Recent Trends in Civil Engineering and Built Environment Vol. 2 No. 1 (2021) 691-699
© Universiti Tun Hussein Onn Malaysia Publisher's Office

RTCEBE

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/rtcebe
e-ISSN :2773-5184

Concrete Strength Prediction Using Linear Regression of Machine Learning Algorithm

Peggy Suenie Anak Achong[1], Nickholas Anting Anak Guntor[1],*

[1]Faculty of Civil Engineering and Built Environment,
Universiti Tun Hussein Onn Malaysia, Batu Pahat,86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/rtcebe.2021.02.01.075
Received 00 Month 2020; Accepted 01 Month 2020; Available online 02 Month 2020

Abstract: Compressive strength is a performance measurement to determine the quality of the concrete. It is one of the crucial parameters used by industries to evaluate the performance of concrete. This study will be more focused on the machine learning model. The Multiple Linear Regression (MLR) algorithm is used to train the machine learning. The model was developed using KNIME analytical platform software. The data set consists of 202 observations with 26 attributes. The dataset was divided into 80% for training set and 20% for testing set. The training set is used to train the algorithm model, while the testing set is used to evaluate the performance of the model. A comparison was made between the MLR model and the Polynomial Regression (PR) model. Fine aggregate (kg/m3) was found to be the most influential contributor to increasing the compressive strength of concrete in the MLR model. The MLR model's R-squared value is 0.589, lower than the R-squared PR model, which is 0.745. The performance of the models was eventually indicated by using the Linear Correlation node. It is clearly indicated that both models are considered good models for the prediction in the machine learning model. Initially, the MLR model was proposed to be used as the machine learning prediction model. Compared with the PR model, the R-squared value of the MLR model was lower than PR model; but, it still indicates the model can still produce a good model. Thus, the PR model is recommended in the machine learning prediction model to predict the concrete compressive strength.

Keywords: Compressive Strength, Prediction, Regression, Performance, R-squared

1. Introduction

The production of concrete is conventionally using the Portland cement type and globally used in the construction field for building construction and infrastructure industries [1]. A better

**Corresponding author: nickholas@uthm.edu.my*

understanding of concrete behaviour can be established through the application of external loading and mix design methodologies. Therefore, the evaluation of concrete strength parameters and study of concrete mechanical properties must be conducted to make sure that the quality of concrete is in guarantee for the use of construction material [2]. Machine learning is the employment of artificial intelligence (AI) in which the systems providing the potential to learn and improve automatically from past experience data without a direct programming. The learning process of machine learning initiates with direct data observations that provide the best decisions in the future by observing the data pattern.

One of the approaches that can be employed for specific function of model training is multiple linear regressions. Multiple linear regressions is one of the technique of machine learning algorithms that serves as a good baseline [3]. Besides, it can be evaluated to provide the best and useful understanding of the data analysis. A model that represented by multiple linear regressions approach is the linear combination of predictor variable that described by the output variable. Thus, this study is purposely conduct to develop a machine learning model that could predict the compressive strength of the concrete mixture based on multiple parameters that obtained through literature.

2. Literature Review

The focus of this chapter is to provide comprehensive literature reviews on current updates of research topics that are related with enhancement of compressive strength of the concrete. The innovation in concrete materials will be explored in this chapter as a review for the study of concrete compressive strength. In order to predict the concrete compressive strength, the data will be obtained through the previous journal. Evaluated parameters such as compressive strength must be done to ensure the concrete is in good quality. Therefore, the mechanical properties such as compressive strength are the factors that influence the true quality of concrete.

2.1 Innovation in concrete materials

Demand in construction materials has increased over the past period of time, hence the construction industries have produced waste from rehabilitation and demolition works [4]. This is a matter of concern that may lead to the production of waste that will impact the environment as well as the human being. Innovation of concrete materials is the solution which helps to increase the efficiency and performance in concrete material, but also, helps to reduce the construction cost in term of production and shorten the time of project. Through the implementation of concrete innovation, construction industries will produce much greener construction process in order to reduce waste, control the pollutions, promote reusable materials, plus to decrease the demand for the raw materials in the process of construction.

2.2 Compressive strength of concrete

Concrete is the combination of elements such as cement, fine and coarse aggregate and water. Cement acts as the binding agent which holds together the aggregates to produce a composite material known as concrete. The nature of concrete can be defined through its strength, durability, and reliability. Consequently, compressive strength of concrete is essential mechanical properties to withstand the load stress that is transmitted from the structure [5]. Concrete strength can be obtained by curing the concrete specimen for approximately 28 days [6].

