



Moment balanced machine: a new supervised inference engine for on-site construction productivity prediction

Min-Yuan Cheng¹ · Riqi Radian Khasani¹

Accepted: 23 March 2024 / Published online: 22 April 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Predicting construction productivity is challenging because of the complexity involved in the construction process and the variability in factors that regularly affect these projects. Machine learning models have the potential to improve the accuracy of construction productivity predictions. This study introduces a model for generating accurate predictions of construction productivity using an AI-based inference engine called Moment Balanced Machine (MBM). Instead of identifying the hyperplane of SVM, MBM considers moments to determine the optimal moment hyperplane. MBM balances the moments, which are the product of force and distance, with force representing the weight assigned to a datapoint and distance indicating its position relative to the moment hyperplane. To obtain the weights for each datapoint within the MBM framework, Backpropagation Neural Network (BPNN) is employed. Moreover, the performance of MBM is benchmarked against five other machine learning models, including SVM, BPNN, K-Nearest Neighbor (KNN), Decision Tree (DT), and Linear Regression (LR). According to the results of the 10-fold cross-validation, MBM consistently outperformed the other models across five evaluation metrics, including RMSE (0.068), MAE (0.054), MAPE (3.42%), R (0.982), and R^2 (0.965). The comprehensive assessment, summarized by the Reference Index (RI), indicates that MBM achieved the highest RI score of 1.000, emphasizing its superior performance. Furthermore, MBM exhibits robustness against data imperfections, including incomplete and noisy datasets. Given these findings, the proposed model could serve as an advanced machine learning decision-support system that improves the prediction accuracy of construction productivity. This reinforces the data-driven approach for improving the efficiency of construction projects.

Keywords Machine learning · Construction productivity · Decision support · Moment hyperplane · Neural network

1 Introduction

Productivity, a ratio of outputs to inputs, is often used to compare the actual man-hours actually expended against the number of man-hours allocated to produce a final product. Productivity is a key factor affecting the success or failure of construction projects. Because the labor-intensive construction industry is a major contributor to gross domestic product in most countries, labor productivity in this industry can impact national economic performance significantly [1]. However, as productivity is influenced by

various factors, some of which can be managed and others which cannot, making accurate assessments of labor productivity is a complex task [2]. Nevertheless, maximizing productivity is crucial to ensuring efficiency and sustainability in the construction industry due to its impact on the final project cost, project implementation schedule, contractor competitiveness, and other project aspects [3]. Because labor expenses, equipment, and overhead costs may be minimized in construction projects completed on time or ahead of schedule, higher productivity results in more cost-efficient projects [4]. Higher productivity enables projects to be completed on time and within budget by minimizing delays and rework [5]. Furthermore, regularly completing projects on time and within budget ensures that client expectations are met and promotes the contractor's reputation in the industry. Contractors known to complete projects efficiently and perform high-quality

✉ Min-Yuan Cheng
myc@mail.ntust.edu.tw

Riqi Radian Khasani
riqi@live.undip.ac.id

¹ National Taiwan University of Science and Technology,
Taipei, Taiwan

work earn a competitive advantage in terms of acquiring new contracts and potentially obtaining more projects [6].

Construction projects face a myriad of risks related to site conditions, weather, budget overruns, scheduling, design changes, material availability, stakeholder coordination, labor issues, and other factors that, while difficult to predict and control, can dramatically affect construction productivity [7]. Moreover, tight budgets and project schedules increase contractor reliance on productivity to help complete tasks using the resources allocated and on time. Construction companies must actively manage and mitigate these risks using effective project planning and management strategies. Construction management is the process used to handle a series of interrelated tasks during a specific period of time and within certain limitations. Identifying and measuring uncertainties early on can facilitate the avoidance / amelioration of potential problems and increase the chance of completing construction projects on time and within budget [8]. Balancing cost and time constraints while maintaining productivity requires effective project management and the appropriate allocation of resources. Inadequately detailed or incomplete planning can result in the need of changes, rework, and delays during the construction phase [9], while inadequate coordination among architects, engineers, and contractors can result in conflicts, design errors, and poor productivity. Managing this complexity requires careful planning, coordination, and collaboration among stakeholders [10].

On-site construction productivity is influenced by various factors affecting the efficiency and effectiveness of construction activities such as manpower skill level, equipment and tools, site layout, and external factors such as weather, labor trouble, and inflation [11]. Access to sufficiently skilled and experienced manpower is crucial to achieving productivity goals, as well-trained workers perform tasks more efficiently, make fewer errors, and minimize the need for rework. Also, access to adequate supplies of properly functioning equipment and tools is crucial to productivity, as well-maintained equipment and tools promote the timely completion of tasks and reduce downtime [12]. In addition, external factors such as inclement weather (e.g., snowstorms, heavy rains, extreme temperatures) can delay project progress and undercut productivity [13].

There is currently strong research interest in implementing various artificial intelligence (AI) methods in the domains of construction engineering and management to use digital evolution to improve project performance and profitability. AI techniques can help construction projects operate more smoothly and efficiently [14]. Moreover, machine learning has been proposed as a potential solution to the abovementioned problems involved in predicting productivity in construction projects [15]. Machine learning algorithms are able to analyze historical project data and

generate predictive models for forecasting project schedules, costs, resource requirements, and construction productivity. These predictions can assist project managers with planning and scheduling, allowing proactive decision making to avoid delays and optimize resource allocation [16]. Machine learning algorithms must be trained using relevant and representative data to ensure the accuracy of predictions and insights. In addition, human expertise and domain knowledge remain important in interpreting and contextualizing the results generated by machine learning models. However, despite the potential of machine learning, the adoption of AI techniques in the construction industry lags behind other industries [17]. Moreover, their application in predicting productivity remains limited, and challenges such as data availability and model complexity remain to be overcome. Nevertheless, developing accurate machine learning models can help improve construction productivity predictions and enhance project planning in the construction industry.

Machine learning techniques have been successfully applied to predict productivity in construction projects. Momade et al. (2022) employed support vector machine (SVM) and Random Forest (RF) to predict construction labor productivity in residential projects, with the results showing that SVM outperformed RF [18]. In addition, Nasirzadeh et al. (2020) used an artificial neural network (ANN) to predict labor productivity for formwork installation in a reinforced concrete building [19]; Gohary et al. (2017) used ANN to predict labor productivity for formwork and steel installation work related to reinforced concrete foundations in residential and commercial buildings [20]; and Sadatnya et al. (2023) used logistic regression, k-nearest neighbors (KNN), decision tree, multilayer perceptron (MLP), SVM, and ensemble methods to estimate work-crew productivity [21]. Based on a literature review, ANN and SVM have been the machine learning models most commonly used in productivity prediction.

The SVM algorithm offers several significant advantages that make it an excellent choice for classification and regression tasks. The primary advantage of SVM is its ability to handle high-dimensional data effectively using kernel functions. SVM produces a well-generalized model that is resistant to overfitting. However, because SVM treats all datapoints equally during model training, in situations where certain datapoints are more reliable or relevant than others, the performance of this algorithm may be suboptimal due to possible reliance on less-reliable data. The challenge to overcome is the limitation of existing algorithms in handling weighted data points, particularly in SVM. The equal treatment of data points can lead to less accurate predictions.

This study introduces a new AI-based inference engine called Moment Balanced Machine (MBM) that aims to enhance the accuracy of productivity predictions in

construction projects. The MBM model extends the SVM concept to consider the moments in determining the optimal moment hyperplane, and backpropagation neural network (BPNN) is employed to obtain the weight of each case used in the MBM model. Furthermore, MBM is applied to predict on-site construction productivity using real-world data to validate the applicability of the MBM model. This study contributes to improving the accuracy and effectiveness of decision making in construction project management by introducing a new AI-based inference engine for predicting construction productivity based on accumulated domain knowledge and experience.

2 Literature review

2.1 Construction productivity and influencing factors

Productivity in the context of construction projects refers to the efficiency and effectiveness of project execution as measured using the ratio of outputs (such as work completed, units installed, and tasks accomplished) to inputs (such as hours of labor, equipment usage, and material consumption) [22]. Productivity measures how well a construction project or contractor has achieved the results and objectives of the project owner given the timeframe and resources allocated. Construction productivity encompasses several factors, including the speed and accuracy of construction activities, resource utilization, project schedule adherence, workmanship quality, equipment availability, technology adoption, and the ability to complete a project on time and within budget [23]. Efforts to improve construction productivity aim to increase project efficiency, reduce related costs, minimize rework and delays, improve safety, and maximize customer satisfaction. Continuously enhancing productivity can help the construction industry realize higher levels of efficiency, profitability, and project success [2].

On-site construction activities refer to tasks performed at construction sites and include excavation work, assembly and installation, finishing work, formwork setup, and other related activities. These tasks impact the progress and completion of construction projects directly [24]. Efficiency and productivity in this activity category is affected by many factors, including resource allocation, coordination, and communications within the construction team, among others. Therefore, optimizing the use of manpower, equipment, and materials is crucial to minimizing waste, rework, and delays [25]. On-site construction productivity is measured by comparing the volume of work completed to the quantity of resources used in a given period. This measure is valuable for project managers, allowing them to assess project performance and further refine their decision-making process [26].

Improving productivity involves using strategies such as task sequencing optimization, ensuring the availability of necessary resources, and adopting efficient construction methods and results in benefits such as faster project completion, cost savings, and improved project quality.

Historical data play an important role in the prediction of construction productivity using machine learning, allowing machine learning models to identify relevant productivity-related patterns and trends. Collected historical data must cover project duration, project size, number of workers employed, equipment used, weather conditions, and other variables that may affect productivity [27]. Machine learning models learn from patterns and relationships in the historical data and their respective impacts on productivity to generate productivity predictions for future projects. As new construction projects are completed and more historical data become available, these machine learning models may be continuously updated and improved. By incorporating the latest data into the training process, these models can adapt to changing industry trends and dynamics, thereby improving predictive capabilities over time and supporting better decision making related to project planning and resource allocation [28].

Amount of labor input, activity type, construction method, location, and external factors are known to significantly influence on-site construction productivity [29]. Level of labor experience and the availability of skilled workers significantly affect productivity, as trained and experienced workers can perform tasks more efficiently [30]. Also, as adverse weather conditions can significantly affect on-site productivity, weather contingency planning and implementing relevant preventive actions are crucial to minimizing construction schedule delays [31]. Twenty articles in the literature (see Table 1) were reviewed in this study to identify the important factors of influence on on-site construction productivity. This review identified an initial list of 29 factors. Factors mentioned in four articles or fewer were subsequently eliminated, leaving a final list of 12 factors that was used in the initial stages of this study.

2.2 Backpropagation neural network

BPNN, a type of ANN, allows the neural network to learn and adjust its weights and biases to make more-accurate predictions by minimizing the difference between predicted and actual outputs. The BPNN algorithm includes forward propagation, error calculations, backward propagation, and weight updates. In forward propagation, after input values are inserted into the neural network and an output is generated, the differences between predicted and actual outputs are compared. In backward propagation, the error is propagated backward through the network to update the weights and biases using a gradient descent process to minimize

errors. The weights are updated continuously until the error is below the acceptable threshold. These steps are repeated either for a set number of iterations or until the desired convergence level has been achieved. Thus, BPNN involves multiple forward and backward propagations through the network that adjust weights and biases iteratively. BPNN is widely used and has been successfully implemented in the civil engineering field on tasks such as predicting rock bursts in tunnels [47], tunnel deformation [48], displacement in concrete-face rockfill dams [49], RC beam flexural capacity [50], strain in wind turbine blades [51], the rate of corrosion in carbon steel [52], the properties of recycled aggregate permeable concrete [53], and labor productivity [36]. The structure of one-hidden-layer BPNNs, which comprise one input layer, one hidden layer, and one output layer, is shown in Fig. 1.

2.3 Support vector machine

SVMs are supervised learning algorithms that can be used for both classification and regression tasks. The main idea behind the SVM technique is to obtain a hyperplane with optimal margins that maximizes the margin between two classes [54]. In regression tasks, SVM aims to find the regression function that best fits the training data while allowing for a certain amount of error within an ϵ -insensitive tube around the predicted values. To handle non-linear problems, a kernel function transforms the training pattern (x_i) through a non-linear transformation ($\varphi: N \rightarrow F$) into a higher dimensional feature space. As shown in Eq. (1), the fitting function is estimated in the higher-dimensional feature space, where (\cdot) denotes the dot product in the feature space. In this transformed space, SVM aims to find a regression

function that best approximates the training data within a specified error tolerance. This function is represented as a hyperplane defined by support vectors. The SVM regression model may be written as a convex optimization problem by requiring Eq. (2). SVM regression involves tuning parameters such as regularization parameter (C), which controls the trade-off between the complexity of $f(x)$ and the amount of deviation larger than the error that can be tolerated as well as the kernel parameters that determine the complexity and flexibility of the regression function. Slack (ξ) measures the distance to points outside the margin, and may be controlled by adjusting the regularization parameter (C). Kernel functions include linear, polynomial, and radial basis functions.

