iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59399

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Comparative Analysis of Ensemble Learning Techniques and ANN for Prediction of Compressive Strength of Concrete Mix

Avani Dedhia[1], Hima Soni[2], Dr. N.K. Arora[3]
[1]*Assistant Professor, Civil Engineering Dept., S.P.B.P.E.C., Saffrony Institute of Technology, India*
[2]*Data Scientist, Bacancy Software LLP, Ahmedabad, Gujarat, India*
[3]*Professor, Applied Mechanics Dept., G.E.C. Gandhinagar, India*

Abstract: ***In this study, non-linear machine learning regressors and an artificial neural network (ANN) were modeled to predict the compressive strength of concrete mixes. A dataset of 1030 mix design data points with a wide range of compressive strength varying from M10 to M80 is taken into account. The data set comprised of 8 input parameters namely- cement, water, fine & coarse aggregate, flyash, slag, superplasticizer, and age. The paper also analyzes effects on the prediction of compressive strength due to waiving of the parameters with lesser importance factors. To improve the performances of models by reducing the variance, biases, and errors, K-fold cross-validation was examined. To examine the performances of the models, metrics such as coefficient of determination (R2) and mean square error (MSE) were employed. Most of the models can predict the compressive strength with good accuracy. Out of the 7-machine learning regressors and ANN models employed, Gradient Boosting Regressor yielded the best prediction accuracy of 0.9561 with the lowest MSE value of 0.0513.***
Keywords: Concrete, Compressive Strength, Non-Linear Machine Learning Regressors, ANN

I. INTRODUCTION

Concrete is the most dominating material in majority of the civil engineering constructions and a very complex composite material. Hence it is imperative to understand its mechanical properties like compressive strength, durability, workability, and many more which are sensitive to changes in raw ingredients depending on the geographical location as well as the skill of the laborers involved, making the accuracy of their predictability very low. Several studies done by Yeh, I.C. & F.A. Oluokun independently have shown that concrete strength development is determined not only by the water-cement ratio but is also influenced by the content of other concrete ingredients [1, 2]. The compressive strength of concrete depends on a wide range of ingredients like cement content, coarse & fine aggregate, water-cement ratio, age, etc. This possesses a major challenge in accurately predicting the compressive strength of concrete.

For over two decades, new algorithms and models, especially those based on soft computing (SC), enabled researchers to solve the most complex systems in different ways. The use of prediction methods based on mathematical models, such as artificial neural network (ANN) and fuzzy logic (FL) methods are becoming widespread in various engineering fields [3]. Neural network (NN) technology, a subfield of Artificial Intelligence, is being used to solve a wide variety of problems in civil and structural engineering [4–6].

J. Bai et al [7] demonstrated the use of NNs to predict concrete workability. Simsek et al [8] proved that contrary to popular belief multiple linear regression (MLR) can achieve high predictive power. Similarly, other researchers have used AI, ML & Deep Learning tools such as Regression Models [9, 10], Neural Networks & Fuzzy Logic (FL) [6, 11], and Fuzzy Inference Systems [12] and concluded that all SC techniques when tweaked correctly are effective in predicting and validating the relations of concrete ingredients with concrete properties such as compressive strength & slump.

Many studies have also compared these techniques and their prediction accuracy like Sarıdemir [13] compared ANN & FL, Bilgehan [14] compared ANN & Neuro-Fuzzy approach, KHADEMI et al [15] have compared MLR, ANN & FL. In [2], the possibilities of adapting ANN at predicting the compressive strength of the high-performance concrete was demonstrated. To generate results and check the reliability of various models, the author assembled experimental data from 17 different sources. The authors in this paper, have worked on the same dataset.

©IJRASET: All Rights are Reserved

This research focuses on using various regressors such as Decision Tree, Random Forest, Gradient Boosting, Extreme Gradient Boosting (XGBoost), Adaptive Boosting (ADA), Bagging Regressor & K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN) to predict the compressive strength of concrete having mixes from the range of M10 to M80. To train the prediction models, exhaustive data of 1030 data points were collected from various sources of literature. The dataset has compressive strength at the age of 1, 3, 7, 14, 28, 56, 90, 91, 100, 120, 180, 270, 360 & 365 with 8 common components. Many researchers have published papers of the same dataset setting a benchmark to compare our results to [1, 16, 17]. In this research, feature engineering has been performed to fine-tune the same data with a different approach compared to the other researchers in [1, 16,]. To minimize the Mean Square Error (MSE), an effective standardization of data was done. Furthermore, k-fold cross-validation had been performed on every model to validate the models and to achieve prediction results above the average k-fold value. Based on the importance of the various input parameters, certain parameters were dropped and/or replaced and the results were compared. In contrast to KHADEMI et al [14], where it was stated that "regression models failed to be reliable and therefore, advances models like ANN & Artificial Neuro Fuzzy based Logic (ANFIS) models are preferred", the researchers have proved that by adjusting the parameters of the regression models, one can achieve better results than ANN.

