https://doi.org/10.48047/AFJBS.6.13.2024. 2771-2779

African Journal of Biological Sciences

Prediction of Concrete Compressive Strength using Machine Learning and Deep Learning Algorithms

D.N.V.S Vamsi
Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
T.Santi Sri, sri_santhi2003@yahoo.com
Professor, Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.

Article History
Volume 6, Issue 13, 2024
Received: 18June 2024
Accepted: 02July 2024
doi:*10.48047/AFJBS.6.13.2024.* 2771-2779

Abstract:-This document presents a comprehensive analysis project focused on predicting the compressive strength of concrete. The study utilizes a dataset comprising various concrete mix constituents and curing conditions. The objective is to employ four distinct machine learning algorithms: Support Vector Machine (SVM), Decision Tree (DT), Linear Regression, Random Forest (RF), and neural networks, to forecast concrete strength accurately. This predictive modeling is of significant importance in the fields of civil engineering and construction materials science, where precise estimation of concrete strength is crucial for ensuring structural integrity and longevity of infrastructure. By reliably predicting concrete strength, engineers and construction professionals can optimize design parameters, enhance material selection processes, and ultimately bolster the resilience and durability of buildings, bridges, roads, and other vital structures. The inclusion of neural networks expands the scope of the analysis, leveraging their ability to capture complex patterns and relationships within the data, thereby potentially improving the accuracy of concrete strength predictions.

Keywords: Linear Regression, Neural Networks, Regression modelling, Support vector machine

D.N.V.S Vamsi /Afr.J.Bio.Sc. 6(13)(2024).2771-2779

1. Introduction

Concrete is the backbone of modern infrastructure, serving as the primary building material for a wide range of structures, from skyscrapers to highways [1]. Its compressive strength, or the ability to withstand pressure, is a critical parameter in determining the safety and longevity of these constructions. Engineers have traditionally relied on empirical formulas and extensive testing to estimate concrete strength [2], but these methods are often timeconsuming, costly, and may lack accuracy.Recent advances in machine learning have transformed predictive modelling in a number of fields, opening up new possibilities for previously unheard-of levels of efficiency and accuracy for estimating concrete strength. Through the examination of extensive datasets that include curing conditions and the components of concrete mixes, machine learning algorithms are able to identify intricate patterns and relationships that impact the strength of concrete.This paper outlines a thorough analysis project that aims to estimate concrete compressive strength by utilising machine learning. The objective is to produce precise predictions of concrete strength by utilising four different machine learning algorithms and one deep learning algorithm: SVM, DT, RF [22] and LR [4] and Artificial Neural Networks (ANN) [3]. These predictions rely on various input parameters, including cement, aggregate, water-cement ratio, temperature, and curing time. [5].This predictive modelling project is extremely important for the domains of building materials science and civil engineering, even beyond its theoretical implications. Engineers can optimise structural designs, choose suitable materials, and guarantee the longevity and safety of infrastructure projects by using precise estimates of concrete strength [6]. Furthermore, this study creates opportunities for innovation and optimisation in the design of concrete mixes and construction techniques by identifying the fundamental elements determining concrete strength. By this project, we hope to improve the sustainability of the built environment while also advancing predictive modelling methods in the field of civil engineering. The importance of this predictive modelling project resonates strongly throughout the fields of civil engineering and building materials science, even beyond its theoretical ramifications [7,8] . By using accurate predictions of concrete strength, engineers can assure the longevity and safety of infrastructure projects, optimise structural designs, and choose the right materials [9]. Furthermore, by identifying the essential components controlling concrete strength, this work opens the door for creativity and optimisation in the design of concrete mixes and building methods [10].

2. Methodology

Data preparation is vital in machine learning, especially in the tasks such as prediction and detection. UCI Concrete Compressive Strength Dataset was used in this project. Initially, the dataset contained 5000 data points [11]. However, to ensure the quality and relevance of our analysis, we conducted extensive preprocessing steps. These steps included data cleaning, outlier removal, and feature selection, resulting in a refined dataset comprising 1031 datapoints.The dataset encompasses various quantities of materials used in concrete production including cement, aggregate and fly ash or slag. Additionally, it includes information on curing time, which has a significant role in the development of concrete strength. Figure-1 shows the work flow diagram of methodology [12].