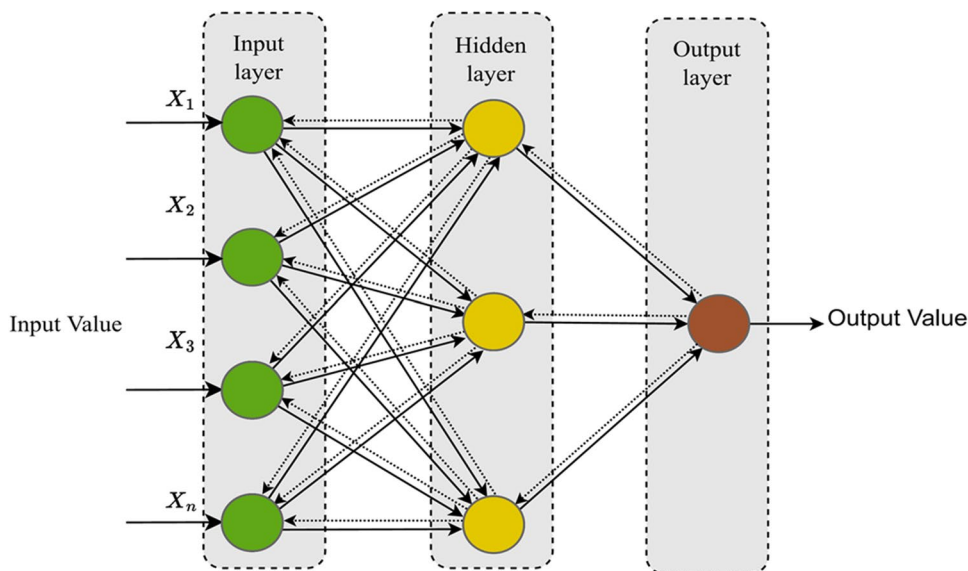
$$f(x) = w \cdot \varphi(x) + b$$

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{1}$$

$$\text{Subject to } \begin{cases} y_i - (w \cdot \varphi(x_i) + b) \leq \epsilon + \xi_i \\ (w \cdot \varphi(x_i) + b) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{2}$$

Expanding on the foundational principles of SVM, previous studies have developed the weighted SVM approach. Zhang et al. (2011) introduced an approach that integrates features and sample weighting within the SVM framework, a method that enhances the flexibility and adaptability of the model to complex datasets [55]. Lapin et al. (2014) introduced SVM+, which utilizes informative features during training, emphasizing the adaptability of SVM in utilizing additional data. It presents an approach to weight assignment in SVM, where the weights are adjusted based on the extra features during the training phase [56]. Moslemnejad &

Fig. 1 Single hidden layer BPNN structure



Hamidzadeh (2021) and Fan et al. (2020) further improved SVM performance by incorporating fuzzy logic into the weighting mechanism [57, 58]. Luo et al. (2020) demonstrated the practical application of weighted SVM ensemble predictor based on AdaBoost to optimize the blast furnace ironmaking process [59]. Scikit-learn also supports SVM with weighted samples, offering extensive documentation and functionality [60]. This study advances the weighted SVM by introducing moment-based weighting to determine the optimal moment hyperplane and employs BPNN to obtain the weight of each case in the MBM model. This unique contribution highlights the potential of the MBM model in predictive modeling.

3 Moment balanced machine for predicting construction productivity

3.1 Model development

The concept of moment balance, which refers to the achievement of balance in structural elements such as beams and columns, is necessary in structural design and analysis. Applying a load to a beam causes bending, which produces a bending moment along that beam. To

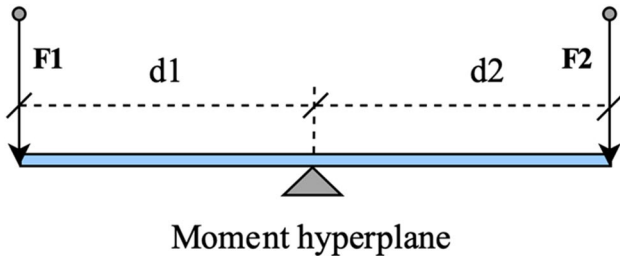


Fig. 2 Moment balance on a simple beam

create balanced moments in the beam, the moment diagram should show a symmetrical distribution where the amount of moment caused by the load on one side of the beam equals the amount of moment on the other side. A simple beam with a given load that must have equal and opposite moments to maintain balance is presented as an example in Fig. 2. These moments, which are a product of the force and the distance of that force from the axis of rotation, may be expressed using Eq. (3).

$$m = \sum(\text{Force} * \text{Distance}) = 0$$

$$m = F_1d_1 + F_2d_2 + \dots + F_i d_i = 0 \tag{3}$$

The concept of moment balance in practice may be observed in the construction of balanced cantilever bridges, especially those with long spans such as segmental and cable-stay bridges. This method involves gradually building a bridge in balanced segments on both sides, supported by pillars in the middle. As segments are added, the bridge gradually extends out on both sides symmetrically from the pillar. This process is repeated until the bridge span is complete. An illustration of the balanced cantilever method and the moment balance concept is shown in Fig. 3.

Unbalanced moments have the potential to lead to structural failures such as bending, torsional, buckling, and overturning. Unbalanced moments can magnify the effects of bending stresses and lead to bending failure. This occurs when the yield strength of the material is exceeded due to excessive loading. In some situations, the structural components subjected to unbalanced moments may experience torsion. Structural failure results when the torsional stress in the material exceeds its yield strength. Unbalanced moments can generate lateral forces that result in buckling, which is manifested by the presence of lateral deflections in structural members. Moreover, in extreme cases, unbalanced moments can result in the overturning of the entire structure.

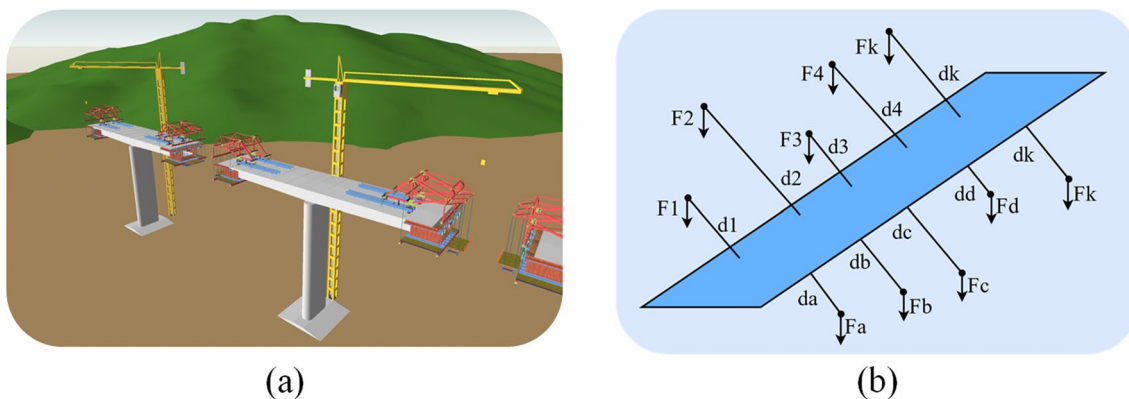


Fig. 3 a Balanced cantilever method; b Concept of balance of moments

Therefore, thorough design and analysis processes are crucial to ensuring the ability of structures to resist unbalanced moments and prevent the danger of structural collapse.

Adapting the concept of moment balance, MBM employs weights to handle balanced moments, the weight multiplied by the distance of the moment from the point or axis of rotation, which reflects the principle of the moment equilibrium. This principle is derived from physics and engineering, particularly from the study of static mechanics. For an object to be in equilibrium, the sum of the moments around any point must be zero. Force, defined as the magnitude of the force acting on an object, equals the weight of that object. Distance refers to the distance from the point to the line of action of the force. The sum of the moments includes all of the forces acting upon the system. For systems in equilibrium, moments working to rotate the system in one direction are perfectly offset (balanced) by other moments working to rotate it in the opposite direction. This concept is analogous to the optimization problem in MBM, where force can be expressed as the weight (F_i) given to a datapoint and distance can be expressed as a slack variable (ξ or ξ^*) that represents the distance of that datapoint from the moment hyperplane. In this model, support vectors act as the points required to balance the model, with any imbalance resulting in a shift of the balance point (i.e., the moment hyperplane).

MBM further introduces the idea of assigning different weights to different datapoints, which can affect the identification of support vectors and the determination of the moment hyperplane. These weights can represent the relative importance of each datapoint in the regression analysis and significantly affect the position and orientation of the moment hyperplane. Similar to the principle of moment equilibrium, MBM balances the product of force and distance (moment). Removing support vectors from this system

is similar to removing forces from a physical system, with the result leading to a shift in balance and altering the position of the moment hyperplane. The balance in the MBM model also comes from a trade-off between model complexity and tolerable errors. This balance, which is controlled by regularization parameters and individual weights for datapoints, determines the optimal position and orientation of the moment hyperplane.

The purpose of this research was to develop a new model for predicting construction productivity using a newly developed, AI-based MBM inference engine. In this model, BPNN is used to obtain the weight of each case, the concept of balance moment is used to find the optimal moment hyperplane, and MBM is utilized to characterize the relationship between the influencing factors of on-site productivity and the desired output (i.e., formwork installation productivity). The architecture of the MBM model is shown in Fig. 4.

Given a regression dataset associated with weights:

$$D = \{(x_1, y_1, F_1), (x_2, y_2, F_2), \dots, (x_i, y_i, F_i)\} \in \mathbb{R}^n \tag{4}$$

The MBM aims to find a moment hyperplane that provides the best fit to the training data, within a specified error tolerance, to achieve optimal generalization capabilities. Where $x_i \in \mathbb{R}^n$ is the input, $y_i \in \mathbb{R}^n$ is the output, and F_i is the weight for (x_i, y_i) . MBM solves and optimizes:

$$\begin{aligned} &\text{Minimize } \frac{1}{2}w \cdot w + C \sum_{i=1}^l m_i \\ &\text{Where : } m = F_i(\xi_i + \xi_i^*) \\ \text{Subject to } &\begin{cases} y_i - (w \cdot x_i + b) \leq \varepsilon + \xi_i \\ (w \cdot x_i + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{5}$$

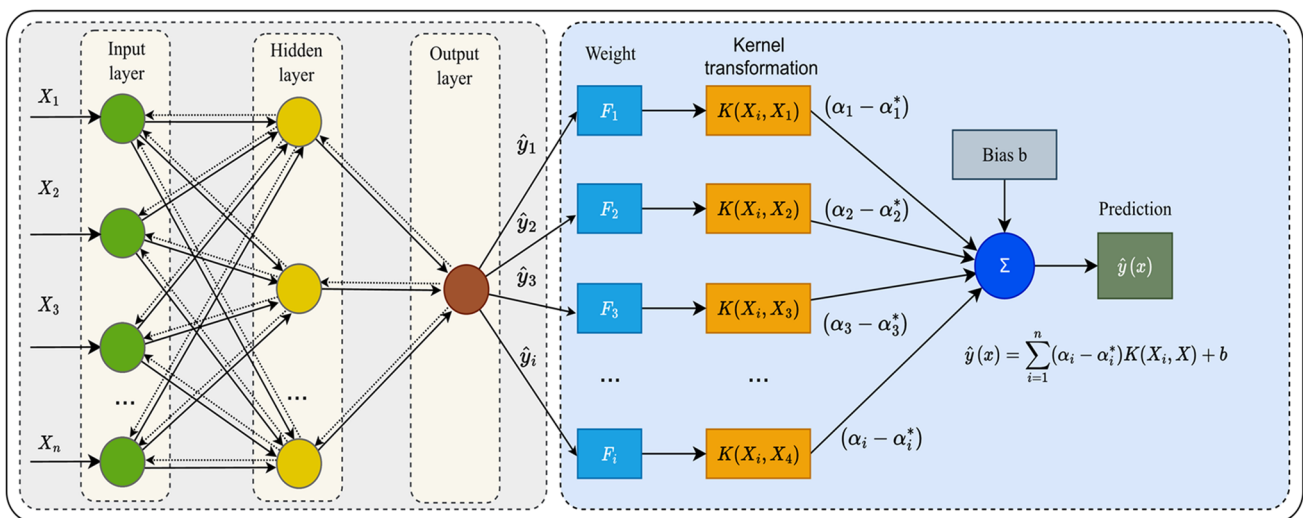


Fig. 4 MBM model architecture

C is a constant that modulates the trade-off between maximizing the margin and minimizing the slack variables; m is the moment; F_i is the weight of each case obtained using the BPNN algorithm; and ξ_i and ξ_i^* respectively denote the upper and lower training errors. Smaller F_i values reduce the effect of parameter ξ_i , indicating a lower importance for the corresponding point x_i . The weight F_i is assigned to datapoint x_i in Eq. (5). The weight of each case (F_i) can be calculated using Eq. (6). Where Y_i and \hat{Y}_i respectively denote actual productivity and predicted productivity. The inverse relationship between weight and forecast error is based on the principle that data points with lower forecast errors are more accurate and therefore should have a higher influence on the moment hyperplane. More reliable data points are assigned higher weights, whereas less reliable data points are assigned lower weights.

