Concrete Compressive Strength Prediction Using Machine Learning Algorithms

Firas Abed. Turkey (PE20917@student.uniten.edu.my)
Universiti Tenaga Nasional
Salmia Bt. Beddu
Universiti Tenaga Nasional
Ali Najah Ahmed
Universiti Tenaga Nasional
Suhair Al-Hubboubi
Building Research Directorate

Article

Keywords: Compressive strength, Machine Learning, Accuracy Metrics

Posted Date: June 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1665395/v1

Concrete Compressive Strength Prediction Using Machine Learning Algorithms

Firas Abed. Turkey[1*], Salmia Bt. Beddu[2,] Ali Najah Ahmed[3], Suhair Al-Hubboubi[4].
[1]PhD. Student in Civil Engineering Department, College of Engineering, UNITEN, Malaysia. PE20917@student.uniten.edu.my
[2] Dr, Department of Civil Engineering, College of Engineering, UNITEN, Malaysia. salmia@uniten.edu.my
[3] Dr, Institute of Energy Infrastructure (IEI), Universiti Tenaga Nasional (UNITEN), Malaysia Mahfoodh@uniten.edu.my.
[4] Dr, Suhair Al-Hubboubi, Building Research Directorate, Iraq. suhairkah@yahoo.com

Abstract: The most widely utilized construction material is concrete. Concrete's physical qualities differ depending on the kind**.** In this paper, we predicted the compressive strength of four types of lightweight aggregate geopolymer concretes (LWAGC), namely, lightweigh expanded clay Leca, recycled foam masonry aggregate RFA, Porcelanite aggregate PA and recycled brick aggregate RBA. For predictions, we used seven models, specifically, Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), Decision Tree (DT), Random Forest (RF), Support Vector Regressor (SVR) and Linear Regression (LR). For evaluation, we employed six metrics, Root Mean Square Error (RMSE), Mean Squared Logarithmic Error (MSLE), Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Logarithmic Error (RMSLE) and Coefficient of Determination R^2. The results of predictions demonstrate that RF offers high accuracy for Leca type (MSE= 2,6697; RMSE= 1,6339; MSLE= 0,0072; RMSLE= 0,0851; MAE= 1,3434; R^2=0,8687), Porcelanite type (MSE= 2,9650; RMSE= 1,7219; MSLE= 0,0065; RMSLE= 0,0805; MAE= 1,5693; R^2=0,8437) and RFA type (MSE= 1,7028; RMSE= 1,3049; MSLE= 0,0059; RMSLE= 0,0768; MAE= 1,1300; R^2=0,8099) because the predictions are closer to the real values, and DT offers better predictions than other models for RBA type (MSE= 3,3069; RMSE= 1,8185; MSLE= 0,0066; RMSLE= 0,0812; MAE= 1,5610; R^2=0,8764). İn order to test the models in predicting new value we gave them a 675° as new temperature and we found that that LR is more accurate than other models in predicting the CS of Leca, RBA and Porcelanite types. However, CNN outperformed other models in predicting RFA type.
Keywords: Compressive strength, Machine Learning, Accuracy Metrics.

1. Introduction

After water, concrete is the most used material on the earth due to many reasons, cheap, high tenacity, available, can be formed to any desired shape [1]. Conventional concrete is identified as the most important contributor to greenhouse gas emissions. Gases that are environmentally harmful due to the cement industry are a principal ingredient of traditional concrete. The cement sector in the world is among the biggest emitters of CO_2. It contributes a significant share of total global greenhouse gas (GHG) emissions [2]. Common Portland cement (PC), as is known, contributes an equal quantity of CO2 emissions to the atmosphere since it has been employed in practically every civil engineering construction exercise [3]. Evidence shows that concrete emits about 8% of carbon dioxide gas into the atmosphere [4]. This improves the impact of the greenhouse effect on the atmosphere. In 1978, Davidovits developed the geopolymer as an innovative alternative to cement and had similar characteristics to ceramic [5]. GPC is a good substitute for the cement industry while the world is taking some measures to decrease gas emissions [6]. As an environmentally friendlier alternative to conventional ordinary PC, geopolymer binder attracts tremendous interest (OPC) [7]. An excellent substitute for OPC is geopolymer concrete (GC). Sometimes concrete is subjected to elevated temperatures, which significantly affect its properties. This impact also depends on the concrete's type and ingredients [8].

In concrete technology, lightweight concrete isn't even a possibility. Lightweight concrete, on the other hand, has been used for over 2000 years and is still being developed [9]. Three of the most prominent lightweight concrete structures in the Mediterranean area are the "Port of Cosa," the Pantheon Dome, and the Coliseum [10]. All of them were built during the early Roman Empire. Lightweight aggregates (LWA) and light-weight concrete (LWC) are defined by the American Concrete Institute (ACI) 213 as "concrete made up of lightweight coarse aggregates and normal weight fine aggregates, potentially with some lightweight fine aggregates" [10]. With a total deadweight of around 2400–2500 kg/m3 and a dry density ranging from 320 kg/m3 to 1920 kg/m3 [10], LWC is 23–80 percent lighter than conventional weight concrete. Reduced density, low thermal conductivity, little shrinkage, good heat resistance, reduced dead load, lower transportation costs, and faster construction speed are all advantages of low-density lightweight concrete in construction [11]. Pumice perlite, porcelanite rocks, expanded clays, shales, and other wastes such as mixed garbage, agricultural waste, plastic or rubber, clay brick sintered fly ash aggregate, and oil palm shell are used to make lightweight concrete. [12]. Compressive strength is the most widely used as a general indicator of the efficiency of concrete in terms of structural work in the case of neutralizing the rest of the factors such as density and durability. Numerous Internal and external factors, influence the CS of the GC after exposure to elevated temperatures. The components of the concrete, as well as the quantity of each component, are some of the internal elements. The degree of temperature, the period of exposure, the rate of increasing and cooling following exposure, and slow or abrupt cooling are some of the external influences. [13]. The commonly utilized approaches are not considered in this study because they are inefficient and take longer to produce results. The most common and efficient strategies are employed in this manuscript to determine the most correct forecast in less time. This research predicts the CS via different ML algorithms, namely, ANN, CNN, LSTM, DT, RF, SVR, and LR. Also, we pretend to compare the results with the experiment results. The objective is to investigate ML techniques' ability to predict the CS of different types of concretes and verify if we can rely on these techniques to know the impact of temperature on CS without doing experiments in the laboratory. İn the next section, we will present our literature review. Then, İn section 3, we expose our methodology. Then, in section 4, we show and discuss the findings. Finally, in section 5, we conclude. Figure (1) shows the divides concrete into harmful and environmentally friendly.

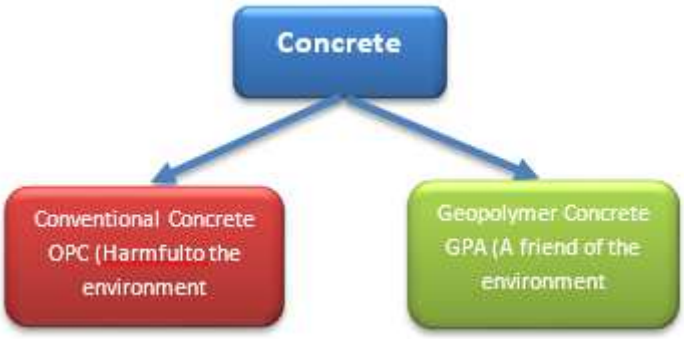

Figure 1: Types of Concrete

2. Literature review

Previous research found various studies that mobilized ML algorithms to predict CS of different types of concretes, and we present below some studies:
Kandiri et al. [14] assessed the CS of recycled aggregates using a modified ANN. The salp swarm algorithm (SSA), genetic algorithm (GA), and grass hop-per optimization algorithm (GOA) techniques were used to improve the outputs of the ANN model.. When compared to other models, the SSA-ANN model was more accurate.Despite the limited data and input parameters, Kaperkiewicz et al. [15] employed a fuzzy-ARTMAP network to predict its CS. They did, however, demonstrate that models can forecast the CS of high-performance concrete (HPC) with a reasonable degree of accuracy.
Yeh [16, 17] has made a significant contribution to the use of machine learning approaches to advanced concretes, particularly HPCs. He confirms the ANN's capacity to forecast the CS of HPCs.
When Tayfur et al [18] examined the capacity of Fland ANN to predict the CS of HPCs, they discovered that ANN outperformed FL by 15%.The weighted genetic pyramid operational tree (GWPOT) was compared to various models such as ANN, SVM, ESIM, GOT, and the weighted operational structure technique by Cheng et al. [19]. Except for ESIM, GWPOT outperformed all other models. Nehdi et al. [20] claim to be the first to predict the CS of self-consolidating concrete using machine learning (SCC). They claimed that ANN could predict not only the CS but also other SCC metrics like segregation, droop, and filling capacity with accuracy. Uysal and Tanyildzi [21] conducted a comparison of two learning algorithms, one of which was the Fletcher conjugate power algorithm and the Levenberg-Marquardt back-propagation algorithm. However, they noted that the Fletcher algorithm was more precise, as its R2 was 0.95, compared to 0.92 for the Lavenberg-Marquanrdt algorithm. Topcu and Saridemir [22,23] identified the FL and ANN models as strong tools for predicting the CS of fly ash and RAC concrete with high accuracy even with the limited data available. They found that ANN surpass FL in both studies. Naderpour et al. [24] identified the CS of FRP confined concrete specimens by utilizing ANN model. İn order to detect the model paramets that optimize the results, they used an iterative approach and they observed that the model needs 11 hidden neurons.
Jalal and Ramezanianpouret [25] employed ANN, LR and NLR to predict the CS of FRP-confined concrete, they identified that ANN model is more accurate than others are. Nehdi et al. [26] employed ANN model to forecast the CS of cellular concrete. As results, they found that their models are able to predict well the CS and the prediction error is lower than the reel results with 47% at least.
Using multivariate adaptive regression splines applying the water cycle algorithm (MARS-WCA), Ashrafian et al [27] calculated the CS of cellular concrete and compared it to several LR models, ANN, standard MARS, and SVR. The researchers discovered that MARS-WCA outperforms other models by 25%. Al-Janabi and Al-Hadithi [28] used multiple ANN models to predict the CS of polymer modified concrete. As a result, the training, test, and validation data sets all had correlation coefficients of 0.89, 0.87, and 0.81, respectively.

