Evaluation of Factors Affecting Compressive Strength of Concrete using Machine Learning

Avinash Kumar Jha
Department of Civil Engineering
Delhi Technological University
New Delhi, India
avinash.jha.6696@gmail.com

Surabhi Adhikari
Department of CSE
Delhi Technological University
New Delhi, India
surabhi_bt2k18@dtu.ac.in

Surendrabikram Thapa
Department of CSE
Delhi Technological University
New Delhi, India
surendrabikram_bt2k17@dtu.ac.in

Sushruti Mishra
Department of CSE
Delhi Technological University
New Delhi, India
mishrasushruti99@gmail.com

Arunish Kumar
Department of Civil Engineering
Delhi Technological University
New Delhi, India
kumararunish@yahoo.com

Abhay Kumar
Department of Civil Engineering
Delhi Technological University
New Delhi, India
abhayshroff1997@gmail.com

***Abstract*—Compressive strength of the concrete is important for analyzing the characteristics of the concrete. The compressive strength is necessary to know if the given mixture of concrete meets the specified requirements. For the sustainability of construction, the compressive strength must meet the required standards. Machine learning models have been really a handy tool for the analysis of a wide range of problems. Machine learning models can find the pattern or trends in the given data. The purpose of the paper is two folds. First, the evaluation of performance of different machine learning models (regression models) is done. In the second fold, the factors affecting compressive strength of the concrete are discussed. Different factors have different degrees of importance for various regressors. The importance of the factors is studied for different regressors and the conclusion is drawn regarding the importance of factors taken in the study.**

Keywords—regression, compressive strength, concrete mixture, machine learning, feature selection

I. INTRODUCTION

Machine learning (ML) algorithms are ubiquitous with its easy implementation over the time. Machine learning models are often used to find trends and patterns in given observations. In quantitative data, regression models are used for finding patterns [1]. The patterns learnt by ML algorithms can be leveraged to build a mathematical model which can do prediction and predictive analysis that are useful for the study of possible outcomes with observations of similar kind [2,3]. Concrete is one of the most used products in construction. The life of buildings, bridges and infrastructures that engineers build depends upon the quality of concrete [4]. The compressive strength of concrete is the property of concrete mixture by virtue of which it can withstand the compression. The compressive strength of concrete must meet the standards laid so that the outcome of construction would come as desired with a longer life of infrastructure [5,6]. In other words, the compressive strength also tells about if the concrete can resist failure or not [7,8]. Various works have been done for predicting the concrete's compressive strength using machine learning algorithms.

Yaseen et. al. [9] discussed an extreme learning model for predicting the compressive strength of the concrete. Similarly, Cheng, Firdausi and Prayogo [10] used Genetic Weighted Pyramid Operation Tree (GWPOT) for compressive strength prediction of high-performance concrete. Pham, Ngo, Nguyen and Truong [11] made use of hybrid ML algorithms for prediction of strength of sustainable concrete. Scientists over the time had used multiple ML techniques for calculation of the compressive strength of the concrete. DeRousseau [12] compared various machine learning models for prediction of the compressive strength of concrete. With advancement in computing methodologies, deep learning models are also used widely for this prediction task. The concrete is a topic of great interest for engineers and scientists and hence a great number of models are being built to study the properties of concrete with the help of computers. Deep learning models [13] can also work well in calculation of the compressive strength of concrete. In this paper, machine learning models are used for our analysis whereas implementation of deep learning models remains as future work of the authors. In the second fold of the paper, the analysis of significance of factors in compressive strength of concrete is done. There are various factors affecting the compressive strength of concrete. Coarse aggregate used [14, 15] for example affects the compressive strength of concrete. There are various other factors affecting the compressive strength. The study is to figure out which factor carries what amount of importance and the paper does mathematical analysis to find out which factors matter the most in compressive strength of the concrete.

II. METHODOLOGY

In this study, different regressors are used to study. Since a single regressor which best fits the given data cannot be known by instinct, multiple regressors are tried to select the one which is the best fit for calculation of the compressive strength of concrete. The implementation of the problem can be shown in the figure 1 below. Each component of the figure is explained in detail below.