$$\text{Weight}(F_i) = \frac{1}{\text{Forecast error}} = \frac{1}{\left(\frac{|Y_i - \hat{Y}_i|}{Y_i}\right)} \tag{6}$$

To solve the optimization problems, the Lagrangian function of the objective function and the corresponding constraints are formulated and the dual problem is derived. The dual variables in Eq. (7) must satisfy the constraints in Eq. (8). Where L is the Lagrangian and $\alpha, \alpha^*, \beta, \beta^*$ are Lagrange multipliers.

$$L := \frac{1}{2}w \cdot w + C \sum_{i=1}^l F_i(\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i(\epsilon + \xi_i - y_i + (w \cdot x_i + b)) - \sum_{i=1}^l \alpha_i^*(\epsilon + \xi_i^* - (w \cdot x_i + b) + y_i) - \sum_{i=1}^l \beta_i \xi_i - \sum_{i=1}^l \beta_i^* \xi_i^* \tag{7}$$

$$\alpha, \alpha^*, \beta, \beta^* \geq 0 \tag{8}$$

Solving the minimization problem involves taking the partial derivative of L with respect to the primal variable (w, b, ξ_i, ξ_i^*), as shown in Eqs. (9)–(12).

$$\frac{\partial L(\alpha, \alpha^*, \beta, \beta^*, w, b, \xi, \xi^*)}{\partial w} = 0 \rightarrow w = \sum_{i=1}^l (\alpha_i - \alpha_i^*)x_i \tag{9}$$

$$\frac{\partial L(\alpha, \alpha^*, \beta, \beta^*, w, b, \xi, \xi^*)}{\partial b} = 0 \rightarrow \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \tag{10}$$

$$\frac{\partial L(\alpha, \alpha^*, \beta, \beta^*, w, b, \xi, \xi^*)}{\partial \xi} = 0 \rightarrow F_i C - \alpha_i - \beta_i = 0 \tag{11}$$

$$\frac{\partial L(\alpha, \alpha^*, \beta, \beta^*, w, b, \xi, \xi^*)}{\partial \xi^*} = 0 \rightarrow F_i C - \alpha_i^* - \beta_i^* = 0 \tag{12}$$

Substituting Eqs. (9)–(12) to Eq. (7) yields the dual optimization problem:

$$\text{Maximize } \begin{cases} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i \cdot x_j \\ + \sum_{i=1}^l y_i(\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) \end{cases} \tag{13}$$

$$\text{Subject to } \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq F_i C \\ K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \end{cases} \tag{14}$$

Regularization (c) and gamma (γ) are the main parameters for tuning the MBM model, while regularization parameter (c) determines the tolerance of the model for errors in the training data. A smaller c value increases the margin but allows more training errors, increasing the risk of mismatch. Conversely, a larger c value reduces the margin but penalizes more training errors. The kernel function is used to transform the data into a higher-dimensional feature space. The choice of the kernel affects the model’s ability to capture complex patterns in the data. Commonly used kernel functions include linear, polynomial, and radial basis functions. In this study, a radial basis function (RBF) kernel was used with the gamma (γ) parameter, which allows MBM to capture the non-linear relationships between input features and targets outputs and allows the modeling of complex patterns that may not be linearly separable in the original feature space. This flexibility allows MBM to handle multiple regression problems using non-linear relationships. The RBF kernel formula is given by Eq. (14). Figure 5 illustrates the procedure used to design the pseudocode, which is used to generate MBM models. The MBM algorithm begins by determining the required data and parameters, including the training and testing data, regularization parameter (C), kernel parameter (γ), and weights (F_i) derived from BPNN. The next step is data preprocessing, including normalization to ensure the features are on the same scale. Subsequently, the algorithm used BPNN to provide initial predictions. The BPNN served as the initial model to generate prediction errors, which were then used to calculate the weights of each data point (F_i). The next stage focuses on training the MBM model. The problem is formulated as a Quadratic Programming (QP) optimization problem at this stage. This involves solving the QP problem to determine the optimal moment hyperplane, computed using Lagrange multipliers. The bias was then calculated. Once the model was trained, it was evaluated using a series of performance metrics on the testing data. If the performance of the model was not acceptable, the algorithm was iterated for further adjustments. This

Fig. 5 Pseudocode for the MBM algorithm

Pseudocode of Moment Balanced Machine algorithm	
Input:	Training Data Testing Data Regularization parameter (C) Kernel parameter (γ) Weights (F_i) for MBM (derived from BPNN)
Process:	Data Preprocessing Train BPNN for Initial Predictions Compute Weights (F_i) based on BPNN Forecast Errors and populate F_i Hyperparameter Selection (Random Search, Grid Search) Train MBM Model: Optimization problem, solve optimization problem use quadratic programming, calculate the bias Model Evaluation: prediction, evaluate performance metric Model Tuning: Update C and γ if needed, retrain the model, Choose optimal C and γ
Output:	Prediction results Optimal parameters Performance metrics

consists of updating the regularization parameter (C) and kernel parameter (γ), and then retraining the model. The tuning process is iterated until the optimal parameter values are obtained, thereby optimizing the performance of the model. Finally, the algorithm outputs prediction results and identifies the optimal parameters. Performance metrics provide a quantitative assessment of the model’s effectiveness, offering insights into how it might be performed on training and testing data. Overall, the MBM shows a unique combination of BPNN and SVM, taking advantage of a weighted approach to combine both strengths.

3.2 Model adaptation

The general framework of this study is illustrated in Fig. 6, which begins with the construction productivity dataset (Step 1). The initial step outlines the systematic literature review and identification of variables, which are used to form the basis of construction productivity database. Then, the input and output variables are identified as preliminary selected features (Step 2), where the inputs for these machine learning models are factors that influence construction productivity. Next, the data pre-processing stage is performed

by processing the raw data before being used in analysis or training models (Step 3). This step includes data normalization and correlation analysis for feature selection, which are essential to ensure data quality and relevance. The processed data are subsequently input into several machine learning models to predict construction productivity (Step 4). For MBM, the hyperparameters are tuned using random search and grid search techniques. Finally, the predicted productivity results are evaluated using performance evaluation metrics, also the sensitivity analysis is applied for analyzing incomplete and noisy data (Step 5). All of the primary processes involved in adapting the MBM model are illustrated in the flowchart in Fig. 7.

Relevant data related to labor productivity were collected and organized for analysis. The collected data were examined to identify the key features and variables that significantly impact construction productivity. The dataset required normalization prior to being used as input variables in the machine learning model. Normalization is a data preprocessing technique used to scale numerical features to a standardized range between 0 and 1, allowing differently valued variables to be normalized to the same scale, eliminating biases that may arise from differences in variable units and

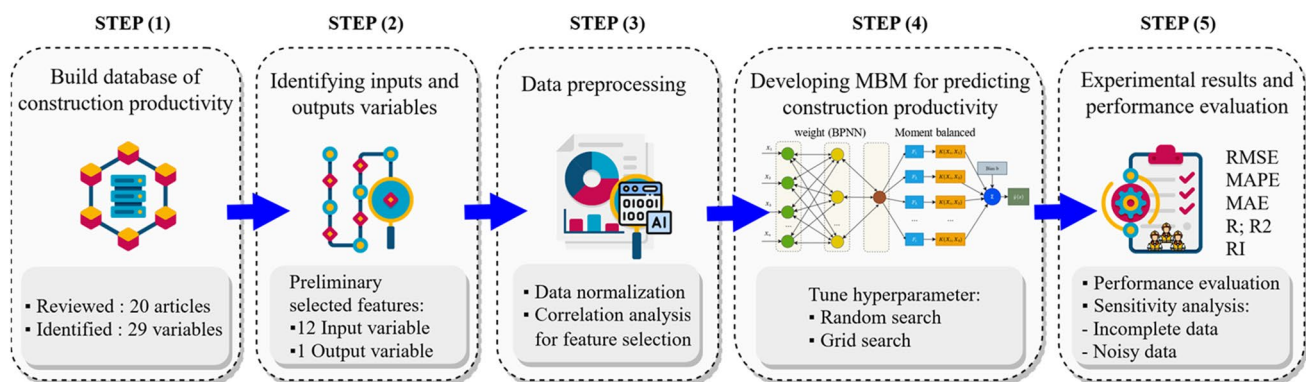


Fig. 6 Proposed framework for developing the construction productivity model

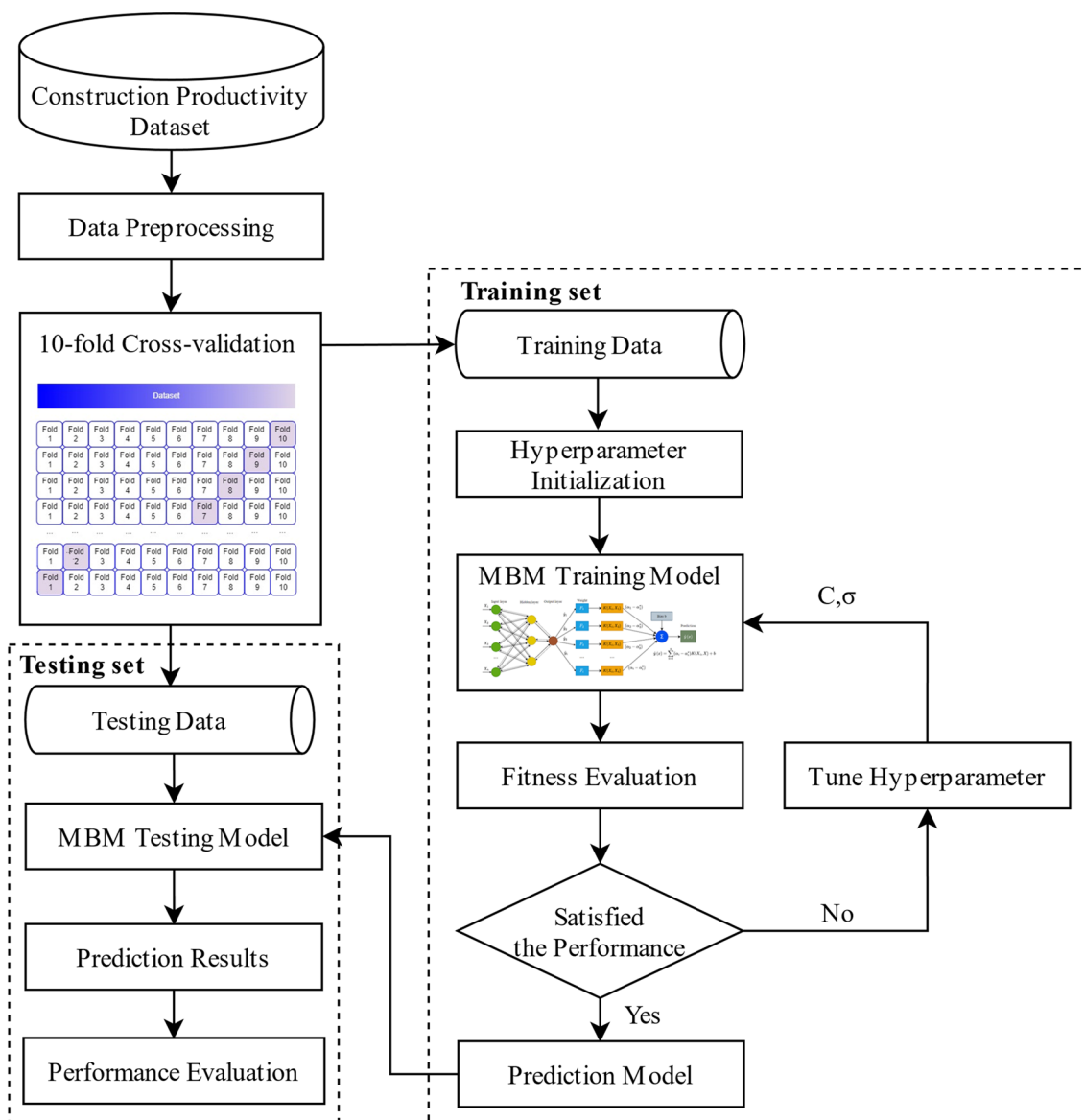


Fig. 7 MBM model adaptation

ranges. The formula used for normalization in this study is shown in Eq. (15)