2. Lightweight Concrete Production:

In order to produce lightweight concrete, there are three methods: creating voids inside the cement paste, which is called aerated (cellular concrete), or using one type of coarse aggregate called concrete (no fine concrete), and the third type is using lightweight coarse aggregate, which is the most common in load-bearing concrete.

2.1. Lightweight Aggregate (LWA):

The ASTM standard supports two types of lightweight aggregate parameters, as shown in Figure 2. Lightweight concrete is constructed mostly from lightweight particles. The ASTM standard covers (i) lightweight aggregates formed by pelletizing, expanding, or sintering natural materials like shale, diatomite, and other natural materials, and (ii) aggregates made by processing natural materials like tuff, pumice, and other natural materials. Lightweight "aggregates made out of final coal or coke products" are also listed in ASTM C331M [29]. The determination of aggregate parameters, notably density, is critical in distinguishing between standard weight aggregate and LWA. ASTM C330M and ASTM C331M define upper limits for loose bulk density, with fine lightweight aggregate weighing 1120 kg/m3, coarse lightweight aggregate weighing 880 kg/m3, and a lightweight aggregate combination weighing 1040 kg/m3.

2.2 Lightweight geopolymer concrete LWAGC:

The structural designer favors a drop in concrete density since concrete weight accounts for more than half of the dead load in a structure. "ASTM C 330" is a lightweight structural concrete. The concrete has a 28-day compressive strength of 17 N/mm2, a density equilibrium of 1120 to 1920 kg/m3, and is made completely of lightweight aggregate or a mix of lightweight and normal-density aggregate." The need to save money on a project, expand functionality, or a mix of the two is frequently used to justify LWC. When employing lightweight concrete, estimating the entire cost of a project is critical because a cubic meter is often more expensive than an equivalent unit of normal concrete. Manufacturers employ raw resources like suitable fly ash, shales, slates, clays, or blast-furnace slag to produce structural-grade lightweight aggregates. Processed lightweight aggregates are becoming increasingly popular, which is a great example of long-term sustainability and environmental planning. By requiring less transportation and the usage of minerals that have limited structural uses in their native state, these products lessen construction industry pressures on rare natural gravel, stone, and sand resources [12]. The scheme of the work technique is shown in figure (2), from the fabrication of four varieties of lightweight geopolymer concrete to the completion of the examination after high temperature exposure.

Figure 2: LWAGC Manufacturing and Testing

3. Methodology:

3.1 Data Collection:

Only one target parameter, concrete compressive strength CS, was investigated in this investigation. The degree of temperature for each type of LWAGC is utilized as an input parameter in this article. The CS data obtained from the experimental test results according to the effect of the elevated temperatures from four kinds of lightweight geopolymer concrete (LWAGC); the strengths records will be taken after being exposed to high temperature and compared with the strength in the normal state (room temperature), strength will be recorded as compressive strength.

3.2 Data Pre-Processing DPP:

DPP (normalization) is useful for modelling applications where the inputs are generally on a widely different scale. Therefore, data splitting is an essential consideration during Machine Learning model development. In this study, the data will be split into two groups. The first test comprised 80% of the data set that was used as the training set, whereas the second test included 20% of the data set that was used as the validation set.

3.3 Model Development

- *Hyper parameter optimization* Optimization of machine learning hyper parameters is a critical work, as the performance of an algorithm can be heavily influenced by the hyper parameter selection. Seven optimization approaches will be studied in this study to identify the optimum value: Artificial Neural Network ANN, Convolu-tional Neural Network CNN, Long Short-Term Memory LSTM, Support Vector Machine SVM, Random Forest RF, Decision Tree DT, and Regression (SPSS) search.*Stopping Criteria*: Usually, the Machine Learning model's training error reduces gradually as the training process progresses.
- *Model Performance Evaluation*: It is essential to identify the criteria by which the model's performance will be judged. In this study, the performances of the models will be evaluated according to many statistical indexes as a Mean Square logarithm Error MSLE, Mean Absolute Error MAE, Root Mean Square Error RMSE, Relative Square Error RSE, Root Mean logarithm Square Error RMSLE, and R^2 Coefficient of Determination.

Sensitivity Analysis: Finally, seven methods will be introduced to check the uncertainties associated with the proposed model, namely sensitivity analysis and uncertainty analysis. In this research work, four types of lightweight aggregates were used included two types were recycled aggregates such as recycled clay bricks aggregate (RBA) and recycled foam masonry block aggregate (RFA), in addition to artificial aggregate type as lightweight expanded clay aggregate (Leca), and natural lightweight rocks as porcelanite. Mixed with the components of geopolymer concrete, which mainly consists of a mixture of fly ash and glass powder in reaction with an activated alkali solution consisting of sodium silicate SS and sodium hydroxide SH with a ratio of SS/SH 2.5 to 1 and molarity of 14. aim to predict CS of various types of Lightweight geopolymer concretes LWAGC using different ML algorithms (See below). The goal is to predict CS for various concrete types using a single characteristic, temperature. Our dataset consists of 25 observations for each type. We are interested in predicting the Leca type, the RBA type, the Porcelanite type, and the RFA type. To evaluate our predictions, we will use the metrics below. The figure below presents our methodology composed of six steps. In the first phase, we import the data. In the second one, we split our data into training testing datasets, 80%, and 20%, respectively. In phase three, we train the data with our models. In the fourth phase, we make predictions of CS. In the fifth step, evaluated the predictions, and in the final stage, we predicted the CS of the different types of concrete under the temperature of 675°C. Figure (3) shows the flow chart methodology, and Table (1) shows the results of the compression test for four types of LWAGC at different temperatures (35, 200, 400, 550, 675, 800°C) for five specimens for each temperature and the standard deviation of the results.

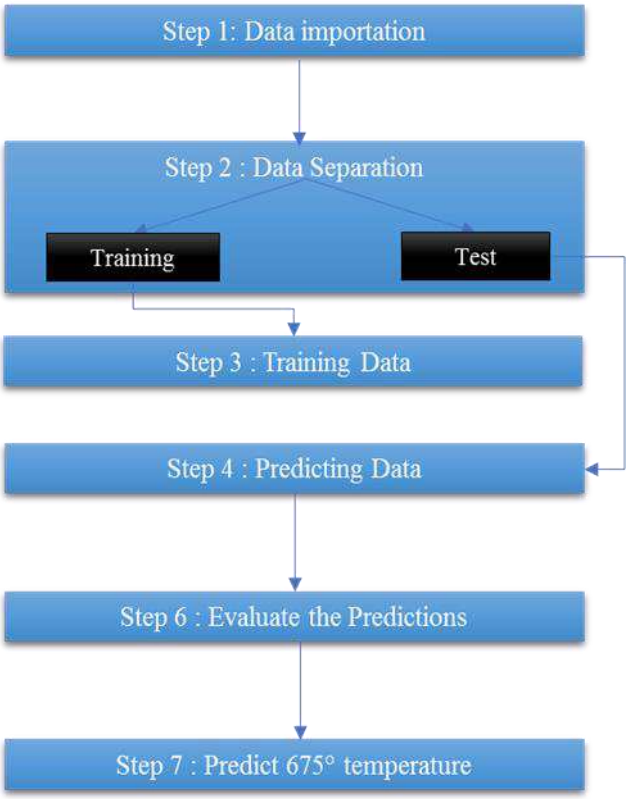

Figure 3: Methodology Diagram

Table (1): Compressive strengths in multi degrees of temperatures for four types of LWAGC

	Leca		*RBA*		*Porcelanite*		*RFA*	
Temp. C	**Comp.**	**SD**	**Comp.**	**SD**	**Comp.**	**SD**	**Comp.**	**SD**
35	20.1		27.2		24.2		17.7	
35	22.3		22.1		23.9		17.5	
35	21.8	1.511	25.5	2.914	21.2	1.486	19.4	1.410
35	23.2		28.1		21.8		16.5	
35	19.6		21.7		24.4		19.9	
200	24.7		25.8		27.1		22.1	
200	22.9		27.0		23.3		17.8	
200	26.8	2.650	28.9	1.685	24.2	2.633	17.4	1.907
200	19.6		25.4		21.7		18.4	
200	24.0		29.0		28.0		19.8	
400	21.7		22.8		19.7		19.0	
400	18.1		20.4		22.8		17.4	
400	19.9	1.656	22.1	1.080	23.2	1.811	16.7	1.294
400	21.4		20.9		23.6		19.1	
400	18.4		22.7		20.2		16.3	
550	15.1		17.8		20.2		13.7	
550	14.6		18.9		17.5		15.2	
550	14.9	0.816	22.2	2.397	16.6	1.887	15.8	1.041
550	15.9		22.1		20.1		13.9	
550	16.6		17.1		16.3		15.9	
675	13.7		16.2		15.9		12.9	
675	12.6		17.1		13.9		13.3	
675	13.0	0.618	15.8	0.750	16.4	1.359	14.0	0.618
675	12.4		15.1		17.2		12.4	
675	12.1		16.5		14.5		12.7	
800	13.0		12.0		15.1		12.1	
800	12.8		14.2		12.5		10.4	
800	9.9	1.621	13.2	1.151	13.2	1.274	12.6	1.394
800	10.4		11.9		12.4		13.3	
800	13.4		14.3		14.8		10.1	

4. Machine Learning Algorithms

In this research work, and to compare the results, seven models of machine learning were used, and the results were compared to them, as detailed later.