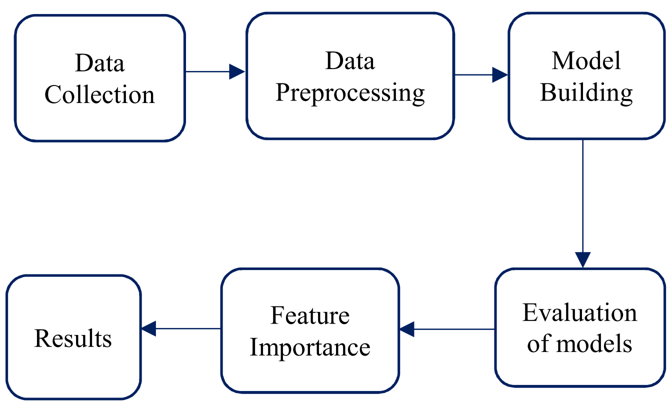

Fig. 1. Flowchart for problem implementation

A. *Data Collection*

The data [16] was retrieved from Kaggle and used for the study. The data had 1030 observations with 9 attributes. The first 8 attributes are the factors affecting the compressive strength of concrete and last attribute is the value of compressive strength of the concrete. The unit of compressive strength of concrete is MegaPascals (MPa). Each observation has 8 attributes namely cement in mixture, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine Aggregate and age of the concrete. All the variables are quantitative variables. TABLE 1 shows the attributes that are taken in our study.

TABLE 1: ATTRIBUTES AND UNITS

S.N.	Attributes	Unit
1	Cement	Kg/m^3 mixture
2	Blast Furnace Slag	Kg/m^3 mixture
3	Fly Ash	Kg/m^3 mixture
4	Water	Kg/m^3 mixture
5	Superplasticizer	Kg/m^3 mixture
6	Coarse Aggregate	Kg/m^3 mixture
7	Fine Aggregate	Kg/m^3 mixture
8	Age	Days

B. *Data Preprocessing*

The data that is used in the study has different ranges for different attributes. For example, the age of concrete (days) and fine aggregate have different ranges. Thus, there is a need to preprocess the data. For the ease of computation to make the model more robust, the data used in the study is standardized using z-score [17]. The formula for z-score is:

$$z = \frac{x - \bar{x}}{\sigma} \tag{1}$$

Where, x is the original feature vector, $\bar{x}$ = average (x) is the mean of that feature vector, and σ is its standard deviation.

C. *Model Building*

The models were built using different regressors. The models used are Linear Regressor [18], Ridge Regressor [19], Lasso Regressor [20], Decision Tree Regressor [21], Random Forest Regressor [22], AdaBoost Regressor [23] and Gradient Boosting Regressor [24]. For prediction of the compressive strength of concrete, all the aforementioned models were used one by one.

D. *Evaluation of Models*

The models were evaluated using RMSE and r^2. Since the data is quantitative data, RMSE is a good measure for measuring error of the model [25]. Five-fold cross-validation technique was used to validate the model.

Mathematically,

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\widehat{y_i} - y_i)^2}{n}} \tag{2}$$

Where, $\widehat{y_1}, \widehat{y_2} \ldots\ldots\ldots\ldots \widehat{y_n}$ are predicted values
$y_1, y_2 \ldots\ldots\ldots\ldots y_n$ are observed values
'n' is the no. of observations

Similarly, R-Squared [26] is another measure for evaluation of regression models. It tells about how close the data fits the regression line.

Mathematically,

$$R - Squared = \frac{Explained\ Variations}{Total\ Variations} \tag{3}$$

E. *Feature Importance*

After calculation of errors and seeing which model fits the best, the feature importance according to different regressors were studied. Different regressors have different importance values for the given features [27,28]. Each of them is analyzed to find out which features are the most significant ones.

III. RESULTS AND DISCUSSION

The various models as discussed in the subsection *C* of section II were used to perform the prediction task for the given dataset. The evaluation of models was done on the basis of RMSE and r^2. The mathematical expressions for RMSE and r^2 are given in *(2)* and *(3)*. The performance can be shown in TABLE 2.