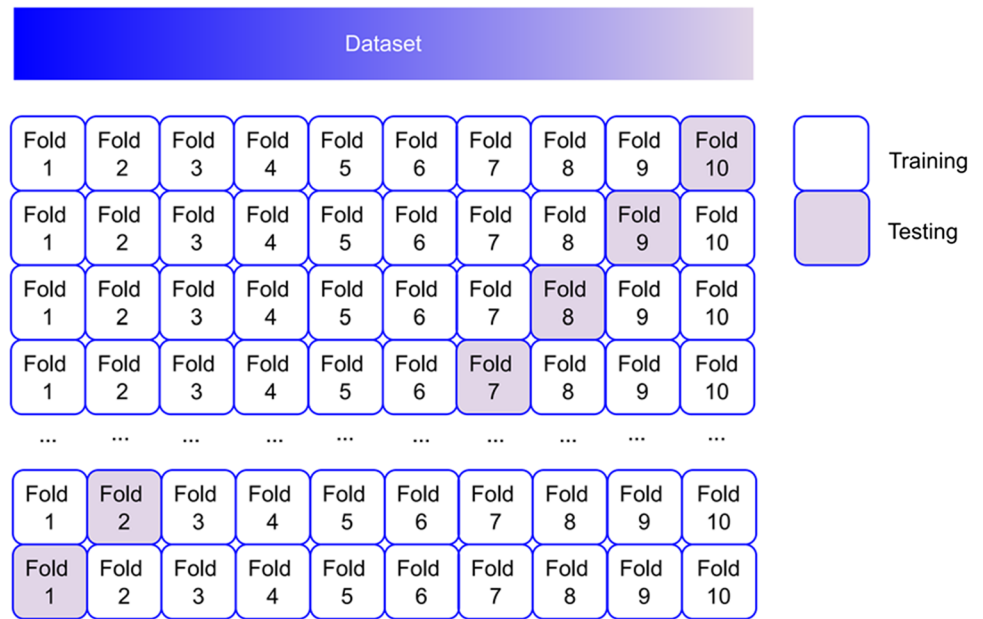
$$X_{norm} = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \tag{15}$$

A 10-fold cross-validation technique was used in this study to partition the dataset into 10 equal folds to ensure model generalization. The model was trained nine times and tested on the remaining fold. This process was repeated ten times with different combinations, and the results were averaged to assess overall model performance. The illustration of the 10-fold cross-validation technique is shown in Fig. 8.

The training data used to train the machine learning model included input and output variables. Machine learning models learn the underlying patterns and relationships between the input and output variables, enabling the model to generalize and accurately predict the output from the testing data. The testing dataset is a separate subset of the case dataset that is not used to train the model but rather used to evaluate the performance of the trained predictive model. The purpose of the testing data was to assess how well the model generalizes to new, unseen cases. Evaluating the model on independent data helps estimate its performance and provides an indication of the generalizability of its results.

The MBM algorithm was trained on the training data. This algorithm learns from historical data to build relationships

Fig. 8 10-fold cross validation



between selected features and productivity outcomes and subsequently to make predictions based on learned patterns. The hyperparameters of MBM were tuned using random search and grid search techniques. In the fitness evaluation, the performance and quality of the MBM training model was evaluated using the root mean square error (RMSE) performance metric. After the desired performance results were obtained, they were used in the predictive model. The predictive model combines MBM with a set of hyperparameters found through a trial-and-error process to make predictions based on the testing dataset. A construction productivity output value may be predicted by inputting features affecting on-site construction productivity. The prediction results are the estimated values generated by the trained predictive model using the testing dataset, and represent the predicted values of the output variable or target based on the input features provided.

Performance was assessed to determine the accuracy and efficacy of the predictive model based on the accuracy of predictions generated using the testing dataset and of estimations of objective variables. The proposed and comparison models were evaluated using performance evaluation metrics including RMSE, mean average percentage error (MAPE), mean absolute error (MAE), R, and R-squared (R²), all of which measure the difference between the actual and predicted values of the objective variables. RMSE is the square root of the average squared difference between predicted and actual values, MAPE is the mean of all absolute percentage errors between predicted and actual values, and MAE is the average absolute error between predicted and actual values. Coefficient of correlation was used to calculate the correlation between two continuous variables and coefficient of determination (R²) was used to identify the

strength of the model. These measures estimate the variance in the predictions explained by the dataset. Furthermore, reference index (RI), which combines the overall results of all measurements by assigning equal weights to each measure, was used as a deterministic measure to summarize all measurements. Normalized values for each measure are required to calculate the RI, the value of which ranges from 0 (worst) to 1 (best). RI was calculated as the average of the normalized values of five performance metrics, namely RMSE, MAE, MAPE, R and R². The RI reaches a value of 1 only if all performance metrics are the best, which is the optimal performance achieved. If any performance metric is less than 1 after normalization, the RI will be less than

Table 2 Performance evaluation

Measure	Formula
RMSE	$\sqrt{\frac{1}{n} \sum_i (y_i - f_i)^2}$
MAPE	$\frac{100}{n} \sum_i \frac{(y_i - f_i)}{y_i}$
MAE	$\frac{1}{n} \sum_i y_i - f_i $
R	$\frac{n \sum_i y_i f_i - (\sum_i y_i)(\sum_i f_i)}{\sqrt{n(\sum_i y_i^2) - (\sum_i y_i)^2} \sqrt{n(\sum_i f_i^2) - (\sum_i f_i)^2}}$
R ²	$1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}$
RI	$\frac{R_{norm} + R_{norm}^2 + (1 - RMSE_{norm}) + (1 - MAE_{norm}) + (1 - MAPE_{norm})}{5}$

1, indicating the proportional average of the performance metrics. The formulas for each performance measurement used in this study are presented in Table 2.

4 Model application

4.1 Data collection and processing

The data employed for this study adopted from Cheng [27], who integrated datasets from Wang [33] and Khan [39]. These original datasets captured the productivity of formwork installation across two construction projects in the Engineering, Computer Science, and Visual Arts Complex at Concordia University, Canada for a period of 30 months. The first project was a 16-story building designed using flat-slab reinforced concrete and shear walls. The second project was a 17-story building designed using reinforced concrete and covering a total floor area of 68,000 square meters. Observations were made using a work sampling method to collect productivity data. Other data were obtained from daily reports and a website providing daily weather data. After building the construction productivity dataset from several sources, the factors influencing on-site productivity were identified and used as input data. The input data included the four main categories of workers, projects, tasks, and weather, and the dataset consisted of 12 input variables and one output variable, as illustrated in Fig. 9.

The correlation coefficient method was employed to select significant correlating parameters as factors for constructing the on-site construction productivity model. Correlations were examined using Pearson's correlation coefficient, Kendall's tau-b, and Spearman's rho, all of which were performed on MATLAB statistics and machine learning toolbox. Figure 10 shows the correlation matrix of Pearson's correlation. In this study, factors showing significant correlations in all three analyses (Pearson's correlation coefficient, Kendall's tau-b, and Spearman's rho) were selected as factors of influence on-site construction productivity. The correlation analysis results for the 12 potentially influential factors on the 220 cases in the case dataset are shown in Table 3. Nine of the twelve factors passed the test and the remaining three variables were rejected as input variables. Wind velocity satisfied the correlation test requirements of Pearson's test only, with the results of Kendall's tau-b and Spearman's rho tests falling short of significance ($\rho > 0.05$). In addition, moisture and crew percentage failed to demonstrate significant correlations ($\rho > 0.05$) in any of the three tests. Therefore, wind velocity, moisture, and crew percentage were dropped as input factors from the model. The other nine factors all showed strong correlations ($\rho < 0.05$) in all three tests and

were retained as input factors in the productivity inference model as shown in Table 4.

The model input variables, all of which were identified as significantly influencing on-site construction labor productivity, included: thermal conditions (F1), rainfall (F2), labor size (F3), activity type (F4), floor level (F5), working method (F6), primary work (F7), secondary work (F8), and non-value added work (F9). The model output variable was the daily productivity of formwork installation operations. The dataset used to build the predictive model consisted of 220 cases, as shown in Table 5.

Descriptive statistics were used to examine data distribution (i.e., minimum, maximum, mean, and standard deviation). As shown in Fig. 11, each box represents a factor of influence on onsite productivity. Mean values are given next to the circle at the center of each box, maximum and minimum values are given next to the beginning and end whiskers, and mean \pm standard deviation values are displayed at the beginning and end of each box. For example, thermal condition (F1) earned maximum and minimum values of 25 and -26 , respectively, a mean of 4.04, and a standard deviation of ± 12.05 . Box plots provide overall statistical information on the dataset. In this study, the data were normalized to [0 1] to handle the large range and different units encompassed by the variables and to prevent certain variables from dominating the model during the training process. Also, ten-fold cross-validation method was used to ensure model generalizability, with all dataset examples used in both training and testing as previously described.

4.2 Model testing

4.2.1 Random search

This study employed a two-phase experimental design, including random search and grid search. Random Search technique is used for tuning the hyperparameters, where various combinations of hyperparameters are randomly selected for training the model. The purpose of this approach is to identify the combination of hyperparameters that maximizes the performance of the model. In the testing process, 10-fold cross-validation technique was used to validate model performance. Table 6 presents the results of 10-fold cross-validation for the two machine learning models, MBM and SVM, using different parameter settings. Random search was used to select the values for the hyperparameters C and γ for each model. With hyperparameters set at $C = 1$ and $\gamma = 1$, MBM demonstrated outperforms performance in terms of RMSE, MAE, and MAPE, with values of 0.083, 0.067, and 4.56%, respectively. Furthermore, while both models demonstrated high R and R^2 values, indicating a good fit with the data, MBM slightly outperformed SVM. This trend was also observed at other hyperparameter combinations, specifically

($C=2, \gamma=0.5$) and ($C=3, \gamma=0.1$), with MBM consistently achieving better results across all evaluation metrics. Lower RMSE and MAE values across all hyperparameter settings indicated that the MBM was more accurate. Similarly, in terms of prediction errors, the MBM is superior to the SVM based on its lower MAPE value. Through various settings and evaluation metrics, the MBM consistently outperformed the SVM.

4.2.2 Grid search

Grid Search is used to identify the optimal hyperparameters across all possible combinations within predetermined parameter ranges. The hyperparameters for the MBM model are regularization parameter (C) and kernel parameter (γ). The choices of C and γ affect the performance of the MBM model. The parameter values of C and γ were configured with four potential values [0.1, 1, 10, 100] for C and [0.1, 0.5, 1, 5] for γ , resulting in 16 different configuration combinations. By exploring 16 possible combinations, the grid search provided a systematic approach for determining the most optimal pair. All combinations were evaluated using training and testing data using 10 fold cross-validation method. The actual and predicted value results of MBM is presented in Table 7. The MBM model showing the differences in the five evaluation criteria between testing and

training phases to be acceptable and indicating the proposed predictive model is able to avoid overfitting. The MBM model also obtained a sufficiently small standard deviation for all criteria, indicating its performance is stable and not affected by random data partitioning. Overall, the performance of the MBM model was reliable and stable.