4.1 Artificial Neural Networks ANN

ANN are the most popular ML technique employed for CS prediction for various kinds of concrete mixtures, both conventional and unconventional. ANNs imitate the biological neural network that forms the brain [28]. An ANN is defined in computer science as an adaptive learning model that can learn from the effect of incoming input to anticipate an output through a training procedure that estimates the weight of each processing component, known as a neuron. First, each input parameter is assigned a weight, as shown in Figure 4. Then, using the weights and biases in combination with the inclusion bias, a basic calculation is performed to provide an input. Finally, it uses a specified activation function to compute the result.

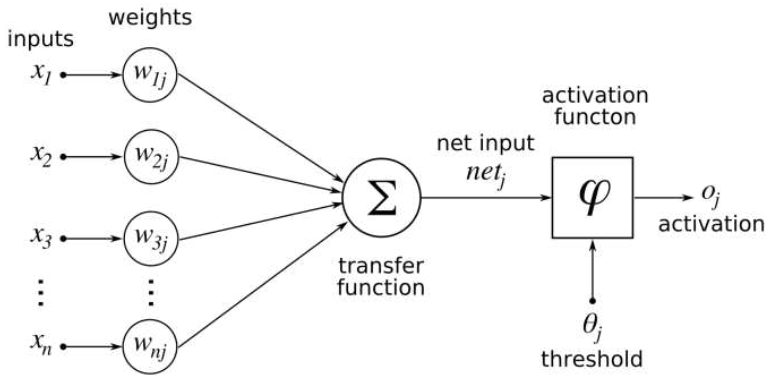

Figure 4: ANN architecture [30]

4.2 Convolutional neural network CNN

LeCun et al. first introduced CNN in 1995. The input layer, the convolutional layer, the pooling layer, the fully connected layer, and the output layer are the layers that make up the CNN as shown in Figure 5. The convolutional layer is responsible for doing the convolution on the data. It is possible to handle the input as a function. It is a different function when a filter is applied to it. The convolution process is a method of calculating the effect of applying a filter to an input signal. The pooling layer handles subsampling the data. This method reduces the computing cost of machine learning while also addressing the problem of overfitting in CNNs. The completely connected layer converts the extracted features from the previous layers to the final output [31].

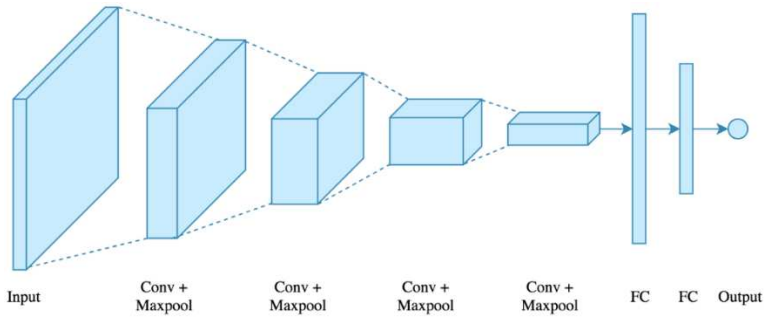

Figure 5: CNN architecture [32]

4.3 Long short-term memory LSTM

The Long short-term memory LSTM model, proposed by Hochreiter and Schmidhuber in 1997, eliminated the gradient problem in RNNs. This problem means that information is not stored for a long time, and the gradient in the deeper levels becomes ineffective. To address this problem, the LSTM model introduces a new cell known as a memory cell "C t," which can hold information from the deepest levels. Every memory cell has three gates: an entrance gate "I t," a forget gate "f t" and an output gate "O t" [33].
I t specifies whether the input should change the cell contents, f t specifies whether the cell contents should be reset, and O t specifies whether the cell contents should provide the neuron output.
The gates are sigmoid functions with binary values of 0 and 1, indicating that 0 allows no transmission and 1 allows all transmission. As shown in Figure 6 below.

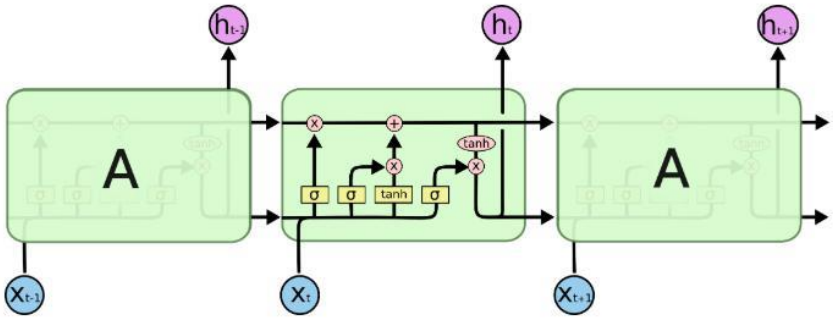

Figure 6: LSTM architecture [34]

4.4 Random Forest RF

Breiman (2001) was the first to introduce RF, which is a supervised machine-learning algorithm. The RF model is a development of the DT model. Forecasts in the RF regression model are based on the average of multiple trees' predictions. We will utilize 100 trees in this scenario. Specifically, each RF tree will be trained on a random subgroup of data using the bagging concept (to reduce the variance of the DT) and a random subgroup of features using the "random projections" method (to decrease the dimensionality of a group of points). The RF dataset is divided into subsets, and the GINI index is a

regularly used division criterion in the RF. Figure 7 shows RF in detail.

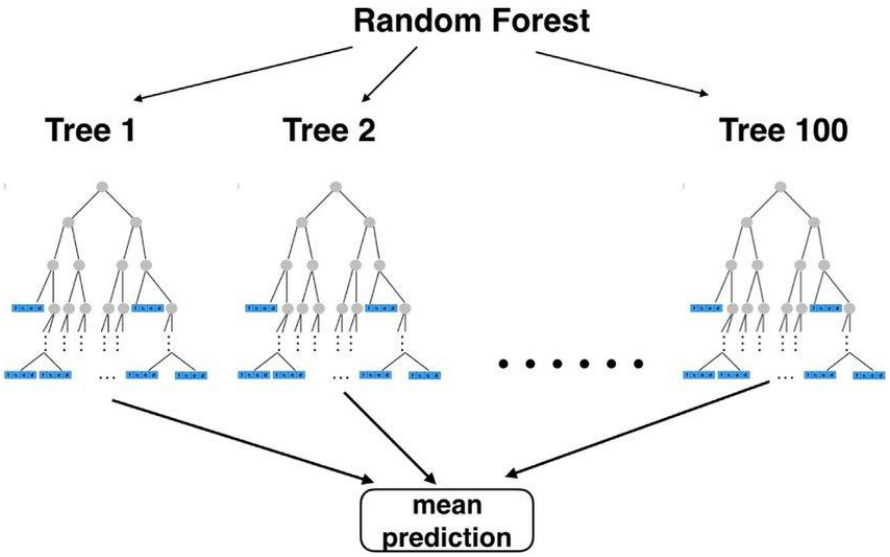

Figure 7: RF architecture[35]

4.5 Decision Tree DT

DT models are commonly used to explore data and generate a tree and associated rules for predictions. CHAID (Chi-squared Automatic Interaction Detection) and CART are two techniques that can be used to create DTs (Classification and Regression Trees). A DT is a tree in which each branch node represents a choice from a set of choices, and each leaf node represents a decision. Every node may contain more than two branches, according to the algorithm. CART, on the other hand, creates trees with only two branches at each node. This is referred to as a bina-ry tree. A multi-branch tree is one that has more than two branches allowed. As shown in Figure 8 below.

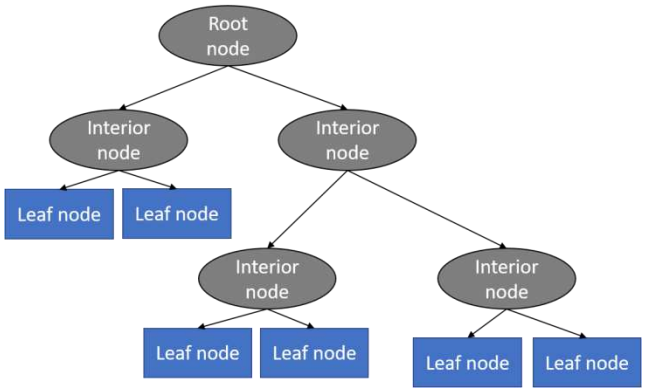

Figure 8: DT architecture [36]

4.6 Support Vector Regression SVM:

SVM is also one of the most well-known and reliable prediction algorithms. For regression, we employ SVR, which is a supervised machine-learning algorithm. It dates back to the 1990s. Its goal is to divide data points into classes using a hyperplane in an N-dimensional space. Figure 9 shows that the SVR to separate two classes, for example, numerous hyperplanes can be employed, but we need someone who can maximize the distance between them.

