model performances for the rebar task

The very wide range of the target variable in the dataset for this activity type makes it difficult for models to produce consistent predictions, especially at high values. Therefore, the high error metrics obtained in the Rebar data should be explained by the structural complexity of the data rather than model failure.

On the other hand, linear regression-based models, SVR variants, and MLPRegressor models produced both low R^2 and high RMSE-MAE error values. Error levels in models such as SVR_RBF and NuSVR were observed to approach 200 units.

The performance of single models such as decision trees (DecisionTree, ExtraTree) remained at a moderate level. This demonstrates that the model alone cannot adequately learn complex patterns, but when used in an ensemble format (e.g., Bagging or RandomForest), it provides a significant performance boost.

In conclusion, the modeling study conducted for the Rebar activity type demonstrated that ensemble learning techniques can provide more accurate and reliable predictions than other models on such datasets.

- *Steel:* R^2 scores for the Steel activities range between 0.60 and 0.50 in most models, with the highest value of 0.64 observed in the ExtraTrees model (Fig. 8). A detailed evaluation of the underlying reasons for this situation reveals the structural characteristics of the dataset.

Firstly, the Steel activity type dataset contains approximately one-quarter of the data compared to other activity types. This limits the model's learning capacity, reducing its generalization ability. Furthermore, there is a significant difference between the minimum and maximum values of the target variable, similar to the Rebar activity type. This wide distribution range makes it difficult for the model to achieve consistent learning, increasing the standard deviation and causing predictions to deviate.

Despite containing enough observations, the Rebar dataset had a negative impact on model performance due to the extreme difference in the target variable. This observation illustrates how outliers and unbalanced distributions within the dataset can impact model accuracy.

Consequently, the relatively low model performance in the Steel activity type is primarily due to the insufficient dataset size and differences in the distribution of the target variable. This demonstrates that not only the algorithm selection but also the data quality and distribution structure play a critical role in model construction.

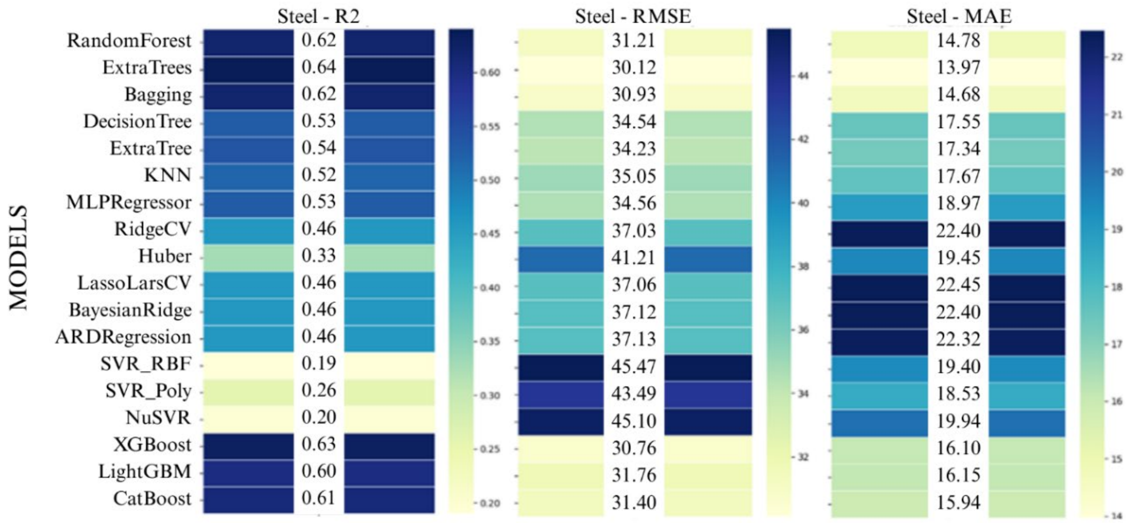

Fig. 8. Comparative heatmaps of model performances for the steel task

This point has been frequently addressed in the literature, emphasizing that model performance largely depends on the quality and size of the datasets used [119-121].

Despite the limitations of the dataset, specifically for the steel task type, ensemble-based models produced better results than other algorithms. In particular, ExtraTrees (R^2 = 0.64), XGBoost (R^2 = 0.63), Bagging, and Random Forest (R^2 = 0.62) achieved high accuracy in terms of R^2 and relatively low error levels in MAE and RMSE. Therefore, despite the extreme value distribution in the dataset, relatively satisfactory results were achieved thanks to the robust structure of these models.

6.1. Performance evaluation of model groups

The regression models tested in this study demonstrated varying levels of accuracy due to the influence of their algorithmic structures and the characteristics of the dataset. The findings revealed significant performance differences among the model groups. The models were generally examined in five main groups, and the strengths and weaknesses of each group were evaluated in consideration of the findings.

- *Superior Performance of Ensemble Models:* Ensemble learning methods have consistently demonstrated the highest performance. Models in this group, such as CatBoost, XGBoost, LightGBM, Random Forest, Extra Trees, and Bagging, stand out with their high R^2 values and low RMSE-MAE scores due to their inclusion of nonlinear and multivariate patterns in the activity types.

This success stems from the ensemble models' ability to balance variance and bias by combining multiple learners. Furthermore, the boosting algorithms' ability to sequentially improve erroneous predictions has yielded highly effective results in complex data structures. The CatBoost and XGBoost models maintained their generalization power, even on unbalanced or small datasets, providing the most consistent solutions for estimating labor productivity.

- *Poor Performance of SVR Models:* SVR models generally performed incompetently. Models such as SVR_RBF, SVR_Poly, and NuSVR exhibited low accuracy and high error metrics. This can be attributed to the SVR software's high simplicity in parameter adjustments and its poor performance with unscaled or heterogeneous data. The high variance and dispersion of data obtained from construction projects limit the predictive power of these models.
- *Limited Effectiveness of Tree-Based Models:* Decision Trees and Extra Trees, on the other hand, have achieved more limited success compared to

their ensemble versions. Their singular structure reduces their accuracy. Furthermore, small changes in decision rules can significantly impact the model structure, negatively affecting the predictive stability of these methods. However, significant performance gains have been observed when the same structures are used in ensemble form.

- *Moderate Performance of Linear Regression Derivatives:* Linear regression variants, on the other hand, performed at an average level. RidgeCV, LassoLarsCV, Huber, Bayesian Ridge, and ARDRegression models generally achieved R^2 values between 0.25 and 0.37. While computationally fast and interpretable, linear regression models fail to adequately capture multifactorial and nonlinear processes such as labor productivity, as expected. This highlights the limitations of linear models, particularly in complex structures where environmental and organizational variables interact.
- *Performance of KNN and ANN Models:* Finally, the KNN and MLPRegressor models were also noted for their low and inconsistent performance. Because the KNN algorithm used the entire dataset as a reference during prediction, it performed particularly poorly on high-dimensional and unscaled data. While MLPRegressor produced acceptable results in some tasks, due to the limited and irregular dataset, it generally produced high errors, and its learning behavior remained unstable.

In light of these evaluations, it is evident that ensemble models yield consistent results across a wide range of accuracies for various activity types. In contrast, other model groups demonstrate limited success with specific data structures. This finding demonstrates that in machine learning-based prediction models, not only the algorithm selection but also the data structure and the degree to which the model adapts to this structure are decisive.

6.2. Comparative analysis of the top three models by activity type

The top three models with the highest R^2 values were identified for the six different activity types and considered in this study. As summarized in Table 5, CatBoost, XGBoost, and Bagging models ranked among the top three across various activity types, demonstrating that these models provide robust and stable solutions for estimating labor productivity. This clearly demonstrates that tree-based ensemble structures can produce generalizable and reliable results across diverse data structures. Furthermore, the consistent performance of these models across both large and limited datasets makes them strong candidates for productivity estimation applications based on field data.