a) Decision Tree

DTs recursively split the dataset depending on feature values, creating a tree-like structure where every node indicates a prediction.In the case of concrete strength prediction, decision trees partition the dataset based on features such as cement content, aggregate size, and curing time, aiming to minimize the variance in compressive strength within each subset. Once the tree is trained, new concrete mixtures can be input into the tree to predict their compressive strength based on the path they take through the tree. Sklearn library is used for importing Decision trees [18].

b) Support Vector Machine

SVM identifies the optimal hyperplane that divides datapoints into distinct classes or makes continous output predictions.In the context of concrete strength prediction, SVM learns a mapping from input features (e.g., cement content, aggregate size, curing time) to the corresponding compressive strength [21].By finding the optimal hyperplane that maximizes the margin between data points representing different concrete mixtures, SVM aims to generalize well to unseen data and make accurate predictions [12,17].

c) Random Forest

Random Forest is and ensemble learning method that uses multiple decision trees to generate predictions. When applied to predicting concrete strength, Random Forest (RF) constructs a collection of decision trees by randomly selecting different subsets of the training data and features for each tree. [13]. Each tree in the forest predicts the compressive strength of a concrete mixture based on its input features independently. The prediction which is obtained by taking average of the predictions of all trees in the forest will be the final prediction, resulting in predictions that are more resilient and accurate than those made by individual decision trees. [15].Random Forest provides a built-in mechanism for feature importance estimation, allowing us to identify which input features are impacting concrete compressive strength prediction significantly. This information can provide valuable insights into the complex interplay of factors that influence the strength of concrete [18].

d) Linear Regression

Linear regression offers simplicity, interpretability, and ease of implementation, making it a valuable tool for concrete strength prediction. However, it operates under the assumption of a linear relationship between independent and dependent variables, which may not always be valid in intricate real-world scenarios. In such cases, more advanced ML techniques, such as Decision trees, random forests, or neural networks, can be use to improved accuracy and predictive performance [5, 7]**.**

e) Neural Networks

These are capable of learning complex patterns and relationships in data through training, where the model adjusts its internal parameters (weights and biases) based on examples from the training data.Concrete strength prediction requires understanding the complex interactions between various mix constituents,curing conditions and environmental factors.Neural networks excel at learning these intricate patterns from large and diverse datasets,capturingnon linear relationships that may be challenging for traditional algorithms [14]. In real world scenarios, finding concrete strength is crucial for ensuring the structutal integrity and longevity of infrastructure projects.Neural networks have demostrated superior performance compared to traditiona; machine learning algorithms. As the training data is small in terms of neural network there is a high chance that overfitting might occur so we tried using generalization techniques to reduce the overfitting [19].

3) Training and Validation

Model training involves the process of feeding prepared data into a machine learning algorithm to allow it to discern various patterns and make predictions.. In his specific training scenario, 824 data points are utilized for updating the model's parameters, while 206 data points are held out for monitoring the model's performance and preventing overfitting. This partitioning strategy helps ensure that the model generalizes well to unseen data. To assess the effectiveness of the trained model, two key evaluation metrics are employed: R^2 Score and Root Mean Squared Error (RMSE).R^2 Score quantifies the proportion of the variance in the target variable that can be explained with input features. A higher R^2 score indicates that it is a better fit of the model to the data, indicating that it can more accurately capture the underlying relationships between the features and the target variable. RMSE measures the square root of the average squared difference between the predicted values and the actual values. It provides a comprehensive measure of the typical magnitude of errors in the model predictions, offering insights into the model's overall performance [20].

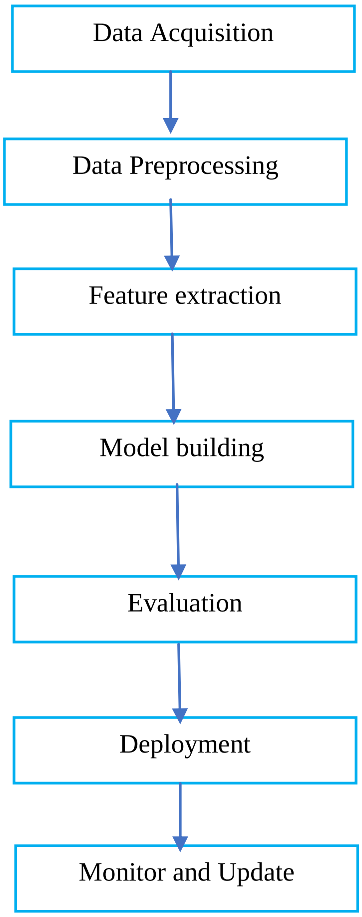

Figure-1 Methodology

4.Model Results using machine learning algorithms:

Training results of models are shown the table -1

Algorithm	Mean SquaredError	R^2 score
Support Vector	89.1999632	0.653
Decision Tree	42.581023	0.834
Random Forest	29.85441712	0.884
Linear Regression	0.83855717	0.014

Table-1 Training Results

The results of neural networks are shown in table-2

Iterations	epochs	mean RMSE	Std_RMSE
50	50	18.741	10.658
50	50(with normalization)	17.252	2.42
50	100	11.846	0.368

Table-2 Neural network training results

The results of neural network after addition of 3-hidden layers is shown in the table-3

Iterations	epochs	Mean RMSE	Std_RMSE
50	50	10.836	0.499

Table-3 Neural network after addition of hidden layers

5. Graphical Results

A graph showing the comparision of Mean Square error and R^2 score in different algorithms is shown in Figure1,2

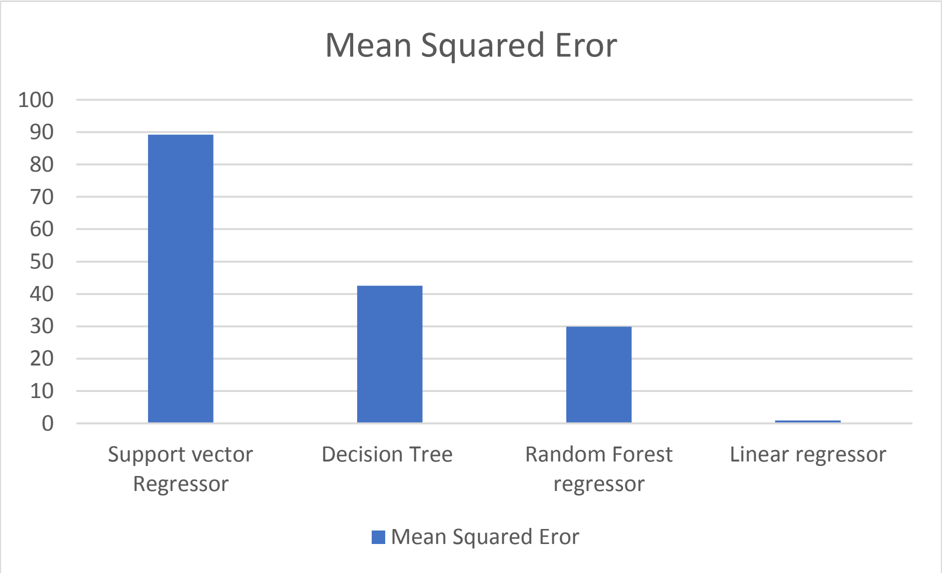

Figure-1 Comparision of Mean Squared Error

1) The Decision Tree Regressor demonstrates impressive performance on the training dataset, achieving a high accuracy of 0.952. Furthermore, its ability to generalize to new data is evident, as it maintains a strong score of 0.801 on the test set. This indicates its reliability in making predictions beyond the examples it was initially trained on, showcasing its practical utility.

2) The Random Forest Regressor exhibits outstanding accuracy on the training data, boasting an impressive R^2 score of 0.982. This underscores its capability to capture complex patterns within the concrete compressive strength during the training phase. Additionally, its robust performance extends to unseen data, as evidenced by a high R^2 score of 0.909 on the test data set. This highlights the model's effectiveness in delivering reliable predictions in real-world scenarios, surpassing its original training data

3) Support Vector Regression [16], employing a linear regression model, yields promising results with an MSE (Mean Squared Error) of 58.79 and an R^2 score of 77.92. These metrics provide insights into the model accuracy and its capability to explain variance in the data. The R^2 value of 0.772 denotes that it is a satisfactory fit of the model to data, capturing approximately 77.2% of the variance in the concrete compressive strength.

4) The linear regressormodel achieves an R2 score of 0.848 on the training dataset and 0.823 on the test dataset, indicating strong predictive performance. Cross-validation further confirms the model's consistency, with an average R2 score of approximately 0.839 across six folds. These results suggest that the model generalizes well to unseen data and effectively captures the relationship between the independent and dependent variables. Further analysis could involve examining the model's coefficients and identifying any outliers to enhance its predictive accuracy.