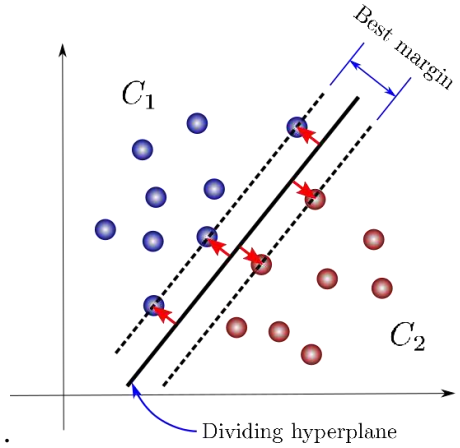

Figure 9: SVR split [37]

4.7 Simple linear regression LR

This model depicts the relationship between two variables, Y and X, via a linear function. Figure 10 and the equation of the model is presented as follows:

$$\boldsymbol{y_i = a_0 + a_1 x_i}$$

With:

$\boldsymbol{a_0}$ Is the intercept.

$\boldsymbol{a_1}$ Is the slope.

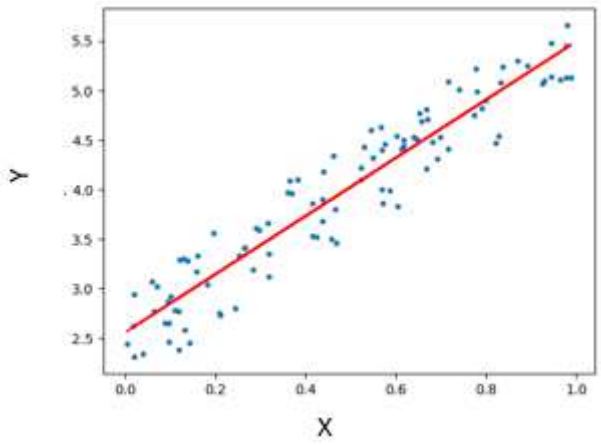

Figure 10: Regression line [38]

5. Forecasting performance measures:

Several metrics to evaluate the predictions in a regression problem; these measures were used.

5.1 Mean Square logarithm Error MSLE:

MSLE, or mean squared logarithmic error, is a measure of the difference between actual and expected values. MSLE is solely interested in the relative difference between the actual and anticipated values since the logarithm was introduced. In other words, the percentage difference between them is all that matters. As an example, see Eq (1) below.

$$\text{MSLE} = \frac{1}{n}\sum_{t=1}^{n}(\log(p_t + 1) - \log(y_t + 1))^2. \quad (1)$$

5.2 Mean Absolute Error MAE:

The MAE is used to calculate the average absolute difference between planned and actual values; a smaller MAE is preferred. In this investigation, the MAE is measured in Newton/mm2 (Mpa). As an Eq (2) below, the MAE is determined:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - p_t| \quad (2)$$

5.3 Mean Square Error MSE:

The arithmetic mean of the squares of the variances between model predictions and observations is called MSE, or mean square error. It's the square of the RMSE, too. The average of the squares of the errors is calculated from the average squared difference between the predicted and observed data by a predicting data indicator [39]. It does, in fact, describe the construction of the set of points along the regression line [40]. As seen in Equation 1, it does this by squaring the distances between the points and the regression line (these distances are the 'errors') (3). To eliminate any negative signals, squaring is required.

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - p_t)^2. \quad (3)$$

5.4 Root Mean Square Error RMSE:

The standard deviation of the residual after prediction is known as the root mean square error, or RMSE. The distance between the regression line and the data points is measured by the residuals in the regression. The residuals can be regarded of as errors that occur after a forecast has been made. The greater error numbers are given more weight by RMSE since it squares the error before adding it. As a result, RMSE is chosen when substantial errors are to be avoided. The Root Mean Square Error (RMSE) is the standard deviation residuals (forecast errors). In residuals, the distance between data points and the regression line is measured. The RMSE is a statistical tool that can be used to determine the distribution of these residuals. It does, in fact, depict the distribution of data around the best fit line [41; 42]. As an Eq (4) below

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - p_t)^2} \quad (4)$$

5.5 Root Mean logarithm Square Error RMSLE:

When there are huge variances in forecasts, RMSLE is used. Furthermore, it is the difference between the expected and actual log-transformed values. RMSLE adds 1 to both actual and projected values before taking the natural logarithm to avoid calculating the natural log of probable 0 (zero) values. As an example, see Eq (5) below.

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\log(p_t + 1) - \log(y_t + 1))^2} \quad (5)$$

5.6 Coefficient of Determination R-Squared:

The coefficient of determination is another approach to assess the accuracy of forecasts. The linear relationship between the expected and actual value is measured here. A high R2 indicates that the model is working well. The R2 is the ratio of the disparity in the parameter predicted by the independent variable to the disparity in the independent variable. It's a metric calculated by regression analysis. In linear regression, the R2 is the square of the correlation (R) between the anticipated f t-value and the y tActual-values; as a result, it has a range of 0 to 1 [42]. If there is no correlation, R2 = 0, suggesting that the independent variable cannot predict the dependent variable. When there is no correlation error, $R^2 = 1$, indicating that the independent variable can accurately forecast the dependent variable. As an Eq (6) below:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - p_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}. \quad (6)$$

Where:
$\boldsymbol{y_t}$: Actual value;
$\boldsymbol{p_t}$: predicted value;
$\boldsymbol{\bar{y}}$ **:** Average value;
$\boldsymbol{n}$: Number of observations.
All the metrics should be closer to zero, except R^2 should be near 1 to provide good predictions [43].

6. Results and discussion

6.1 Leca-Type Concrete:

With the exception of ANN and CNN, all models provide good forecasts for the Leca type. Table (2) and Figure (11) indicate the link between experimental and predicted values for various models. All models show a positive association between experimental and predicted values, with the exception of ANN and CNN, which exhibit a negative relationship. This demonstrates that the. The observed and projected values of the CS of Leca Type are shown in Figure (12).

Table 2. Performance measures for Leca type

	ANN	DT	RF	SVR	LSTM	CNN	LR
MSE	178,2266	2,7255	2,6697	2,7712	5,2004	2,7712	5,2828
RMSE	13,3502	1,6509	1,6339	1,6647	2,2804	1,6647	2,2984
MSLE	1,3978	0,0073	0,0072	0,0074	0,0121	0,0074	0,0126
RMSLE	1,1823	0,0856	0,0851	0,0861	0,1102	0,0861	0,1120
MAE	11,3251	1,3530	1,3434	1,3543	1,8389	1,3543	1,8145
R^2	0,6848	0,8660	0,8687	0,8637	0,7443	0,8637	0,6478