6.3. Feature importance analysis

Feature importance quantifies the contribution of each variable to the prediction process in a machine learning model. This concept helps us understand the extent to which variables the model relies on and the information it uses to make its decisions. Variables with high importance play a particularly decisive role in model output [122].

Fig. 9 shows the relative importance of feature groups used in machine learning models designed for construction labor productivity prediction. The graph is generated by averaging the feature importance values of the RandomForest, ExtraTrees, DecisionTree, ExtraTree, XGBoost, LightGBM, and CatBoost algorithms. This allows for direct comparison of the relative impact of feature groups on the overall modeling process.

All 14 feature groups in the image were analyzed. The results show that Complexity and Temperature stand out as the features with the highest average impact on the models. These two variables are followed by Crew Experience, Precipitation, and Humidity, respectively. On the other hand, the contribution of feature groups such as Local Ratio and Country is relatively low.

This analysis provides a model-independent overall assessment that can guide decision-makers in prioritizing the factors affecting construction labor productivity. Furthermore, these results can serve as a basis for improving model explainability and revising field data collection strategies.

Table 5. Comparative evaluation of the three most successful models by activity types

Activity Type	1. Model	2. Model	3. Model
Formwork	XGBoost R^2=0.63 RMSE=11.89 MAE=5.04	Bagging R^2=0.62 RMSE=12.08 MAE=4.65	CatBoost R^2=0.62 RMSE=12.11 MAE=5.05
Rebar	Bagging R^2=0.70 RMSE=120.98 MAE=56.21	RandomForest R^2=0.70 RMSE=121.19 MAE=59.01	CatBoost R^2=0.69 RMSE=122.64 MAE=58.97
Concrete	ExtraTrees R^2=0.68 RMSE=5.84 MAE=2.61	CatBoost R^2=0.68 RMSE=5.86 MAE=2.63	Bagging R^2=0.63 RMSE=6.34 MAE=2.57
Steel	ExtraTrees R^2=0.64 RMSE=30.12 MAE=13.97	XGBoost R^2=0.63 RMSE=30.76 MAE=16.10	Bagging R^2=0.62 RMSE=30.93 MAE=14.68
Prefabrication	CatBoost R^2=0.73 RMSE=5.80 MAE=2.94	XGBoost R^2=0.72 RMSE=6.16 MAE=2.98	Bagging R^2=0.71 RMSE=6.29 MAE=2.91
Alignment	CatBoost R^2=0.73 RMSE=3.83 MAE=1.99	XGBoost R^2=0.69 RMSE=4.10 MAE=2.07	LightGBM R^2=0.64 RMSE=4.40 MAE=2.29

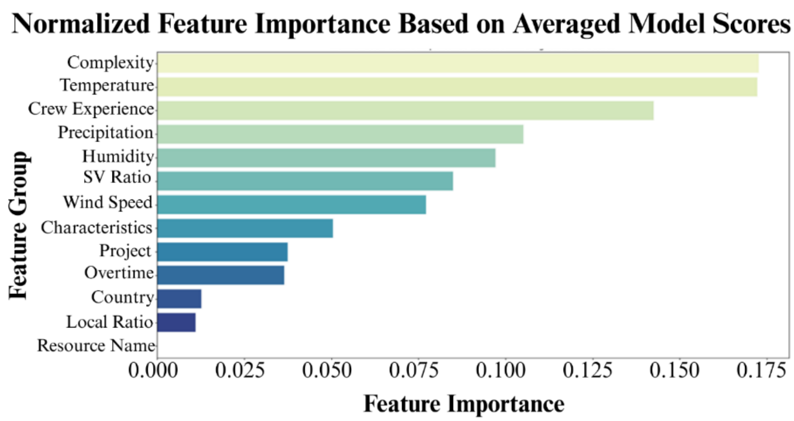

Fig. 9. Relative impact of attribute groups in labor productivity estimation

7. Contribition to the Body of Knowledge

This study makes four main contributions to the literature on construction labor productivity by explicitly positioning its findings against prior research.

First, while most empirical studies rely on single-country datasets (e.g., studies conducted in Egypt [18], New Zealand [19], Middle East [25], China [28, 29], Turkey [35], India [40], Iran [44], Uganda [47]), this research integrates data from multi-project sites in Senegal, Uzbekistan, and Iraq. Although a few cross-regional studies exist, they remain limited in scope. By combining heterogeneous site conditions and managerial contexts, this study enhances the external validity of model results and broadens their generalizability to diverse construction settings.

Second, prior studies often focus on a single activity type, which restricts insights to task-specific patterns. For example, research has investigated formwork productivity [16, 32, 68, 72], masonry tasks [43, 57], or rebar works [75]. In addition, several studies have addressed specific scopes such as marble finishing [49], concrete pouring (pumped and skipped) [62], steel drafting and fabrication [59], or combined tasks like concrete pouring, formwork, and finishing [58]. While these studies provide valuable findings, they remain limited in their scope. Some works have considered multiple tasks simultaneously, such as formwork and rebar [36], or multiple trades including concreting, masonry, and HVAC [84], as well as bricklaying, rebar, and carpentry [85]. More generally, survey-based studies have analyzed a broad range of construction activities without distinguishing specific trades [18, 26, 60, 61, 65-67, 70, 71, 73, 74, 76, 79, 80, 82]. In contrast, this study covers six distinct activity types—formwork, rebar, concrete, steel structure installation, prefabrication, and alignment—allowing performance comparison across a wider spectrum of construction processes. This design helps uncover task-specific differences in algorithmic behavior that are often overlooked in single-activity or survey-based studies.

Third, most machine learning applications in this domain test only a narrow set of algorithms (typically ANN, regression, or SVM [18, 36, 37, 43, 58, 59, 64, 68, 72, 74, 75, 78]). By systematically benchmarking 18 regression models with robust procedures (GridSearchCV, 10-fold CV, multiple error metrics), this research provides one of the most comprehensive comparative analyses to date. The results confirm findings from earlier works that ensembles often outperform single learners, but they also highlight nuances—such as the unstable performance of SVR and the dataset sensitivity of linear models—that are rarely reported in detail.

Finally, some prior studies have examined the relative importance of factors influencing labor productivity, often within the scope of a single model or limited activity type—for example, ANN sensitivity analyses for formwork and rebar [36], ensemble-based comparisons for formwork [16], or explainable ML techniques such as SHAP applied to multi-trade and task-level datasets [78, 84, 85]. While these studies provide valuable insights, comprehensive cross-model evaluations of feature importance remain limited. By aggregating importance scores across multiple ensemble methods, the present study highlights consistently dominant predictors (e.g., task complexity, temperature, crew experience) and demonstrates their practical implications for workforce planning and risk management. This approach complements earlier efforts and extends them by offering a broader and more systematic perspective on feature importance.

Taken together, these contributions extend existing research by moving beyond narrow case studies and limited methodological scopes, and instead providing a multi-country, multi-activity, and multi-model framework that is both data-driven and comparative. This broader approach not only advances theoretical understanding but also strengthens the practical applicability of productivity estimation in construction management.

8. Conclusion

This study aimed to analyze the multidimensional factors affecting labor productivity in the construction industry using machine learning algorithms. Comparatively, it evaluated the performance of different model groups on various activity types. Ensemble-based algorithms such as CatBoost, XGBoost, and Bagging stood out for their high generalization capabilities for the nonlinear and complex data structures, producing consistent and successful results across activity types. The performance of these models in terms of accuracy and error metrics proves them to be reliable tools for labor productivity estimation.

Model outputs revealed that labor productivity is directly related to multidimensional variables, including activity complexity, team experience, climatic conditions, and resource-load balance. This demonstrates the effectiveness of data-driven decision support systems not only in predicting historical data but also in identifying critical factors that drive the process. The findings also indicate that model selection in planning processes should be based not only on technical accuracy criteria but also on practical criteria such as interpretability, flexibility, and ease of implementation. In this context, CatBoost stands out for its ability to directly utilize categorical variables without preprocessing and its resistance to overfitting, while algorithms such as XGBoost and Bagging offer alternatives suitable for different production types, boasting high predictive accuracy.

