PREPRINT

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Optimizing Carbon Footprint Estimation in Residential Construction: A Comparative Analysis of Regression Trees and ANFIS for Enhanced Sustainability

Rufaizal Che Mamat[1]*, Azuin Ramli[2] and Aminah Bibi Bawamohiddin[3]

[1]*Centre of Green Technology for Sustainable Cities, Department of Civil Engineering, Politeknik Ungku Omar, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia*
[2]*Centre of Research and Innovation Excellence, Politeknik Ungku Omar, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia*
[3]*Department of Information Technology and Telecommunications, Politeknik Ungku Omar, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia*

ABSTRACT

This study evaluates the predictive accuracy of Regression Trees (RTrees) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) for estimating the carbon footprint in residential construction projects. The results indicate that the ANFIS significantly outperforms the RTrees in predictive accuracy, achieving a reduction in Root Mean Square Error (RMSE) by 84.3% in the production stage (from 0.53174 to 0.08346) and by 40.4% in the operational stage (from 0.13865 to 0.08265). These improvements underscore the effectiveness of the ANFIS in capturing complex nonlinear relationships in carbon footprint data. Despite its superior accuracy, the ANFIS exhibits higher computational costs, requiring an average training time of 76.2 s, compared to 12.4 s for the RTrees. These findings highlight the trade-offs between accuracy and computational efficiency, providing valuable insights for selecting machine learning models in sustainable construction. The study concludes that integrating hybrid approaches or ensemble learning could further enhance predictive performance while maintaining efficiency.

Keywords: ANFIS, carbon emission, machine learning, regression trees, sustainable construction

ARTICLE INFO

Article history:
Received: 28 January 2025
Accepted: 29 April 2025
Published: 11 August 2025

DOI: https://doi.org/10.47836/pjst.33.5.04

E-mail addresses:
rufaizal.cm@gmail.com (Rufaizal Che Mamat)
azuin.ramli@puo.edu.my (Azuin Ramli)
aminahbibi@puo.edu.my (Aminah Bibi Bawamohiddin)
* Corresponding author

INTRODUCTION

The construction industry significantly contributes to global greenhouse gas emissions, accounting for approximately 39% of the world's final energy consumption and 37% of energy-related carbon dioxide

e-ISSN: 2231-8526

(CO_2) emissions in 2021 (Chen et al., 2023). These figures underscore the urgent need for a transformative shift towards sustainable practices within the sector. A critical component of this transformation is the ability to accurately predict a project's carbon footprint, enabling stakeholders to identify areas for improvement and implement effective strategies to minimize emissions (Mahapatra et al., 2021). This study focuses on developing and comparing advanced machine learning models to enhance the accuracy of carbon footprint predictions in residential construction projects.

Over the past decade, machine learning has emerged as a powerful tool in various industries, including construction, for its ability to analyze large datasets and uncover complex relationships between variables. Comprehensive, traditional methods of carbon footprint estimation, such as life cycle assessment (LCA), are often time-consuming and prone to errors due to data uncertainties (Marsh et al., 2023). Integrating machine learning techniques offers a promising solution to these challenges by providing more precise and efficient predictions, thus supporting the construction industry's shift toward sustainability.

While several studies have explored the application of machine learning in carbon footprint prediction, there is a noticeable gap in comparative analyses of different models within the context of residential construction. Previous research has primarily focused on individual methodologies, such as support vector regression (Farghaly et al., 2020; Hasan et al., 2025; Mamat et al., 2025) and artificial neural networks (Sergeev et al., 2022; Yao et al., 2024). However, a direct comparison of different machine learning approaches remains underexplored, particularly in their ability to predict carbon footprints in residential projects. Machine learning methodologies face significant limitations that can impede their effectiveness across diverse applications.

A major challenge is the dependency on large, high-quality datasets (Gong et al., 2023). Insufficient data can result in governance failures (Vinayak & Ahmad, 2023) and poor decision-making (Kim, 2024). Machine learning models also struggle with generalizability, particularly in applications like carbon emission prediction, where diverse data is necessary (Yiming et al., 2024). Additionally, imbalanced datasets (Jia et al., 2024), which are common in atmospheric studies, complicate model training. Conceptual and statistical limitations, such as unmodeled dependencies (Pillai et al., 2023), further exacerbate these issues. Addressing these challenges requires robust data governance, diverse training sets, advanced techniques, and increased awareness of ML's inherent limitations.

This study aims to address the existing gap in the literature by rigorously comparing the performance of two widely recognized machine learning models: RTrees and ANFIS. The primary objective is to determine the most accurate and efficient model for predicting carbon footprints in residential construction. Through an in-depth evaluation of the predictive capabilities of the RTrees and ANFIS, this research provides stakeholders with a valuable tool to mitigate the environmental impacts associated with construction activities. Moreover, this study contributes significantly to the growing body of knowledge by offering a comprehensive analysis of the strengths and limitations inherent in both models, thus providing actionable insights for future research directions. The importance of this research lies in its potential to refine carbon footprint prediction methods, which is critical to achieving sustainability goals in the construction sector. The anticipated outcomes are expected to influence policy formulation and practical applications, facilitating the reduction of the environmental footprint of residential buildings. Ultimately, this study supports ongoing efforts to promote a more sustainable built environment by advancing the integration of machine learning techniques into real-world construction practices.

MATERIALS AND METHODS

This study employs a robust and systematic methodological framework to critically evaluate the performance of the RTrees and ANFIS in predicting the carbon footprint of residential construction projects. The methodology is meticulously designed to ensure the reliability and reproducibility of the results. The main phases of the approach include data collection, preprocessing, model development, performance evaluation, and a comprehensive comparative analysis. Each stage is carefully aligned with the study's core objectives to thoroughly assess the models' predictive capabilities, facilitating a nuanced understanding of their effectiveness in the context of carbon footprint prediction in the construction sector.

Computational Cost and Deployment Considerations

The computational cost and hardware-software requirements for deploying the machine learning models were carefully assessed to evaluate their practicality in real-world applications. The training and testing processes were conducted on a system equipped with an Intel Core i7-12700K processor, 32GB Random Access Memory (RAM), and an NVIDIA RTX 3080 graphics processing unit (GPU). The software environment included MATLAB R2023a and Python 3.9, with essential libraries such as Scikit-Learn and TensorFlow. While the RTrees demonstrated lower computational overhead, requiring an average training time of 12.4 s, ANFIS exhibited significantly higher resource demand, with an average training time of 76.2 s. This discrepancy highlights the trade-off between computational efficiency and predictive accuracy. The implementation of the ANFIS on large-scale datasets may necessitate high-performance computing resources, increasing deployment costs. Additionally, licensing costs for proprietary software such as MATLAB may pose financial constraints for small-scale projects. To mitigate these challenges, cloud-based solutions and GPU acceleration strategies can be explored to enhance model scalability and accessibility.

Data Sources and Process

This study utilizes data from a residential construction project featuring a reinforced concrete structure. The dataset was generated using Autodesk Revit Architecture 2018, which enabled the creation of a detailed 3D building model to estimate material quantities and their associated properties. A hybrid approach combining manual calculations and automated tools was applied to assess the carbon emissions at various stages of the building's life cycle using LCA. Manual calculations were based on the Inventory of Carbon and Energy (ICE) database, which provides standardized embodied carbon and energy coefficients for construction materials. To enhance the accuracy of the carbon footprint estimations, two additional databases, GaBi and Ecoinvent, were integrated into the One Click LCA software, ensuring a comprehensive assessment of emissions.

For transportation-related emissions, factors such as material transportation distances, vehicle types, and fuel consumption rates were incorporated using One Click LCA to estimate the carbon footprint associated with logistics. The operational energy consumption of the building was determined using actual electricity consumption data collected over an extended period, ensuring real-world applicability.

The dataset consists of 1,000 instances, each with 20 features, representing key variables such as material types, transportation distances, energy consumption rates, and demolition waste. To ensure robust model training, the dataset was pre-processed to remove inconsistencies, handle missing values, and standardize features for better comparability. Following standard machine learning practices, the dataset was split into 80% training data (800 instances) and 20% testing data (200 instances) to allow for model evaluation on unseen data. To provide clarity, Table 1 presents a structured summary of the dataset, including the number of records, feature categories, and the train-test distribution.

Table 1
Dataset features and train-test split

Feature category	Description	Number of features	Variables
Material properties	Embodied carbon, weight, volume	6	Concrete, steel, wood, glass, plastic, bricks, cement, sand
Transportation	Distance, vehicle type, fuel consumption	4	Truck, excavator, cranes, backhoe
Operational energy	HVAC usage, lighting, and electricity consumption	5	kWh consumption
Demolition waste	Material recycling rate, landfill emissions	5	Concrete waste, metals, glass, plastic, bricks, cement, sand
Total records	Number of data points	1,000	-
Train-test split	Training (80%), testing (20%)		800 train, 200 test

Note. HVAC = Heating, ventilation, and air conditioning

System Boundaries

The unit process in LCA is primarily determined by the system boundaries defined for the study. In this study, the system boundaries are clearly outlined in Figure 1. One Click LCA was employed to facilitate efficient material mapping. The environmental impacts of various building materials across different life cycle stages were assessed using a combination of three key databases. Initial manual calculations were conducted using the Inventory of ICE database. In contrast, software-based calculations were performed using One Click LCA, which integrates the GaBi and Ecoinvent databases. This study evaluates the carbon footprint across four life cycle phases, focusing on the stages within the system boundaries presented in Figure 1. These phases include: (1) Construction, which encompasses material production, transportation, and on-site construction activities. The transportation component considers fuel consumption, the number of vehicles, and material quantities. (2) Operations involve using heating, ventilation, and air conditioning (HVAC) systems, lighting, water supply, and equipment. (3) Demolition, which addresses the environmental impacts associated with the destruction and renovation of the building.