2.3 Machine learning approach

Machine learning is one of the artificial intelligence (AI) subsets that could be used, for example, to predict the compressive strength of concrete, for data classification and recognition [7]. The machine learning model can predict the concrete compressive strength through the targeted variable such as types and quantities, as well as through input variable such as the concrete mixture component [3].

Feng et al. (2020) stated that several algorithms can be taught based on input data and will provide accurate output data results through the adoption of machine learning [7]. DeRousseau et al. (2019) also claimed that, due to intuitive understanding, the collected data set can be instructed and processes the relationship between the different targeted input data in the absence of restriction [3]. Machine learning models can be trained by different methods that are typical to be classified into three key categories such as supervised, unsupervised and reinforcement machine learning.

2.4 KNIME software

KNIME is an applicable open-source of analysis tool to analyse statistics of Univariant and Multivariant. The analytical tool are provides of network, social media analysis, web analysis, plus, to decide time series and image processing. KNIME platform is available in API integration and natural language processing.

In an experiment by Fillbrunn et al. (2012) indicate that KNIME is an analytical platform focus to analyse data through a platform of graphical connecting tools that would show the result on different operating systems [8]. As KNIME is an open-source of analytical platform, it will assist programmers to perform their analytical pipelines which combine the data retrieval visualization, analysis and exploration. Thus, the researcher conclude that the tools of workflow comes with variety of different usage and it is the assignment for the programmers to select the tool which suitable for the certain situation and there are precisely overlaps for each tool that will perform at its purposes.

3. Methodology

The predictions of concrete compressive strength are demonstrated through the selected machine learning approach. Initially, the process will involve data collection, data preparation, exploration and pre-processing, and model performance evaluation. It is to create a prediction model based on the data collected from the previous study literature on concrete innovation study. The parameters such as cement, fine and coarse aggregate, water-to-cement ratio, temperature, humidity, curing period, concrete compressive strength, and so on are selected to form a new set of data. KNIME software will be proposed in the development of a concrete compressive strength prediction model.

Data preparation is a procedure of converting raw data, train the data, and make predictions through the algorithm of machine learning. The procedure is also known as data pre-processing which transforms the collected raw data into data that can be interpreted simply by using the algorithm. The well-organized data based on sixteen (16) literature journals were prepared before proceeding with the machine learning prediction model.

Data exploration and pre-processing is the initial phase of data analysis, in which an unstructured data set is analyzed to discover the initial trends, features, and points of interest. It is to uncover trends and points of interest, the aim is to explore and understand relationships between different data variables, the structure of the data collection, the occurrence of outliers, missing value and duplicated data, and the distribution of data values in order to gain greater insight into the raw data.

The data is partitioned or well-divided into 80% for training set and 20% for test set. A multiple linear regression algorithm is proposed to train and test the partitioned data set. The model's evaluation will be represented by the coefficient value, denoted as R^2. The evaluation also will be confirmed using the correlation coefficient value to measure the correlation between the Actual Value and Predicted Value of the model, denoted as R. The performance of the model will be valid if the coefficient value is near 1. Figure 1 below shows the framework of methodology for this study.