However, the study also has some limitations. The limited dataset, particularly in the Steel dataset, resulted in limited model performance. Furthermore, the sensitivity of some algorithms to hyperparameter structures prevented them from achieving similar performance across cases. However, this directly aligns with the primary objective of the study, as its aim is not only to compare absolute accuracy values but also to reveal the behavior of different model types across various contextual situations.

In this respect, these limitations also provided valuable ground for observing the context-sensitive nature of model performance.

Future studies should focus on developing more dynamic and context-sensitive modeling approaches. Working with datasets that include a time dimension is critical for monitoring changes in labor productivity throughout the process. In this regard, it is recommended that time-series-based data collected within the context of construction site conditions, project schedules, and work packages be integrated into the model. This will enable early detection of productivity fluctuations and the development of preventative planning scenarios. Moreover, the use of model-independent explainability techniques such as SHAP (SHapley Additive Explanations) will contribute to greater transparency in model decision-making mechanisms. This method not only reveals which variables are influential but also the direction of the effects and their impact on a case-by-case basis, enabling a more robust analytical interpretation.

Furthermore, it is recommended that future research evaluate model success not only through accuracy metrics but also through its impact on project performance. Performance indicators such as planning accuracy, resource allocation efficiency, workforce balancing, and time-cost variance reduction will be crucial in demonstrating the practical benefits of machine learning-based systems.

Testing model performance across different project types (e.g., infrastructure, industrial facilities, residential buildings) and construction sites of varying sizes will contribute to improving overall validity. Furthermore, model robustness can be tested using data collected from different geographic regions and climatic conditions.

In conclusion, this study presents a data-driven, systematic, and interpretable model evaluation for estimating labor productivity in the construction industry. The presented framework makes an original contribution to the literature at both theoretical and practical levels, providing not only technical accuracy but also a strong foundation that supports planning and management decisions.

Declaration

Funding

The authors acknowledge the support of the Scientific and Technological Research Council of Türkiye (TÜBİTAK). Ahmet Esat Keser received a scholarship from TÜBİTAK (2210-A Graduate Scholarship Program) during the course of this research.

Author Contributions

A.E. Keser: Methodology, Data curation, Software, Validation, Formal analysis, Writing – Original Draft. Ş.Ş. Uçak: Validation, Formal analysis, Writing – Original Draft, Writing – Review & Editing, Visualization. M. Kuruoğlu: Conceptualization, Resources, Supervision. O.B. Tokdemir: Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Project administration.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request, subject to restrictions.

Ethics Committee Permission

Not applicable.

Conflict of Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Chassiakos AP, Sakellaropoulos SP (2005) Time-cost optimization of construction projects with generalized activity constraints. J Constr Eng Manag 131(10):1115–1124. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:10(1115).

[2] Sakr H, Naderpour H, Sharbatdar MK (2024) Critical factors influencing the labor productivity of mass housing construction. J Struct Des Constr Pract 30(1). https://doi.org/10.1061/JSDCCC.SCENG-1546.

[3] Thomas HR (2012) Benchmarking construction labor productivity. Pract Period Struct Des Constr 20(4). https://doi.org/10.1061/(ASCE)SC.1943-5576.0000141.

[4] Silpa SS (2023) Improving labour productivity in construction industries. Int J Adv Trends Eng Manag 962–975. https://doi.org/10.59544/flup8114/ngcesi23p108.

[5] Koç K, Budayan C, Ekmekçioğlu Ö, Tokdemir OB (2023) Predicting cost impacts of nonconformances in construction projects using interpretable machine learning. J Constr Eng Manag 150(1). https://doi.org/10.1061/JCEMD4.COENG-13857.

[6] Shehata ME, El-Gohary KM (2011) Towards improving construction labor productivity and projects' performance. Alexandria Eng J 50(4):321–330. https://doi.org/10.1016/j.aej.2012.02.001.

[7] Yi W, Chan APC (2013) Critical review of labor productivity research in construction journals. J Manag Eng 30(2). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000194.

[8] Ozola A (2023) Factors affecting labour productivity in the construction sector. Rural Environ Educ Pers 16. https://doi.org/10.22616/reep.2023.16.006.

[9] Kazerooni M, Raoufi M, Fayek AR (2020) Framework to analyze construction labor productivity using fuzzy data clustering and multi-criteria decision-making. In: Construction Research Congress 2020: Computer Applications. American Society of Civil Engineers.

[10] Alagbe OA, Aina-Badejo TF (2019) Exploring prefabricated construction principles for smart and fast housing delivery in Abuja, Nigeria. Int J Eng Res Technol 8(6). https://doi.org/10.17577/ijertv8is060488.

[11] Utkucu D, Sözer H (2022) Building performance optimization throughout the design–decision process with a holistic approach. J Archit Eng 29(1). https://doi.org/10.1061/(ASCE)AE.1943-5568.0000573.

[12] Shubham S, Saloni S, Sidra-Tul-Muntaha (2023) Optimizing construction processes and improving building performance through data engineering and computation. World J Adv Res Rev 18(1):390–398. https://doi.org/10.30574/wjarr.2023.18.1.0614.

[13] Sundarrajan N, Prabakaran PA (2023) Review on decision making in construction project management. Int J Res Publ Rev 4(5):3357–3359. https://doi.org/10.55248/gengpi.4.623.41927.

[14] Tsehayae AA, Fayek AR (2014) Data-driven approaches to discovering knowledge gaps related to factors affecting construction labor productivity. In: Construction Research Congress 2014: Construction in a Global Network. American Society of Civil Engineers. https://doi.org/10.1061/9780784413517.086.

[15] Mostofi F, Tokdemir OB, Toğan V (2024a) Generating synthetic data with variational autoencoder to address class imbalance of graph attention network prediction model for construction management. Adv Eng Inform 62(A):102606. https://doi.org/10.1016/j.aei.2024.102606.

[16] Karataş I, Budak A (2022) Development and comparative of a new meta-ensemble machine

learning model in predicting construction labor productivity. Eng Constr Archit Manag. https://doi.org/10.1108/ECAM-08-2021-0692.

[17] Mostofi F, Tokdemir OB, Toğan V (2024b) A decision-support productive resource recommendation system for enhanced construction project management. Adv Eng Inform 62(C):102793. https://doi.org/10.1016/j.aei.2024.102793.

[18] El-Gohary KM, Aziz RF (2013) Factors influencing construction labor productivity in Egypt. J Manag Eng 30(1). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000168.

[19] Ghodrati N, Yiu TW, Wilkinson S, Shahbazpour M (2018) Role of management strategies in improving labor productivity in general construction projects in New Zealand: Managerial perspective. J Manag Eng 34(6). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000641.

[20] Naeimi Rad M, Maghrebi M (2024) Bi-level integration of discrete-event simulation (DES) and social force modeling (SFM) for construction labor productivity. Autom Constr 160:105317. https://doi.org/10.1016/j.autcon.2024.105317.

[21] Jian F, Liu Q, Feng C, Hu Q, Yu Q, Guo Q (2025) Critical factors affecting construction labor productivity: A systematic review and meta-analysis. Buildings 15(14):2463. https://doi.org/10.3390/buildings15142463.

[22] Verhaeghen P, Salthouse TA (1997) Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. Psychol Bull 122(3):231–249. https://doi.org/10.1037/0033-2909.122.3.231.

[23] Murman DL (2015) The impact of age on cognition. Semin Hear 36(3):111–121. https://doi.org/10.1055/s-0035-1555115.

[24] Skirbekk V (2004) Age and individual productivity: A literature survey. Vienna Yearb Popul Res 2:133–153. https://www.jstor.org/stable/23025440.

[25] Adebowale OJ, Agumba JN (2023) A meta-analysis of factors affecting construction labour productivity in the Middle East. J Constr Dev Ctries 28(1):193–220. https://doi.org/10.21315/jcdc-12-21-0192.