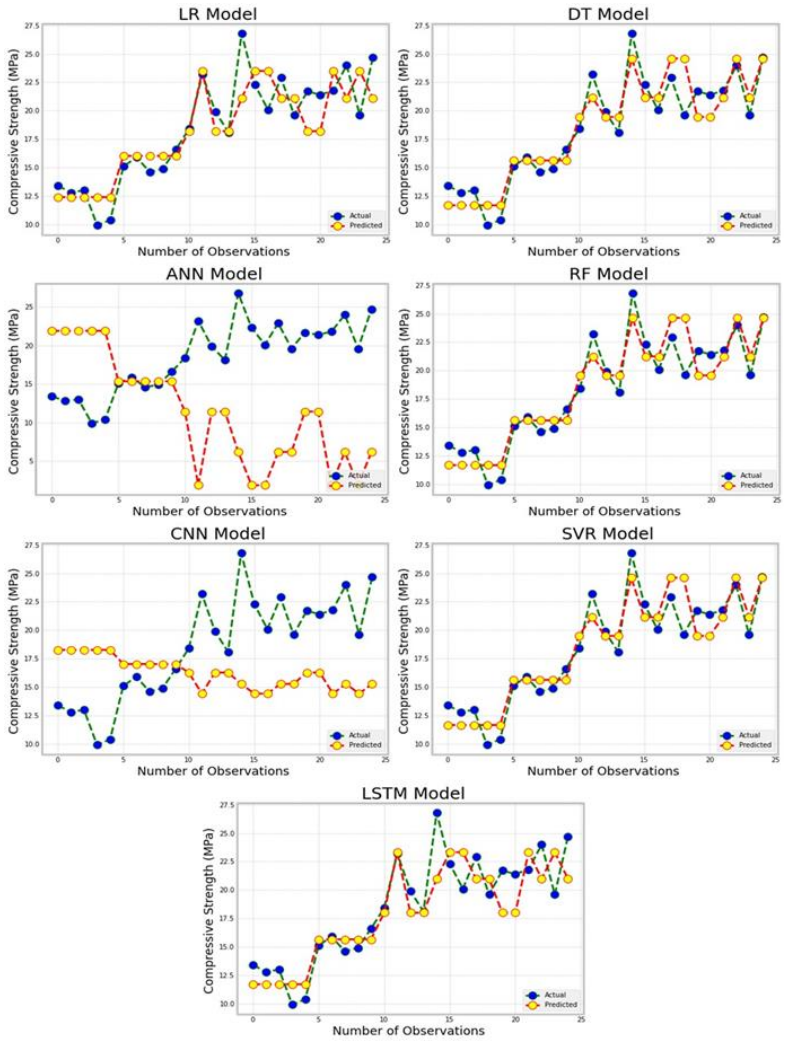

Figure 11: Actual VS predictions of Leca type

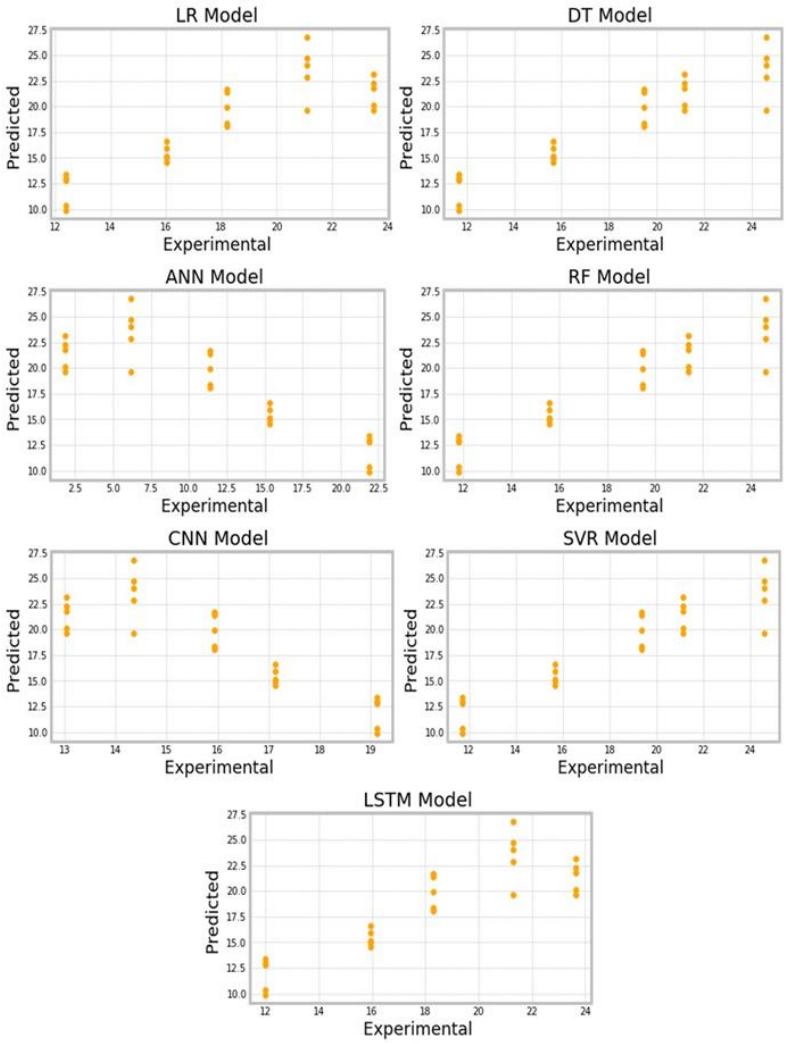

Figure 12: The link between the experimental and predicted values of the CS of Leca Type

6.2 RBA-Type Concrete:

For the RBA type, all models except ANN and CNN provide good forecasts. Table (3) and Figure (13) illustrate the relationship between experimental and predicted values for the various models; all models show a positive association between experimental and predicted values, except for ANN and CNN, which exhibit a negative relationship. This confirms that the projected values do not differ much from the values. Table (2) shows that the models for RBA type outperform other models in terms of prediction accuracy. Figure 14 depicts the relationship between experimental and expected RBA Type CS values.

Table 3: Performance measures for RBA type

	ANN	DT	RF	SVR	LSTM	CNN	LR
MSE	235,0804	3,3069	3,4117	3,3431	6,3950	3,3431	6,3364
RMSE	15,3323	1,8185	1,8471	1,8284	2,5288	1,8284	2,5172
MSLE	1,3782	0,0066	0,0068	0,0066	0,0117	0,0066	0,0121
RMSLE	1,1740	0,0812	0,0823	0,0814	0,1083	0,0814	0,1099
MAE	13,2336	1,5610	1,5741	1,5631	1,8275	1,5631	1,9098
R^2	0,6848	0,8764	0,8725	0,8750	0,7610	0,8750	0,7020

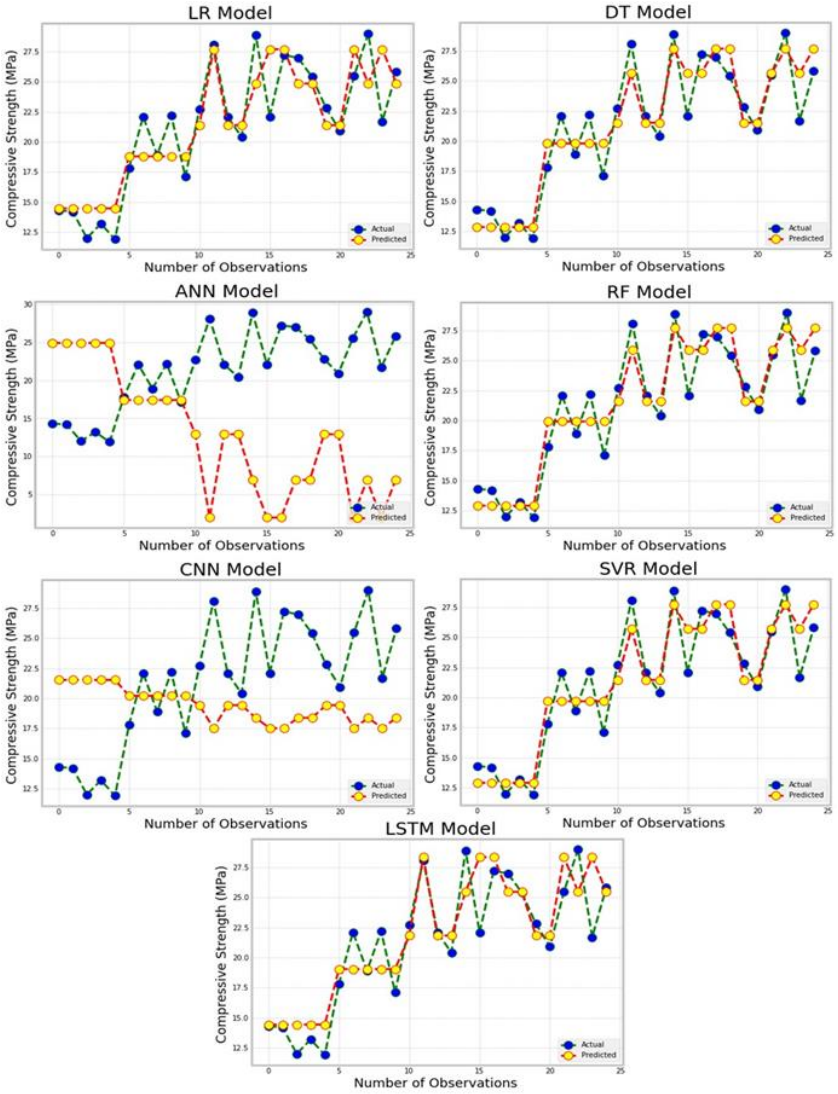

Figure 13: Actual VS predictions of RBA type

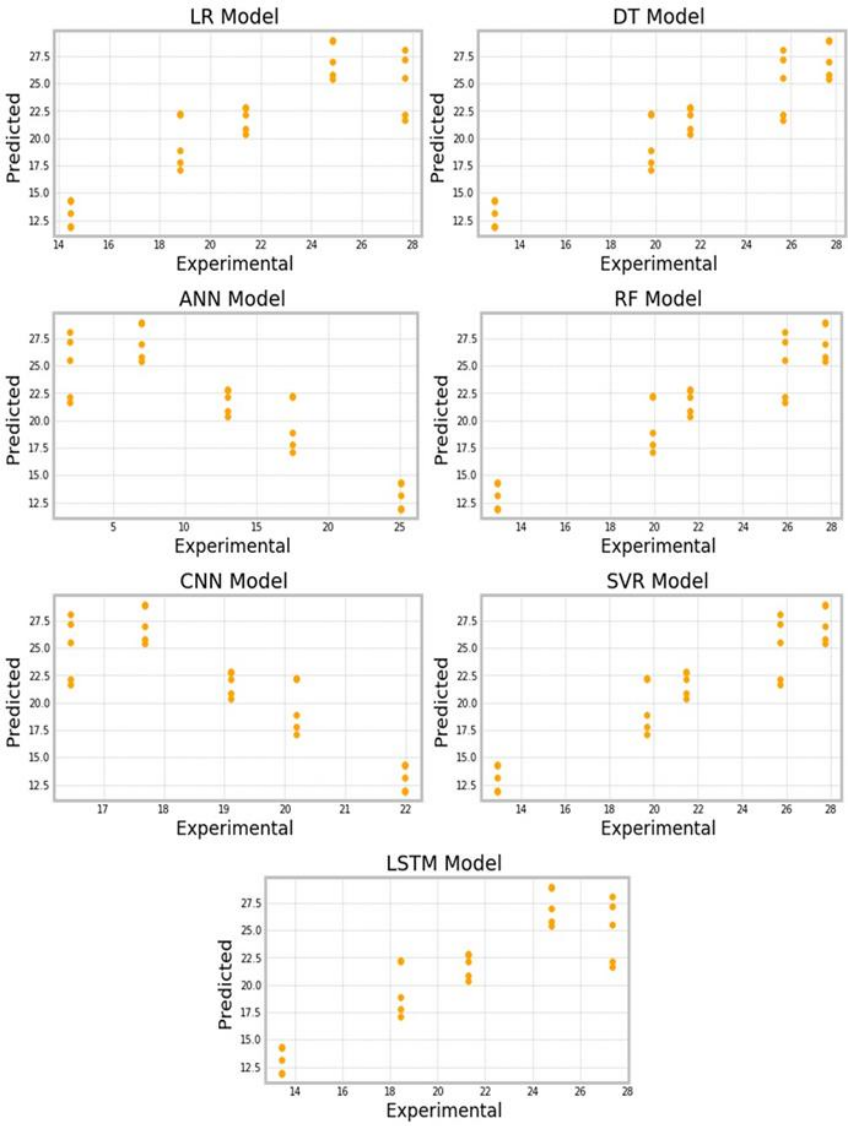

Figure 14: The link between the experimental and predicted values of the CS of RBA Type

6.3 Porcelanite-Type Concrete:

In the instance of Porcelanite type, we can see in Figure (15) that all of the models, with the exception of ANN and CNN, provide good forecasts. Figure (16) shows the relationship between experimental and anticipated values for several models; all models show a positive association between experimental and predicted values, with the exception of ANN and CNN, which exhibit a negative relationship. This demonstrates that the projected values do not match the experimental results. Table (4) shows the prediction accuracy of the models for Porcelanite, indicating that RF outperforms the other models..