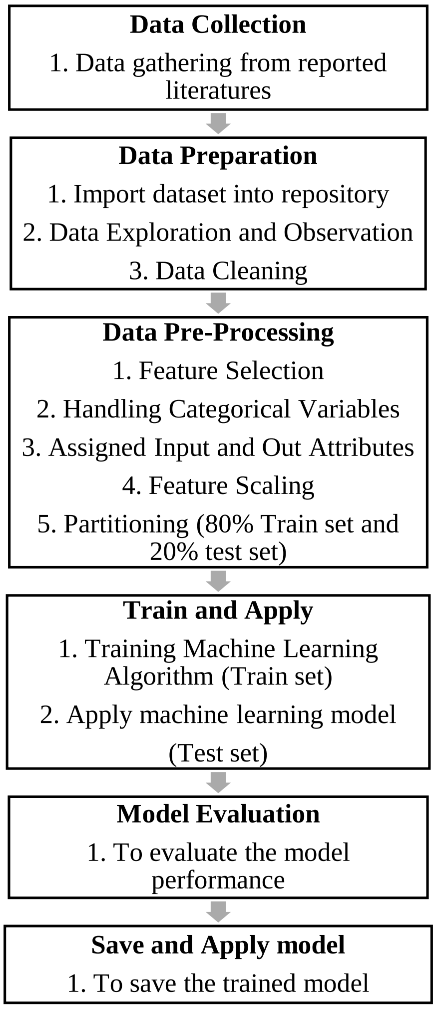

Figure 1: Framework of methodology

4. Results and Discussion

4.1 Data Exploration of models

The descriptive statistical of variables are presented in the form of statistical table in KNIME interactive view. The data distribution shows the maximum and minimum value of the dependent variable is 104.70 MPa and 14.70 MPa, respectively. The value of the mean and median are 43.222 MPa and 40.205 MPa, respectively. The value of the standard deviation for the dependent variable is 19.489 MPa, while the value of variance is 379.815 MPa. In Figure 2, the histogram shows that a positively skewed distribution occurs with the tail on its right side. The value of the skewed distribution is 1.262 indicates the value is larger than 0. Therefore, it determines that the distribution of the skewness is likely on the right. It also means that the value of the mean is the highest followed by the median value.

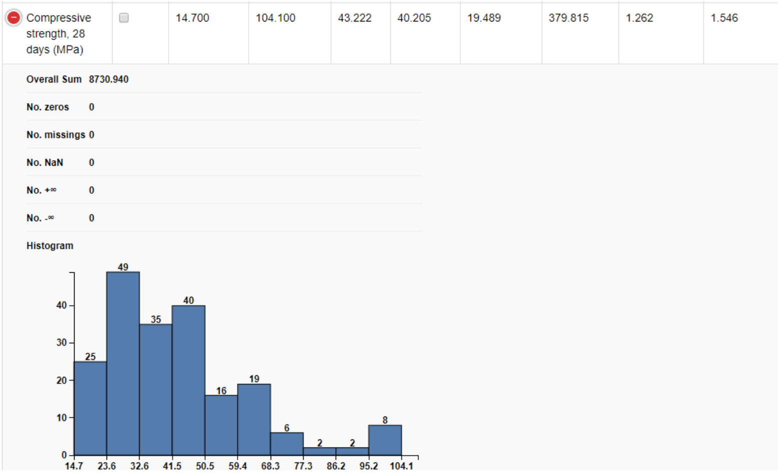

Figure 2: Data distribution of concrete compressive strength (MPa) histogram 28 days

4.2 Correlation of models

In the Figure 3, the correlation coefficient can be measure by the Linear Correlation node. The analysis of linear correlation, denoted as R, was represented in the form of correlation matrix visualization. The visualization of the correlation matrix managed to show that the darker the colour contrast, the stronger the correlation between the two variables. Correlation values were shown by highlighting the mouse cursor on the colour pallet. The darkest colour contrast between two variables that have an R-value of more than 0.9 was considered to be excluded from the feature selection to train the data for the algorithm.

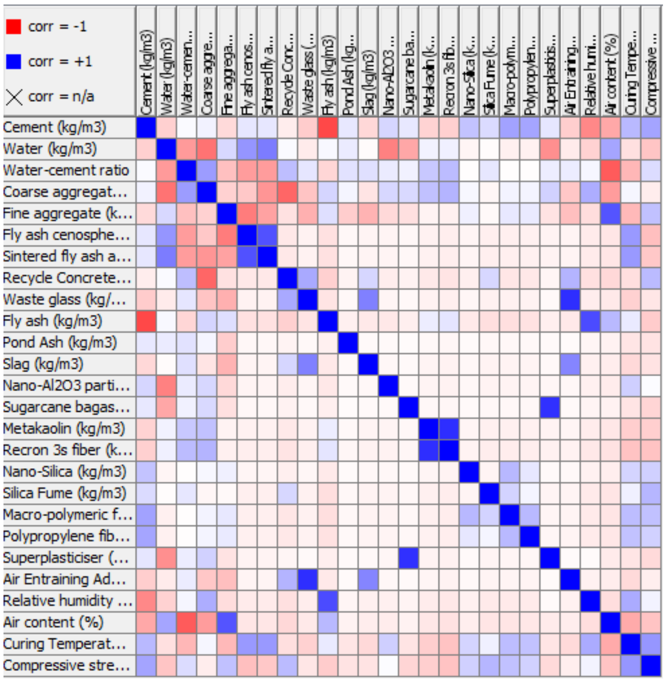

Figure 3: Linear Correlation matrix between all variables

4.3 Comparison between MLR model and PR model performance evaluation

A strong correlation between prediction and actual compressive strength value can be assumed if the R-squared value near to the value of 1. Based on Figure 4 and Figure 5, the MLR model shows the value of the R-squared is 0.589. The PR model with degree 2 in Figure 6 and Figure 7, on the other hand, has the highest R-squared value compared to the R-squared value of the MLR model. Even though the correlation coefficient result indicates that the MLR model has a good fit model, the PR model eventually produces much better performance. Table 1 and Table 2 indicates the five (5) top and least independent variable contributors.