[26] Almamlook R, Bzizi M, Al-Kbisbeh M, Ali T, Almajiri E (2020) Factors affecting labor productivity in the construction industry. Am J Environ Sci Eng 4(2):24. https://doi.org/10.11648/j.ajese.20200402.13.

[27] Assaad RH, El-Adaway IH, Hastak M, LaScola Needy K (2023) Key factors affecting labor productivity in offsite construction projects. J Constr Eng Manag 149(1). https://doi.org/10.1061/jcemd4.coeng-12654.

[28] Li X, Chow KH, Zhu Y, Lin Y (2016) Evaluating the impacts of high-temperature outdoor working environments on construction labor productivity in China: A case study of rebar workers. Build Environ 95:42–52. https://doi.org/10.1016/j.buildenv.2015.09.005.

[29] Zhang S, Wang Y, Wang Y, Fan W (2016) Factors affecting construction labor productivity in China: A case study of Chongqing. https://doi.org/10.29173/mocs12.

[30] Kazaz A, Manisali E, Ulubeyli S (2008) Effect of basic motivational factors on construction workforce productivity in Turkey. J Civ Eng Manag 14(4):95–106. https://doi.org/10.3846/1392-3730.2008.14.4.

[31] Dabirian S, Moussazadeh M, Khanzadi M, Abbaspour S (2023) Predicting the effects of congestion on labour productivity in construction projects using agent-based modelling. Int J Constr Manag 23(4):606–618. https://doi.org/10.1080/15623599.2021.1901330.

[32] Cheng MY, Cao MT, Jaya Mendrofa AY (2021) Dynamic feature selection for accurately predicting construction productivity using symbiotic organisms search-optimized least square support vector machine. J Build Eng 35:101973. https://doi.org/10.1016/j.jobe.2020.101973.

[33] Ebrahimi S, Kazerooni M, Sumati V, Fayek AR (2022) Predictive model for construction labour productivity using hybrid feature selection and principal component analysis. Can J Civ Eng 49(8):1366–1378. https://doi.org/10.1139/cjce-2021-0248.

[34] Sevim D, Kuruoğlu M (2012) Verimliliklerin mevsime göre değişiminin analizi. e-J NWSA Eng Sci 7(2):544–553.

[35] Kazar G, Doğan NB, Ayhan BU, Tokdemir OB (2022) Quality failures–based critical cost impact factors: Logistic regression analysis. J Constr Eng Manag 148(12). https://doi.org/10.1061/(ASCE)CO.1943-7862.0002412.

[36] El-Gohary KM, Aziz RF, Abdel-Khalek HA (2017) Engineering approach using ANN to improve and

predict construction labor productivity under different influences. J Constr Eng Manag 143(8). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001340.

[37] Ezeldin AS, Sharara LM (2016) Neural networks for estimating the productivity of concreting activities. J Constr Eng Manag 132(6):650–656. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:6(650).

[38] Chaturvedi S, Thakkar JJ, Shankar R (2018) Labor productivity in the construction industry: An evaluation framework for causal relationships. Benchmarking Int J 25(1):334–356. https://doi.org/10.1108/BIJ-11-2016-0171.

[39] Mahamid I (2020) Study of relationship between rework and labor productivity in building construction projects. Rev Constr 19(1):30–40. https://doi.org/10.7764/RDLC.19.1.30-41.

[40] Thomas AV, Sudhakumar J (2014) Factors influencing construction labour productivity: An Indian case study. J Constr Dev Ctries 19(1):53–68. https://core.ac.uk/download/pdf/199244557.pdf.

[41] Doğan NB, Ayhan BU, Kazar G, Saygili M, Ayözen YE, Tokdemir OB (2022) Predicting the cost outcome of construction quality problems using case-based reasoning (CBR). Buildings 12(11):1946. https://doi.org/10.3390/buildings12111946.

[42] Florez-Perez L, Song Z, Cortissoz JC (2022a) Using machine learning to analyze and predict construction task productivity. Comput-Aid Civ Inf Eng 37(12):1602–1616. https://doi.org/10.1111/mice.12806.

[43] Florez-Perez L, Song Z, Cortissoz JC (2022b) Predicting construction productivity with machine learning approaches. Autom Constr 138:104280. https://doi.org/10.22260/ISARC2022/0017.

[44] Ghoddousi P, Hosseini MR (2012) A survey of the factors affecting the productivity of construction projects in Iran. Technol Econ Dev Econ 18(1):99–116. https://doi.org/10.3846/20294913.2012.661203.

[45] Gunduz M, Abu-Hijleh A (2020) Assessment of human productivity drivers for construction labor through importance rating and risk mapping. Sustainability 12(20):8614. https://doi.org/10.3390/su12208614.

[46] Heravi G, Eslamdoost E (2015) Applying artificial neural networks for measuring and predicting construction-labor productivity. J Constr Eng Manag. https://doi.org/10.1061/(ASCE)CO.1943-7862.

[47] Alinaitwe HM, Mwakali JA, Hansson B (2007) Factors affecting the productivity of building craftsmen: Studies of Uganda. Technol Econ Dev Econ 13(3):169–176. https://doi.org/10.3846/13923730.2007.9636434.

[48] Nasirzadeh F, Kabir HMD, Akbari M, Khosravi A, Nahavandi S, Carmichael DG (2020) ANN-based prediction intervals to forecast labour productivity. Eng Constr Archit Manag 27(9):2335–2351. https://doi.org/10.1108/ECAM-08-2019-0406.

[49] Al-Zwainy FMS, Rasheed HA, Ibraheem HF (2012) Development of the construction productivity estimation model using artificial neural network for finishing works for floors with marble. ARPN J Eng Appl Sci 7(6):201–208. https://www.researchgate.net/publication/288719846.

[50] Sadatnya A, Sadeghi N, Sabzekar S, Khanjani M, Tak AN, Taghaddos H (2023) Machine learning for construction crew productivity prediction using daily work reports. Autom Constr 152:104891. https://doi.org/10.1016/j.autcon.2023.104891.

[51] Naoum SG (2016) Factors influencing labor productivity on construction sites: A state-of-the-art literature review and a survey. Int J Prod Perform Manag 65(3):401–421. https://doi.org/10.1108/IJPPM-03-2015-0045.

[52] Momade MH, Shahid S, Hainin MRB, Nashwan MS, Umar AT (2022) Modelling labour productivity using SVM and RF: A comparative study on classifiers performance. Int J Constr Manag 22(10):1924–1934. https://doi.org/10.1080/15623599.2020.1744799.

[53] Mahamid I (2013) Contractors' perspective toward factors affecting labor productivity in building construction. Eng Constr Archit Manag 20(4):446–460. https://doi.org/10.1108/ECAM-08-2011-0074.

[54] Orlova EV (2021) Innovation in company labor productivity management: Data science methods application. Appl Syst Innov 4(3):68. https://doi.org/10.3390/asi4030068

[55] Jacobsen EL, Teizer J, Wandahl S, Brilakis I (2024) Probabilistic forecasting of construction labor productivity metrics. J Inf Technol Constr 24:58–83. https://doi.org/10.36680/j.itcon.2024.004.

[56] Pourrahimian E, Eltahan A, Salhab D, Crawford J, AbouRizk S, Hamzeh F (2024) Integrating expert insights and data analytics for enhanced

construction productivity monitoring and control: A machine learning approach. Eng Constr Archit Manag. https://doi.org/10.1108/ECAM-02-2024-0268.

[57] Thomas HR, Sanders SR, Bilal S (1992) Comparison of labor productivity. J Constr Eng Manag 118(4):635–650. https://doi.org/10.1061/(ASCE)0733-9364(1992)118:4(635).

[58] Sonmez R, Rowings JE (1998) Construction labor productivity modeling with neural networks. J Constr Eng Manag 124(6):498–504. https://doi.org/10.1061/(ASCE)0733-9364(1998)124:6(498).