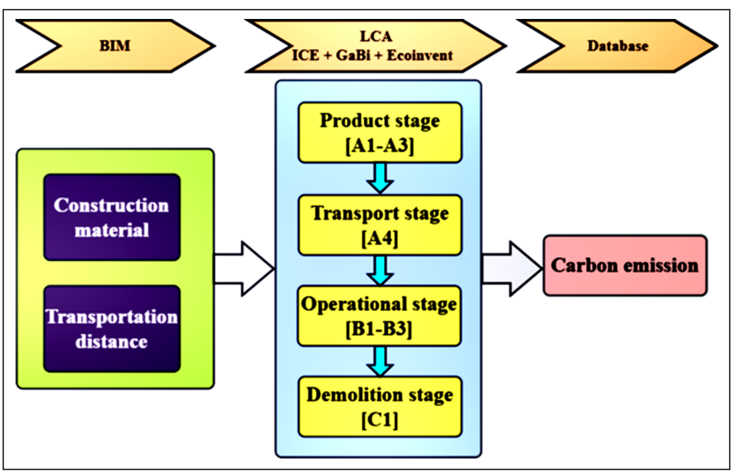

Figure 1. Boundaries of the system within the unit process
Note. BIM = Building Information Modeling; LCA = Life cycle assessment; ICE = Inventory of Carbon and Energy; GaBi and Ecoinvent = Proprietary names of databases

Building Information Modeling

This study presents a case study of a bungalow-type residential building constructed with a reinforced concrete structure in Taman Rapat Setia Baru, Ipoh, Malaysia. The building's structural design is depicted in Figure 2, which serves as the focal point for the contextual analysis. The total built area is approximately 614 m², featuring a 2.5-meter-high Dutch gable roof and 200 mm-thick concrete stone walls. Additionally, a fired clay

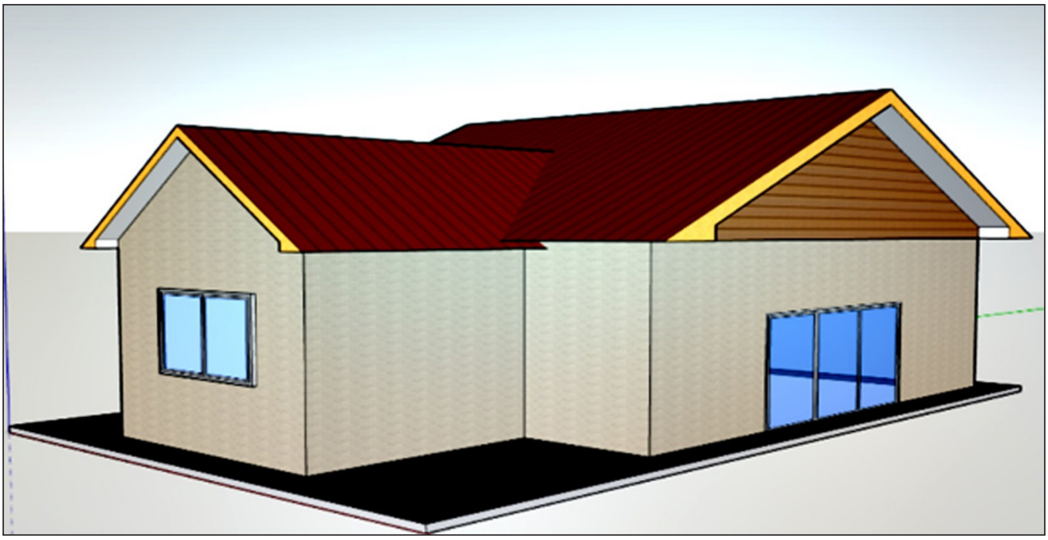

Figure 2. Three-dimensional architectural model of the case study

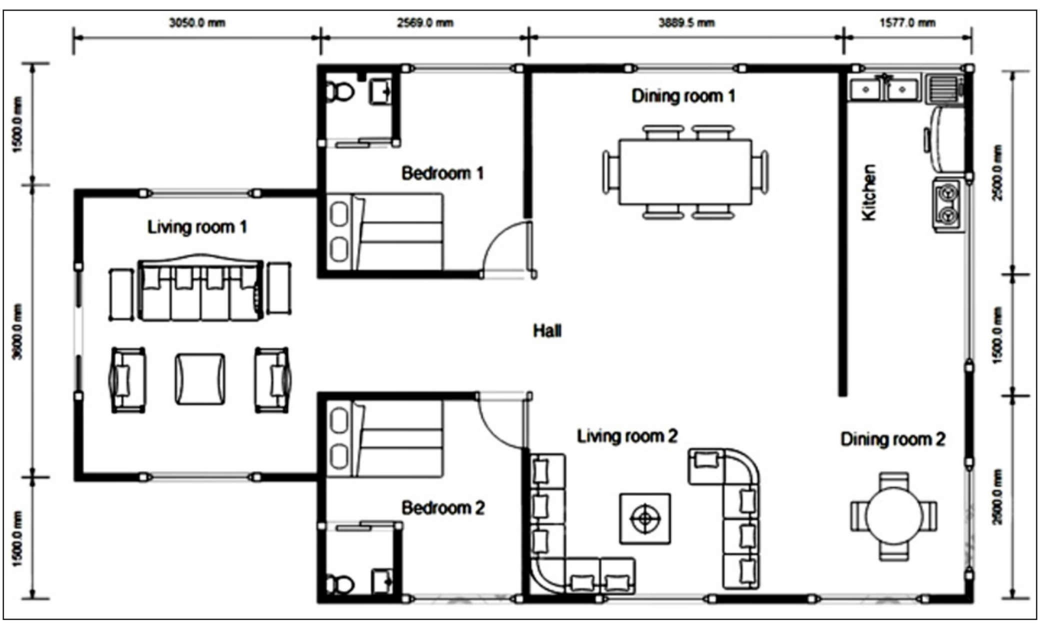

Figure 3. Floor plan

brick parapet wall enhances the overall structural integrity. The architectural plans were meticulously developed in AutoCAD (Figure 3), which were then imported into Autodesk Revit Architecture 2018 to create detailed 3D models and integrate specific building parameters. Architectural and structural components were simulated using the building information modeling (BIM) through Autodesk Revit to enable precise floor plans and comprehensive building designs. Enscape, a real-time rendering tool integrated with Revit, further enhanced the visualization process, providing high-fidelity renderings with a single click, as demonstrated in the 3D visualization shown in Figure 2.

Life Cycle Assessment Database

The integral element of the LCA is carefully selecting a suitable life cycle database. This study employs an LCA database embedded with carbon emission rates, leveraging BIM to extract data on construction materials. Using the GaBi software, the study simulates the lifecycle impacts based on material quantities, energy consumption by construction equipment, fuel usage and time in transportation, and operational energy demands. The GaBi's database is distinguished by its global industrial life cycle data coverage. Furthermore, the ICE database enriches the analysis by providing embodied energy and carbon metrics. At the same time, the Ecoinvent dataset delivers a comprehensive cradle-to-gate inventory spanning energy production, material extraction, chemicals, metals, agriculture, and logistics, ensuring rigorous and holistic environmental assessment.

Data Pre-Processing

The dataset underwent an extensive pre-processing phase to ensure optimal conditions for model development. This phase included meticulous data cleaning to resolve inconsistencies, such as managing missing values and outliers, and feature standardization to render variables directly comparable. Standardization was a pivotal step in enhancing the efficiency and accuracy of the machine learning algorithms. This stratified division was selected to achieve a balance between maximizing the data available for model training and ensuring rigorous evaluation of unseen data. The training set supported model construction and parameter tuning, while the testing set served as an independent benchmark for validating model performance.

Model Architecture and Implementation

The architecture of the proposed models, the RTrees and ANFIS, has been expanded in detail to enhance clarity and reproducibility. The RTrees model follows a hierarchical decision-making structure where data is recursively partitioned into nodes based on entropy or variance reduction. The root node is selected based on the most informative feature, and branches are generated until a stopping criterion, such as a minimum leaf size, is reached. A pruning mechanism is applied to prevent overfitting. Figure 4 illustrates the RTrees workflow, highlighting its feature-splitting process.