For the MBM model, the training performance was consistently robust across all folds, with average RMSE, MAE, and MAPE values of 0.068, 0.057, and 3.76%, respectively. The high R and R^2 values, averaging 0.981 and 0.962, respectively, indicate an excellent fit with the training data. In the testing phase, MBM's performance remained comparable to its training metrics, with average RMSE, MAE, and MAPE values of 0.068, 0.054, and 3.42%, respectively. The R and R^2 values are 0.982 and 0.965, respectively. The best hyperparameters, C and γ , varied across different folds, indicating that the model is sensitive to these settings, yet its performance remains strong. Overall, the results indicate that the MBM performed well across various metrics in the training and testing sets.

In the comparative analysis, both models demonstrated impressive training performance, with minor variations in the five evaluation metrics. Although the SVM exhibits slightly better RMSE and MAPE during the training phase, the marginally lower error rates do not necessarily indicate better generalization. In contrast, MBM

Fig. 9 Variables influencing on-site construction productivity

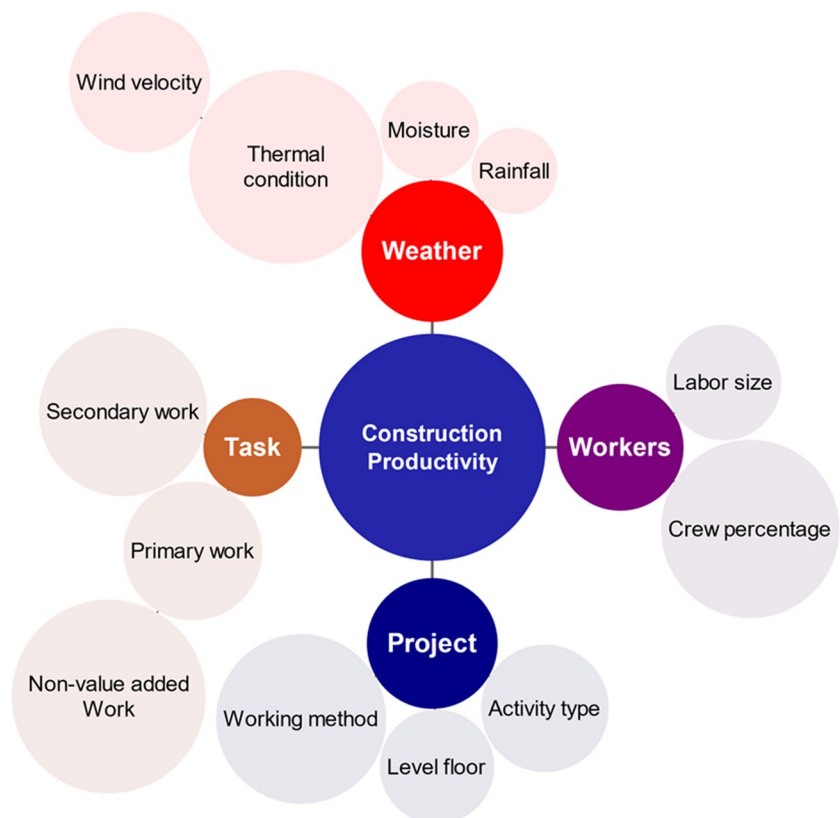
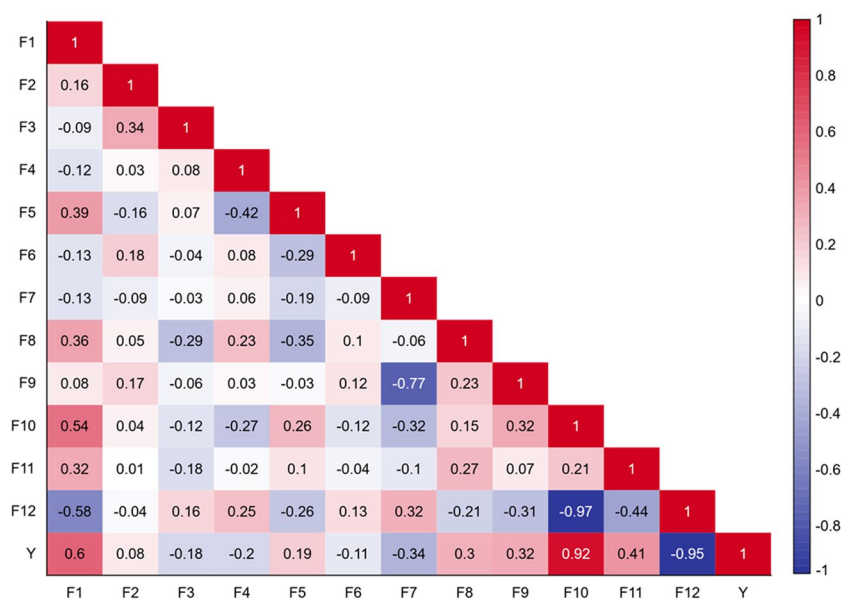


Fig. 10 The result of Pearson's correlation method



demonstrated more minor differences between training and testing errors, indicating robust performance on training data and effective generalization to unseen data. Specifically, MBM outperformed the SVM on several key metrics during the testing phase. MBM showed average RMSE, MAE, and MAPE values of 0.068, 0.054, and 3.42%, respectively, lower than the SVM values of 0.070, 0.058, and 3.74%, respectively. These results indicate that MBM possesses more significant potential for improved generalization.

Next, the predictive performance of the MBM model was compared against the four other machine learning models, such as backpropagation neural network (BPNN), decision tree (DT), linear regression (LR), and k-nearest neighbors (KNN). This comparative analysis aimed to evaluate the relative effectiveness of the developed model in predicting formwork installation productivity. SVM and BPNN are machine learning techniques widely used to predict construction productivity and DT, KNN, and LR have been widely used to solve regression problems [61–63]. All of the comparative models were run in MATLAB. MBM and SVM used the results of grid search values for the regularization parameter (c) and kernel parameter (γ) to produce comparative performance. The hyperparameters of BPNN, including number of hidden neurons, number of neuron layers, and iterations, as well as the hyperparameters of the DT, KNN,

and LR models were set manually using the default settings from MATLAB.

The five performance evaluations were used to measure the accuracy of the proposed predictive model. The low values obtained for RMSE, MAPE, and MAE combined with the high values obtained for R and R^2 indicate the performance of the proposed model in predicting formwork installation productivity. In addition, the normalized RI was used to obtain an overall performance comparison by combining the five performance metrics. The performance evaluations obtained using 10-fold cross-validation technique are shown in Table 8.

MBM obtained the lowest values for RMSE, MAPE, MAE and the highest values for R and R^2 (almost equal to 1). In the context of formwork installation productivity prediction, MAPE measures the percentage difference between predicted and actual productivity. For MBM, the respective MAPE values of 3.76% and 3.42% for training and testing indicate a small average percentage error in the dataset. The second-best performance was given by SVM, with respective training and testing MAPE values of 3.72% and 3.74%, followed by BPNN, LR, DT, and KNN, with MAPE values, with respective testing MAPE values of 4.59%, 5.17%, 5.45%, and 6.18%. In terms of RMSE, MBM achieved the lowest prediction-error averages during training and testing of 0.068 and 0.068, respectively. The second-lowest values were obtained SVM with 0.067 and 0.070, respectively,

Table 3 Results of correlation analysis

Code	Factor	Pearson	Kendall's tau-b	Spearman's rho
F1	Thermal condition	●	●	●
F2	Moisture			
F3	Rainfall	●	●	●
F4	Wind Velocity	●		
F5	Labor Size	●	●	●
F6	Crew Percentage			
F7	Activity Type	●	●	●
F8	Level Floor	●	●	●
F9	Working Method	●	●	●
F10	Primary Work	●	●	●
F11	Secondary Work	●	●	●
F12	Non-value added Work	●	●	●

● Significant correlation at the 0.05 level (two-tailed)

followed by BPNN, LR, DT, and KNN, which earned testing RMSE values that were, respectively, 0.088, 0.102, 0.108, and 0.122 higher than that of MBM.

In terms of MAE, MBM achieved values of 0.057 and 0.054, respectively, for the training and testing phases, indicating low average absolute errors between predicted and actual values. Despite its relatively good performance in predicting formwork installation productivity with a small RMSE testing value of 0.070, SVM earned an MAE testing value of 0.058, higher than that of MBM. Furthermore, the R and R² testing values of, respectively, 0.981 and 0.962 obtained by SVM were lower than those of MBM (i.e., 0.982 and 0.965, respectively), while BPNN, LR, DT, and KNN obtained MAE testing values that were, respectively, 0.071, 0.083, 0.084 and 0.093 higher than that of MBM. Of the compared models, MBM obtained the highest R and R² values in both training and testing phases, supporting that MBM is able to predict the output variable most accurately using the input variables. The linear correlation between actual and predicted productivity using the MBM model is shown in Fig. 12. These results demonstrate MBM closely captures the relationship between input and output parameters in formwork installation productivity.

The RI value was used in this study to perform an overall performance comparison that combined all of the five performance measures. MBM earned the highest RI value of 1.000, followed by SVM, BPNN, LR, DT, and KNN with RI values of 0.934, 0.615, 0.355, 0.229 and 0.000, respectively. This supports that MBM delivers the best overall performance of the compared models. The parallel diagram in Fig. 13 provides an easily interpretable, visual representation of the five error metrics, their correlations, and their combinations (RI) along with six machine learning models. Each row connects the error metrics (means) related to a specific machine learning model. Three of the five metrics (RMSE, MAPE, and MAE) performed better for smaller values, whereas the other two (R and R²) performed better for larger values. In the combined performance evaluation, RI performed better for larger values. Visually, the MBM outperformed the other models.

4.3 Sensitivity analysis phase

To further evaluate the performance of the MBM model, an additional test was conducted to assess the robustness and generalization capabilities of the model, such as incomplete data and noisy data conditions. A comparative analysis of these scenarios provides a better understanding of the strengths and limitations of each model. The SVM algorithm serves as a benchmark for evaluating the performance of the MBM model, which modifies the SVM concept by assigning different weights to each data point.

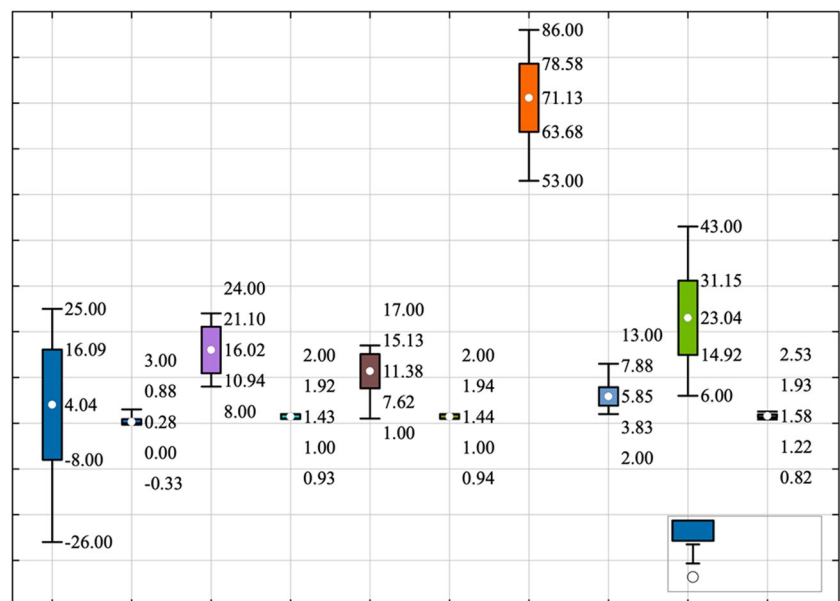
Table 4 Factor identified in the correlation analysis

Code	Factor	Unit	Description
F1	Thermal condition (°C)	°C	Average during eight-hour workday
F2	Rainfall	type	Including: (0) No rainfall, (1) Mild rain, (2) Rain, and (3) Snow
F3	Labor Size	workers	Total number of workers
F4	Activity Type	type	Including: (1) Slab and (2) Walls
F5	Level Floor	number	Floor number
F6	Working Method	type	Including: (1) Wooden form and (2) Flying form
F7	Primary Work	%	Labor-related activities carried out using tools
F8	Secondary Work	%	Work done in support of other workers
F9	Non-value added Work	%	Unproductive labor activities

Table 5 Dataset of on-site construction productivity

No	F1	F2	F3	F4	F5	F6	F7	F8	F9	Y
1	3.50	1.00	24.00	1.00	6.00	2.00	73.00	5.00	22.00	1.39
2	4.50	1.00	24.00	1.00	6.00	2.00	75.00	3.00	22.00	1.41
3	4.50	1.00	22.00	2.00	6.00	1.00	70.00	4.00	26.00	1.26
4	2.50	0.00	23.00	1.00	6.00	2.00	75.00	2.00	23.00	1.48
5	-4.50	0.00	19.00	1.00	7.00	2.00	70.00	4.00	26.00	1.36
...
216	-3.00	0.00	10.00	1.00	17.00	2.00	69.00	5.00	26.00	1.51
217	-1.00	0.00	12.00	1.00	17.00	2.00	72.00	6.00	22.00	1.58
218	1.00	0.00	12.00	1.00	17.00	2.00	71.00	8.00	21.00	1.65
219	1.00	0.00	11.00	1.00	17.00	2.00	70.00	6.00	24.00	1.49
220	1.00	0.00	11.00	1.00	17.00	2.00	65.00	6.00	29.00	1.32

Fig. 11 Dataset statistical measurement



4.3.1 Comparison of different incomplete datasets

The first sensitivity analysis phase was conducted to evaluate the performance of the model to handle incomplete data. It

is common in real-world applications to contain incomplete or missing data. To simulate the missing values, three different scenarios of missing data were considered: 5%, 10%, and 20% of the original dataset was randomly set to zero.