II. MIX DESIGN AND DATA SETS

Experimental mix design data from M10 to M80 from 17 different sources were used to check the reliability of the various strength models. This data set was originally collected by I.C. Yeh [1]. The components of the concrete mix or the input variables controlling the compressive strength were as follows:

1) Cement (kg/m3)
2) Fly ash (kg/m3)
3) Blast furnace slag (kg/m3)
4) Water (kg/m3)
5) Superplasticizer (kg/m3)
6) Coarse aggregate (kg/m3)
7) Fine aggregate (kg/m3)
8) Age of testing (days)

There were 1030 rows in the data set and 9 columns - 8 input values listed above and the 9th being compressive strength. Apart from the component types, the properties of concrete are influenced by the mixing proportions and by the mixing preparation technique [1]. These 1030 data points have a variety of ranges and they have been tabulated in Table 1. Another variation was attempted where water was replaced with water-cementitious material ratio as one of the input parameters. Though using this parameter didn't yield any significant improvement in the accuracy.

The input parameters that contain outliers, had in some cases been replaced by the mean values and in some cases, have been dropped completely - keeping in mind the nuances of civil as well as computer engineering.

For every regressor model and neural network applied, the data set of 1030 data points were randomly sampled into training, testing, and validation sets. Effective shuffling of the data set was done to allocate unique data points in each set.

TABLE I
RANGES OF MIX OF DATA SETS AND THEIR STANDARD DEVIATION

Component	Unit	Min	Max	SD
Cement	(kg/m^3)	102.00	540.00	104.51
Water	(kg/m^3)	121.80	247.00	21.35
Fly ash	(kg/m^3)	0.00	200.10	64.00
Blast furnace slag	(kg/m^3)	0.00	359.40	86.28

Superplasticizer	(kg/m^3)	0.00	32.20	5.97
Coarse aggregate	(kg/m^3)	801.00	1145.00	77.76
Fine aggregate	(kg/m^3)	594.00	992.60	80.17
Age	days	1	365	63.17
Water-cementitious ratio	-	0.235	0.9	0.1271
Compressive Strength	(N/mm^2)	2.33	82.6	16.71

III.METHODOLOGY

The initial steps in the program modeling include data featurization, multivariate analysis, and treating outliers. After this exhaustive process of data cleaning, the steps that follow can be divided into 3 parts - standardization, ensemble learning techniques, and lastly verification & validation of the results, all of which have been explained in detail in the sections below.

A. Standardization

Z-score standardization was applied to the input parameters to normalize and fit them in a Gaussian Curve. This technique was used to ensure rescaling of mean and standard deviation to 0 and 1 respectively.

B. Ensemble Learning

To leverage better predictive performance on the data, ensemble learning techniques were applied. Ensemble learning accompanies multiple base learners to improve the performance and accuracy of the machine learning regressors. These base learners are utilized to generate the final decisions. Having learned complementary information, the base learners can be implemented to reach greater accuracies and performance. In this work, various ensemble learning techniques were implemented along with K-nearest neighbors (KNN) algorithm and Artificial Neural Network (ANN).

1) *Bagging:* The bagging technique or also commonly bagging aggregate technique is a homogenous weak learning algorithm to learn key information from individual learners to draw an average decision for the output. This algorithm can be used for both statistical classification and regression. Best suited for low bias and high variance, the technique decreases the variance aiding to avoid over fitting. Various popularly known bagging algorithms such as decision tree regressor, random forest regressor, and classical bagging regressor were implemented and discussed to enhance the model generalization towards the data.
2) *Boosting:* Similar to bagging, boosting is a weak-learner based ensemble technique. But unlike bagging, the algorithm uses sequential and adaptive learning to improve the performances of the base learner. This adaptive approach towards learning and understanding new information results in a decrease in error in every step of the training, till the information is correctly predicted or the maximum number of steps are reached. Boosting techniques such as AdaBoosting, Gradient boosting, and Xtreme gradient boosting techniques were experimented with and discussed to improve the model accuracy.
3) *K-nearest neighbours (KNN):* KNN is one of the most elementary algorithms for both classification and regression. The predictions assume that objects grouped or having minimal distance potentially belong to the same class. K-neighbour regressor was implemented for each query point, where k is an integer value specified by the user. Uniform weights were introduced in such a way that each point in the local neighbourhood uniformly to the classification of a query point. It was observed to be advantageous to weigh points, as nearby points positively contribute more to the regression than points further away.
4) *Artificial Neural Networks (ANN):* ANN is a popular approach for both classification and regression of data. A commonly used ANN architecture with a multi-layer feed-forward approach was applied which consisted of a single input layer, three hidden layers, and one output layer. These layers consisted of fully connected neurons. The input values traverse through neurons consisting of suitable activation functions in addition to a specifying weight. Multiple experimental approaches were used to find an optimum number of hidden layers; 300 epochs were performed along with different types of activation functions to find the best possible accuracy.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Verification and Validation