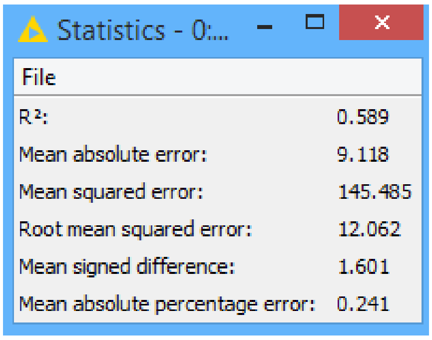

Figure 4: Numeric Scorer result for MLR model

Table 1: 5 top independent contributory variables

Independent Variables	Coefficient Value	P-Value
Fine Aggregate (kg/m3)	35.336	0.000
Recycle Concrete Aggregate (kg/m3)	20.625	0.000
Coarse Aggregate (kg/m3)	18.781	0.000
Cement (kg/m3)	17.140	0.000
Fly Ash Cenosphere (kg/m3)	12.756	0.000

Table 2: 5 least independent contributory variables

Independent Variables	Coefficient Value	P-Value
Nano-Silica (kg/m3)	-0.077	0.913
Metakaolin (kg/m3)	1.545	0.247
Silica Fume (kg/m3)	3.944	0.000
Water-cemen ratio	5.061	0.004
Fly Ash (kg/m3)	5.851	0.011

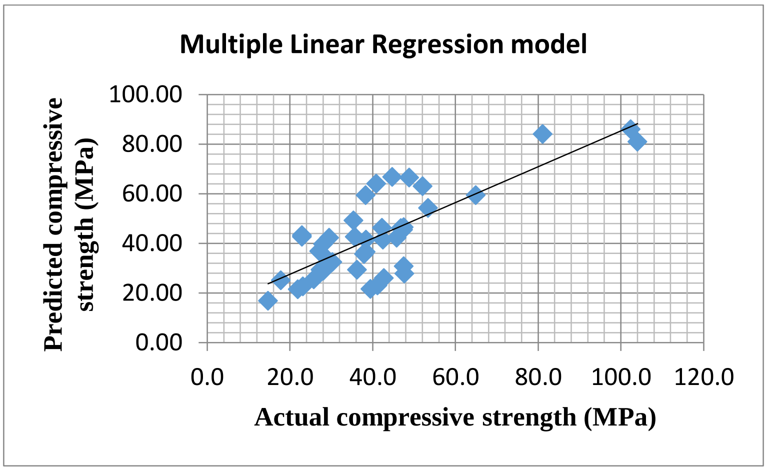

Figure 5: Relationship between predictd compressive strength and actual predicted compressive strength (MPa)

Based on Table 1, Fine Aggregate (kg/m3) is the most influential contributor to the increment of the concrete compressive strength. Surprisingly, Recycle Concrete Aggregate (kg/m3) is the second most influential contributor after fine aggregate. According to Maier & Durham (2012), fifty percent (50%) of the recycled materials were found to be beneficial to a concrete mixture [9]. On the opposite, Table 2 showed the 5 least independent contributory variables. It was found in Table 2, that Nano-silica has a negative coefficient value of -0.077. It means that the direction of the independent variables and dependent variable indicated by the sign of each coefficient value.