In contrast, the ANFIS model integrates fuzzy logic with a multi-layer artificial neural network (ANN) to capture nonlinear relationships within the dataset. The five-layer

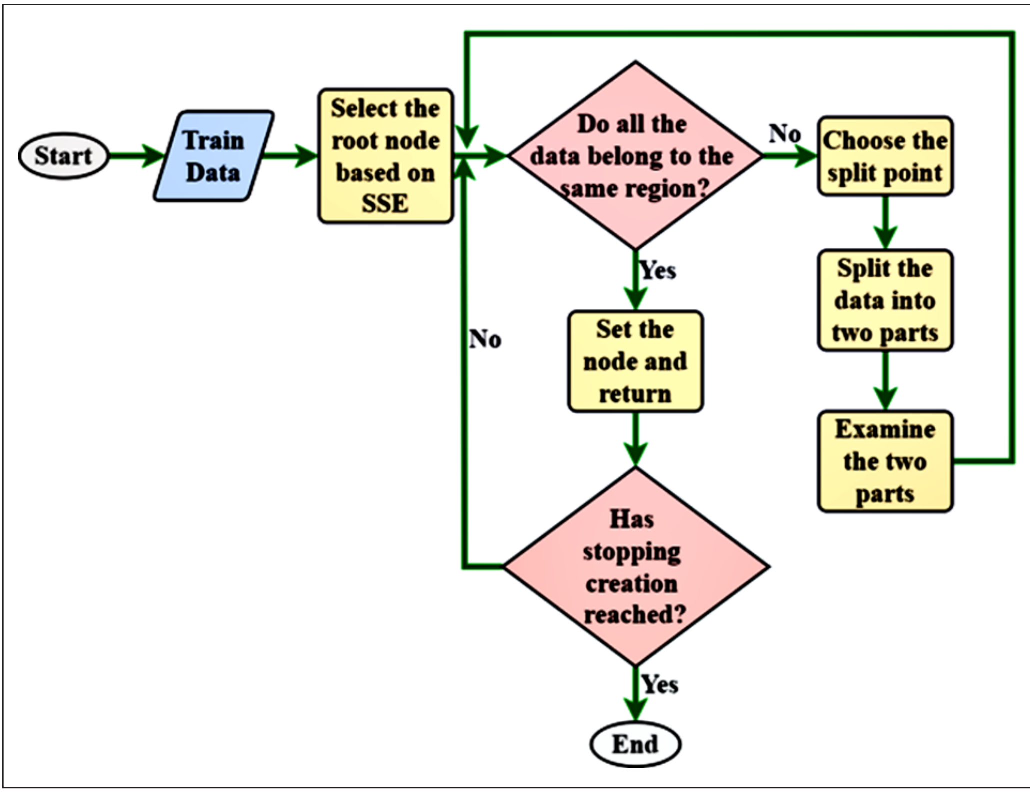

Figure 4. Regression tree flow diagram

architecture of the ANFIS consists of (1) fuzzification, where crisp input variables are converted into fuzzy membership functions, (2) rule inference, where Sugeno-type fuzzy rules are applied, (3) normalization, where rule strength is adjusted, (4) defuzzification, which maps fuzzy results to a continuous output, and (5) final summation, which aggregates rule-based outputs to provide the final prediction. The training process employs a hybrid optimization approach combining the gradient descent method and the least squares estimator (LSE) to adjust membership function parameters iteratively. Figure 5 presents a structured visualization of the ANFIS model, illustrating the connectivity between layers. These expanded descriptions provide a deeper understanding of the predictive models used in this study, supporting their effective implementation in carbon footprint estimation.

To enhance clarity and reproducibility, we have included a detailed step-by-step explanation of the RTrees model used for carbon footprint estimation in residential construction, as shown in Figure 6. The implementation begins with data preprocessing, where missing values are handled, numerical variables are normalized, and categorical variables are encoded. The dataset is divided into 80% training and 20% testing sets to ensure a fair evaluation. Once the data is prepared, the regression tree model is constructed

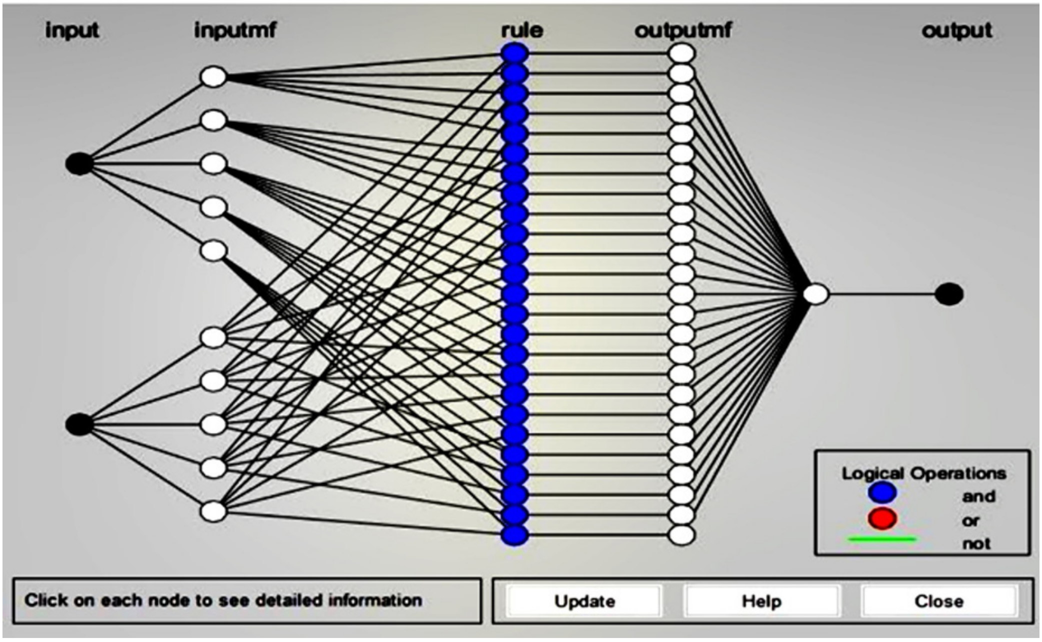

Figure 5. The foundational principle of the Adaptive Neuro-Fuzzy Inference Systems (ANFIS) architecture

using a recursive partitioning approach. The root node is selected based on the optimal splitting criterion, typically minimizing variance within child nodes. The dataset is progressively divided into smaller subsets until a stopping condition is met, such as reaching a minimum leaf size. After the initial tree construction, a pruning process is applied to eliminate branches that do not contribute to improved performance, reducing overfitting and enhancing model generalization. Following the model training, the RTrees model is evaluated using key performance metrics, including the RMSE, Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). The final trained model is then applied to predict carbon footprint values across different life cycle stages of construction, providing insights into emission trends.

The ANFIS model integrates fuzzy logic with artificial neural networks to enhance predictive accuracy in carbon footprint estimation. The ANFIS implementation begins with data preprocessing, where input variables such as material usage, transportation distances, energy consumption, and demolition waste are normalized to ensure consistency, as shown in Figure 7. Unlike RTrees, which rely on decision trees, ANFIS employs a fuzzy inference system (FIS) to model nonlinear relationships within the dataset. During the model development phase, triangular membership functions are used to define fuzzy rules, allowing the system to interpret input variables more effectively. The training employs a hybrid optimization technique, combining the gradient descent method with the least squares estimator to adjust membership function parameters iteratively. This adaptive

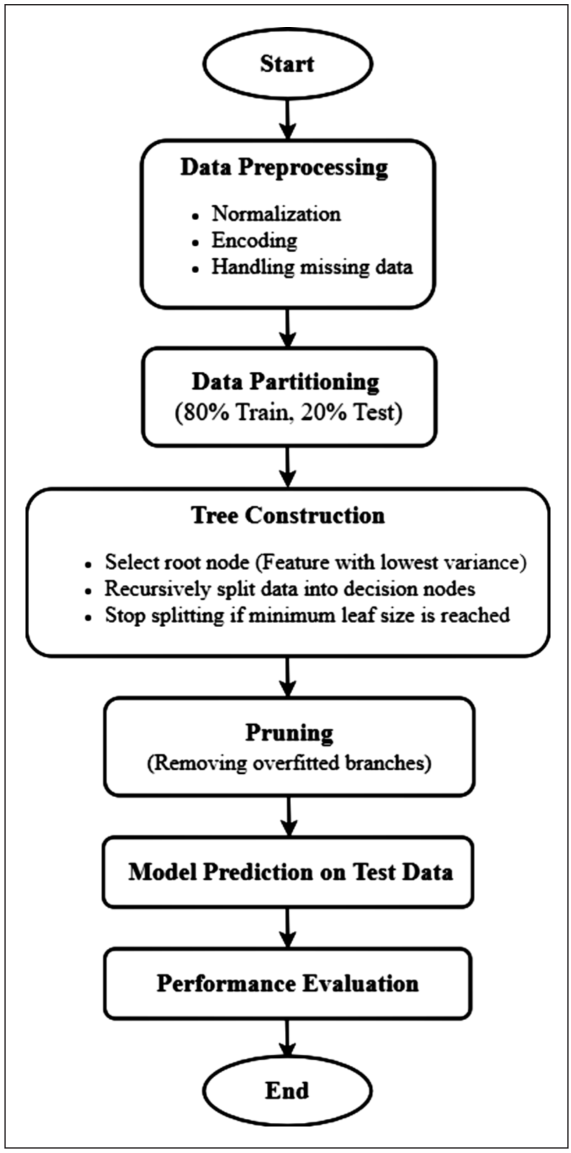

Figure 6. Regression trees implementation

learning process refines the model's generalization ability across different construction scenarios. After training, the ANFIS model is tested on the reserved dataset, and its predictive accuracy is assessed using RMSE, MSE, and MAPE. The ability of ANFIS to capture complex patterns in carbon emissions makes it particularly useful for analyzing environmental impacts at various construction stages. By leveraging the strengths of fuzzy logic and neural networks, the model provides a robust framework for estimating carbon footprints with high precision.