Evaluation the performance of the model using incomplete data can provide insights into the extent to which the model has become less accurate. Table 9 presents the performance metrics under these incomplete data scenarios based on 10-fold cross-validation. MBM is superior in terms of robustness, particularly when the proportion of missing data increases. With 5% missing data in the training phase, MBM showed slightly higher RMSE and MAE values, but maintained nearly identical R and R² values compared to SVM. In the testing phase, however, MBM outperformed SVM across all metrics, suggesting enhanced robustness against outliers in such scenarios. Furthermore, under 10% and 15% missing data conditions, the MBM model continued to outperform SVM, maintaining slightly superior performance across all evaluated metrics.

4.3.2 Comparison of different noisy dataset

The next sensitivity analysis phase includes synthetic Gaussian noise into the dataset to evaluate the performance of the model under less ideal conditions. The experiment was conducted at three distinct noise levels: 5%, 10%, and 20%. Furthermore, the performance metrics in the 10-fold cross-validation were reevaluated. Maintained performance levels indicate the robustness of the model to data noise, whereas a significant decline in these metrics may indicate sensitivity to data variability. Table 10 presents the performances of the models under varying noise conditions. Both MBM and SVM demonstrated robust performance across different levels of noise. In the low-noise scenario of 5%, MBM outperformed SVM in the testing phase, as indicated by lower

Table 6 Performance evaluation using different parameter settings

Fold	Param		Training					Testing				
	C	γ	RMSE	MAE	MAPE	R	R ²	RMSE	MAE	MAPE	R	R ²
MBM	1	1	0.067	0.058	3.80%	0.983	0.965	0.083	0.067	4.56%	0.975	0.950
SVM			0.073	0.062	4.02%	0.979	0.958	0.093	0.074	4.95%	0.968	0.937
MBM	2	0.5	0.065	0.056	3.69%	0.983	0.966	0.077	0.062	4.20%	0.977	0.955
SVM			0.070	0.059	3.82%	0.980	0.960	0.084	0.067	4.43%	0.973	0.946
MBM	3	0.1	0.069	0.058	3.80%	0.980	0.961	0.077	0.061	3.96%	0.977	0.955
SVM			0.078	0.064	4.09%	0.976	0.952	0.084	0.069	4.36%	0.973	0.947

Table 7 Performance evaluation for the MBM using grid search

Fold	Best		Training					Testing				
	C	γ	RMSE	MAE	MAPE	R	R ²	RMSE	MAE	MAPE	R	R ²
1	10	0.1	0.069	0.058	3.84%	0.981	0.962	0.059	0.048	2.94%	0.989	0.978
2	100	0.1	0.063	0.052	3.45%	0.983	0.967	0.075	0.059	3.41%	0.983	0.967
3	1	0.5	0.067	0.056	3.70%	0.982	0.964	0.066	0.054	3.40%	0.988	0.975
4	100	0.1	0.062	0.052	3.43%	0.984	0.969	0.091	0.071	4.22%	0.974	0.949
5	10	0.1	0.069	0.059	3.79%	0.980	0.961	0.066	0.048	3.12%	0.989	0.979
6	1	0.1	0.071	0.060	3.89%	0.979	0.959	0.085	0.062	3.87%	0.977	0.955
7	1	0.1	0.072	0.059	3.82%	0.980	0.960	0.063	0.052	3.80%	0.977	0.955
8	0.1	0.1	0.080	0.065	4.20%	0.973	0.947	0.057	0.048	3.34%	0.993	0.986
9	10	0.5	0.065	0.055	3.66%	0.984	0.968	0.057	0.050	3.01%	0.983	0.966
10	1	1	0.067	0.058	3.82%	0.984	0.968	0.057	0.047	3.04%	0.969	0.938
Average			0.068	0.057	3.76%	0.981	0.962	0.068	0.054	3.42%	0.982	0.965

Table 8 Comparison of model performance

Model	Training					Testing						
	RMSE	MAE	MAPE	R	R ²	RMSE	MAE	MAPE	R	R ²	RI	Rank
MBM	0.068	0.057	3.76%	0.981	0.962	0.068	0.054	3.42%	0.982	0.965	1.000	1
SVM	0.067	0.057	3.72%	0.982	0.964	0.070	0.058	3.74%	0.981	0.962	0.934	2
BPNN	0.079	0.062	3.97%	0.975	0.951	0.088	0.071	4.59%	0.970	0.941	0.615	3
LR	0.099	0.080	5.19%	0.959	0.920	0.102	0.083	5.17%	0.961	0.923	0.355	4
DT	0.085	0.066	4.23%	0.970	0.942	0.108	0.084	5.45%	0.954	0.910	0.229	5
KNN	0.101	0.075	4.97%	0.962	0.925	0.122	0.093	6.18%	0.947	0.897	0.000	6

RMSE, MAE, and MAPE values. Although the R and R² values were nearly identical for both models, MBM maintained a slight edge. This performance remains consistent at higher noise levels. Under 10% and 20% noise conditions, MBM continued to exhibit modest improvements during the testing phase. These findings suggest that MBM provides slightly better generalization when applied to unseen data. Overall, MBM provides robust alternative to the SVM. MBM consistently exhibited robust performance at various noise levels during testing. The resilience and effective generalization of the model make it an appealing alternative to the SVM model.

4.4 Evaluating model performance in diverse domains

Three open-source datasets from the UCI Machine Learning Repository were utilized to further evaluate the performance of the proposed MBM model, namely the High

Performance Concrete (HPC) compressive strength dataset [64], the Boston housing price dataset [65], and the energy efficiency under heating load dataset [66]. These datasets were selected for their diversity and relevance to real-world applications, allowing for a comprehensive assessment of the MBM capabilities in various domains. The HPC dataset, focusing on factors affecting concrete compressive strength, includes 1030 instances with eight input variables, such as cement, blast furnace slag, fly ash, and age. It is divided into 824 training and 206 testing datasets. The Boston dataset, which focuses on housing values, consists of 506 instances with 13 input variables divided into 379 training and 127 testing samples. The energy efficiency under heating load dataset encompasses 768 instances, with eight input variables affecting the building heating efficiency, split into 614 training and 154 testing samples. The inputs and outputs for each dataset are presented in Table 11. Each dataset was normalized to a range of 0–1 to ensure uniformity in scale. In the

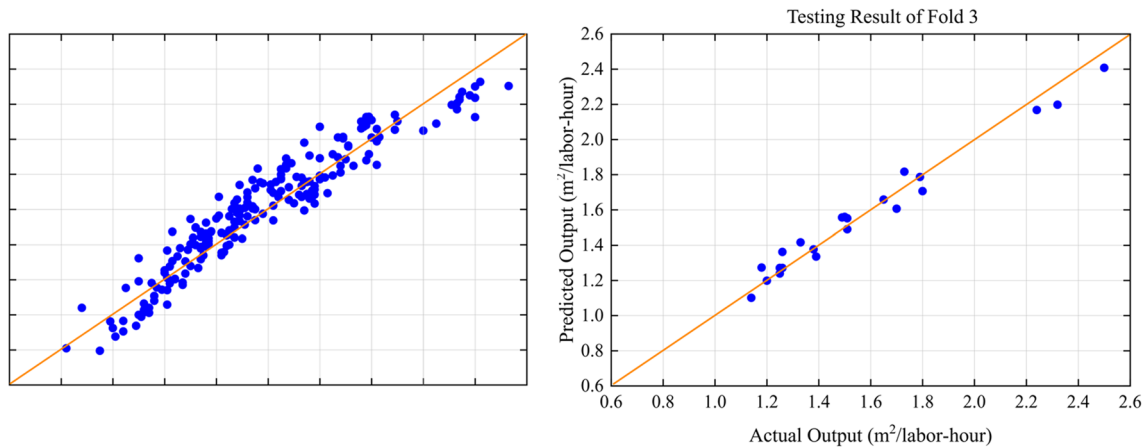


Fig. 12 Actual and Predicted Results for Fold 3

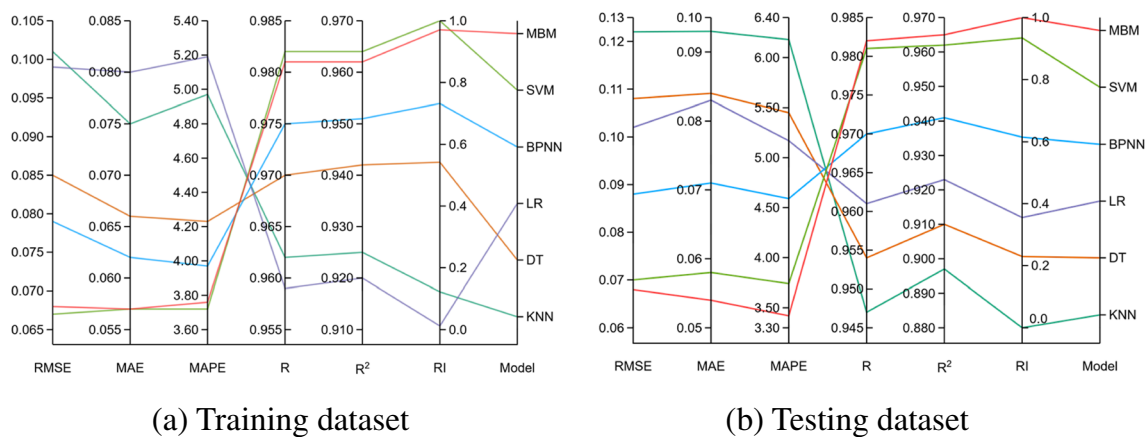


Fig. 13 Performance Evaluation

sensitivity analysis, this study compared MBM with SVM. For consistency, MBM was also compared with SVM for performance evaluation. The comparison results provide a rigorous benchmarking verifying the excellent performance of MBM in various predictive tasks. This approach offers an in-depth analysis of the predictive accuracy, generalization ability, and reliability of the models, which forms a robust foundation for evaluating the potential of MBM to address regression problems in diverse fields.

The performance evaluations presented in Table 12 provide a comparative analysis of the MBM against the SVM across three diverse datasets. Across all datasets, MBM demonstrates superior generalization capabilities and achieves lower testing error metrics (RMSE, MAE, and MAPE) comparing with SVM. MBM consistently exhibits competitive testing performance, highlighting its flexibility in handling diverse data types. This underlines its potential as a reliable tool for various predictive tasks across varied domains. The findings not only support the capability of MBM as an effective alternative approach for predicting construction productivity, but also for its applications in various fields.

4.5 Discussion and limitations

4.5.1 Discussion

The main contribution of this study is the introduction of Moment Balanced Machine (MBM), an AI-based inference engine to address regression problems. As an illustration of its application, the MBM was tested in a case

study to predict the formwork installation productivity in construction projects. The MBM combines the principles of SVM to determine the optimal moments hyperplane and utilizes BPNN to assign weights to each case. The integration of these two methods provides a distinct advantage to MBM in the development of its predictive model. An important aspect of this research approach is the two-stage experimental design that utilizes random search and grid search techniques. This was necessary to ensure a comprehensive assessment and fine-tuning of the hyperparameters, resulting in optimal model performance.