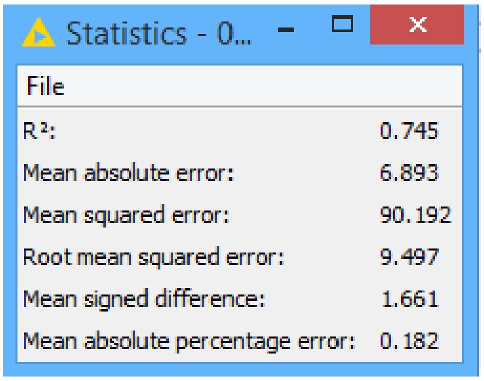

Figure 6: Numeric Scorer result for PR model with degree 2

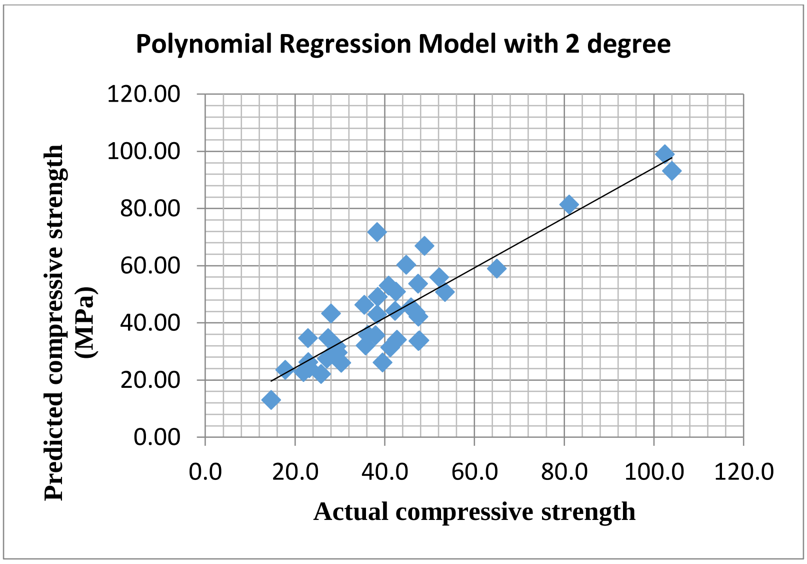

Figure 7: Relationship between predictd compressive strength and actual predicted compressive strength (MPa) of degree 2 PR model

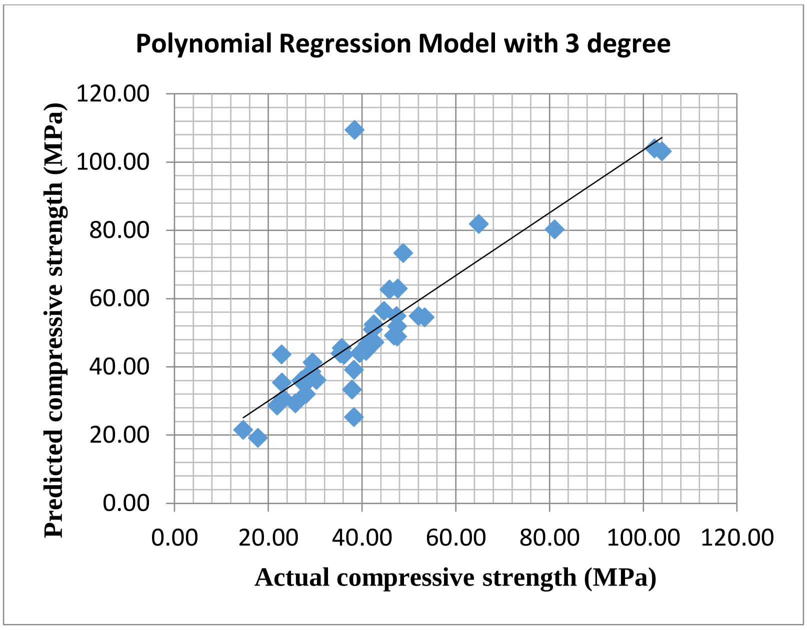

Figure 8: Relationship between predictd compressive strength and actual predicted compressive strength (MPa) of degree 3 PR model