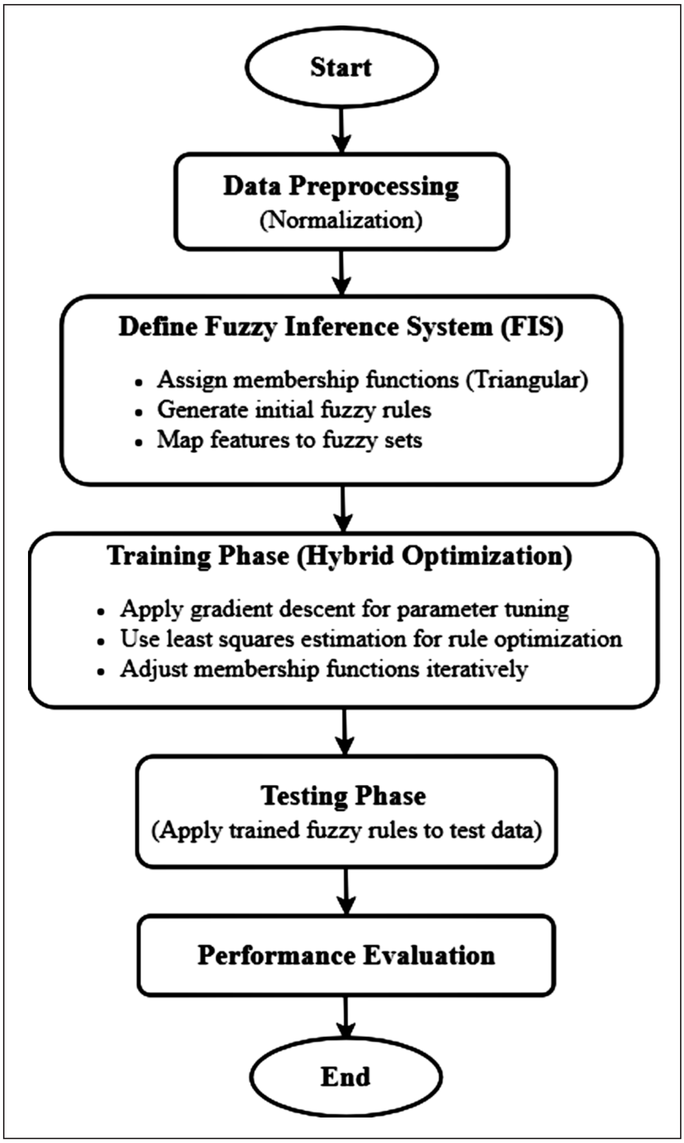

Figure 7. Adaptive neuro-fuzzy inference system implementation

Regression Trees

The divergence between statistical modeling and machine learning lies in their methodologies and objectives. The statistical models employ explicit mathematical equations to represent relationships between variables, estimating population parameters from sample data (Selvan & Balasundaram, 2021). In contrast, the machine learning autonomously extracts predictive patterns from data, bypassing the need for predefined rules or assumptions (Ramon et al., 2024). Both frameworks utilize training and testing datasets, but machine learning incorporates hyperparameter tuning via validation datasets, enhancing its capacity to analyze both small and large datasets with superior predictive accuracy. Its ability to learn directly from historical data makes it a powerful tool for data-driven predictions. Regression trees, among the diverse machine learning methodologies, especially the classification and regression trees (CART) technique, are recognized for their transparency, minimal preprocessing requirements, and resilience to outliers and incomplete data (Mienye & Jere, 2024). Unlike traditional approaches, CART models can analyze relationships without data normalization, making them particularly suitable for complex, multifaceted datasets.

This study leverages the regression tree analysis to estimate carbon emissions within LCA boundaries, focusing on parameters such as construction material usage and transportation distances. The model was implemented in MATLAB using the fine tree algorithm, which systematically partitions the data based on key lifecycle features of construction projects. Figure 4 visualizes the structure and decision-making process of the regression tree. To optimize the model's performance, hyperparameter tuning was conducted, with particular attention in refining the minimum leaf size for terminal nodes, thereby enhancing the model's predictive capabilities.

The regression trees provide a robust alternative to linear methods, excelling in modeling non-linear and intricate relationships by leveraging decision trees to predict continuous response variables. The methodology partitions data recursively into smaller, homogeneous regions, assigning each region a constant value equal to the average response. This recursive partitioning, performed in a top-down greedy manner, optimizes splits to minimize the sum of squares error, as formalized in Equation 1:

$$minimize\left\{SSE = \sum_{i\in R_1}\left(x_i - \hat{y}_{R_1}\right)^2 + \sum_{i\in R_2}\left(x_i - \hat{y}_{R_2}\right)^2 + \right\} \quad [1]$$

In this context, x_i represents the actual carbon emissions, and $\hat{y}_{R_1}$ and $\hat{y}_{R_2}$ signify the predicted emissions for regions 1 and 2, respectively. The regression trees are particularly useful for exploratory data analysis due to their simplicity and interpretability. However, they are prone to overfitting, which can limit their generalization capabilities and predictive accuracy when compared to more advanced machine learning techniques.

Adaptive Neuro-Fuzzy Inference Systems

The adaptive neuro-fuzzy inference system is a cutting-edge neuro-fuzzy architecture that models complex and nonlinear systems (Chopra et al., 2021). The ANFIS uniquely integrates the adaptive learning capabilities of neural networks with the interpretability of fuzzy logic, leveraging Sugeno's first-order adaptive fuzzy inference model (Mamat, Kasa, & Razali, 2019). Its hybrid optimization methodology, combining gradient descent and least-squares algorithms, enables precise parameter adjustment, making it a powerful tool for capturing and modeling sophisticated system behaviors (Karaboga & Kaya, 2019).

This study implemented the ANFIS model using MATLAB to address the complexities and uncertainties associated with carbon footprint data. By integrating the strengths of fuzzy logic and neural networks, ANFIS is particularly effective for modeling nonlinear and uncertain systems such as carbon footprint prediction. The model was meticulously fine-tuned by adjusting the membership functions, which are crucial for defining the degree of membership of data points within specific fuzzy sets.

A triangular membership function was selected due to its simplicity and computational efficiency, which are especially advantageous when working with large datasets or real-time applications. Unlike more complex functions, such as Gaussian or bell-shaped, triangular membership functions require fewer parameters, thereby reducing computational overhead and preserving flexibility for modeling uncertainties. This simplicity also enhances the interpretability of fuzzy rules, a key factor in ensuring the model's practicality in the construction industry. To improve predictive accuracy, the ANFIS model employed an iterative learning process to optimize the parameters of the triangular membership functions. The model refined these functions through repeated training cycles to better capture the complex patterns in carbon footprint data, leading to more reliable and accurate predictions. As shown in Figure 5, the ANFIS architecture comprises five distinct layers. The following equations determine the outputs of the first layer:

$$L1_i = \mu_{Ai}(x), \qquad i = 1,2 \qquad [2]$$

$$L1_i = \mu_{Bi}(y), \qquad i = 3,4 \qquad [3]$$

where, x is the input to node i, and A_i represents the linguistic function label. The terms $\mu_{Ai}(x)$ and $\mu_{Bi}(y)$ are Gaussian membership functions defined as:

$$\mu_{Ai}(x) = \exp\left[-\left(\frac{x - c_i}{a_i}\right)^2\right] \qquad [4]$$

where, a_i and c_i are the deviation and centre parameters of the membership function, respectively. The outputs of the second layer are computed as:

$$L2_i = w_i = \mu_{Ai}(x)\mu_{Bi}(y), \qquad i = 1,2 \qquad [5]$$

where w_i represents the strength of the rules. The third layer outputs are given by:

$$L3_i = \overline{w_i} = \frac{w_i}{w_1 + w_2}, \qquad i = 1,2 \qquad [6]$$

where, $\overline{w_i}$ are the normalized weights. In the fourth layer, the outputs are calculated as:

$$L4_i = \overline{w_i}(p_i x + q_i y + r_i), \qquad i = 1,2 \qquad [7]$$

where, p_i, q_i and r_i are known as consequent parameters. Finally, the output of the fifth layer, which represents the model's final output, is computed by summing the inputs as:

$$L5_i = \sum_{i=1}^{2} \overline{w_i}(p_i x + q_i y + r_i), \qquad i = 1,2 \qquad [8]$$

The ANFIS structure thus involves two types of parameters: premise parameters $\{a_i, c_i\}$ and consequent parameters $\{p_i, q_i, r_i\}$. These parameters are jointly optimized. ANFIS employs a hybrid algorithm for parameter optimization: the least-squares method is used to optimize the consequent parameters while keeping the premise parameters fixed, and the gradient descent method then tunes the premise parameters using the previously optimized consequent parameters.

Model Parameters and Hyperparameter Settings

To ensure reproducibility and enhance the interpretability of the models used in this study, Table 2 presents the key parameters and hyperparameter settings applied in the RTrees and ANFIS models. These parameters were selected based on an extensive hyperparameter tuning process to optimize prediction accuracy while maintaining computational efficiency. For the RTrees model, main parameters such as minimum leaf size, maximum tree depth, and pruning strategy were adjusted to minimize overfitting and enhance generalization. In contrast, the ANFIS model was fine-tuned by selecting appropriate membership functions, optimization algorithms, and epoch iterations to achieve the best trade-off between accuracy and computational efficiency.

These parameters were determined through experimental tuning to achieve optimal performance across all construction lifecycle stages. The results indicate that while RTrees excel in computational efficiency, ANFIS provides superior accuracy due to its adaptive learning process. The selected configurations allow for a balanced evaluation of both

Table 2
Model parameters and hyperparameter settings

Parameter	RTrees value	ANFIS value
Minimum leaf size	5	Not applicable
Maximum tree depth	15	Not applicable
Pruning method	Cost complexity	Not applicable
Membership function	Not applicable	Triangular
Optimization algorithm	Not applicable	Gradient descent and least squares
Number of epochs	Not applicable	100
Learning rate	Not applicable	0.01

Note. RTrees = Regression trees; ANFIS = Adaptive Neuro-Fuzzy Inference Systems

models, ensuring that the findings contribute valuable insights into machine learning applications for carbon footprint prediction in construction.

Model Evaluation Metrics

To evaluate the predictive accuracy of the machine learning models, this study utilized four widely recognized evaluation metrics: RMSE, MSE, MAPE, and mean squared logarithmic error (MSLE), as defined in Equations 9–12. These metrics are well-documented in the literature, and their detailed formulations and interpretations can be found in the corresponding references (Liu et al., 2022; Mamat, Kasa, Razali, et al., 2019). The evaluation metrics are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N}} \quad [9]$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \quad [10]$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{x_i - y_i}{y_i}\right| \times 100\% \quad [11]$$

$$MSLE = \frac{1}{N}\sum_{i=1}^{N}(\log(y_i + 1) - \log(x_i + 1))^2 \quad [12]$$

where, x_i are the simulated carbon emission values, y_i are the predicted carbon emission values, and N is the number of simulations.