Table 4: Performance measures for Porcelanite type

	ANN	DT	RF	SVR	LSTM	CNN	LR
MSE	**196,0019**	**3,0313**	**2,9650**	**3,0401**	**5,2227**	**3,0401**	**5,3366**
RMSE	**14,0001**	**1,7411**	**1,7219**	**1,7436**	**2,2853**	**1,7436**	**2,3101**
MSLE	**1,2457**	**0,0066**	**0,0065**	**0,0066**	**0,0105**	**0,0066**	**0,0106**
RMSLE	**1,1161**	**0,0810**	**0,0805**	**0,0812**	**0,1023**	**0,0812**	**0,1030**
MAE	**12,0988**	**1,5710**	**1,5693**	**1,5702**	**1,8041**	**1,5702**	**1,8730**
R^2	**0,6848**	**0,8402**	**0,8437**	**0,8397**	**0,7246**	**0,8397**	**0,6300**

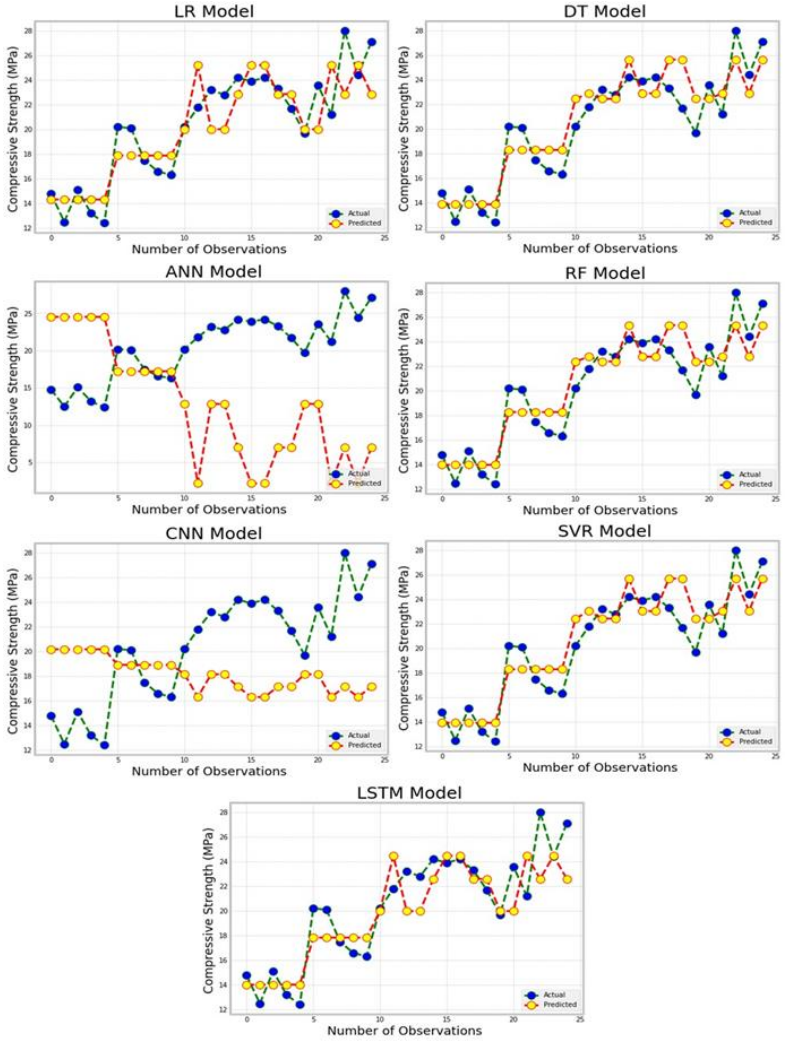

Figure 15: Actual VS predictions of Porcelanite type

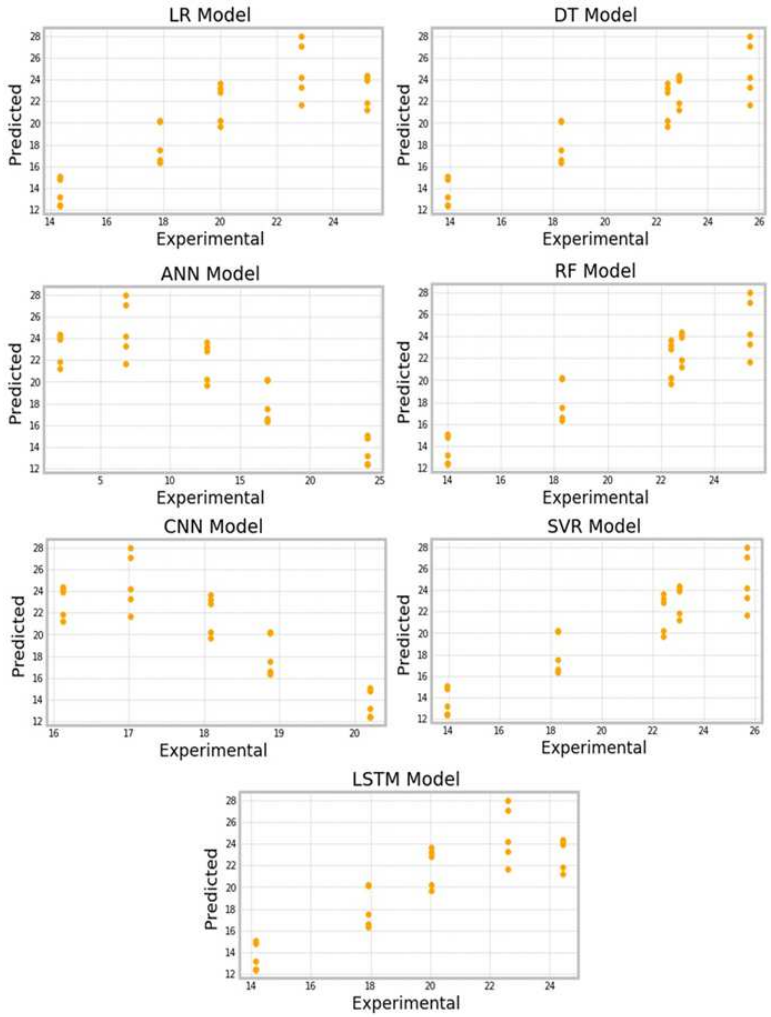

Figure 16: The link between the experimental and predicted values of the CS of PorcelaniteType

6.4 RFA-Type Concrete:

In the instance of the RFA type, we can see in Figure (17) that all of the models, with the exception of ANN and CNN, provide good forecasts. Figure 18 depicts the relationship between experimental and anticipated values for several models; all models exhibit a positive association between experimental and predicted values, with the exception of ANN and CNN, which show a negative relationship. This means that the anticipated values are not nearly as accurate as the experimental values. Finally, Table (5) shows the prediction accuracy of the models for the RFA type, indicating that RF outperforms the other models.

Table 5: Performance measures for RFA type

	ANN	DT	RF	SVR	LSTM	CNN	LR
MSE	116,5421	1,7092	1,7028	1,7185	2,8679	1,7185	2,9857
RMSE	10,7955	1,3074	1,3049	1,3109	1,6935	1,3109	1,7279
MSLE	1,1175	0,0059	0,0059	0,0060	0,0095	0,0060	0,0098
RMSLE	1,0571	0,0771	0,0768	0,0774	0,0975	0,0774	0,0991
MAE	9,3062	1,1450	1,1300	1,1442	1,3229	1,1442	1,3492
R^2	0,6848	0,8092	0,8099	0,8082	0,6799	0,8082	0,4512