The models were validated using 10-fold cross-validation. The comparative analysis was expanded to include other machine learning models, such as SVM, BPNN, LR, DT, and KNN. Various performance metrics were used to evaluate the capability of each model during the training and testing phases. MBM demonstrates superior accuracy and robustness, substantiated by its lower error metric values and higher correlation coefficient (R) and coefficient of determination (R²). The training and testing phases of the MBM model generated good performance scores for all five performance evaluations. The average results of the 10-fold cross-validation technique showed that the MBM obtained the best scores for all five evaluation criteria and achieved the highest rank in the RI calculation. The comparative performance values obtained by MBM support that the proposed predictive model best predicts formwork installation productivity.

Furthermore, sensitivity analysis involving missing data at proportions of 5%, 10%, and 20% were conducted to test the robustness of the model. The performance of

Table 9 Performance evaluation in incomplete data scenarios

Fold	Missing data	Training					Testing				
		RMSE	MAE	MAPE	R	R ²	RMSE	MAE	MAPE	R	R ²
MBM	5%	0.069	0.058	3.84%	0.981	0.962	0.081	0.065	4.32%	0.974	0.949
SVM		0.063	0.052	3.45%	0.983	0.967	0.083	0.068	4.48%	0.974	0.948
MBM	10%	0.067	0.056	3.70%	0.982	0.964	0.086	0.071	4.64%	0.973	0.946
SVM		0.062	0.052	3.43%	0.984	0.969	0.089	0.073	4.81%	0.971	0.942
MBM	20%	0.069	0.059	3.79%	0.980	0.961	0.089	0.072	4.62%	0.967	0.936
SVM		0.071	0.060	3.89%	0.979	0.959	0.096	0.078	5.07%	0.965	0.931

Table 10 Performance evaluation in noisy data scenarios

Fold	Noisy	Training					Testing				
		RMSE	MAE	MAPE	R	R ²	RMSE	MAE	MAPE	R	R ²
MBM	5%	0.078	0.066	4.31%	0.975	0.951	0.068	0.056	3.62%	0.981	0.963
SVM		0.077	0.066	4.31%	0.976	0.953	0.070	0.057	3.69%	0.981	0.962
MBM	10%	0.099	0.083	5.42%	0.959	0.919	0.086	0.069	4.40%	0.973	0.948
SVM		0.100	0.084	5.43%	0.959	0.919	0.088	0.072	4.62%	0.973	0.946
MBM	20%	0.145	0.115	7.47%	0.907	0.825	0.114	0.089	5.77%	0.956	0.914
SVM		0.137	0.111	7.13%	0.919	0.845	0.115	0.092	5.91%	0.957	0.915

Table 11 Input and output variables for different datasets

No	HPC [64]	Boston Housing [65]	Energy [66]
	Input:		
1	Cement	Per capita crime rate	Compactness
2	Blast furnace slag	Proportion of residential	Surface Area
3	Fly ash	Proportion of non-retail business	Wall Area
4	Water	Charles River dummy variable	Roof Area
5	Superplasticizer	Nitric oxides concentration	Height
6	Coarse aggregate	Average number of rooms	Orientation
7	Fine aggregate	Proportion of occupied units	Glazing Area
8	Age	Weighted distances	Glazing Area Distribu- tion
9		Index of accessibility	
10		Full-value property-tax rate	
11		Pupil-teacher ratio	
12		Proportion of blacks	
13		% lower status of the population	
	Output:		
1	Compressive Strength	Housing prices	Heating Load

MBM was found to be consistently stable and superior to that of SVM, particularly as the proportion of missing data increased. These findings indicate that MBM is accurate and robust under suboptimal conditions. Another important aspect of this research is the robustness of the model to noise, which is a common problem in empirical datasets. This study involved synthetic Gaussian noise at various levels (5%, 10%, and 20%) to test the robustness of the model. Both MBM and SVM maintained robust performance under these conditions. However, MBM maintained a slight advantage during the testing phase under multiple noise conditions. Reinforcing the MBM's good generalization capabilities makes it a robust and reliable model for complex tasks.

The performance of MBM model was evaluated using three diverse UCI Machine Learning Repository datasets, namely High Performance Concrete (HPC), Boston housing prices, and energy efficiency under heating load. These datasets were selected for their variety and real-world applicability, allowing for an extensive assessment of the MBM capabilities. Across all domains, MBM showed superior generalization and lower testing errors (RMSE, MAE, and MAPE) comparing with SVM, highlighting its adaptability and reliability for various predictive tasks. This study supports the potential of MBM not only in construction productivity prediction but also in broader applications, establishing it as an effective tool in predictive modeling.

The MBM model was introduced as an alternative to existing machine learning models. Its superior performance, robustness to incomplete data, and synthetic noise make it suitable for real-world applications. Project managers can use the

resulting predictions to distribute resources more accurately, identify potential delays, and develop and implement data-driven plans to maximize productivity. Therefore, productivity predictions provide an overview of the possible outcomes and strategic guidance for achieving efficiency and effectiveness in construction projects.

4.5.2 Limitations

Despite the promising results, this study has several limitations. First, it presents the prediction of construction productivity using point predictions rather than an interval or a probability distribution, which is beyond the scope of this study. Future studies may explore integrating MBM with the Gaussian Process to deal with the prediction of confidence intervals. Second, the model demonstrated sensitivity to hyperparameter configurations. Although the MBM demonstrates consistent performance, the optimal hyperparameter settings can differ across the cross-validation folds. Future work could explore more advanced hyperparameter tuning techniques, possibly integrating metaheuristic methods, such as genetic algorithms (GA) [67], particle swarm optimization (PSO) [68], and symbiotic organisms search (SOS) [69], etc., to enhance the model performance and potentially overcome its sensitivity to parameter variations. Finally, the current study mainly compares MBM with other machine learning models, such as SVM, BPNN, LR, DT, and KNN. Future research could benefit from including a more diverse set of models, such as hybrid approaches, to comprehensively evaluate their strengths and weaknesses.

Table 12 Performance evaluation of the model across the different datasets

Dataset	Fold	Best	C	γ	Training					Testing				
					RMSE	MAE	MAPE	R	R ²	RMSE	MAE	MAPE	R	R ²
HPC[64]	MBM	100	5	3.188	1.497	6.06%	0.981	0.963	6.668	4.145	12.75%	0.926	0.857	
	SVM	100	5	3.945	2.200	8.65%	0.971	0.943	6.836	4.496	15.10%	0.922	0.851	
Boston [65]	MBM	10	1	2.314	1.276	6.87%	0.965	0.930	2.821	2.028	10.76%	0.964	0.929	
	SVM	100	1	2.074	1.094	5.48%	0.973	0.947	2.894	2.117	11.20%	0.958	0.919	
Energy [66]	MBM	100	1	0.170	0.118	0.68%	1.000	1.000	0.900	0.618	3.72%	0.996	0.993	
	SVM	100	5	0.096	0.095	0.53%	1.000	1.000	0.954	0.642	3.93%	0.996	0.993	

5 Conclusion

The research has successfully developed the Moment Balanced Machine (MBM), an AI-based inference machine, designed for accurate prediction in regression cases. As a case study, this model was tested on the productivity of formwork installation in construction projects. In its development, MBM utilizes Backpropagation Neural Networks (BPNN) to determine the weights of each case and adopts the principles of Support Vector Machines (SVM) to determine the optimal moment hyperplane. The integration of these two approaches provides a significant advantage to MBM in the creation of its predictive model. The generated productivity predictions provide a support tool for making and implementing data-driven planning and resource allocation decisions in construction projects. The new, AI-based inference engine for predicting construction productivity developed and tested in this study contributes substantially to improving decision-making in construction engineering and management processes by leveraging accumulated domain knowledge and experience. Project managers may utilize model-generated predictions to allocate resources, plan project-completion schedules, and identify delays more accurately and to make informed decisions that significantly enhance productivity.

The robustness of MBM was tested through a two-stage experimental design, utilizing both random and grid searches for hyperparameter optimization. The real-world formwork installation productivity in construction projects, including nine input features that influence on-site construction productivity, were used to construct the predictive model. MBM performance was validated through 10-fold cross-validation, and it demonstrated superior performance across key metrics like RMSE, MAE, MAPE, R, and R2 when compared to five other machine learning models (SVM, BPNN, KNN, DT, and LR). Specifically, MBM achieved the lowest RMSE (0.068), MAE (0.054), and MAPE (3.42%) along with the highest R (0.982) and R2 (0.965) values. Additionally, it scored the highest Reference Index (RI) of 1.000, further confirming its predictive accuracy and overall performance. These results support that MBM outperforms the other compared prediction methods and facilitates accurate predictions of on-site construction productivity. Furthermore, MBM also showed resilience in handling missing or incomplete data and was found to be robust against synthetic Gaussian noise. The comparisons of MBM with SVM using diverse datasets verified the robust generalization and adaptability of MBM. This effectiveness across various domains highlights the potential of the model for predictive modeling beyond the construction productivity. This makes it a reliable and robust model for real-world applications.

While the MBM demonstrates superior performance compared to other models, this study has limitations that need to be considered in future research. First, it predicts construction productivity using point predictions. Future studies may be enhanced by integrating MBM with the Gaussian Process to address confidence interval predictions. Second, the model displays sensitivity to the hyperparameter settings. To enhance its performance, the integration of metaheuristic methods could be explored in upcoming research. Lastly, the comparative analysis in this study was limited to a specific set of machine learning models, suggesting an opportunity for more extensive model comparisons in future research.

Data availability The code and dataset for this study is available at: <https://github.com/MomentBalancedMachine/MBM>.