TABLE 2: PERFORMANCE OF DIFFERENT REGRESSORS

S.N.	Regressor	RMSE	R-Squared
1	Linear Regressor	10.29	0.61
2	Ridge Regressor	10.29	0.61
3	Lasso Regressor	10.91	0.57
4	Decision Tree Regressor	7.5	0.79
5	Random Forest Regressor	5.3	0.90
6	AdaBoost Regressor	7.75	0.77
7	Gradient Boosting Regressor	5.2	0.91

From the table above, the Gradient Boosting Regressor and Random Forest Regressor have performed well in the given dataset in terms of RMSE and r^2. Now, the importance of given factors for different regressors were calculated as shown in Fig.2-8.

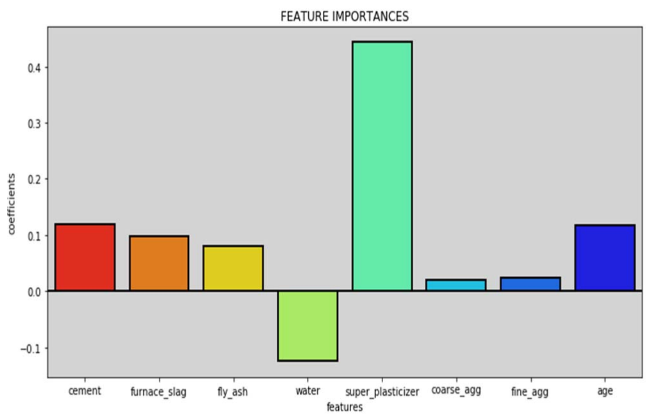

Fig. 2. Feature-Importance for Linear Regression

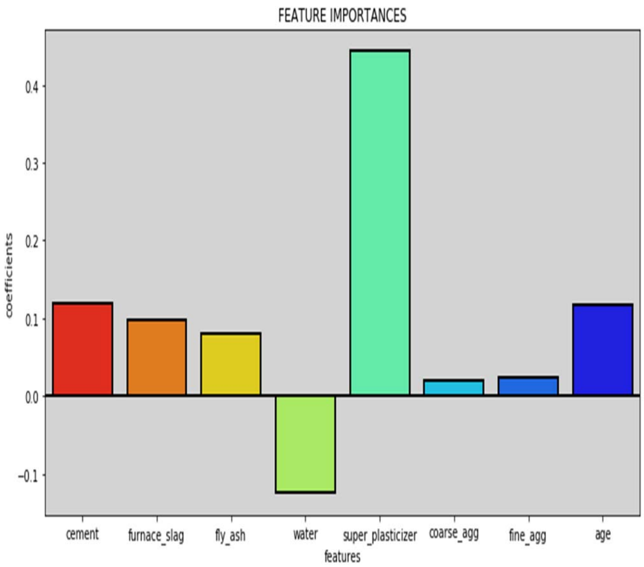

Fig. 3. Feature-Importance for Ridge Regressor

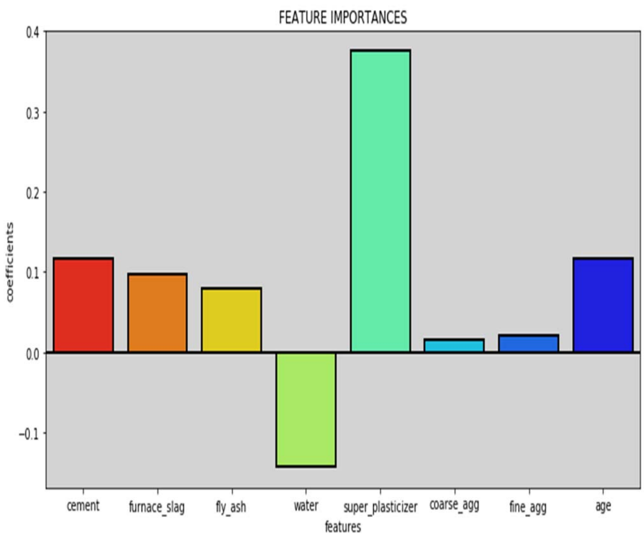

Fig. 4. Feature-Importance for Lasso Regressor

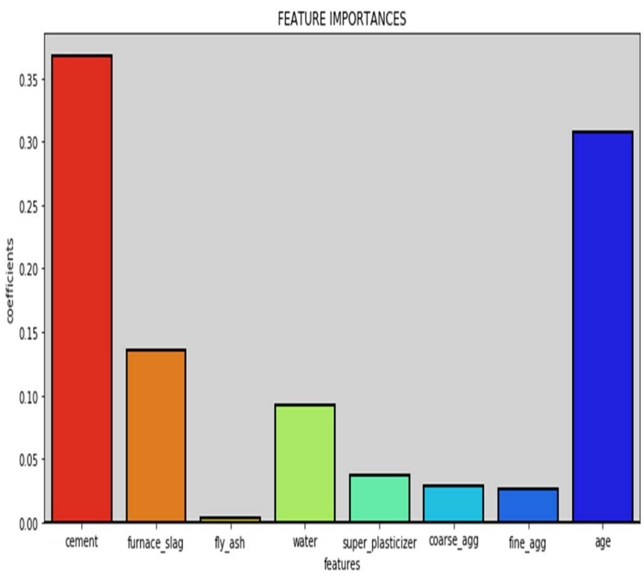

Fig. 5. Feature-Importance for Decision Tree Regressor

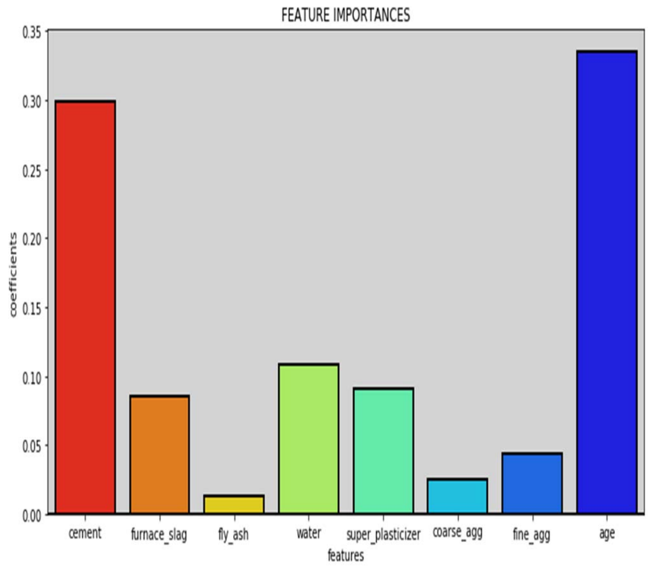

Fig. 6. Feature-Importance for Random Forest Regressor

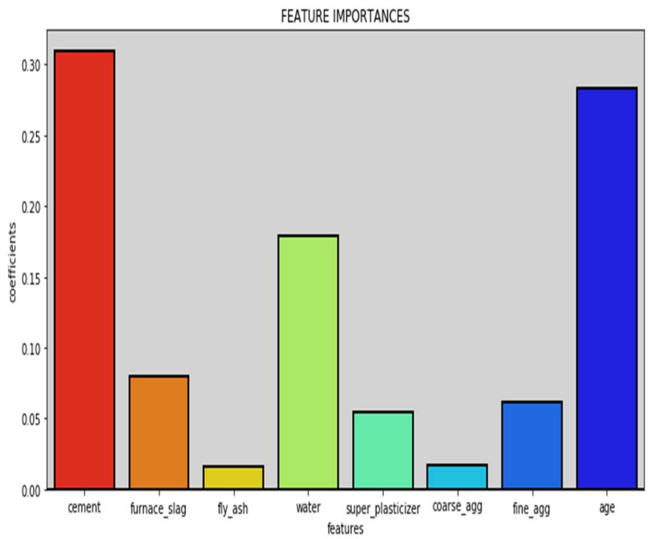

Fig. 7. Feature-Importance for AdaBoost Regressor

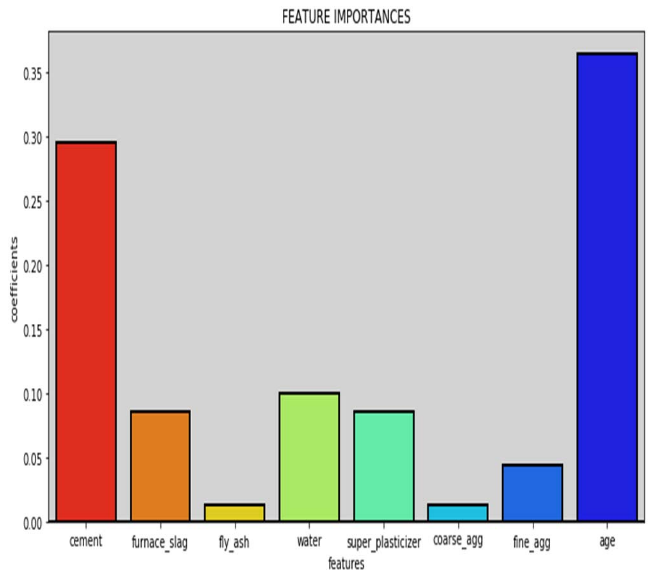

Fig. 8. Feature-Importance for Gradient Boosting Regressor

The index and columns no in the fig 2-8 represents the features as shown in TABLE 3.

TABLE 3: INDEX AND FEATURE NAME

S.N.	Index	Features (Factors)
1	cement	Cement
2	Furnace_slag	Blast Furnace Slag
3	fly_ash	Fly Ash
4	water	Water
5	super_plasticizer	Superplasticizer
6	coarse_agg	Coarse Aggregate
7	fine_agg	Fine Aggregate
8	age	Age