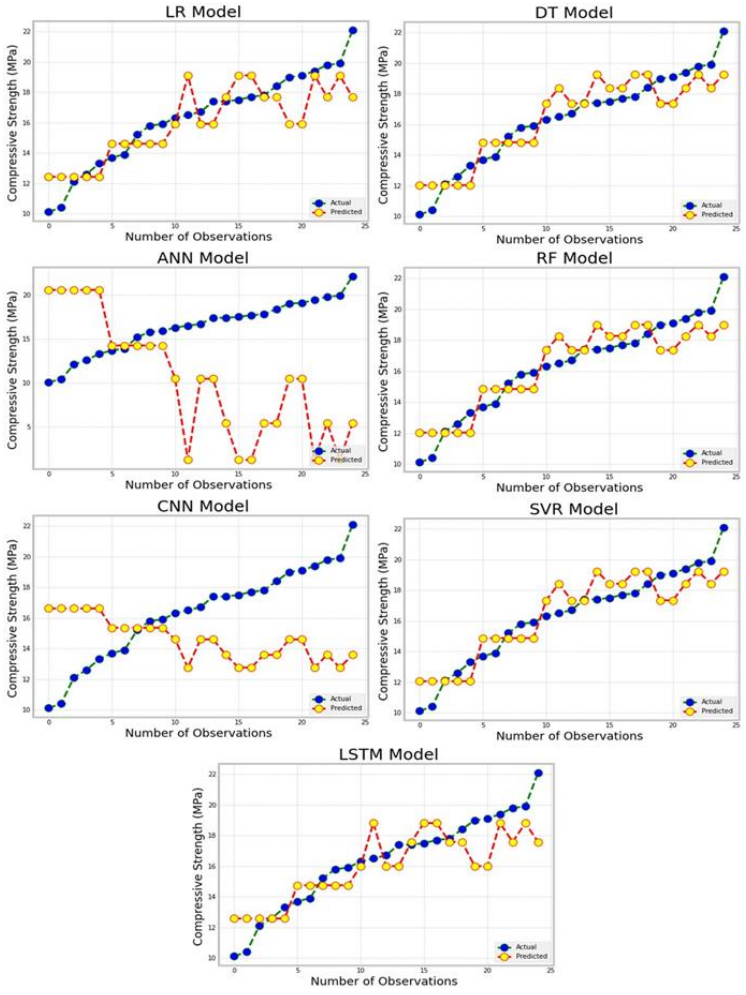

Figure 17: Actual VS predictions of RFA type

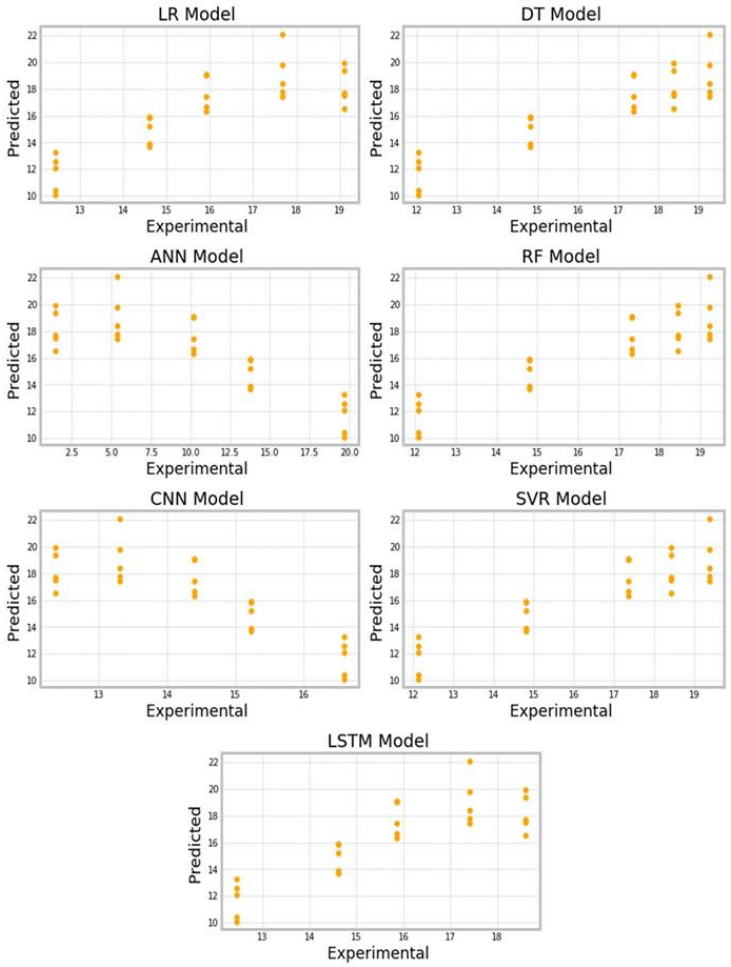

Figure 18: The link between the predicted and experimental values of the CS of RFA Type

The expected outcomes from the numerical models were compared with the actual results, as shown in the table below, and the model in forecasting CS based on other temperature values, to ensure that the models generated using machine learning were accurate. We change the temperature to 675 degrees Fahrenheit in the various models and compare the results to the experiment data. The results of the comparison are shown in Table 6, with the first three categories well predicted using LR and the fourth type well predicted using CNN. The comparison of the results is shown in Figure 19.

Table 6: Prediction of CS in 675°

	ANN	DT	RF	SVR	LSTM	CNN	LR	Actual
Leca	18,14	15,63	15,58	15,61	23,84	14,63	14,21	12,76
RBA	21,27	19,80	19,93	19,62	28,35	17,65	16,63	16,14
Porcelanite	20,68	18,30	18,27	18,21	24,66	14,98	16,10	15,58
RFA	16,77	14,83	14,84	14,81	18,72	12,70	13,51	13,06

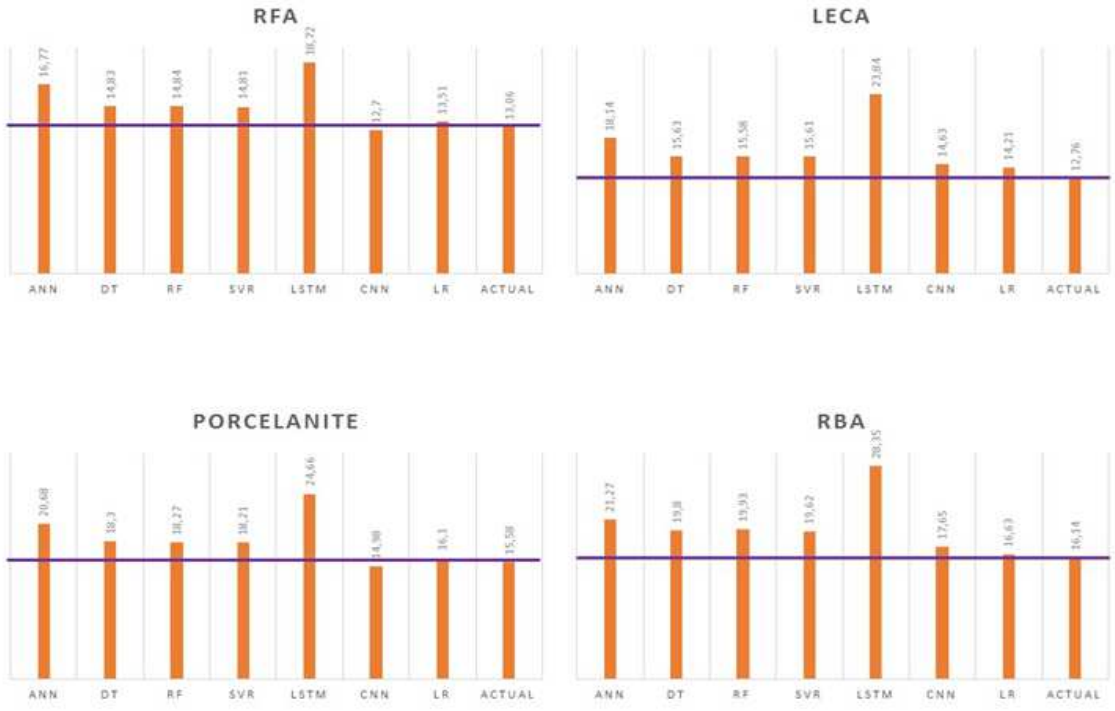

Figure 19: Prediction of CS in 675°

7. Comparison of ML Models

This study examines the performance of ANN, CNN, LSTM, DT, RF, SVR and LR models to evaluate its performance in predicting the CS of 4 types of concretes.

In Leca type RF provides the highest correlation and lowest performance metrics (MSE= 2,6697;

RMSE= 1,6339; MSLE= 0,0072; RMSLE= 0,0851; MAE= 1,3434; R^2=0,8687). In RBA type DT provides the highest correlation and lowest performance metrics (MSE= 3,3069; RMSE= 1,8185; MSLE= 0,0066; RMSLE= 0,0812; MAE= 1,5610; R^2=0,8764). In Porcelanite type, RF has the highest correlation and lowest performance metrics (MSE= 2,9650; RMSE= 1,7219; MSLE= 0,0065; RMSLE= 0,0805; MAE= 1,5693; R^2=0,8437). In RBA type RF has highest correlation and lowest performance metrics (MSE= 1,7028; RMSE= 1,3049; MSLE= 0,0059; RMSLE= 0,0768; MAE= 1,1300; R^2=0,8099).

The prediction of the CS using the new temperature shows that LR gives the best results for Leca, RBA, and Porcelanite types, and CNN provides the best prediction for RFA type.

The inability of other models to predict the CS for the different types can be explained by two factors. The first one is the limited data, and the second one is that we used only one feature.

Compared to other works, our work predicts the CS with seven models; however, other studies use 2 until three models ([15], [16], [17], [29], [20], [21], [22], [23], [24], [25], [26], [27], [28]). In addition, in the majority of cited works above, they used ANN and found good results; in contrast, ANN gives the worst results among the models.

8. Conclusion

We proposed seven models, namely ANN, CNN, LSTM, DT, RF, SVR, and LR, to anticipate the CS of different concrete types. The results demonstrate that RF offers high accuracy for Leca type, Porcelanite type, and RFA type because the predictions are closer to the actual values, and DT offers better predictions than other models for the RBA type. Furthermore, the comparison between experimental results and the predictions of CS when concrete is subjected to a temperature equal to 675° demonstrates that LR is more accurate than other models for Leca, RBA, and Porcelanite types. However, CNN outperformed other models in predicting RFA type.