RESULTS AND DISCUSSION

This study sought to evaluate and compare the performance of the RTrees and ANFIS in predicting carbon footprints across various stages of residential construction. Given the construction sector's significant contribution to global greenhouse gas emissions, developing precise and efficient carbon footprint estimation tools is imperative for advancing sustainability initiatives. This analysis provides a comprehensive assessment of each model's accuracy and robustness, highlighting their potential for enhancing predictive capabilities in the construction industry.

Performance of the RTrees in the Carbon Footprint Estimation

This section evaluates the RTrees model's performance across different LCA stages in residential construction. The primary focus is on the model's ability to predict carbon footprints with particular attention to the RMSE and MSE metrics across production, transportation, operation, and demolition stages (Table 3). During the production stage, the RTrees showed a training RMSE of 0.53174 and MSE of 0.51424, but test errors were significantly higher (RMSE = 4.57221, MSE = 4.44606), indicating possible overfitting issues during model training. This tendency for overfitting was observed in complex datasets, particularly where training data diversity is lacking (Heydari & Stillwell, 2024; Mamat et al., 2021). The notable discrepancy in test errors highlights difficulties in accurately predicting carbon emissions, suggesting the necessity for refining training methodologies to improve reliability across all stages.

In contrast, the transportation stage exhibited a more consistent generalization with lower discrepancies between the training and testing errors (training RMSE = 0.49223, test RMSE = 3.65314). Operational and demolition stages, however, reflected the model's limitations in capturing variable emissions effectively, with the operational stage showing a rise in test RMSE to 2.10647 despite a low training RMSE of 0.13865. The demolition

Table 3
The statistical error metric for the RTrees model

LCA stages	Cross validation fold	Leaf size	Training		Test	
			RMSE	MSE	RMSE	MSE
Production	50	2	0.53174	0.51424	4.57221	4.44606
Transportation	30	4	0.49223	0.44707	3.65314	3.30173
Operational	15	4	0.13865	0.11793	2.10647	3.90404
Demolition	10	8	0.06312	0.06104	3.00541	4.05205
Total carbon emission	50	4	0.17177	0.16086	3.13147	4.12267

Note. RTrees = Regression trees; LCA = Life cycle assessment; RMSE = Root Mean Square Error; MSE = Mean Squared Error stage presented a similar pattern with a low training RMSE of 0.06312 but a higher test RMSE (3.00541), underscoring challenges in predicting variability inherent in demolition processes. These findings highlight the RTrees' tendency to overfit training data, especially in the LCA stages characterized by high variability. Future enhancements could include integrating ensemble methods or hybrid models to address these limitations and boost predictive accuracy (Seghetta & Goglio, 2020; Yıldız & Beyhan, 2025).

ANFIS Model Performance in the Carbon Footprint Estimation

This section examines the performance of the ANFIS across various LCA stages in residential construction. Performance metrics, including the RMSE and MSE, were utilized for both training and testing phases, as detailed in Table 4. The ANFIS demonstrated high accuracy during training, particularly in the production and operational stages (RMSE = 0.08346 and 0.08265, respectively). However, it exhibited significant increases in test errors across all stages, with the operational stage test RMSE escalating to 5.55423. This pronounced discrepancy between the training and testing performance suggests potential overfitting, which is a common issue with models trained on highly variable data sets (Srivastava et al., 2023; Yelghi, 2024).

Table 4

The error statistic for ANFIS using triangular membership functions

LCA stages	Training		Test	
	RMSE	MSE	RMSE	MSE
Production	0.08346	0.03974	3.90782	4.33421
Transportation	0.03172	0.03557	4.74549	5.04436
Operational	0.08265	0.00644	5.55423	1.39193
Demolition	0.02485	0.03573	4.14777	3.50365
Total of carbon emission	0.05213	0.02618	4.32182	3.80569

Note. ANFIS = Adaptive Neuro-Fuzzy Inference Systems; LCA = Life cycle assessment; RMSE = Root Mean Square Error; MSE = Mean Squared Error

The model's limited ability to generalize beyond the training data highlights the need for refining the ANFIS framework to improve its robustness. Proposed methods include incorporating regularization techniques and expanding the training dataset to enhance the model's exposure to diverse construction scenarios, thereby improving its predictive accuracy for carbon emissions (Dosdoğru, 2019). Moreover, the consistent overperformance of ANFIS in training relative to testing across stages indicates the model's sensitivity to the specific characteristics of the training data. This necessitates further investigation into the model's parameter tuning and the adoption of ensemble techniques, which could mitigate the observed overfitting and support more reliable carbon footprint predictions.

Comparative Performance of the RTrees and ANFIS in the Carbon Footprint Prediction

This section presents a comparative analysis of the RTrees and ANFIS models in estimating the carbon emissions during various stages of the residential construction life cycle. The comparison utilizes the RMSE and MAPE values, detailed in Figures 9 and 10, to assess the models' relative strengths and weaknesses. ANFIS consistently outperforms RTrees in percentage-based accuracy (MAPE), especially notable in the production and transportation stages. As illustrated in Figure 8, this superior performance in the production stage is clearly visible, with ANFIS showing a consistently lower MAPE than RTrees. This superior performance is attributed to ANFIS's ability to handle complex nonlinear relationships within the dataset, providing a more accurate prediction of carbon emissions (Rajab, 2019). Conversely, RTrees demonstrate better adaptability in managing logarithmic errors (MSLE) in larger datasets, suggesting a potential advantage in scenarios where precise handling of small logarithmic discrepancies is crucial (Gu et al., 2016).

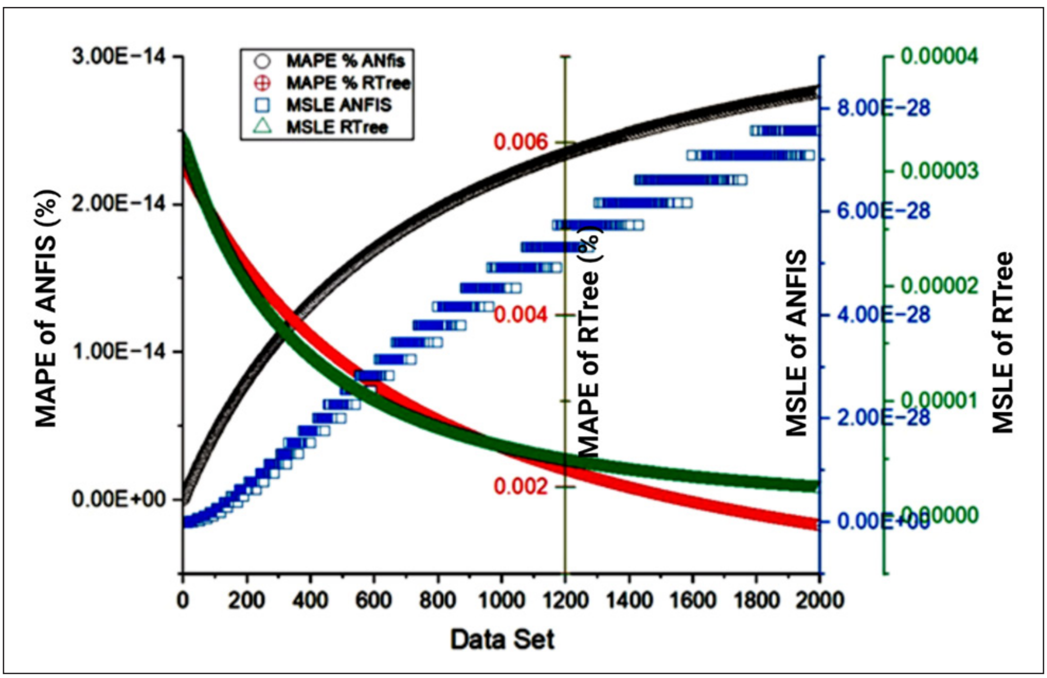

Figure 8. Performance of ANFIS and RTrees in production stage
Note. ANFIS = Adaptive Neuro-Fuzzy Inference Systems; RTree = Regression tree; MAPE = Mean Absolute Percentage Error; MSLE = Mean Squared Logarithmic Error

During the transportation stage, ANFIS showed a significant decline in MAPE, stabilizing at a lower value compared to RTrees, which gradually improved but stabilized