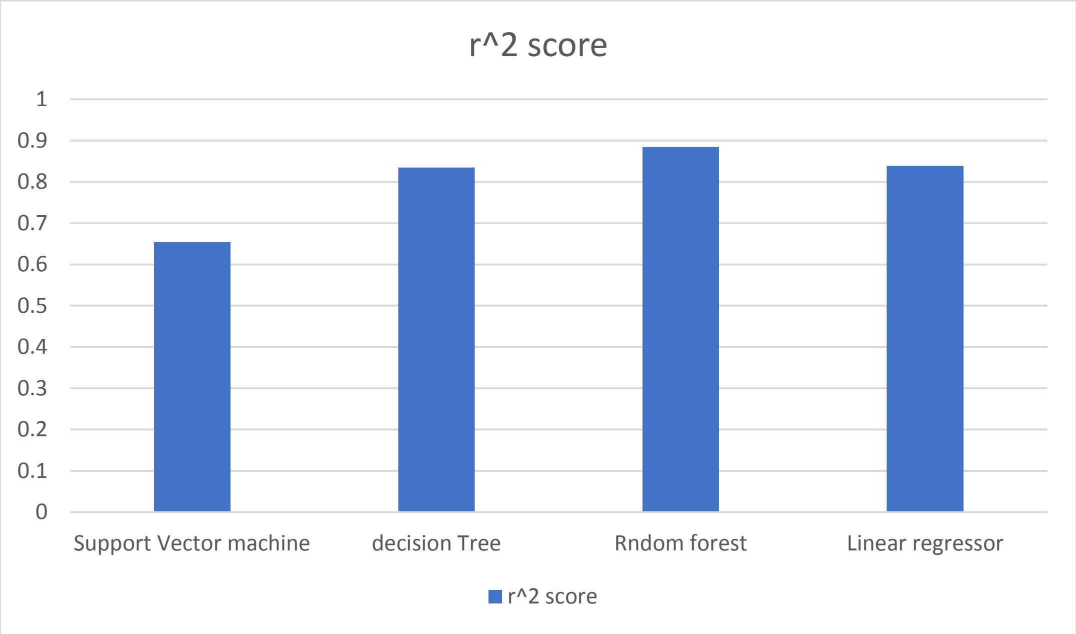

Figure-2 Comparision of r^2 score

6. Neural Network Implementation

The implementation of the neural network model using TensorFlow-Keras framework with a Sequentialarchitecture involves several key components and techniques to achieve accurate predictions of concrete compressive strength**.**

a) Model Architecture:

The model architecture consists of a Sequential model, which allows for a straightforward layer-by-layer construction of the neural network.Three hidden layers are incorporated, each containing 10 neurons. This architecture provides the model with the capacity to capture complex relationships within the data.ReLU (Rectified Linear Unit) activation function is applied to each neuron in the hidden layers. ReLU is preferred for its ability to introduce non-linearity into the model, enabling it to learn and represent intricate patterns in the data.A single neuron is utilized in the output layer since the task is a regression problem, aiming to predict a continuous numerical value (concrete compressive strength).

b) Regularization:

To mitigate overfitting, a dropout layer with a dropout rate of 0.2 is added after the last hidden layer. Dropout randomly deactivates a fraction of neurons during training, thereby preventing the network from relying too heavily on specific neurons and features.

c) Optimization and Loss Function:

The Adam optimizer is employed for model compilation. Adam optimizer is well-suited for training neural networks as it dynamically adjusts the learning rate during training, leading to faster convergence and improved performance.Mean Squared Error (MSE) serves as the loss function. MSE computes the average squared difference between the predicted values and the actual values, providing a measure of the model's performance in minimizing prediction errors.

d) Training Strategy:

Early stopping callback is utilized to monitor the validation accuracy during training. If the validation accuracy fails to improve after a specified number of iterations (e.g., 50 epochs), training is halted to prevent overfitting and unnecessary computational expense.The model is trained over multiple epochs (e.g., 100 epochs) to iteratively update the model parameters and improve its predictive capabilities.

e) Normalization:

To enhance training stability and convergence, data normalization is applied before training the model. Normalization ensures that input features have a similar scale, preventing certain features from dominating the learning process and enabling more efficient optimization.

f) Model Evaluation:

Mean RMSE (Root Mean Squared Error) and standard deviation RMSE are computed to evaluate the model's performance across multiple iterations. RMSE measures the typical magnitude of errors in the model's predictions, with lower values indicating better performance.The average and standard deviation of RMSE values provide insights into the consistency and accuracy of the model's predictions across different epochs.

g) Model Improvement:

Additional hidden layers are introduced to the model to further enhance its predictive capabilities. By increasing the depth of the neural network, the model gains the capacity to capture more intricate patterns in the data, leading to improved performance.The iterative process of experimentation and refinement aims to minimize RMSE and improve the accuracy of concrete compressive strength predictions.