[59] Song L, AbouRizk SM (2008) Measuring and modeling labor productivity using historical data. J Constr Eng Manag 134(10):786–794. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:10(786).

[60] Dai J, Goodrum PM, Maloney WF, Srinivasan C (2009) Latent structures of the factors affecting construction labor productivity. J Constr Eng Manag 135(5):397–406. https://doi.org/10.1061/(ASCE)0733-9364(2009)135:5(397).

[61] Durdyev S, Mbachu J (2011) On-site labour productivity of New Zealand construction industry: Key constraints and improvement measures. Australas J Constr Econ Build 11(3):18–33. https://doi.org/10.5130/ajceb.v11i3.2120.

[62] Jarkas AM (2011) Buildability factors influencing concreting labor productivity. J Constr Eng Manag 138(1):89–97. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000404.

[63] Bonham DR, Goodrum PM, Littlejohn R, Albattah MA (2017) Application of data mining techniques to quantify the relative influence of design and installation characteristics on labor productivity. J Constr Eng Manag 143(8). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001347.

[64] Ponmalar VVA, K N (2018) Study on factors influencing labour productivity in residential buildings in Indian scenario. Int J Eng Technol Manag Res 5(2):239–248. https://doi.org/10.29121/ijetmr.v5.i2.2018.168.

[65] Alyew A, Reta A, Bassa M, Tora M (2020) A study on factors affecting labour productivity on construction projects in Wolaita Zone, Ethiopia. Int J Eng Res Technol 8(12). https://doi.org/10.17577/IJERTV8IS120385.

[66] Ali MZ (2019) Impact of various aspects on efficiency of labour productivity in building construction project. SSRN. https://doi.org/10.2139/SSRN.3375933.

[67] Palikhe S, Kim S, Kim JJ (2019) Critical success factors and dynamic modeling of construction labour productivity. Int J Civ Eng 17:427–442. https://doi.org/10.1007/s40999-018-0282-3.

[68] Golnaraghi S, Zangenehmadar Z, Moselhi O, Alkass S (2019) Application of artificial neural network(s) in predicting formwork labour productivity. Adv Civ Eng 2019:1–11. https://doi.org/10.1155/2019/5972620.

[69] Agarwal KS, Tatekar SS, Makwana NV, Sharma AA (2018) Analysis of labor efficiency and determination of the factors governing labor productivity in Mumbai Metropolitan Region. https://doi.org/10.1109/ICSCET.2018.8537276.

[70] Jatinkumar SB, Patel AS (2020) Factor affecting labour productivity in construction industry and its improvement. J Emerg Technol Innov Res 7(5):249–254. https://www.jetir.org/view?paper=JETIR2005347.

[71] Jain A, Ali N, Sharma A (2020) Improve the factors affecting labour productivity in Indian construction industry. J Emerg Technol Innov Res. https://www.jetir.org/view?paper=JETIR2002480.

[72] Golnaraghi S, Moselhi O, Alkass S, Zangenehmadar Z (2020) Modelling construction labour productivity using evolutionary polynomial regression. Int J Product Qual Manag 31(1):207. https://doi.org/10.1504/IJPQM.2020.10023977.

[73] Jalal MP, Shoar S (2019) A hybrid framework to model factors affecting construction labour productivity: Case study of Iran. J Financ Manag Prop Constr 24(3):630–654. https://doi.org/10.1108/JFMPC-10-2018-0061.

[74] Krishnan B (2022) Analysis of labour productivity in construction industry using Analytical Hierarchy Process and linear regression method. Int J Sci Technol Eng 10(7):1770–1776. https://doi.org/10.22214/ijraset.2022.45365.

[75] Saini A, Thomas A (2023) Development of a machine learning based-labour productivity prediction tool to practice lean construction. In: Proceedings of the 31st Annual Conference of the International Group for Lean Construction (IGLC31), pp. 1326–1336. https://doi.org/10.24928/2023/0209.

[76] Vigneshwar R, Shanmuga Priya S (2022) Investigating the factors affecting construction site productivity – a case of India. Eng Constr Archit Manag. https://doi.org/10.1108/ecam-06-2022-0529.

[77] Latief RU, Anditiaman NM, Rahim I, Arifuddin R, Tumpu M (2023) Labor productivity study in construction projects viewed from influence factors. Civ Eng J 9(3):583–595. https://doi.org/10.28991/cej-2023-09-03-07.

[78] Rana B, Sharma SB, Neupane U, Shrestha SK (2023) Analysis of labor productivity in bricklaying operation of building construction in Surkhet, Nepal. Adv Eng Technol Int J. https://doi.org/10.3126/aet.v3i1.60628.

[79] Ardila R, Durán Padra MY, Vides Martinez KY, Mejía Aguilar G (2024) Factors affecting labor productivity in the global construction industry: A critical review, classification and ranking. Scient Technica 29(01):18–33. https://doi.org/10.22517/23447214.25546.

[80] Bahamnia H, Kumar R (2024) Exploring the multifaceted dimensions of worker productivity and satisfaction in construction site environments. In: 2024 OPJU Int Technol Conf (OTCON) on Smart Comput Innov Adv Ind 4.0, IEEE, pp. 5–7 June 2024. https://doi.org/10.1109/OTCON60325.2024.10687471.

[81] Jiang L, Shen H, Zhang J (2024) Study on labor productivity improvement based on situational awareness and improved value stream mapping. Build 14(5):1253. https://doi.org/10.3390/buildings14051253.

[82] Srikanth B, Raut A, Charpe AU, Reddy RJ (2024) Factors affecting improvements in labour productivity in building construction projects—India. In: Springer Nature, pp. 573–584. https://doi.org/10.1007/978-981-99-3557-4_42.

[83] Ahn S, Chettupuzha AJA, Ekyalimpa R, Hague S, AbouRizk S, Stylios CD (2015) Fuzzy Cognitive Maps as a tool for modeling construction labor productivity. In: North Am Fuzzy Inf Process Soc, pp. 1–7. https://doi.org/10.1109/NAFIPS-WCONSC.2015.7284150.

[84] Elbashbishy T, El-adaway IH (2025) Labor productivity losses across construction trades: A machine learning approach. J Manag Eng 41(5). https://doi.org/10.1061/JMENEA.MEENG-6671.

[85] Zong H, Yi W, Chan APC, Yang H, Wu P, Xiao B (2025) Developing a fatigue model for construction workers: An interpretable machine learning approach. J Manag Eng 41(4). https://doi.org/10.1061/JMENEA.MEENG-6662.

[86] Taha FF, Ahmed MA, Aldhamad SHR, Imran H, Bernardo LFA, Nepomuceno MCS (2025) Explainable machine learning-based estimation of labor productivity in rebar-fixing tasks. Eng 6(9):219. https://doi.org/10.3390/eng6090219.

[87] Shreya, Pathak H (2025) Explainable artificial intelligence credit risk assessment using machine learning. arXiv. https://doi.org/10.48550/arXiv.2506.19383.

[88] Refaeilzadeh P, Tang L, Liu H (2009) Cross-validation. In: Liu L, Özsu MT (eds) Encyclopedia of Database Systems. Springer, pp. 1–? https://doi.org/10.1007/978-0-387-39940-9_565.

[89] SAP SE (2024) SAP ERP system for construction project management. https://www.sap.com/index.html.

[90] Oracle Corporation (2025) Oracle Primavera P6 Enterprise Project Portfolio Management (EPPM). https://www.oracle.com/industries/construction-engineering/primavera-p6/.

[91] Candaş AB, Tokdemir OB (2022) Automating coordination efforts for reviewing construction contracts with multilabel text classification. J Constr Eng Manag 148(6):04022052. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002275.

[92] Bichri H, Chergui A, Hain M (2024) Investigating the impact of train/test split ratio on the performance of pre-trained models with custom datasets. Int J Adv Comput Sci Appl 15(2). https://doi.org/10.14569/IJACSA.2024.0150235.