To evaluate the performance of the models, evaluation metrics such as Mean Square Error (MSE) and Coefficient of Determination (R2) were analysed and compared. MSE is a risk function that corresponds to the anticipated value of the squared error loss as shown in equation 1.

$$MSE\ (y_i, y_i^p) = \frac{1}{n}\sum_{i=0}^{n-1} (y_i - y_i^p)^2 \ldots (1)$$

Some cases are noise-free where y_i equal to y_i^p which is the desired output. When noise is present, y_i^p becomes the estimation of the desired output or semi-desired output. R^2 is the degree of success in reducing the standard deviation (SD) and is widely used in the sciences as a measure of the strength of linear dependence between two variables. Equation 2 presents the R^2 value as follows.

$$R^2(y_i, y_i^p) = 1 - \frac{\sum_{i=1}^{n} (y_i - y_i^p)^2}{\sum_{i=1}^{n} (y_i - \underline{y})^2} \ldots (2)$$

where, $\underline{y} = \frac{1}{n}\sum_{i=0}^{n-1} y_i$ and $\sum_{i=1}^{n} (y_i - y_i^p)^2 = \sum_{i=1}^{n} \varepsilon_1^2$

To evaluate how well-observed outcomes are replicated by the models, the R^2 coefficient which varies from 0 to 1 is provided as accuracy in Table 2. It is worth mentioning that the higher the quantity of R^2 is the better the model would be.

Moreover, to test the extent of over fitting k-fold (10 fold) cross-validation was performed on the experimented models to attain a generalized model with combinations of data split and analyse the standard deviation in the accuracies of the trained model.

IV. RESULTS & DISCUSSION

Following are some of the observations and discussions from the modelling:

1) Figure 1 shows the importance factors of the 8 input parameters according to the various algorithms that were implemented.