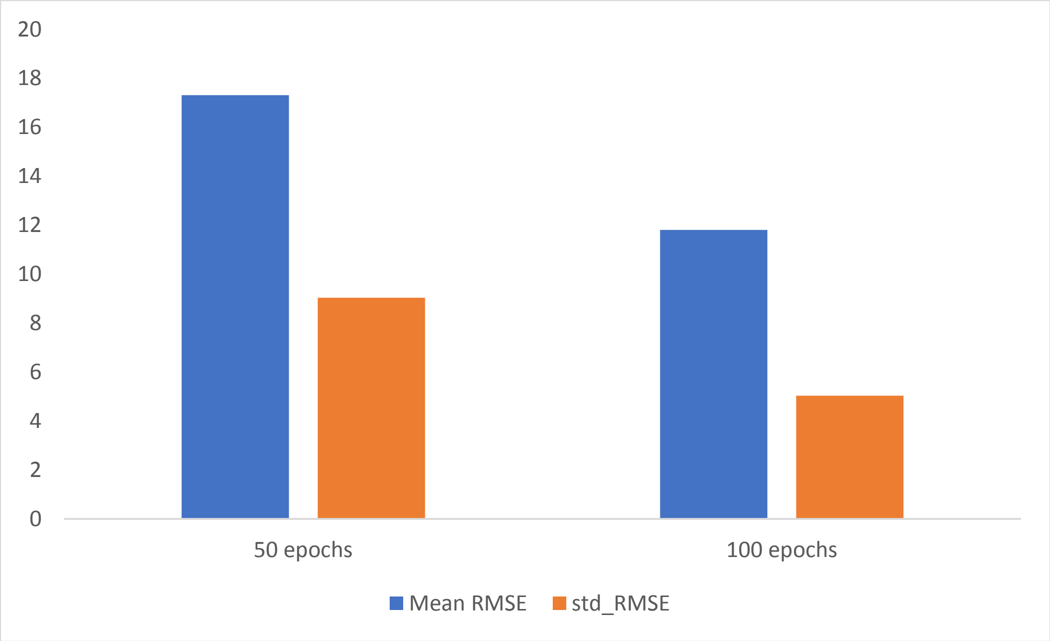

Figure-3 Neural network training results

Conclusion

The paper focuses on the dangers of using poor-quality construction materials and proposes a solution: a predictive model to assess material quality. Data from reputable sources like Kaggle and UCI archives are used to train the model. Neural networks are chosen for their ability to learn complex patterns. As the model is trained over more epochs, its performance improves. This highlights the iterative nature of neural network training. The model's enhanced performance demonstrates its effectiveness in assessing material quality, aiding decision-making in the construction industry and improving safety standards.

Futurescope

Moving forward, the project holds immense potential for further advancement and expansion. One avenue for exploration involves integrating advanced machine learning techniques, such as neural networks (NNs), which offer advantages over regular models due to their ability to capture complex patterns and relationships within the data. Neural networks excel in learning from large and diverse datasets, as they can adapt and develop more sophisticated representations of the underlying data structure with increasing data volume. Additionally, continued refinement of feature engineering and selection methods, as well as the incorporation of additional data sources such as environmental factors, could enrich the predictive model and its applicability in real-world scenarios. Deployment of the predictive models in practical settings, collaboration with industry partners, and expansion to include other construction materials beyond concrete also present exciting opportunities for innovation and impact in the construction industry. By embracing these future scopes, the project can continue to drive advancements in predictive modeling and contribute to safer, more sustainable infrastructure development.