[93] Nguyen QH, Ly H-B, Ho LS, Al-Ansari N, Le HV, Tran VQ, Prakash I, Pham BT (2021) Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Adv Civ Eng 2021:4832864. https://doi.org/10.1155/2021/4832864.

[94] Bühlmann P (2012) Bagging, boosting and ensemble methods. In: Gentle J, Härdle W, Mori Y (eds) Handbook of Comput Stat (pp. 985–1022). Springer. https://doi.org/10.1007/978-3-642-21551-3_33.

[95] Mostofi F, Toğan V, Ayözen YE, Tokdemir OB (2022) Predicting the impact of construction rework cost using an ensemble classifier. Sustainability 14(22):14800. https://doi.org/10.3390/su142214800.

[96] Kondratieva T, Chepurnenko A, Poliakova K, Rodionov K (2024) CatBoost algorithms to predict the load-bearing capacity of centrally compressed short CFST columns of circular cross-section. E3S Web Conf 583:06009. https://doi.org/10.1051/e3sconf/202458306009.

[97] Hancock J, Khoshgoftaar T (2020) CatBoost for big data: An interdisciplinary review. J Big Data 7(1):54. https://doi.org/10.1186/s40537-020-00369-8.

[98] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2019) CatBoost: Unbiased boosting with categorical features. arXiv preprint. https://doi.org/10.48550/arXiv.1706.09516.

[99] Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min, pp 785–794. https://doi.org/10.1145/2939672.2939785.

[100] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: A highly efficient gradient boosting decision tree. In: Adv Neural Inf Process Syst, pp 3146–3154. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[101] Breiman L (2001) Random forests (Technical report). University of California, Berkeley. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf.

[102] Haddouchi M, Berrado A (2024) A survey and taxonomy of methods interpreting random forest models. arXiv. https://doi.org/10.48550/arXiv.2407.12759.

[103] Kumari R, Sharma M, Unnisa S (2024) Airline price prediction using ExtraTrees Regressor. Int J Inf Technol Res Appl 3(3):7–14. https://doi.org/10.59461/ijitra.v3i3.97.

[104] Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140. https://doi.org/10.1007/BF00058655.

[105] Song Y-Y, Lu Y (2015) Decision tree methods: Applications for classification and prediction. Shanghai Arch Psychiatry 27(2):130–135. https://doi.org/10.11919/j.issn.1002-0829.215044.

[106] Dar U (2020) Lineer (doğrusal) regresyon. https://bookdown.org/ugurdar/dogrusalregresyon/_main.pdf.

[107] scikit-learn (2024a) sklearn.linear_model.RidgeCV [Online documentation]. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeCV.html.

[108] scikit-learn (2024b) sklearn.linear_model.LassoLarsCV [Online documentation]. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoLarsCV.html.

[109] scikit-learn (2024c) sklearn.linear_model.HuberRegressor [Online documentation]. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.HuberRegressor.html.

[110] scikit-learn (2024d) sklearn.linear_model.BayesianRidge [Online documentation]. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.BayesianRidge.html.

[111] scikit-learn (2024e) sklearn.linear_model.ARDRegression [Online documentation]. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ARDRegression.html.

[112] Xu K (2024) Industrial steam quantity prediction model based on grid search-optimized support vector regression. In: 2024 IEEE 3rd Int Conf Electr Eng, Big Data Algorithms (EEBDA), pp 1088–1094. https://doi.org/10.1109/eebda60612.2024.10485882.

[113] Drucker H, Burges CJC, Kaufman L, Smola AJ, Vapnik V (1996) Support vector regression machines. Adv Neural Inf Process Syst 9:155–161.

[114] Beskopylny AN, Stel'makh SA, Shcherban' EM, Mailyan LR, Meskhi B, Razveeva I, Chernil'nik A, Beskopylny N (2022) Concrete strength prediction using machine learning methods: CatBoost, k-nearest neighbors, support vector regression. Appl Sci 12(21):10864. https://doi.org/10.3390/app122110864.

[115] scikit-learn (2024f) sklearn.neural_network.MLPRegressor [Online documentation]. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html.

[116] Sudhamathi T, Perumal K (2024) Ensemble regression based Extra Tree Regressor for hybrid crop yield prediction system. Measurement: Sensors 35:101277. https://doi.org/10.1016/j.measen.2024.101277.

[117] Sharma DK, Chatterjee M, Kaur G, Vavilala S (2022) Deep learning applications for disease

diagnosis. In: Gupta D, Kose U, Khanna A, Balas VE (eds) Deep learning for medical applications with unique data, pp 31–51. Academic Press. https://doi.org/10.1016/B978-0-12-824145-5.00005-8.

[118] Belyadi H, Haghighat A (2021) Supervised learning. In: Belyadi H, Haghighat A (eds) Machine learning guide for oil and gas using Python, pp 169–295. Gulf Professional Publishing. https://doi.org/10.1016/B978-0-12-821929-4.00004-4.

[119] Mostofi F, Tokdemir OB, Toğan V (2024c) Leveraging variational autoencoder for improved construction progress prediction performance. In: Proc 10th Int Conf Civil Eng, pp 538–545. Lecture Notes in Civil Engineering. Springer. https://doi.org/10.1007/978-981-97-4355-1_51.

[120] Barbierato E, Della Vedova ML, Tessera D, Toti D, Vanoli N (2022) A methodology for controlling bias and fairness in synthetic data generation. Appl Sci 12(9):4619. https://doi.org/10.3390/app12094619.

[121] Althnian A, Yahya A, Abdullahi M, Al-Zahrani A, Al-Badarneh A, Algarni A (2021) Impact of dataset size on classification performance: An empirical evaluation in the medical domain. Appl Sci 11(2):796. https://doi.org/10.3390/app11020796.

[122] Ewald FK, Bothmann L, Wright MN, Bischl B, Casalicchio G, König G (2024) A guide to feature importance methods for scientific inference. In: Longo L, Lapuschkin S, Seifert C (eds) Explainable Artificial Intelligence. xAI 2024, Communications in Computer and Information Science, vol 2154, pp 369–384. Springer. https://doi.org/10.1007/978-3-031-63797-1_22.

Appendix A

Comparative summary of studies on construction labor productivity: Factors, methods, and key findings.

Ref	Activity Types	Factors (classified by our 4-category framework)				Methodology	ML Models Used	Key Findings / Notes
		Individual	Task-Specific	Environmental	Managerial			
57	Masonry Works	crew composition, absenteeism, labor source	wall type, cutting complexity	temperature, humidity, weather	material/tool availability, storage, interference, incentive schemes, scheduling	Field study across 13 international projects, Factor Model validation, separation of normal/disrupted days	None	Cross-country comparison (286 days) revealed minimal variation. Main loss drivers were management-related: storage, interference, and tool delays.
58	Concrete Pouring, Formwork, Finishing	crew size, % laborer	quantity, job type, concrete pump	temperature, humidity, precipitation	% overtime	Regression + ANN + Cross-validation	Neural Networks (ANN)	ANN-based models were developed for three activity types. Most influential factor was completed quantity. ANN outperformed regression in some tasks. Weekly data were collected from 8 construction sites.
59	Steel Drafting and Fabrication	fitter skill level, draftsperson qualification, crew size	project type, scope, piece complexity, number of cutouts/fittings	dynamic structure, fireproofing, shift type	client index, engineer firm index, admin %, overtime %, subcontract %	Historical data modeling + ANN + Discrete-Event Simulation	ANN (PINN), Simulation (Simphony)	PINN predicted drafting productivity with 75% accuracy (17 factors). ANN-based simulation modeled fabrication (120 parts, 95% CI). The study pioneered a variable selection framework for steel works.
60	General Works	craft worker qualification, training	tool & consumables, construction equipment, engineering drawing management	material availability	project management, direction & coordination, foreman/superintendent competency	Focus groups + National survey (n=1996) + Factor Analysis + Regression	None	83 factors were grouped into 10 latent variables. Most negative impact came from equipment inadequacy, poor management, and low worker qualification. Regression analysis tested their effect on productivity.