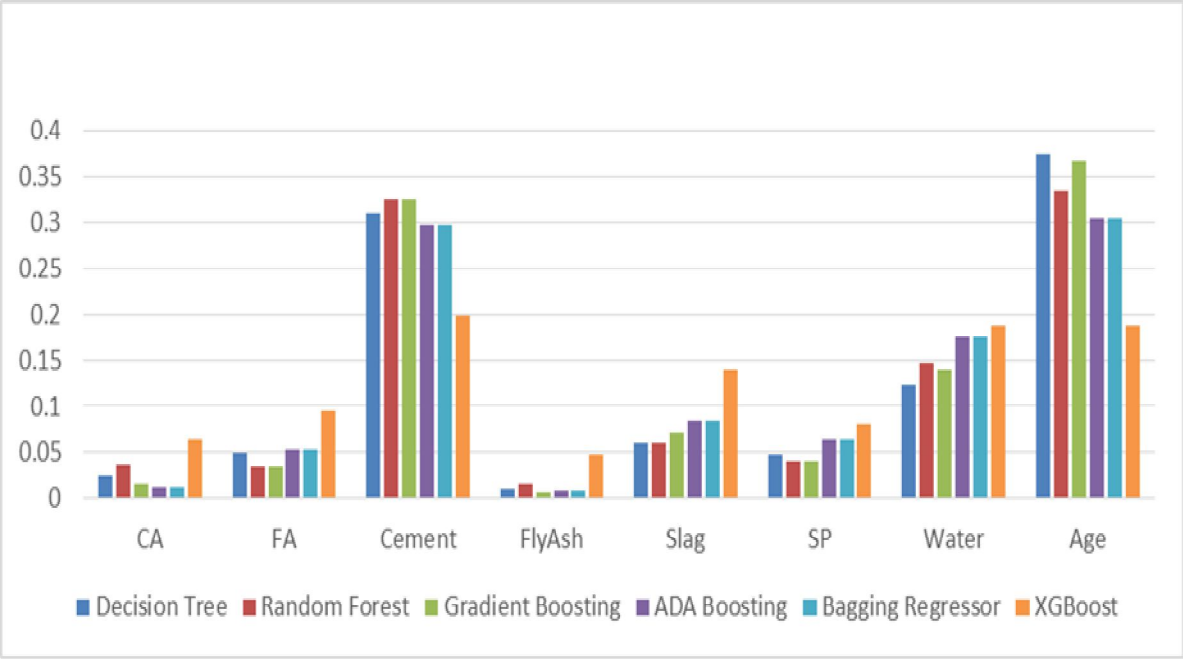


Fig. 1 Importance factors of Input Parameters for different models

Many practitioners of civil engineering might criticize the minuscule importance factor of coarse aggregate presented by various models towards the governance of compressive strength of the concrete. This unprecedented observation can be witnessed due to highly monotonous data for coarse aggregate over a huge variance of cement and age. Rafat et al [18] pointed out the same observation in their work for the prediction of compressive strength of SCC. Moreover, the practice of removing the lower importance input parameters in the research showcased very little to no changes, just as observed by Rafat et al [18].

2) The evaluation metrics corresponding to the applied algorithms are enlisted in Table 2 below.

TABLE III
ACCURACY, MSE, AND STANDARD DEVIATION OF THE MATHEMATICAL MODELS

Sr. No.	Mathematical Model	Accuracy (R^2)	MSE	SD
1	Decision Tree	0.8622	0.1621	0.0000
2	Decision Tree k-fold	0.6113	0.0000	0.6740
3	Random Forest	0.9311	0.0891	0.0000
4	Random Forest k-fold	0.9139	0.0000	0.0334
5	Gradient Booster	0.9561	0.0513	0.0000
6	Gradient Booster k-fold	0.9429	0.0000	0.0262
7	XGBoost Regressor	0.9315	0.0836	0.0000
8	XGBoost Regressor k-fold	0.9096	0.0000	0.0285
9	ADA Booster	0.9060	0.1206	0.0000
10	ADA Booster k-fold	0.8748	0.0000	0.0255
11	Bagging Regressor	0.9284	0.1019	0.0000
12	Bagging Regressor k-fold	0.9040	0.0000	0.0337
13	KNN	0.8647	0.1665	0.0000
14	KNN k-fold	0.8408	0.0000	0.0387
15	ANN	0.9376	0.0723	0.0000

3) Based on the predicted and actual values of compressive strength of concrete, the scatter plots of each of the listed models are plotted as below:

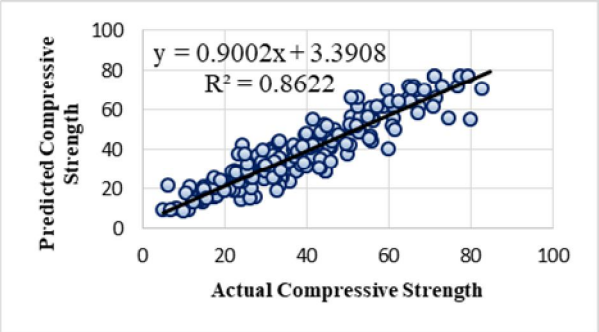


Fig. 2: Scatter Plot for Decision Tree

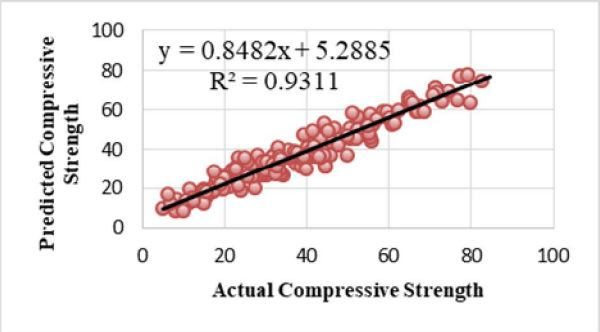


Fig. 3: Scatter Plot for Random Forest

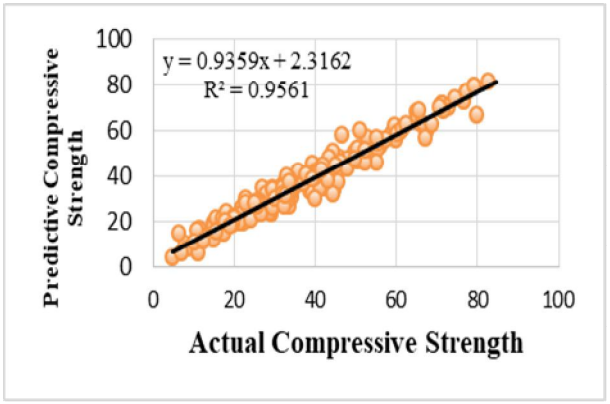


Fig. 4: Scatter Plot for Gradient Boosting

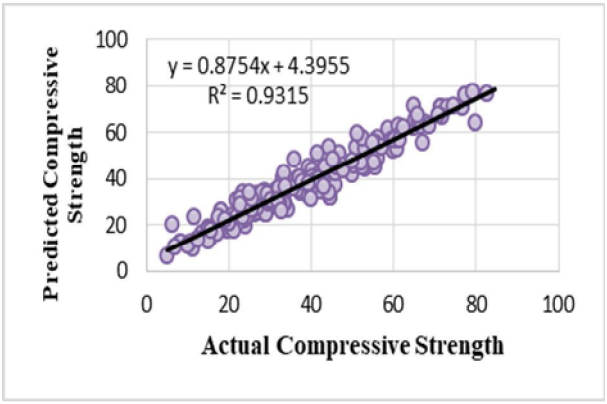


Fig. 5: Scatter Plot for XGBoost Regressor

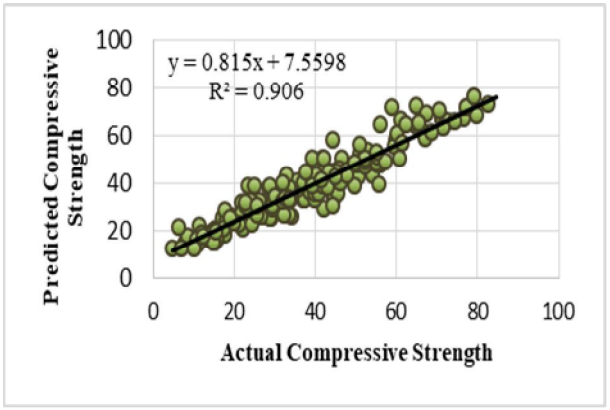


Fig. 6: Scatter Plot for ADA

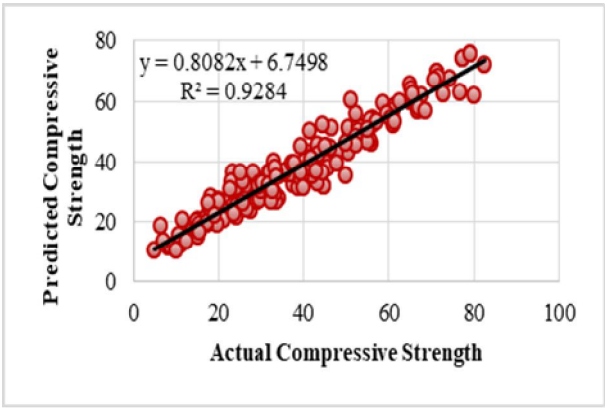


Fig. 7: Scatter Plot for Bagging Regressor

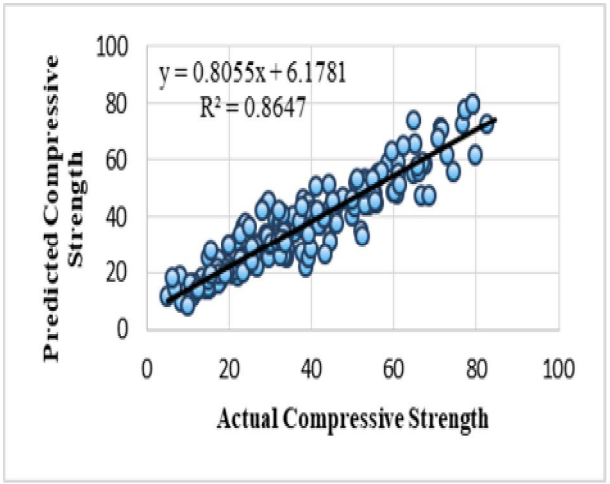


Fig. 8: Scatter Plot for KNN

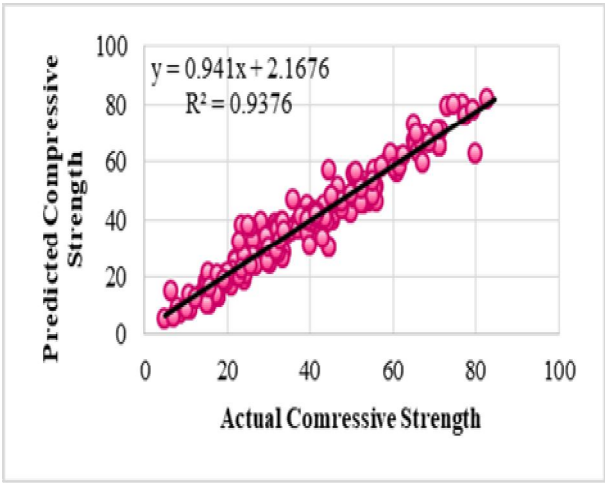


Fig. 9: Scatter Plot for ANN

As can be seen, from Table 2 and the plots, Gradient Boosting has the best line of regression for the prediction of compressive strength having R2 = 0.9561 & MSE = 0.0513.

4) Performance of the Neural Network was also plotted in terms of the MSE graph for the training and testing data as shown in figure 10.

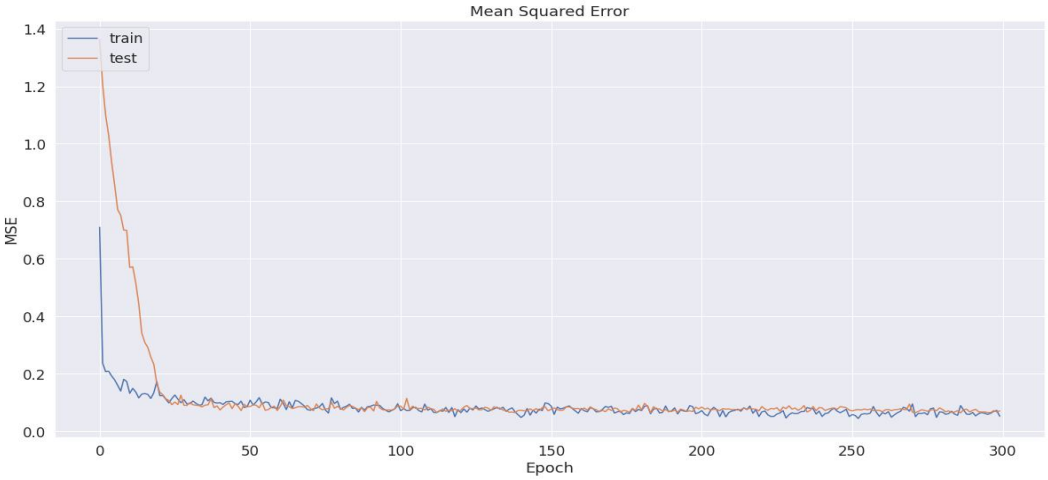


Fig. 10: MSE Plot for Neural Networks

The best performance of the neural network was achieved at 300 epochs, with 3 hidden layers. The MSE value was 0.0723 with an accuracy of 0.9376.

V. CONCLUSION

This research has examined and compared the use of machine learning & deep learning methods to predict the compressive strength of concrete as a function of its mix proportions. A large data set of 1030 data points having 8 mix design parameters and 9th compressive strength were collected from literature to train and test the models. The concrete mixes ranged from M10 to M80 with their compressive strength being measured from Day 1 to Day 365. Given the complex nature of concrete ingredients which depend on various factors like their origin, their shape and size, their chemical composition, etc., predicting the compressive strength of concrete becomes a complex task too. Thus even an accuracy of 70% can be considered good enough. In this paper, models were generated to predict the compressive strength using Machine Learning Regressors namely - Decision Tree, Random Forest, Gradient Boosting, XGBoosting, ADA Boosting, Bagging Regressor, KNN Regressor, and Artificial Neural Networks. All the models achieved an accuracy i.e. correlation coefficient well above the acceptable limit. This indicates that mathematically the models can very well predict the physical state of the material. Therefore, the following points can be deduced:

1) Machine Learning Regressors & Artificial Neural Networks can be used for the prediction of compressive strength of concrete. Featurization of data is very important to enhance the prediction results. In contrast to KHADEMI et al [14], where they stated “regression models failed to be reliable and therefore, advances models like ANN & ANFIS models are preferred”, the researchers of this paper have proved that by adjusting the parameters of the non-linear regression models, they can be trained to achieve better results than ANN.