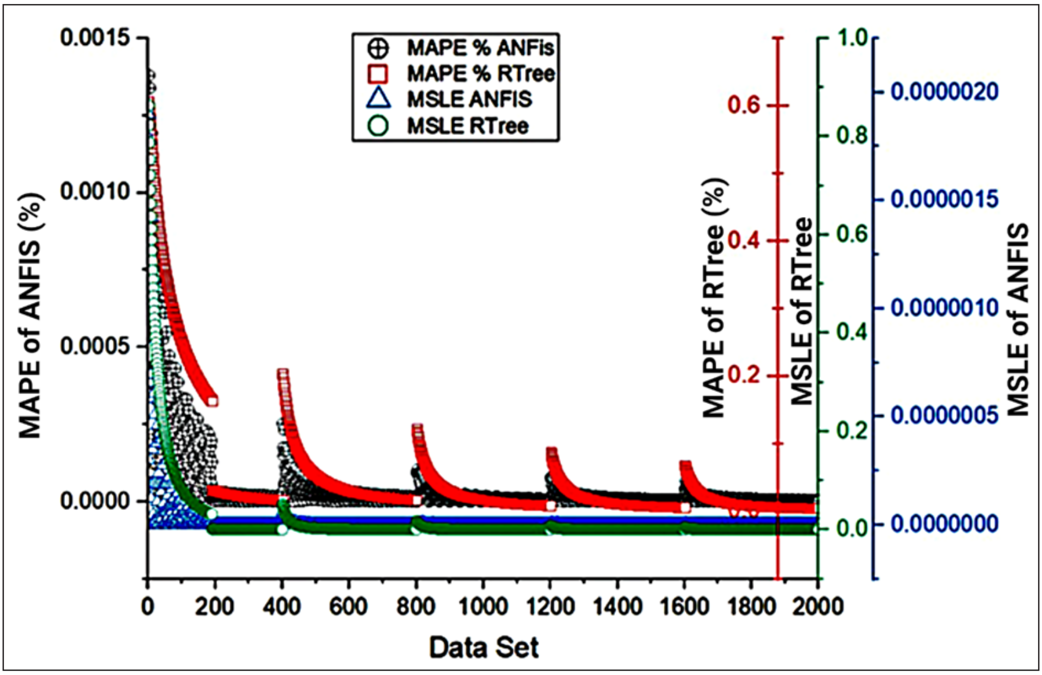

Figure 9. Comparative analysis of ANFIS and RTree in the transportation stage
Note. ANFIS = Adaptive Neuro-Fuzzy Inference Systems; RTree = Regression tree; MAPE = Mean Absolute Percentage Error; MSLE = Mean Squared Logarithmic Error

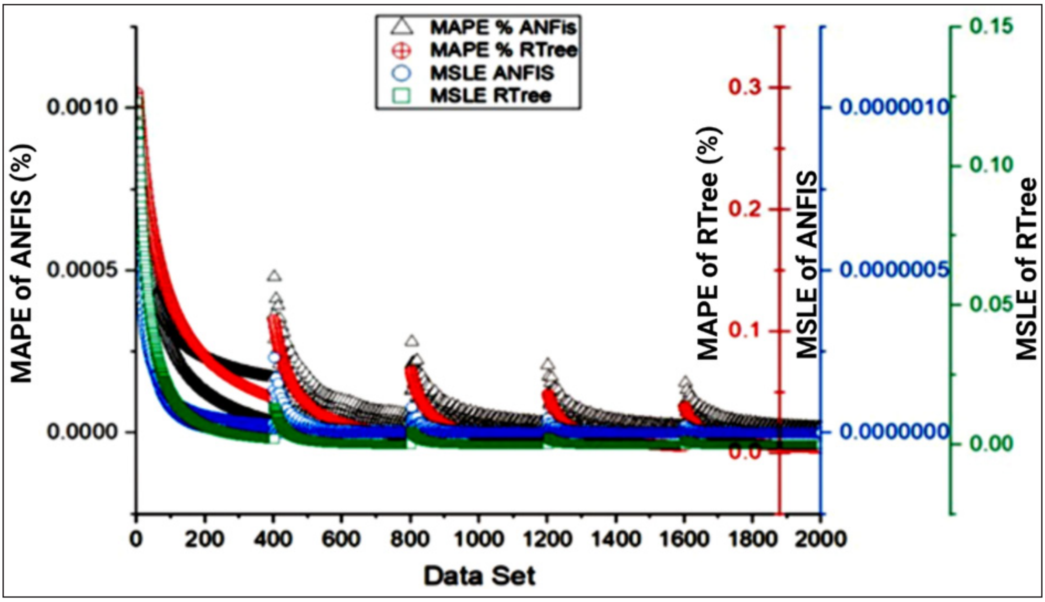

Figure 10. Operational stage performance of ANFIS and RTree
Note. ANFIS = Adaptive Neuro-Fuzzy Inference Systems; RTree = Regression tree; MAPE = Mean Absolute Percentage Error; MSLE = Mean Squared Logarithmic Error

at a higher MAPE (Figure 9). This trend underscores ANFIS's enhanced ability to minimize percentage-based prediction errors as the dataset size increases. Both models exhibited relatively low and comparable MSLE, indicating their effectiveness in managing small logarithmic errors in carbon emissions related to transportation. In the operational stage, while both models' MAPE values converged as dataset size increased, indicating an overall improvement in predictive accuracy with larger training datasets, RTrees marginally outperformed ANFIS, suggesting a slightly better capability in minimizing percentage-based errors in this specific phase (Figure 10).

Given these observations, the choice between the RTrees and ANFIS should be guided by the specific requirements of the project. The ANFIS is recommended for scenarios requiring high accuracy in percentage error minimization, whereas the RTrees may be preferred for its simplicity and effectiveness in managing logarithmic errors, particularly in large-scale applications where computational efficiency and interpretability are prioritized.

Impact of Dataset Size on Model Accuracy

This section investigates how the size of the dataset influences the accuracy of the RTrees and ANFIS models in estimating the carbon emissions, as depicted in Figures 11 and 12. Figure 12 provides a granular view of this trend, showing how both the MAPE and the MSLE for the ANFIS and RTrees stabilize as the dataset size increases, confirming that larger datasets contribute to more consistent model performance. Both models demonstrate improved accuracy with the expansion of the dataset, though they exhibit distinct performance trends. RTrees showed a rapid improvement in the MAPE as dataset size increased, indicating enhanced performance with larger data volumes (Figure 11). This trend suggests that the RTrees are particularly effective when processing extensive datasets, making them suitable for large-scale carbon emission estimation projects where data abundance can significantly influence predictive accuracy.

Conversely, while the ANFIS excels with smaller datasets, its generalization does not significantly improve as data volume increases. This finding is particularly relevant for scenarios demanding fine-grained accuracy in small-scale environments, where its hybrid neuro-fuzzy architecture effectively captures subtle nonlinear patterns in the carbon emission dynamics (Dzakiyullah et al., 2018). However, this strength also highlights a potential limitation in scaling to larger datasets without compromising predictive accuracy, a crucial factor for applications involving diverse and extensive inputs. Interestingly, both models maintained relatively low and stable MSLE values regardless of dataset size, suggesting their robust ability to manage small logarithmic errors effectively. This consistency is crucial for applications that require precise logarithmic error handling, such as detailed environmental impact assessments, where even minor inaccuracies can lead to significant deviations in policy planning and environmental compliance.

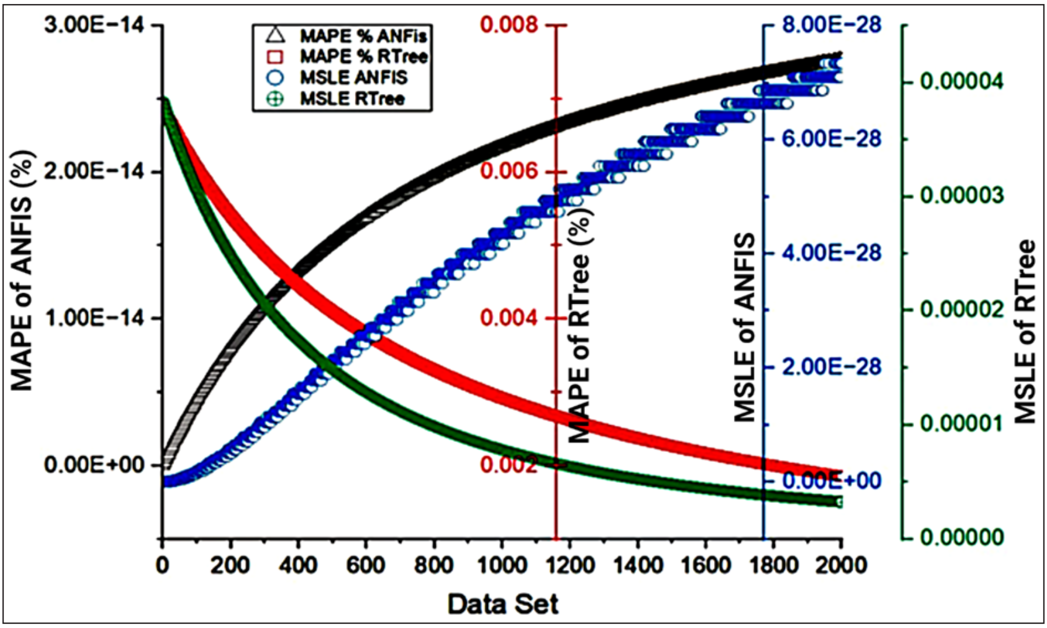

Figure 11. Performance of ANFIS and RTree in demolition stage
Note. ANFIS = Adaptive Neuro-Fuzzy Inference Systems; RTree = Regression tree; MAPE = Mean Absolute Percentage Error; MSLE = Mean Squared Logarithmic Error