61	Not Specified (General Onsite Construction)	workforce skill & experience, motivation, absenteeism	buildability, method of construction, equipment adequacy	weather, ground conditions, market fluctuations	coordination, supervision, project team competency, rework, client interference, planning	Descriptive survey + statistical analysis	None	56 sub-factors were identified. Internal factors (e.g., rework, worker skill, planning) had a significantly higher impact on productivity than external factors such as weather or market conditions.
49	Marble Finishing (Floors)	age, experience, health status	number of assist labor, height of floor, tile size	weather condition, site condition, material availability, security condition	—	Field observation + work sampling across 10 projects (150 obs.)	BPNN	A simple ANN model predicted marble flooring productivity with 90.9% accuracy. Main factors: labor age, experience, and assist workers. $R^2 = 80.19\%$, MAPE = 9.1%. A rare case study from Iraq.
62	Concrete Works (Pumped & Skipped)	—	concrete workability, steel congestion ratio	height above ground level	volume of pour	Field observation on 39 sites (19 months), Categorical Regression (PHStat)	None	Labor productivity increased by up to 125%. Volume of pour was the strongest positive factor. Steel congestion and height reduced efficiency. Site design recommendations provided.
18	General Building Construction	craftsmen absenteeism, lack of experience, turnover, poor pay, low motivation, safety neglect	material delays, drawing errors, unavailable tools	political strikes (hartals), harsh weather, site congestion	project coordination, rework, poor planning, scheduling, lack of drawings, frequent design changes	Survey (RII analysis, 489 responses)	ANN	Management-related factors were found to be most critical. Top 5 factors: labor skill, incentive schemes, material availability, construction management competency, and supervision.

40	General High-Rise Construction	craftsmen absenteeism, lack of experience, turnover, poor pay, low motivation, safety neglect	material delays, drawing errors, unavailable tools	political strikes (hartals), harsh weather, site congestion	project coordination, poor planning, scheduling, lack of drawings, frequent design changes, rework	Questionnaire (Severity Index on 44 factors)	None	Key issues: material shortages, delayed deliveries, political disruptions (hartals), frequent design changes, and drawing delays. Highlights region-specific factors.
36	Formwork, Rebar (Foundations)	xperience, age, education, rest time, incentive	constructability, method & materials, planning, work interruptions, available workload, overtime, work at height	weather, site access, material availability, site distance, surrounding events	leadership, labor supervision, clarity of instructions, management type	ANN with 640 data points from 8 real projects + sensitivity analysis	ANN (Tanh)	An ANN model predicted labor productivity with 91.6–96.4% accuracy. Top factors: experience, incentives, material access, leadership, and supervision. Sensitivity analysis showed that ANN significantly outperformed traditional models.
63	Mechanical piping	—	pipe size, schedule, connection type, floor, fitting vs. spool installation	installation location, prefabrication vs. field conditions	(not explicitly defined but indirectly related to scope definitions and code structures)	Data mining on 2.5M+ estimating records (MCAA, RICH, RS Means); CART-based post hoc classification to develop productivity data collection codes	CART (Classification and Regression Tree)	A data-driven productivity metric was developed using industry estimating catalogs. CART-based feature selection identified 22 data codes, with five variables explaining 77.9% of the variance. The study demonstrated that machine learning–based post hoc classification can outperform traditional productivity models.
64	Residential Masonry	experience, absenteeism, personal problems, age	design complexity, equipment/tools, storage location, site layout	weather, material shortage, site access	payment delays, overtime, insufficient supervision, lighting, safety violations	RII + SPSS + Regression analysis	Linear Regression	35 factors evaluated in masonry work. Top three: material shortage (RII = 0.780), tool/equipment shortage (0.760), and poor site conditions (0.751). Regression yielded $R^2 = 0.94$.

65	Building Works	poor labor experience, poor workmanship, lack of inspiration	material shortages, equipment shortages, change orders	site accidents	delays in decision making, payment delays, shortage of meetings	Questionnaire survey + RII analysis	None	RII analysis identified 10 key issues. Top-ranked was "decision delays" (RII = 0.899), followed by material shortages, payment issues, change orders, and poor labor performance.
66	Building Works	employee motivation, training/skills, welfare, human factors	equipment management, access issues	temperature/humidity, materials/tools management	management skills, safety management, communication, schedule management	Expert survey (Delphi) + weighted scoring	None	12 factors were scored by experts. Most influential were management skills (17.39%), schedule control (13.22%), and safety oversight (10.35%).
67	Building Works	lack of motivation, trait of labor, fatigue, shortage of skilled labor	clarity of technical specs, congestion, access, inspection delay, design complexity	tool/material unavailability, procurement delay, unsafe conditions	lack of leadership, lack of supervision, no monetary incentive, no periodic meetings, salary delays, rework	FAHP + Factor Analysis + SD Modeling	None	30 factors were evaluated using FAHP. Leading factors: tool/material shortages and lack of incentive. A system dynamics model was built to visualize CLP improvement pathways in Nepal.
68	Formwork	crew size, labor percentage	work type (slab, wall, column), work method (BIP, flying form), floor level	temperature, humidity, wind speed, precipitation	—	AI-based modeling (BNN, RBFNN, GRNN, ANFIS); 221 data points; ANN evaluation with MAE, RMSE, R^2	BNN, RBFNN, GRNN, ANFIS (no classical regression model used)	BNN-based AI model predicted formwork productivity using 221 data points. Achieved 83% accuracy. Key drivers: floor level and temperature. BNN outperformed other AI models.
69	Block Work	skill level, experience, age	construction method, availability of materials/tools	electricity availability (indirect), site layout	leadership, labor supervision, communication, scheduling	Questionnaire (RII) + Site Observations	None	10 productivity factors in block work were ranked. Top issues: skilled labor, equipment/material access, and leadership. Productivity rose 41% post-intervention; delays dropped by 25%.

26	Not Specified (General Construction)	labor skill and experience	construction tech, design errors, RFI delays	—	lack of supervision, coordination, incentive scheme, rework, manager's leadership	Literature review + expert opinion	None	30 labor productivity factors were identified. Most influential factors include poor supervision, low-skilled labor, and obsolete construction technologies. Emphasized the need for training and strategic policy interventions in Libya.
70	Building Projects	skill level, motivation, communication, technical understanding	construction method, material/equipment shortages	none explicitly discussed	decision delays, clarity of specifications, payment delays	Questionnaire survey + RII analysis	None	72 factors were grouped into 8 categories. Top-ranked were: skill-to-unskilled labor ratio (RII: 0.81), payment delays, and communication problems.
71	Building Projects	Individual: lack of experience, absenteeism, age, accidents, motivation	site access, tool/equipment shortages, stacking, late crew buildup	weather, daylight hours, holidays	planning delays, safety enforcement, miscommunication, reassignments	Questionnaire survey + RII analysis	None	30 factors were analyzed. Most critical ones were: worker inexperience, safety violations, material shortages, and poor site conditions.
72	Formwork	floor level (learning curve), labor percentage	work type, work method, gang size	temperature, humidity, wind speed, precipitation	—	EPR (Evolutionary Polynomial Regression) vs. GRNN, Best Subset, Stepwise Regression	EPR, GRNN, Linear Regression	EPR was applied to formwork productivity using 221 observations. Top predictors: temperature and crew size. EPR outperformed GRNN and regression (R^2 = 52.69%). Managerial aspects were excluded.