2) Random Forest Regressor which is a collection of many Decision Trees showed better performance than Decision Tree. Hence whenever a Decision Tree regressor is used, Random Forest may also be checked as it has been proven to provide better results.

3) Modeling of K-fold CV has proven helpful as it gives us an average performance of the particular model for which it is implemented. This can help tweak the parameters to achieve accuracy higher than the k-fold value.

4) Dropping columns with the least importance factors does not cause a major difference. The same was observed when water was replaced with water-cementitious material ratio. Authors believe the reason is that when the data set is fed into the model, it’s no longer relying on the law of physics but the mathematical correlation between the input and output parameters. Hence taking a call of which parameters to use and whether or not to drop the columns can be decided based on the computation time and budget.

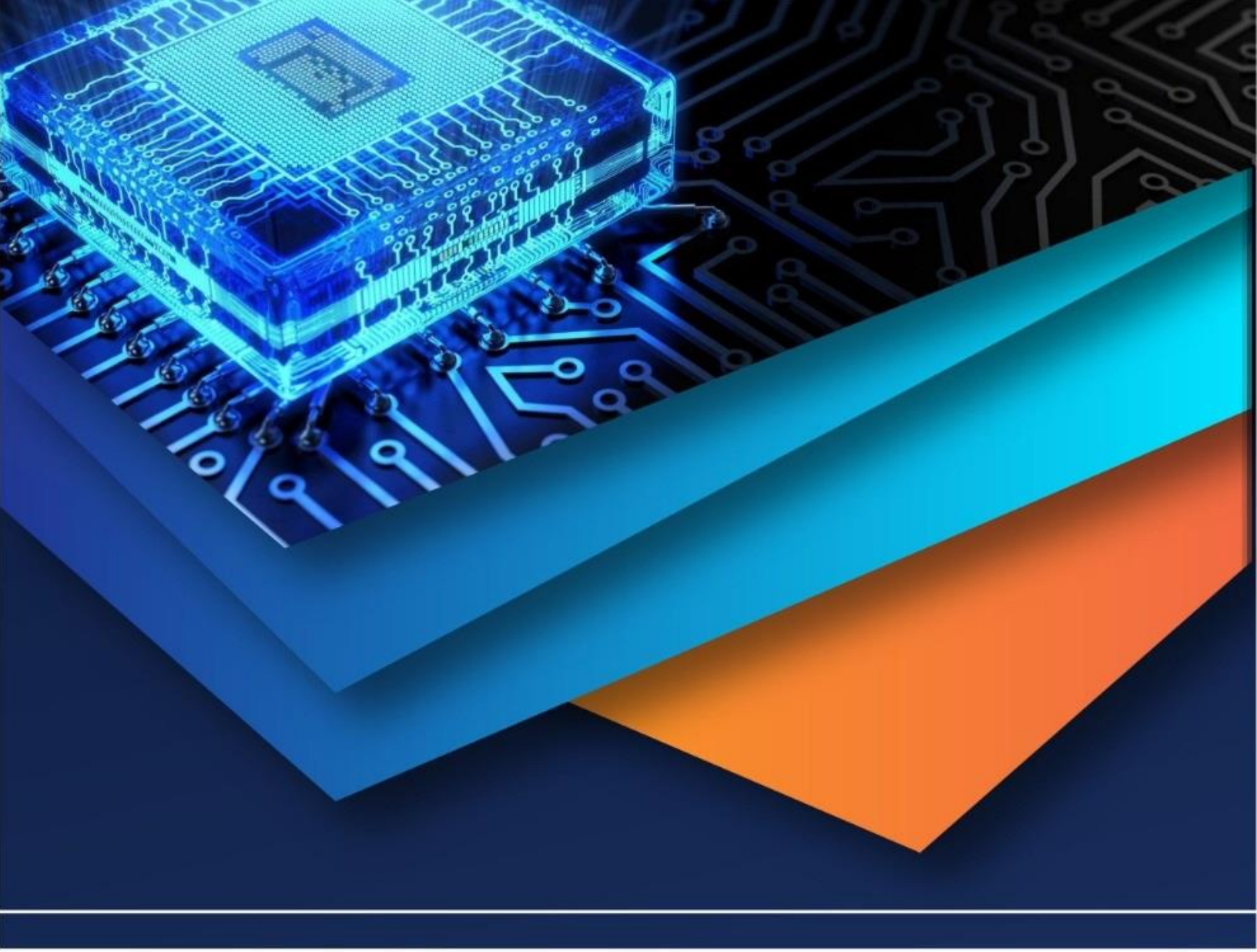
Lastly, the authors would like to suggest that the readers explore a wide range of feature engineering and tweaking parameters of every model, by understanding their working and functionality in depth. Then the calculated trial and errors will be more meaningful and give better results in less time. This can in turn produce an effective tool for the prediction of compressive strength which shall help predict the strength before the concrete is cast and experimentally tested.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Oluokun, Francis A. Fly ash concrete mix design and the water-cement ratio law. Materials Journal 91, no. 4., 1994: 362-371.
[2] Yeh, I.C., MODELING OF STRENGTH OF HIGH-PERFORMANCE CONCRETE USING ARTIFICIAL NEURAL NETWORKS. Cement and Concrete Research (Elsevier Science Ltd) 1998: 12-24.
[3] Mahesh Kothari, K D Gharde. Application of ANN and fuzzy logic algorithms for streamflow modelling of Savitri catchment. J. Earth Syst. Sci. (Indian Academy of Sciences), 2015: 933-943.
[4] Rguig Mustapha, EL Aroussi Mohamed. High-Performance Concrete Compressive Strength Prediction Based Weighted Support Vector Machines. Int. Journal of Engineering Research and Application, 2017: 68-75.
[5] De-Cheng Feng, Zhen-Tao Liu, Xiao-Dan Wangc, Yin Chen, Jia-Qi Chang, Dong-Fang Wei,. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Construction and Building Materials (Elsevier), 2020.
[6] Mustafa Sarıdemir, Ilker Bekir Topçu, Fatih Özcan, Metin Hakan Severcan. Prediction of long-term effects of GGBFS on compressive strength of concrete by artificial neural networks and fuzzy logic. Construction and Building Materials (Elsevier), 2006: 1279-1286.