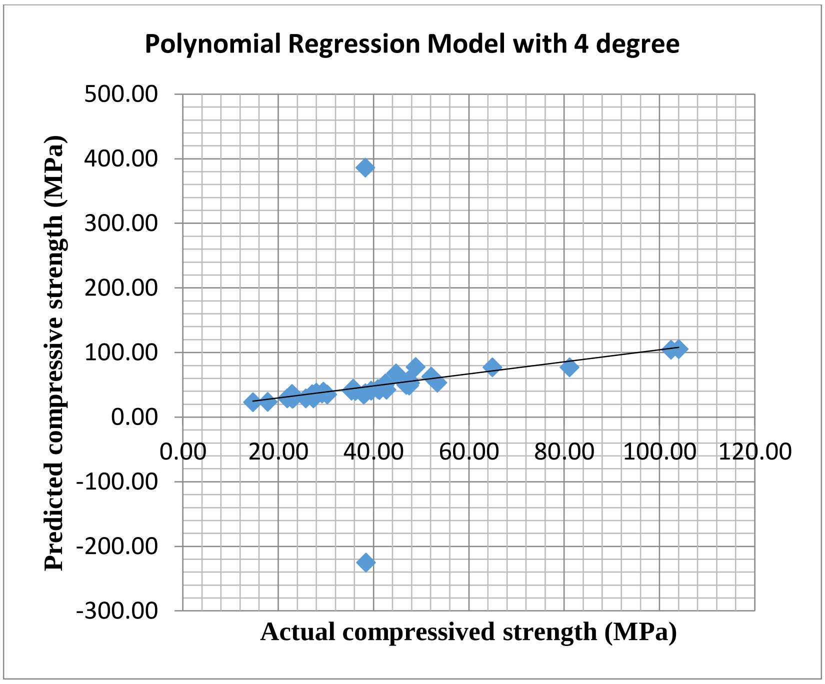

Figure 9: Relationship between predictd compressive strength and actual predicted compressive strength (MPa) of degree 4 PR model

Figure 8 and Figure 9 shows that the predicted compressive strength value is increased as the degree level rises to degree 3. Unfortunately, as the degree begins to increase up until degree 4, the predicted value of compressive strength started to decrease as well. It showed that the model is not possible to be trained from degree 4 onwards. Means, as the predicted compressive strength value decreases, the model's performance also tends to decrease.

4. Conclusion

In line with these findings, the objectives of the study were achieved. A concrete dataset comprising concrete compressive strength as a dependent variable was established through data preparation and preprocessing. The dataset was collected based on 16 literature journals and listed in Chapter 3. With 26 attributes, the data set consists of 202 observations. It consists the result of the 28 days compressive strength. Therefore, the prediction of the machine learning model was conducted based on the establishment of the compressive strength 28 days dataset.

A comparison of performance evaluation between MLR model and PR model was conducted. The result value of the R-squared of MLR model is 0.589. It was found that the PR model with degree 2 has a good performance of the model. The correlation value of the PR model with degree 2 is 0.876. PR model with degree 3 comes in second with a value of the correlation coefficient is 0.826. At a degree of 4, the model's performance tends to decline gradually. Initially, the MLR model was proposed to be used as the machine learning prediction model. Compared with the PR model, the R-squared value of the MLR model was lower than PR model; it still indicates the model can still produce a good model. Thus, the PR model is recommended in the machine learning prediction model to predict the concrete compressive strength.

References

[1] Cui, J., Hao, H., and Shi, Y. (2018). Study of concrete damage mechanism under hydrostatic pressure by numerical simulations. *Construction and Building Materials*, *160*, 440–449.

[2] Buller, A. H., Oad, M., and Memon, B. A. (2018). *Relationship between Cubical and Cylindrical Compressive Strength of Recycled Aggregate Concrete*. (September).

[3] DeRousseau, M. A., Laftchiev, E., Kasprzyk, J. R., Rajagopalan, B., and Srubar, W. V. (2019). A comparison of machine learning methods for predicting the compressive strength of field-placed concrete. *Construction and Building Materials*, *228*, 116661.

[4] Matias, D., De Brito, J., Rosa, A., and Pedro, D. (2013). Mechanical properties of concrete produced with recycled coarse aggregates - Influence of the use of superplasticizers. Construction and Building Materials, 44, 101–109.

[5] Prakash, M., Manikandan, S., Surenther, I., Aswin Kumar, M. N., Ilakkiya, S., and Menaka, D. (2019). Speculation of compressive strength of concrete in real-time. *International Journal of Recent Technology and Engineering*, *7*(6), 988–992.

[6] Ajagbe, W., and Tijani, M. A. (2018). Compressive Strength of Concrete Made from Aggregates of Different Sources Compressive Strength of Concrete Made from Aggregates of Different Sources. *USEP Journal of Research Information in Civil Engineering*, *15*(1), 1963–1974.

[7] Feng, D. C., Liu, Z. T., Wang, X. D., Chen, Y., Chang, J. Q., Wei, D. F., and Jiang, Z. M. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Construction and Building Materials, 230, 117000.

[8] Fillbrunn, A., Dietz, C., Pfeuffer, J., Rahn, R., Landrum, G. A., and Berthold, M. R. (2017). KNIME for reproducible cross-domain analysis of life science data. Journal of Biotechnology,

[9] Maier, P. L., and Durham, S. A. (2012). Beneficial use of recycled materials in concrete mixtures. *Construction and Building Materials*, *29*, 428–437.