References

- Liang W, Yang M (2019) Urbanization, economic growth and environmental pollution: evidence from China. *Sustain Comput Inf Syst* 21:1–9. <https://doi.org/10.1016/j.suscom.2018.11.007>
- Kianto A, Shujahat M, Hussain S, Nawaz F, Ali M (2019) The impact of knowledge management on knowledge worker productivity. *Baltic J Manage* 14:178–197. <https://doi.org/10.1108/BJM-12-2017-0404>
- Alaloul WS, Liew MS, Zawawi NAWA, Kennedy IB (2020) Industrial revolution 4.0 in the construction industry: challenges and opportunities for stakeholders. *Ain Shams Eng J* 11:225–230. <https://doi.org/10.1016/j.asej.2019.08.010>
- Zou X, Zhang L (2020) A constraint programming approach for scheduling repetitive projects with atypical activities considering soft logic. *Autom Constr* 109:102990. <https://doi.org/10.1016/j.autcon.2019.102990>
- Kavuma A, Ock J, Jang H (2019) Factors influencing time and cost overruns on freeform construction projects. *KSCE J Civ Eng* 23:1442–1450. <https://doi.org/10.1007/s12205-019-0447-x>
- Zhang N, Deng X, Zhao X, Chang T (2019) Exploring the sources of contractors' competitive advantage on international HSR construction projects. *Int J Civil Eng* 17:1115–1129. <https://doi.org/10.1007/s40999-018-0373-1>
- Javed AA, Pan W, Chen L, Zhan W (2018) A systemic exploration of drivers for and constraints on construction productivity enhancement, built environment project and asset. *Management* 8:239–252. <https://doi.org/10.1108/BEPAM-10-2017-0099>
- Safapour E, Kermanshachi S (2019) Identifying early indicators of manageable rework causes and selecting mitigating best practices for construction. *J Manag Eng* 35:04018060. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000669](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000669)
- Abbasi O, Noorzai E, Gharouni Jafari K, Golabchi M (2020) Exploring the causes of delays in construction industry using a cause-and-effect diagram: case study for Iran. *J Archit Eng* 26:05020008. [https://doi.org/10.1061/\(ASCE\)AE.1943-5568.0000431](https://doi.org/10.1061/(ASCE)AE.1943-5568.0000431)
- Tariq J, Gardezi SSS (2023) Study the delays and conflicts for construction projects and their mutual relationship: a review. *Ain Shams Eng J* 14:101815. <https://doi.org/10.1016/j.asej.2022.101815>
- Alaghbari W, Al-Sakkaf AA, Sultan B (2019) Factors affecting construction labour productivity in Yemen. *Int J Constr Manage* 19:79–91. <https://doi.org/10.1080/15623599.2017.1382091>
- Hasan A, Baroudi B, Elmualim A, Rameezdeen R (2018) Factors affecting construction productivity: a 30 year systematic review. *Eng Constr Archit Manag* 25:916–937. <https://doi.org/10.1108/ECAM-02-2017-0035>
- Cui P, Ge Y, Li S, Li Z, Xu X, Zhou GGD, Chen H, Wang H, Lei Y, Zhou L, Yi S, Wu C, Guo J, Wang Q, Lan H, Ding M, Ren J, Zeng L, Jiang Y, Wang Y (2022) Scientific challenges in disaster risk reduction for the Sichuan–Tibet railway. *Eng Geol* 309:106837. <https://doi.org/10.1016/j.enggeo.2022.106837>
- Regona M, Yigitcanlar T, Xia B, Li RYM (2022) Opportunities and adoption challenges of AI in the construction industry: a PRISMA Review. *J Open Innov: Technol Market Complex* 8:45. <https://doi.org/10.3390/joitmc8010045>
- Khanh HD, Kim SY, Linh LQ (2022) Construction productivity prediction through Bayesian networks for building projects: case from Vietnam. *Eng Constr Archit Manag* 30:2075–2100. <https://doi.org/10.1108/ECAM-07-2021-0602>
- Taboada I, Daneshpajouh A, Toledo N, de Vass T (2023) Artificial intelligence enabled project management: a systematic literature review. *Appl Sci (Switzerland)* 13. <https://doi.org/10.3390/app13085014>
- Pan Y, Zhang L (2021) Roles of artificial intelligence in construction engineering and management: a critical review and future trends. *Autom Constr* 122:103517. <https://doi.org/10.1016/j.autcon.2020.103517>
- Momade MH, Shahid S, bin Hainin MR, Nashwan MS, Tahir Umar A (2022) Modelling labour productivity using SVM and RF: a comparative study on classifiers performance. *Int J Constr Manage* 22:1924–1934. <https://doi.org/10.1080/15623599.2020.1744799>
- 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* 27:2335–2351. <https://doi.org/10.1108/ECAM-08-2019-0406>
- 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* 143:1–10. [https://doi.org/10.1061/\(asce\)co.1943-7862.0001340](https://doi.org/10.1061/(asce)co.1943-7862.0001340)
- 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>
- Ghodrati N, Wing Yiu T, 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* 34:04018035. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000641](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000641)
- Oppong GD, Chan APC, Dansoh A (2017) A review of stakeholder management performance attributes in construction projects. *Int J Project Manage* 35:1037–1051. <https://doi.org/10.1016/j.ijproman.2017.04.015>
- Agrawal A, Halder S (2020) Identifying factors affecting construction labour productivity in India and measures to improve productivity. *Asian J Civil Eng* 21:569–579. <https://doi.org/10.1007/s42107-019-00212-3>
- Luangcharoenrat C, Intrachooto S, Peansupap V, Sutthinarakorn W (2019) Factors influencing construction waste generation in building construction: Thailand's perspective. *Sustainability (Switzerland)* 11. <https://doi.org/10.3390/su11133638>
- García de Soto B, Agustí-Juan I, Hunhevicz J, Joss S, Graser K, Habert G, Adey BT (2018) Productivity of digital fabrication in construction: cost and time analysis of a robotically built wall. *Autom Constr* 92:297–311. <https://doi.org/10.1016/j.autcon.2018.04.004>
- Cheng MY, Cao MT, 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>
28. Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2:1–21. <https://doi.org/10.1007/s42979-021-00592-x>
 29. Karatas 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>
 30. Chan APC, Zhang Y, Wang F, Wong FFK, Chan DWM (2017) A field study of the effectiveness and practicality of a novel hybrid personal cooling vest worn during rest in Hong Kong construction industry. *J Therm Biol* 70:21–27
 31. Ballesteros-Pérez P, Smith ST, Lloyd-Papworth JG, Cooke P (2018) Incorporating the effect of weather in construction scheduling and management with sine wave curves: application in the United Kingdom. *Constr Manage Econ* 36:666–682. <https://doi.org/10.1080/01446193.2018.1478109>
 32. Elwakil E (2018) Construction productivity fuzzy knowledge base management system. *Can J Civ Eng* 45:329–338. <https://doi.org/10.1139/cjce-2017-0540>
 33. Wang F (2005) On-Site labor productivity estimation using neural networks. Concordia University. <https://spectrum.library.concordia.ca/id/eprint/8516/>. Accessed 14 Apr 2023
 34. 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>
 35. Elshaboury N (2022) Training adaptive neuro fuzzy inference system using genetic algorithms for Predicting Labor Productivity. Springer International Publishing. https://doi.org/10.1007/978-3-030-97255-4_24
 36. Golnaraghi S, Zangenehmadar Z, Moselhi O, Alkass S, Vosoughi AR (2019) Application of artificial neural network(s) in predicting formwork labour productivity. *Adv Civ Eng* 2019. <https://doi.org/10.1155/2019/5972620>
 37. Shahtaheri M, Nasir H, Haas CT (2015) Setting baseline rates for on-site work categories in the construction industry. *J Constr Eng Manag* 141:1–11. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000959](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000959)
 38. Liu X, Song Y, Yi W, Wang X, Zhu J (2018) Comparing the random forest with the generalized additive model to evaluate the impacts of outdoor ambient environmental factors on scaffolding construction productivity. *J Constr Eng Manag* 144:4018037. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001495](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001495)
 39. Khan ZU (2005) Modeling and parameter ranking of construction labor productivity. Concordia University. [https://concordiauniversity.on.worldcat.org/search?queryString=ot:\(Spectrum\)+8615](https://concordiauniversity.on.worldcat.org/search?queryString=ot:(Spectrum)+8615). Accessed 14 Apr 2023
 40. Senarath Jayasinghe R, Fernando NG (2017) Developing labour productivity norms for aluminium system formwork in Sri Lanka. *Built Environ Project Asset Manag* 7:199–211. <https://doi.org/10.1108/BEPAM-09-2015-0049>
 41. Al Refaie AM, Alashwal AM, Abdul-Samad Z, Salleh H (2021) Weather and labor productivity in construction: a literature review and taxonomy of studies. *Int J Product Perform Manage* 70:941–957. <https://doi.org/10.1108/IJPPM-12-2019-0577>
 42. Wandahl S, Pérez CT, Salling S, Neve HH (2023) Correlation of construction workers' movement and direct work rates. *J Eng Project Prod Manage* 13:125–137. <https://doi.org/10.32738/JEPPM-2023-0013>
 43. Chan APC, Guo YP, Wong FKW, Li Y, Sun S, Han X (2016) The development of anti-heat stress clothing for construction workers in hot and humid weather. *Ergonomics* 59:479–495. <https://doi.org/10.1080/00140139.2015.1098733>
 44. Jarkas AM, Al Balushi RA, Raveendranath PK (2015) Determinants of construction labour productivity in Oman. *Int J Constr Manage* 15:332–344. <https://doi.org/10.1080/15623599.2015.1094849>
 45. Ghoddousi P, Poorafshar O, Chileshe N, Hosseini MR (2015) Labour productivity in Iranian construction projects perceptions of chief executive officers. *Int J Product Perform Manage* 64:811–830. <https://doi.org/10.1108/IJPPM-10-2013-0169>
 46. Rivas RA, Borcherding JD, González V, Alarcón LF (2011) Analysis of factors influencing productivity using craftsmen questionnaires: case study in a Chilean construction company. *J Constr Eng Manag* 137:312–320. [https://doi.org/10.1061/\(asce\)co.1943-7862.0000274](https://doi.org/10.1061/(asce)co.1943-7862.0000274)
 47. Li G, Xue Y, Qu C, Qiu D, Wang P, Liu Q (2023) Intelligent prediction of rockburst in tunnels based on back propagation neural network integrated beetle antennae search algorithm. *Environ Sci Pollut Res*: 33960–33973. <https://doi.org/10.1007/s11356-022-24420-8>
 48. Fei J, Wu Z, Sun X, Su D, Bao X (2021) Research on tunnel engineering monitoring technology based on BPNN neural network and MARS machine learning regression algorithm. *Neural Comput Appl* 33:239–255. <https://doi.org/10.1007/s00521-020-04988-3>
 49. Han B, Geng F, Dai S, Gan G, Liu S, Yao L (2020) Statistically optimized back-propagation neural-network model and its application for deformation monitoring and prediction of concrete-face Rockfill dams. *J Perform Constr Facil* 34:1–8. [https://doi.org/10.1061/\(asce\)cf.1943-5509.0001485](https://doi.org/10.1061/(asce)cf.1943-5509.0001485)
 50. Cai B, Pan G, Fu F (2020) Prediction of the postfire flexural capacity of RC beam using GA-BPNN machine learning. *J Perform Constr Facil* 34:1–11. [https://doi.org/10.1061/\(asce\)cf.1943-5509.0001514](https://doi.org/10.1061/(asce)cf.1943-5509.0001514)
 51. Liu Z, Liu X, Wang K, Liang Z, Correia JAFO, De Jesus A (2019) Neural network-based strain prediction in full-scale static testing of wind turbine blades. *Energies* 12:1026. <https://doi.org/10.3390/en12061026>
 52. Song X, Wang K, Zhou L, Chen Y, Ren K, Wang J (2022) Multi-factor mining and corrosion rate prediction model construction of carbon steel under dynamic atmospheric corrosion environment. *Eng Fail Anal* 134:105987. <https://doi.org/10.1016/j.engfailanal.2021.105987>
 53. Chen S, Associate D, Zhao Y (2020) The prediction analysis of properties of recycled aggregate permeable concrete based on back-propagation neural network. *J Clean Prod* 276:124187. <https://doi.org/10.1016/j.jclepro.2020.124187>
 54. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
 55. Zhang Q, Liu D, Fan Z, Lee Y, Li Z (2011) Feature and sample weighted support vector machine. *BMC Bioinformatics* 8:365–371. https://doi.org/10.1007/978-3-642-25661-5_47
 56. Lapin M, Hein M, Schiele B (2014) Learning using privileged information: SV M + and weighted SVM. *Neural Netw* 53:95–108. <https://doi.org/10.1016/j.neunet.2014.02.002>
 57. Moslemnejad S, Hamidzadeh J (2021) Weighted support vector machine using fuzzy rough set theory. *Soft Comput* 25:8461–8481. <https://doi.org/10.1007/s00500-021-05773-7>
 58. Fan Z, Chiong R, Hu Z, Lin Y (2020) A fuzzy weighted relative error support vector machine for reverse prediction of concrete components. *Comput Struct* 230:106171. <https://doi.org/10.1016/j.compstruc.2019.106171>
 59. Luo S, Dai Z, Chen T, Chen H, Jian L (2020) A weighted SVM ensemble predictor based on AdaBoost for blast furnace ironmaking process. *Appl Intell* 50:1997–2008. <https://doi.org/10.1007/s10489-020-01662-y>

60. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R (2011) Dubourg, others, Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
61. Gohari M, Eydi AM (2020) Modelling of shaft unbalance: Modelling a multi discs rotor using K-Nearest Neighbor and Decision Tree Algorithms, *Measurement: Journal of the International Measurement Confederation*. 151:107253. <https://doi.org/10.1016/j.measurement.2019.107253>
62. Bansal M, Goyal A, Choudhary A (2022) A comparative analysis of K-Nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decis Anal J* 3:100071. <https://doi.org/10.1016/j.dajour.2022.100071>
63. Chang CC, Lin CJ (2011) LIBSVM: A Library for support vector machines. *ACM Trans Intell Syst Technol* 2. <https://doi.org/10.1145/1961189.1961199>
64. Yeh I-C (2007) Concrete compressive strength. UCI Mach Learn Repository. <https://doi.org/10.24432/C5PK67>
65. Harrison D, Rubinfeld D (2016) Boston Housing. UCI Mach Learn Repository. <https://kaggle.com/competitions/boston-housing>
66. Tsanas A, Xifara A (2012) Energy efficiency. UCI Mach Learn Repository. <https://doi.org/10.24432/C51307>
67. Ding C, Xia Y, Yuan Z, Yang H, Fu J, Chen Z (2023) Performance prediction for a fuel cell air compressor based on the combination of backpropagation neural network optimized by genetic algorithm (GA-BP) and support vector machine (SVM) algorithms. *Therm Sci Eng Prog* 44:102070. <https://doi.org/10.1016/j.tsep.2023.102070>
68. Rosso MM, Cucuzza R, Di Trapani F, Marano GC (2021) Non-penalty machine learning constraint handling using PSO-SVM for structural optimization. *Adv Civil Eng* 2021:1–17. <https://doi.org/10.1155/2021/6617750>
69. Cheng M-Y, Prayogo D (2014) Symbiotic organisms search: a new metaheuristic optimization algorithm. *Comput Struct* 139:98–112. <https://doi.org/10.1016/j.compstruc.2014.03.007>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.