References

1. Khan, M. A., Zafar, A., Akbar, A., Javed, M. F., & Mosavi, A. (2021). Application of Gene Expression Programming (GEP) for the prediction of compressive strength of geopolymer concrete. *Materials*, *14*(5), 1106.
2. Khan, M. A., Memon, S. A., Farooq, F., Javed, M. F., Aslam, F., & Alyousef, R. (2021). Compressive strength of fly-ash-based geopolymer concrete by gene expression programming and random forest. *Advances in Civil Engineering*, *2021*, 1-17.
3. Ahmad, A., Chaiyasarn, K., Farooq, F., Ahmad, W., Suparp, S., & Aslam, F. (2021). Compressive strength prediction via gene expression programming (GEP) and artificial neural network (ANN) for concrete containing RCA. *Buildings*, *11*(8), 324.
4. Wu, H., Okutani, I., & Xu, J. (2009, August). Identification of band reflectance and land-cover classification using fuzzy linear regression analysis. In *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery* (Vol. 3, pp. 114-118). IEEE.
5. Goetschalckx, R., Driessens, K., & Sanner, S. (2008, December). Cost-sensitive parsimonious linear regression. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 809-814). IEEE.
6. Dagdanpurev, S., Sun, G., Choimaa, L., Abe, S., & Matsui, T. (2018, July). Clinical application of multiple vital signs-based infection screening system in a mongolian hospital: Optimization of facial temperature measurement by thermography at various ambient temperature conditions using linear regression analysis. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5313-5316). IEEE.
7. Feng, Q., Yuan, C., Huang, J., & Li, W. (2015, September). Center-based weighted kernel linear regression for image classification. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 3630-3634). IEEE.
8. Sravani, B., & Bala, M. M. (2020, June). Prediction of student performance using linear regression. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-5). IEEE.
9. Momot, M., Momot, A., & Piekar, E. (2015, June). Robust estimation of respiratory rate based on linear regression. In *2015 Signal Processing Symposium (SPSympo)* (pp. 1-5). IEEE.
10. Zehtab-Salmasi, A., Feizi-Derakhshi, A. R., Nikzad-Khasmakhi, N., Asgari-Chenaghlu, M., & Nabipour, S. (2023). Multimodal price prediction. *Annals of Data Science*, *10*(3), 619-635.

11. Andi, H. K. (2021). An accurate bitcoin price prediction using logistic regression with LSTM machine learning model. *Journal of Soft Computing Paradigm*, *3*(3), 205-217.
12. Zheng, W., & Ye, Q. (2009, November). Sentiment classification of Chinese traveler reviews by support vector machine algorithm. In *2009 Third International Symposium on Intelligent Information Technology Application* (Vol. 3, pp. 335-338). IEEE.
13. Naderpour, H., Rafiean, A. H., & Fakharian, P. (2018). Compressive strength prediction of environmentally friendly concrete using artificial neural networks. *Journal of Building Engineering*, *16*, 213-219.
14. Kewalramani, M. A., & Gupta, R. (2006). Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks. *Automation in Construction*, *15*(3), 374-379.
15. BKA, M. A. R., Ngamkhanong, C., Wu, Y., & Kaewunruen, S. (2021). Recycled aggregates concrete compressive strength prediction using artificial neural networks (ANNs). Infrastructures 6: 1–20.
16. Huang, Y., Zhang, J., Ann, F. T., & Ma, G. (2020). Intelligent mixture design of steel fibre reinforced concrete using a support vector regression and firefly algorithm based multi-objective optimization model. *Construction and Building Materials*, *260*, 120457.
17. Mohammadi, M., Khorrami, M. K., Vatani, A., Ghasemzadeh, H., Vatanparast, H., Bahramian, A., & Fallah, A. (2021). Genetic algorithm based support vector machine regression for prediction of SARA analysis in crude oil samples using ATR-FTIR spectroscopy. *Spectrochimica Acta part A: Molecular and biomolecular spectroscopy*, *245*, 118945.
18. Idowu, O. S., & Lam, K. C. (2020). Conceptual quantities estimation using bootstrapped support vector regression models. *Journal of Construction Engineering and Management*, *146*(4), 04020018.
19. Ahmad, M. S., Adnan, S. M., Zaidi, S., & Bhargava, P. (2020). A novel support vector regression (SVR) model for the prediction of splice strength of the unconfined beam specimens. *Construction and building materials*, *248*, 118475.
20. Ly, H. B., Nguyen, T. A., & Tran, V. Q. (2021). Development of deep neural network model to predict the compressive strength of rubber concrete. *Construction and Building Materials*, *301*, 124081.
21. Ma, H., Liu, J., Zhang, J., & Huang, J. (2021). Estimating the compressive strength of cement-based materials with mining waste using support vector machine, decision tree, and random forest models. *Advances in civil engineering*, *2021*, 1-10.
22. Vamsi, D. N. V. S., & Mehrotra, S. (2023, April). Comparative Analysis of Machine Learning Algorithms for Predicting Mobile Price. In *International Conference on Information and Communication Technology for Intelligent Systems* (pp. 607-615). Singapore: Springer Nature Singapore.