From TABLE 2, Decision Tree Regressor, Random Forest Regressor, AdaBoost Regressor and Gradient Boosting Regressor are the best performing algorithms. Ensemble learners always tend to outperform the conventional machine learning algorithms and, in this experiment too, ensemble learners have shown good performance in terms of RMSE and r^2. For those algorithms, the most important features are cement and age. For example, from the laboratory experiments, it is also found experimentally that the compressive strength of concrete increases with increment in the number of days. Hence, the results given by machine learning models are in compliance with the laboratory results. Importance of other features can be discussed and learnt with the other machine learning algorithms as well as deep learning algorithms.

IV. CONCLUSION AND FUTURE WORKS

The study of algorithms for predicting compressive strength of the concrete are of high importance. Since the life of infrastructures depend upon the compressive strength of concrete, the factors affecting the concrete should be studied well. There might be some features which may have been neglected, and they might have an effect on the compressive strength of concrete. More robust algorithms can be developed to overcome such limitations. Also, the feature importance depends upon algorithm to algorithm. In future, feature selective methods like wrapper methods can be used to select the features while building predictive models. Similarly, filter-based selection techniques can be used to filter out the features of less importance. For example, CFS is a filter-based technique that removes highly correlated features. Other deep learning methods can be explored for this purpose. Similarly, CNN based models can also be built which would use the images of concrete to predict the compressive strength of concrete. The future study of different predictive models and analysis of the importance of factors affecting compressive strength of the concrete would make the construction of infrastructures more sustainable.