[7] J. Bai, S. Wild, J.A. Ware, B.B. Sabir. Using neural networks to predict workability of concrete incorporating metakaolin and fly ash. Advances in Engineering Software, Elsevier, 2003: 663-669.
[8] SerhatSimsek, Mehmet Gumus, Mohamed Khalafalla & Tahir Bachar Issa. A hybrid data analytics approach for high performance concrete compressive strength prediction. Journal of Business Analytics (Taylor & Francis), 2020.
[9] S. Wu, B. Li, J. Yang, and S. Shukla, Predictive modeling of high-performance concrete with regression analysis, in Proceedings of the 2010 IEEE International Conference on Industrial Engineering and Engineering Management, (2010): 1009–1013
[10] M. F. M. Zain and S. M. Abd, Multiple regression model for compressive strength prediction of high performance concrete, Journal of Applied Sciences, (2009):155–160.
[11] Fatih Özcan, Cengiz D. Atis, OkanKarahan, ErdalUncuog˘lu, Harun Tanyildizi. Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete. Advances in Engineering Software (Elsevier), 2008: 856-863.
[12] BURAGOHAIN, MRINAL. Adaptive Network based Fuzzy Inference System (ANFIS) as a Tool for System Identification with Special Emphasis on Training Data Minimization. Guwahati, India: Indian Institute of Technology Guwahati, 2008.
[13] Mustafa Sarıdemir, IlkerBekirTopçu, FatihÖzcan, MetinHakanSevercan. Prediction of long-term effects of GGBFS on compressive strength of concrete by artificial neural networks and fuzzy logic. Construction and Building Materials (Elsevier), 2006: 1279-1286.
[14] Bilgehan, Mahmut. A comparative study for the concrete compressive strength estimation using neural network and neuro-fuzzy modelling approaches. Nondestructive Testing and Evaluation, 2013: 35-55.
[15] Faezehossadat KHADEMI, Mahmoud AKBARI, Sayed Mohammadmehdi JAMAL, Mehdi NIKOO. Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete. Front. Struct. Civ. Eng., 2017: 90-99.
[16] Susom Dutta, PijushSamui, Dookie Kim. Comparison of machine learning techniques to predict compressive strength of concrete. Computers and Concrete, Vol. 21, No. 4, 2018: 463-470.
[17] De-Cheng Feng, Zhen-Tao Liu, Xiao-Dan Wang, Yin Chen, Jia-Qi Chang, Dong-Fang Wei, Zhong-Ming Jiang. Machine learning-based compressive strength prediction for concrete:An adaptive boosting approach. Construction and Building Materials (Elsevier), 2019.
[18] Rafat Siddique, Paratibha Aggarwal, Yogesh Aggarwal. Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks. Advances in Engineering Software (Elsevier), 2011: 780-786.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)