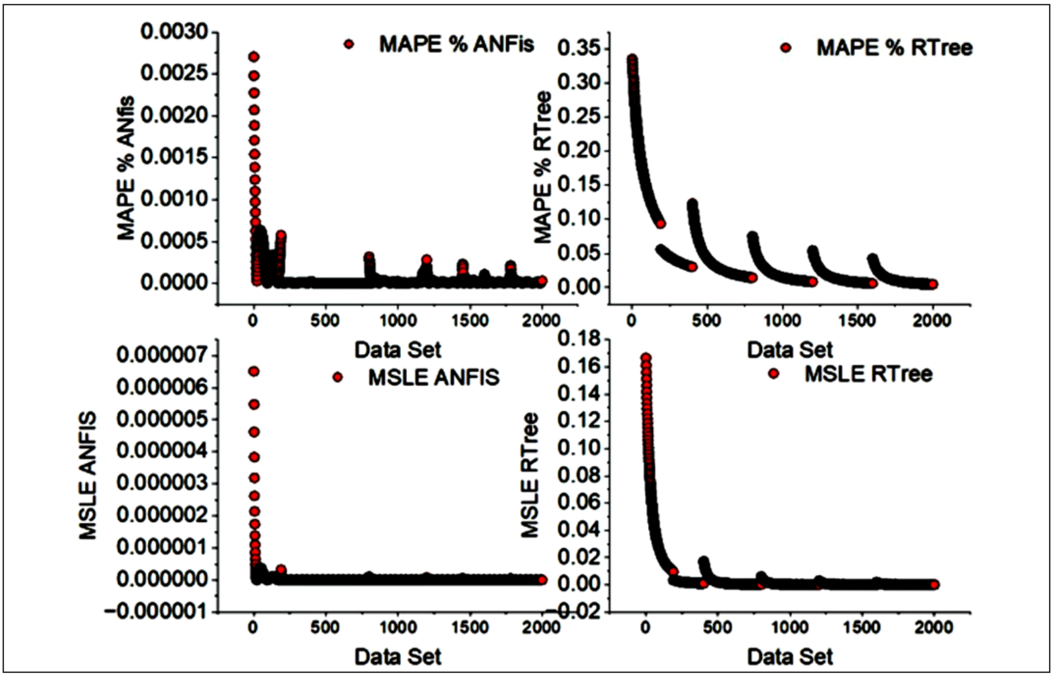

Figure 12. Overall carbon emission prediction accuracy of models
Note. ANFIS = Adaptive Neuro-Fuzzy Inference Systems; RTree = Regression tree; MAPE = Mean Absolute Percentage Error; MSLE = Mean Squared Logarithmic Error

The findings from this analysis emphasize the importance of selecting a machine learning model that aligns with specific data characteristics and project requirements. While the ANFIS may be preferred for smaller, more detailed datasets where high precision is necessary, the RTrees offers a robust option for broader applications where data volume and complexity are higher. These insights should guide the strategic selection of predictive models in construction projects to optimize carbon footprint assessments across various life cycle stages.

Computational Time Analysis

This section evaluates the computational efficiency of the RTrees and ANFIS models by measuring the time required for training and testing across various construction lifecycle stages, as detailed in Table 5. This analysis is crucial for assessing the practicality of deploying these models in real-world applications. The RTrees demonstrated a significantly lower training time than the ANFIS, indicating its computational efficiency and suitability for scenarios where quick model deployment is essential. However, despite the ANFIS's higher training times, which are attributable to its complex iterative optimization process, it consistently provided more accurate predictions, as shown in the enhanced carbon footprint estimation accuracy.

Table 5
Computational time for model training and testing

Model	Training time (s)	Testing time (s)
RTrees	12.4	1.8
ANFIS	76.2	2.5

Note. RTrees = Regression trees; ANFIS = Adaptive Neuro-Fuzzy Inference Systems

The testing times for both models were also analyzed, with the ANFIS showing greater stability across different datasets. This suggests that while the ANFIS requires more time to train, its predictions are reliable and consistent, making it suitable for applications where prediction accuracy is prioritized over computational speed. These results highlight a fundamental trade-off between the computational efficiency and predictive accuracy. The RTrees, with its quicker training times, offers an advantageous solution for applications needing rapid updates with less computational resource usage. Conversely, the higher accuracy of the ANFIS justifies its longer computational times, particularly in complex predictive tasks where precision is crucial.

Comprehensive Performance Evaluation

This section delivers a thorough evaluation of the RTrees and ANFIS models using multiple performance metrics: the RMSE, MSE, MAPE, and MSLE, which provide a comprehensive view of prediction accuracy, error distribution, and their capabilities in the carbon footprint estimation across various construction lifecycle stages (Table 6). The analysis demonstrated that the ANFIS consistently outperforms the RTrees in all error measures, achieving lower

Table 6
Performance metrics of RTrees and ANFIS models

Model	RMSE	MSE	MAPE (%)	MSLE
RTrees	0.5317	0.5142	8.23	0.0451
ANFIS	0.0834	0.0826	2.17	0.0598

Note. RTrees = Regression trees; ANFIS = Adaptive Neuro-Fuzzy Inference Systems; RMSE = Root Mean Square Error; MSE = Mean Squared Error; MAPE = Mean Absolute Percentage Error; MSLE = Mean Squared Logarithmic Error

RMSE, MSE, and MAPE values, which are particularly notable in the production and operational stages. This superior accuracy of the ANFIS underscores its effectiveness in precise carbon footprint predictions, where handling complex nonlinear relationships is essential for achieving high precision in sustainable construction practices.

Despite its lower performance in certain error metrics, the RTrees showed relatively better management of small logarithmic errors (MSLE), indicating its potential advantage in applications where handling of logarithmic discrepancies is critical. This makes the RTrees a viable alternative, offering computational efficiency with stable logarithmic error management, which is beneficial for large-scale applications or scenarios where rapid predictions are essential.

Comparison with Existing Studies

To assess the effectiveness of the proposed machine learning models, this study compares its results with previous research on carbon footprint estimation in construction. Several studies have investigated machine learning-based approaches for predicting the carbon emissions, yet direct comparisons between the RTrees and ANFIS models remain limited. For instance, Mamat et al. (2025) applied the Gaussian process regression (GPR), achieving an RMSE of 0.211 and a MAPE of 3.01%, which is significantly higher than the ANFIS model in this study (RMSE = 0.0834, MAPE = 2.17%). This demonstrates the superior accuracy of the ANFIS model in handling nonlinear relationships in carbon footprint estimation. Similarly, Kwon and Kim (2023) employed the deep neural networks (DNNs), obtaining an RMSE of 0.195 with relatively lower computational costs (16.5 sec training time), indicating that deep learning models can also achieve competitive performance in this domain.

However, when considering computational efficiency, the RTrees model in this study offers a significant advantage. With a training time of only 12.4 s, the RTrees outperform the ANFIS (76.2 s) and other deep learning models, such as the LSTMs (145 s) (Shao & Ning, 2023), in terms of speed, making them suitable for applications where rapid predictions are required. Nevertheless, this efficiency comes at the cost of higher error rates (RMSE

= 0.5317, MAPE = 8.23%), suggesting that while the RTrees provide a fast alternative, they may not be the best choice for high-accuracy applications.

Table 7 presents a comparative summary of various models used in previous studies alongside the results obtained in this research. These comparisons highlight that while the ANFIS achieves superior accuracy, it requires longer training time compared to the RTrees and DNN-based approaches. Meanwhile, the RTrees exhibit faster training and prediction times but with slightly higher errors compared to deep learning-based methods. Overall, this study contributes to the literature by demonstrating that the ANFIS offers a highly accurate solution for the carbon footprint estimation, particularly for complex, nonlinear datasets. However, the RTrees remain a viable alternative for applications requiring faster computation with reasonable accuracy. Future work should explore hybrid models that integrate the strengths of both the ANFIS and RTrees to achieve both efficiency and accuracy.

Table 7
Comparison of model performance with existing studies

Study	Model used	RMSE	MAPE (%)	Computational time (sec)
This study	ANFIS	0.0834	2.17	76.2 (train)
This study	RTrees	0.5317	8.23	12.4 (train)
Mamat et al. (2025)	GPR	0.211	3.01	110 (train)
Kwon and Kim (2023)	DNN	0.195	3.77	16.5 (train)
Shao and Ning (2023)	LSTM	4.984	0.024	145 (train)

Note. ANFIS = Adaptive Neuro-Fuzzy Inference Systems; RTrees = Regression trees; GPR = Gaussian Process Regression; DNN = Deep neural networks; LSTM = Long Short-Term Memory; RMSE = Root Mean Square Error; MAPE = Mean Absolute Percentage Error

Limitations and Future Considerations

While this study demonstrates the effectiveness of the RTrees and ANFIS in predicting the carbon footprint of residential construction, several limitations exist. The dataset, though comprehensive, is limited to a specific region, potentially affecting generalizability. Additionally, the ANFIS requires high computational power, making real-time applications challenging. The RTrees, though computationally efficient, exhibit lower predictive accuracy in complex scenarios. Another limitation is the lack of interpretability in the ANFIS, which may hinder industry adoption. Moreover, policy and environmental factors are not explicitly integrated into the models, limiting their adaptability to evolving sustainability regulations. Future research should focus on expanding datasets, optimizing computational efficiency, incorporating explainable AI techniques, and integrating policy-based factors to enhance model reliability. Additionally, real-time data from smart construction monitoring systems could further refine predictions. Addressing these challenges will improve machine learning applications for sustainable construction.

CONCLUSION

This study critically evaluated the predictive performance of the RTrees and ANFIS in estimating the carbon emissions across the life cycle stages of residential construction. The results demonstrate that the ANFIS consistently delivers higher predictive accuracy compared to the RTrees, with RMSE improvements exceeding 80% in key life cycle stages such as the production and operation. This superior performance makes the ANFIS an effective tool for precise carbon footprint estimation, aiding decision-makers in developing sustainable mitigation strategies. However, the higher computational demands of the ANFIS highlight the importance of balancing accuracy and processing efficiency in real-world applications. The findings emphasize the need for hybrid modeling approaches that combine the strengths of the RTrees and ANFIS to achieve both accuracy and efficiency. Future research should explore ensemble learning techniques, cloud-based deployment strategies, and real-time carbon footprint monitoring to enhance model applicability in the construction sector. These improvements align with evolving sustainability goals, paving the way for more data-driven decision-making in reducing environmental impacts.

ACKNOWLEDGEMENTS

The project has been financially supported by the TVET Applied Research Grant Scheme (Grant No. T-ARGS/2024/BK01/00076) from the Department of Polytechnic and Community College Education, Malaysian Ministry of Higher Education.