73	General Works	fatigue, lack of motivation, skill gaps, unfamiliarity with techniques	improper feasibility study, design errors, use of new methods, execution errors	extreme weather conditions	schedule pressure, poor project management, payment delays, overtime, contractor financing, rework	System Dynamics (SD) + DEMATEL	None	60 CLP factors were ranked using DEMATEL and expert judgment. Top influencers were fatigue, lack of motivation, schedule stress, and skill shortages. Cause-effect loops were constructed.
32	Formwork Installation	direct work (%), delay work (%)	floor level, work type, work method	temperature, precipitation	crew size, work type, work method	10-fold CV + feature selection + SOS optimization	SOS-LSSVM-FS	SOS-LSSVM-FS model was applied to 220 observations. 8 out of 12 variables were auto-selected. Key factors: temperature, precipitation, crew size, method, and direct/delay work ratios.
74	Building Projects	motivation, physical fatigue, skill of labor, absenteeism	construction method, site layout	rain, temperature, humidity, site access, wind	supervision, payment delay, coordination, unrealistic scheduling, rework	Questionnaire + AHP + RII + Regression	Linear Regression	27 factors were assessed via AHP and RII. Key factors: skill level, payment delays, and accidents. Regression for plastering tasks showed high accuracy ($R^2 = 0.972$).
27	Offsite Construction	unskilled labor, fatigue, absenteeism	alternating work shifts, overcrowding	rain/weather disruptions, ripple effects	insufficient coordination, poor logistics, out-of-sequence work, inadequate supervision, rework	Survey (n=100), Risk Rating (L×I), Hierarchical Clustering, Kendall's Concordance, ICC, Cronbach's α, statistical tests	None	20 risk factors in offsite construction were ranked. Top 5: unskilled labor, poor logistics, rework/errors, out-of-sequence work, and lack of coordination. Based on surveys and statistical tests.

43	Masonry Tasks	experience, age, personality compatibility	task difficulty	min/max temperature, humidity	crew size	Supervised ML with standardization, balanced dataset, 1977 data points, feature selection (permutation)	KNN, DNN, Logistic Regression, SVM, ResNet18	KNN (K=10) achieved 97.7% accuracy with feature selection. Top variables: temperature, task difficulty, experience, team size, personality fit. Deep models (e.g. ResNet18) supported fine-grained classification.
16	Formwork Installation	gang size, labor percentage	floor level, work type	temperature, humidity, wind speed, precipitation	work method	10-fold CV + GridSearchCV + meta-ensemble modeling (Voting + Stacking)	Voting: ET + GBM + LightGBM + CatBoost + MLPStacking ET (meta-learner) + GBM + XGBoost + CatBoost + MLP	A stacking meta-ensemble model achieved the best results (R^2 = 0.7967, RMSE = 0.0658). Top factors: floor height and weather. Ensemble methods outperformed single models in formwork tasks.
75	Reinforcement (Rebar)	crew size, % skilled worker	rebar diameter, weight, cuts, bends, time per activity	crane path, height, inventory buffer use	lean principles (Kanban, 5S, JIT), idle time minimization	Work sampling + regression + VSM	Multivariate Linear Regression	VSM-based redesign guided by ML regression improved reinforcement productivity by 13.7%. Sub-processes like cutting and bending were modeled independently. Lean principles enhanced cycle efficiency.
76	General Site Works	labor experience, skill, behavior, fatigue	equipment/tools, buildability, method selection	weather, site layout, safety, amenities, regulations	supervision, planning & scheduling, rework, communication, payment delays	Questionnaire (n=204) + RII + ANOVA	None	28 factors across 7 themes analyzed. Most impactful were equipment/material planning, realistic scheduling, communication, and supervision.

77	Road Maintenance	motivation, training, skill, health, experience	construction method, execution flow	site access, field conditions, temperature, resources	time scheduling, payment	Questionnaire + SEM (SmartPLS) + Field Observation	SEM	In SEM models across 3 regions, four key drivers were found: internal labor (most dominant), time management, site conditions, and finances. Internal labor alone explained over 90% of variation.
78	Bricklaying	absenteeism, behavior, PPE use, expectations, incentives, skill	drawing complexity, changes, congestive workspace, material storage	social conditions, access, strikes, floor layout/height	planning, coordination, leadership, supervisor-inspector relations	RII + MRA + SPSS + Observation + ANN (MLP)	ANN	RII and ANN consistently highlighted material factors (100%), leadership (84.4%), and manpower (84%) as top predictors. ANN model achieved RMSE of 0.035 on test data; $\alpha = 0.976$.
79	Not Specified (Global Scope)	labor skill, training, attitude, motivation, absenteeism	availability of materials/tools, technical complexity, construction method	weather, political conditions	planning, scheduling, resource coordination, supervision, rework, payment	Systematic literature review (97 docs) + IVI + clustering + statistical preprocessing (outlier removal, imputation)	None	From 97 studies, 267 factors were extracted and reduced to 30 using IVI. 83.3% of impactful factors were internal. Top issues: material/tool availability, worker skill, supervision, and scheduling.
80	General Building Construction	worker skill, age, motivation, absenteeism, communication issues	tools/materials access, training gaps, overtime, BIM, lean principles	weather, regulatory delays, transport access	planning, unclear objectives, supervision, inspection delays, rework	Survey (473 responses), RII analysis	None	45+ productivity-related factors were grouped into 5 domains. Key challenges include communication gaps, unclear goals, poor logistics, outdated methods, and material shortages.
81	Auxiliary window frame installation	worker situational awareness (SA), participation motivation	VSM simplification, flow efficiency	on-site workflow and layout adjustments	lean training, incentives	Case study + time-motion study + improved VSM with SA framework	None	A VSM-enhanced flow system improved labor productivity by 24.07%. Flow efficiency accounted for 88.46% of the gains. Study demonstrates usability of lean tools (VSM + SA) for task-level optimization.

82	Building Works	clear instructions, repeating crews	equipment/tools, constructability	industrialization, supply chain management	communication, leadership, site meetings, BIM, lean construction, clear roles/responsibility	Questionnaire survey + RII analysis	None	25 factors were evaluated. Top factors were: clear instructions (RII: 75.58%), lean construction, BIM integration, well-defined roles, and effective communication.
83	General Works	fatigue, morale, worker competency	interruptions, work complexity	adverse work environment	supervision quality, overtime, rework, work pressure, organizational management	Fuzzy Cognitive Maps (FCM) from experts	FCM (neuro-fuzzy)	A fuzzy cognitive model based on 12 concepts simulated productivity under scenarios like low skill and harsh weather. Useful for decision testing and predictive planning.
84	Multiple trades (e.g., concreting, masonry, HVAC)	absenteeism, skill gaps, crew mismatch	poor instructions, tool/material quality	weather, jobsite congestion, project location	lack of communication, poor planning, lack of employment system, rework	Expert survey (n=106) + SHAP interpretation Analysis across 10 trades	XGBoost (per trade), compared with ANN, RF, GBDT, DT	SHAP-based inefficiency analysis across 10 trades. Key factors: direct work loss, skill gaps, rework, high turnover. Strategy proposals grouped under 5 themes: labor, safety, training, communication, and site logistics.
85	Bricklaying, Rebar Work, Carpentry	age, BMI, sleep time, alcohol/smoking habits	job nature, work duration, work session	WBGT, temperature, humidity, radiation, wind velocity	—	Field study + Wearable sensors + SHAP analysis	Random Forest + 8 others incl. XGBoost, CatBoost	SHAP-based fatigue prediction model highlighted top predictors: work duration, shift (AM/PM), WBGT, and worker age. Model aids safety and task optimization.

86	Rebar-Fixing Works	crew skill/size	rebar diameter and quantity (M16–M40, hoop rebars), work type	temperature, humidity, wind speed, precipitation	—	Field study in Hong Kong public housing projects (326 samples); work sampling, stratified train-test split, GridSearchCV with 5-fold CV	RF, XGBoost, KNN, SVR	RF achieved best performance (R^2=0.874, RMSE≈22.3). SHAP analysis showed temperature (negative) and large-diameter rebars (positive) as key drivers. Demonstrated value of explainable ML for construction productivity forecasting.