Two main factors might explain the inefficiency of other models in predicting the CS for different types. The first one is the limited data, and the second one is using one feature. These limits constitute the basics of our future works, where we will use more features and data and intend to use hybrid models such as ANN-LSTM or CNN-LSTM.

Acknowledgements:

Authors are thankful to Universiti Tenaga Nasional (UNITEN), Malaysia for providing financial supports for this study under BOLD Publication Fund 2022 (J510050002 - IC-6 BOLDREFRESH2025 - CENTRE OF EXCELLENCE).

Data availability:

The datasets used and analyzedduring the current study available from the corresponding author on reasonable request

9. References:

1. Hosan, S. Haque, F. Shaikh, Compressive behaviour of sodium and potassium activators synthetized fly ash geopolymer at elevated temperatures: a comparative study, Journal of Building Engineering 8 (2016) 123–130.
2. Chandrakanth, V., & Koniki, S. (2020). Effect of Review elevated temperature on geopolymer concrete–A Review. In E3S Web of Conferences (Vol. 184, p. 01090). EDP Sciences.
3. H. Mikulcˇic´ , J.J. Klemeš, M. Vujanovic´ , K. Urbaniec, N. Duic´ , Reducing greenhouse gasses emissions by fostering the deployment of alternative raw materials and energy sources in the cleaner cement manufacturing process, Journal of cleaner production 136 (2016) 119–132.
4. K.C. Onyelowe, D. Bui Van, O. Ubachukwu, C. Ezugwu, B. Salahudeen, M. Nguyen Van, B. Ugorji, Recycling and reuse of solid wastes; a hub for ecofriendly, ecoefficient and sustainable soil, concrete, wastewater and pavement reengineering, International Journal of Low-Carbon Technologies 14 (3) (2019) 440–451.
5. A.A. Patil, H.S. Chore, P.A. Dode, Effect of curing condition on strength of geopolymer concrete, Advances in concrete construction 2 (1) (2014) 029.
6. F.N. Okoye, S. Prakash, N.B. Singh, Durability of fly ash based geopolymer concrete in the presence of silica fume, Journal of cleaner Production 149 (2017) 1062–1067.
7. Bosoaga, O. Masek, J.E. Oakey, CO2 capture technologies for cement industry, Energy procedia 1 (1) (2009) 133–140.
8. P. Duxson, A. Fernández-Jiménez, J.L. Provis, G.C. Lukey, A. Palomo, J.S. van Deventer, Geopolymer technology: the current state of the art, Journal of materials science 42 (9) (2007) 2917–2933.
9. Thienel, K. C., Haller, T., & Beuntner, N. (2020). Lightweight concrete–from basics to innovations. Materials. 13, 1120.
10. ACI Committee 213. ACI 213R-14 Guide for Structural Lightweight-Aggregate Concrete; American Concrete Institute: Farmington Hills, MI, USA, 2014.
11. Wongkeo, W., Thongsanitgarn, P., Pimraksa, K., & Chaipanich, A. (2012). Compressive strength, flexural strength and thermal conductivity of autoclaved concrete block made using bottom ash as cement replacement materials. *Materials & Design*, *35*, 434-439.
12. Kumar, A., Arora, H. C., Kapoor, N. R., Mohammed, M. A., Kumar, K., Majumdar, A., & Thinnukool, O. (2022). Compressive strength prediction of lightweight concrete: machine learning models. *Sustainability*, *14*(4), 2404.
13. G. Mallikarjuna Rao, T.D. Gunneswara Rao, A quantitative method of approach in designing the mix proportions of fly ash and GGBS-based geopolymer concrete, Australian Journal of Civil Engineering 16 (1) (2018) 53–63.
14. Kandiri, A., Sartipi, F., & Kioumarsi, M. (2021). Predicting compressive strength of concrete containing recycled aggregate using modified ANN with different optimization algorithms. *Applied Sciences*, *11*(2), 485.
15. J. Kasperkiewicz, J. Racz, A. Dubrawski, HPC strength prediction using artificial neural networks, Comput. Civ. Eng. 4 (1995) 279–284

16. I.C. Yeh, Modeling of strength of high-performance concrete using artificial neural networks, Cem. Concr. Res. 28 (1998) 1797–1808
17. I.C. Yeh, Design of high-performance concrete mixture using neural networks and nonlinear programming, J. Comput. Civ. Eng. 13 (1999) 36–42.
18. G. Tayfur, T.K. Erdem, K. Onder, Strength prediction of high-strength concrete by fuzzy logic and artifficial neural networks, J. Mater. Civ. Eng. 26 (2014)
19. M.-Y. Cheng, P.M. Firdausi, D. Prayogo, High-performance concrete CS prediction using genetic weighted pyramid operation tree, Eng. Appl. Artif. Intell. 29 (2014) 104–113.
20. M. Nehdi, H. El Chabib, H. El Naggar, Predicting performance of self-compacting concrete mixtures using artificial neural networks, ACI Mater. J. 98 (2001) 394–401.
21. M. Uysal, H. Tanyildzi, Predicting the core concrete strength of self compacting concrete mixtures with mineral additives using artificial neural network, Constr. Build. Mater. 25 (2011) 4105–4111.
22. I.B. Topcu, M. Saridemir, Prediction of CS of concrete containing fly ash using artificial neural networks and fuzzy logic, Comput. Mater. Sci. 41 (2008) 305–311.
23. I.B. Topçu, M. Saridemir, Prediction of mechanical properties of recycled aggregate concretes containing silica fume using artificial neural networks and fuzzy logic, Comp. Mater. Sci. 42 (1) (2008) 74–82.
24. H. Naderpour, A. Kheyroddin, G.G. Amiri, Prediction of FRP-confined CS of concrete using artificial neural networks, Compos. Struct. 92 (12) (2010) 2817–2829.
25. M. Jalal, A.A. Ramezanianpour, Strength enhancement modeling of concrete cylinders confined with CFRP composited using artificial neural networks, Compos. Part B 43 (2012) 2990–3000.
26. M. Nehdi, Y. Djebbar, A. Khan, Neural network model for preformed-foam cellular concrete, ACI Mater. J. 98 (5) (2001) 402–409.
27. A. Ashrafian, F. Shokri, M.J. Taheri Amiri, Z.M. Yaseen, M. Rezaie-Balf, CS of foamed cellular lightweight concrete simulation: new development of hybrid artificial intelligence model, Constr. Build. Mater. 230 (2020), 117048.
28. K.R. Mahmood Al-Janabi, A.I. Abdulwahab Al-Hadithi, Modeling of Polymer Modified-Concrete Strength with Artificial Neural Networks, 2008, pp. 47–68.
29. ASTM C330M-17a. Standard Specification for Lightweight Aggregates for Structural Concrete; ASTM International: West Conshohocken, PA, USA, 2017.
30. https://bdtechtalks.com/2019/08/05/what-is-artificial-neural-network-ann/
31. Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables. Expert Syst. Appl., 129, 273-285.
32. https://viblo.asia/p/ung-dung-convolutional-neural-network-trong-bai-toan-phan-loai-anh-4dbZNg8ylYM.
33. Ifleh, A., & El Kabbouri, M. (2021). Moroccan Stock Market Prediction Using LSTM Model on a Daily Data. In Intelligent Systems (pp. 11-17). Springer, Singapore.
34. Achkar, R., Elias-Sleiman, F., Ezzidine, H., & Haidar, N. (2018, August). Comparison of BPA-MLP and LSTM-RNN for stocks prediction. In 2018 6th International Symposium on Computational and Business Intelligence (ISCBI) (pp. 48-51). IEEE.
35. Nedjati-Gilani, G. L., Schneider, T., Hall, M. G., Cawley, N., Hill, I., Ciccarelli, O. & Alexander, D. C. (2017). Machine learning based compartment models with permeability for

white matter microstructure imaging. NeuroImage, 150, 119-135.
36. https://mobillegends.net/tree-based-algorithms-implementation-in-python-r.
37. https://ankitnitjsr13.medium.com/math-behind-svm-support-vector-machine-864e58977fdb.
38. https://medium.com/@shafaqshaikh10/linear-regression-8628af6861cc.
39. Najah, A., Lam, T. V., Duy, N., Thieu, N. V., & Kisi, O. (2021). A comprehensive comparison of recent developed meta-heuristic algorithms for streamflow time series forecasting problem. Applied Soft Computing, 105, 107282
40. Sami, B. H. Z., Sami, B. F. Z., Fai, C. M., Essam, Y., Ahmed, A. N., & El-Shafie, A. (2021). Investigating the reliability of machine learning algorithms as a sustainable tool for total suspended solid prediction. Ain Shams Engineering Journal,
41. Chong, K. L., Lai, S. H., Ahmed, A. N., Zaafar, W. Z. W., Rao, R. V., Sherif, M., Sefelnasr, A., & El-Shafie, A. (2021). Review on dam and reservoir optimal operation for irrigation and hydropower energy generation utilizing meta-heuristic algorithms. IEEE Access, 9, 19488–19505
42. Teo, A. N. F. Y., Huang, M. F. C. Y. F., Abdullah, S. D. L. S., & Shafie, A. E. (2021). Surface water quality status and prediction during movement control operation order under COVID – 19 pandemic: Case studies in Malaysia. International Journal of Environmental Science and Technology, 0123456789. https://doi.org/10.1007/s13762-021-03139-y.
43. Nur Adli Zakaria, M., Abdul Malek, M., Zolkepli, M., & Najah Ahmed, A. (2021). Application of artificial intelligence algorithms for hourly river level forecast: A case study of Muda River, Malaysia. Alexandria Engineering Journal