REFERENCES

Chen, L., Zhao, Y., Xie, R., Su, B., Liu, Y., & Renfei, X. (2023). Embodied energy intensity of global high energy consumption industries: A case study of the construction industry. *Energy*, *277*, 127628. https://doi.org/10.1016/j.energy.2023.127628

Chopra, S., Dhiman, G., Sharma, A., Shabaz, M., Shukla, P., & Arora, M. (2021). Taxonomy of Adaptive Neuro-Fuzzy Inference System in modern engineering sciences. *Computational Intelligence and Neuroscience*, *2021*(1), 6455592. https://doi.org/10.1155/2021/6455592

Dosdoğru, A. T. (2019). Improving weather forecasting using de-noising with maximal overlap discrete wavelet transform and GA based neuro-fuzzy controller. *International Journal on Artificial Intelligence Tools*, *28*(3), 1950012. https://doi.org/10.1142/S021821301950012X

Dzakiyullah, N. R., Saleh, C., Rina, F., & Fitra, A. R. (2018). Estimation of carbon dioxide emission using Adaptive Neuro-Fuzzy Inference System. *Journal of Engineering and Applied Sciences*, *13*(6), 5196 - 5202.

Farghaly, H. M., Ali, A. A., & Abd El-Hafeez, T. (2020). Building an effective and accurate associative classifier based on support vector machine. *Sylwan*, *164*(3), 39-56.

Gong, Y., Liu, G., Xue, Y., Li, R., & Meng, L. (2023). A survey on dataset quality in machine learning. *Information and Software Technology*, *162*, 107268. https://doi.org/10.1016/j.infsof.2023.107268

Gu, Y., Wylie, B. K., Boyte, S. P., Picotte, J., Howard, D. M., Smith, K., & Nelson, K. J. (2016). An optimal sample data usage strategy to minimize overfitting and underfitting effects in regression tree models based on remotely-sensed data. *Remote Sensing*, *8*(11), 943. https://doi.org/10.3390/rs8110943

Hasan, M. S., Tarequzzaman, M., Moznuzzaman, M., & Juel, M. A. A. (2025). Prediction of energy consumption in four sectors using support vector regression optimized with genetic algorithm. *Heliyon*, *11*(2), e41765. https://doi.org/10.1016/j.heliyon.2025.e41765

Heydari, Z., & Stillwell, A. S. (2024). Comparative analysis of supervised classification algorithms for residential water end uses. *Water Resources Research*, *60*(6), e2023WR036690. https://doi.org/10.1029/2023WR036690

Jia, W., Lu, M., Shen, Q., Tian, C., & Zheng, X. (2024). Dual generative adversarial networks based on regression and neighbor characteristics. *PLOS One*, *19*(1), e0291656. https://doi.org/10.1371/journal.pone.0291656

Karaboga, D., & Kaya, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: A comprehensive survey. *Artificial Intelligence Review*, *52*, 2263-2293. https://doi.org/10.1007/s10462-017-9610-2

Kim, E.-S. (2024). Can data science achieve the ideal of evidence-based decision-making in environmental regulation? *Technology in Society*, *78*, 102615. https://doi.org/10.1016/j.techsoc.2024.102615

Kwon, O. I., & Kim, Y. I. (2023). Review of machine learning for building energy prediction. *Journal of the Architectural Institute of Korea*, *39*(5), 133-140. https://doi.org/10.5659/JAIK.2023.39.5.133

Liu, Y., Zhang, W., Yan, Y., Li, Z., Xia, Y., & Song, S. (2022). An effective rainfall–ponding multi-step prediction model based on LSTM for urban waterlogging points. *Applied Sciences*, *12*(23), 12334. https://doi.org/10.3390/app122312334

Mahapatra, S. K., Schoenherr, T., & Jayaram, J. (2021). An assessment of factors contributing to firms' carbon footprint reduction efforts. *International Journal of Production Economics*, *235*, 108073. https://doi.org/10.1016/j.ijpe.2021.108073

Mamat, R. C., Kasa, A., & Razali, S. F. M. (2019). The applications and future perspectives of adaptive neuro-fuzzy inference system in road embankment stability. *Journal of Engineering Science and Technology Review*, *12*(5), 75-90.

Mamat, R. C., Kasa, A., Razali, S. F. M., Samad, A. M., Ramli, A., & Yazid, M. R. M. (2019). Application of artificial intelligence in predicting ground settlement on earth slope. In *AIP Conference Proceedings* (Vol. 2138, No. 1, p. 040015). AIP Publishing. https://doi.org/10.1063/1.5121094

Mamat, R. C., Ramli, A., & Bawamohiddin, A. B. (2025). A comparative analysis of gaussian process regression and support vector machines in predicting carbon emissions from building construction activities. *Journal of Advanced Research in Applied Mechanics*, *131*(1), 185 - 196. https://doi.org/10.37934/aram.131.1.185196

Mamat, R. C., Ramli, A., Samad, A. M., Kasa, A., Razali, S. F. M., & Omar, M. B. H. C. (2021). Artificial neural networks in slope of road embankment stability applications: A review and future perspectives. *International Journal of Advanced Technology and Engineering Exploration*, *8*(75), 304-319. https://doi.org/10.19101/IJATEE.2020.762127

Marsh, E., Allen, S., & Hattam, L. (2023). Tackling uncertainty in life cycle assessments for the built environment: A review. *Building and Environment*, *231*, 109941. https://doi.org/10.1016/j.buildenv.2022.109941

Mienye, I. D., & Jere, N. (2024). A survey of decision trees: Concepts, algorithms, and applications. In *IEEE Access* (Vol. 12, pp. 86716-86727). IEEE. https://doi.org/10.1109/ACCESS.2024.3416838

Pillai, R., Sharma, N., & Gupta, R. (2023). Fine-tuned EfficientNetB4 transfer learning model for weather classification. In *2023 3rd Asian Conference On Innovation In Technology* (pp. 1-6). IEEE. https://doi.org/10.1109/ASIANCON58793.2023.10270698

Rajab, S. (2019). Handling interpretability issues in ANFIS using rule base simplification and constrained learning. *Fuzzy Sets and Systems*, *368*, 36-58. https://doi.org/10.1016/j.fss.2018.11.010

Ramon, Y., Martens, D., Evgeniou, T., & Praet, S. (2024). Can metafeatures help improve explanations of prediction models when using behavioral and textual data? *Machine Learning*, *113*, 4245-4284. https://doi.org/10.1007/s10994-021-05981-0

Seghetta, M., & Goglio, P. (2020). Life cycle assessment of seaweed cultivation systems. In K. Spilling (Ed.), *Biofuels from algae: Methods and protocols* (Vol. 1980, pp. 103-119). Humana. https://doi.org/10.1007/7651_2018_203

Selvan, C., & Balasundaram, S. R. (2021). Data analysis in context-based statistical modeling in predictive analytics. In *Handbook of research on engineering, business, and healthcare applications of data science and analytics* (pp. 96-114). IGI Global. https://doi.org/10.4018/978-1-7998-3053-5.ch006

Sergeev, A., Buevich, A., Shichkin, A., Baglaeva, E., Subbotina, I., & Sergeeva, M. (2022). Comparing the types of artificial neural networks to predict the carbon dioxide concentration changes. In *AIP Conference Proceedings* (Vol. 2425, No. 1, p. 110007). AIP Publishing. https://doi.org/10.1063/5.0081641

Shao, C., & Ning, J. (2023). Construction and application of carbon emission prediction model for China's textile and garment industry based on improved WOA-LSTM. *Journal of Beijing Institute of Fashion Technology*, *43*(4), 73-81. https://doi.org/10.16454/j.cnki.issn.1001-0564.2023.04.010

Srivastava, A. K., Rathore, J. S., & Shrivastava, S. (2022). Efficacy of ANN and ANFIS as an AI technique for the prediction of COF at finger pad interface in manipulative tasks. In A. Kumar, M. Zunaid, K. A. Subramanian, & H. Lim (Eds.), *Recent Advances in Materials, Manufacturing and Thermal Engineering* (pp. 13-21). Springer. https://doi.org/10.1007/978-981-19-8517-1_2

Vinayak, N., & Ahmad, S. (2023). Sample size estimation for effective modelling of classification problems in machine learning. In I. Woungang, S. K. Dhurandher, K. K. Pattanaik, A. Verma, & P. Verma (Eds.), *Advanced Network Technologies and Intelligent Computing* (Vol. 1798, pp. 365-378). Springer. https://doi.org/10.1007/978-3-031-28183-9_26

Yao, X., Zhang, H., Wang, X., Jiang, Y., Zhang, Y., & Na, X. (2024). Which model is more efficient in carbon emission prediction research? A comparative study of deep learning models, machine learning models, and econometric models. *Environmental Science and Pollution Research*, *31*, 19500-19515. https://doi.org/10.1007/s11356-024-32083-w

Yelghi, A. (2024). Estimation single output with a hybrid of ANFIS and MOPSO_HS. *Sakarya University Journal of Computer and Information Sciences*, *7*(1), 112-126. https://doi.org/10.35377/saucis...1414742

Yıldız, M. E., & Beyhan, F. (2025). Prediction of cooling load via machine learning on building envelope design parameters. *Journal of Building Engineering*, *100*, 111724. https://doi.org/10.1016/j.jobe.2024.111724

Yiming, L., Junhan, Y., Zhang, Z., Peiqi, X., & Nianxiong, L. (2024). A comparative study of machine learning algorithm models for predicting carbon emissions of residential buildings in cold zones. *Journal of Tsinghua University*, *64*(10), 1734-1745. https://doi.org/10.16511/j.cnki.qhdxxb.2024.22.031