ACKNOWLEDGMENT

The authors would like to thank the Department of Civil Engineering, Delhi Technological University and Department of CSE, Delhi Technological University, New Delhi, India for providing the resources that were needed for the study. Also, we are really grateful to Central Library, Delhi Technological University for providing access to all the research papers that were referred during the study.

REFERENCES

[1] A. Saxena et al., "A review of clustering techniques and developments," Neurocomputing, vol. 267, pp. 664-681, 2017.

[2] J. Verrelst et al., "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and-3," Remote Sensing of Environment, vol. 118, pp. 127-139, 2012.

[3] E. Formisano, F. De Martino, and G. Valente, "Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning," Magnetic resonance imaging, vol. 26, no. 7, pp. 921-934, 2008.

[4] M. Shannag, "High strength concrete containing natural pozzolan and silica fume," Cement and concrete composites, vol. 22, no. 6, pp. 399-406, 2000.

[5] J. Sobhani, M. Najimi, A. R. Pourkhorshidi, and T. Parhizkar, "Prediction of the compressive strength of no-slump concrete: A comparative study of regression, neural network and ANFIS models," Construction and Building Materials, vol. 24, no. 5, pp. 709-718, 2010.

[6] A. Öztaş, M. Pala, E. a. Özbay, E. Kanca, N. Caglar, and M. A. Bhatti, "Predicting the compressive strength and slump of high strength concrete using neural network," Construction and Building Materials, vol. 20, no. 9, pp. 769-775, 2006.

[7] C. S. Poon, Z. Shui, and L. Lam, "Effect of microstructure of ITZ on compressive strength of concrete prepared with recycled aggregates," Construction and Building Materials, vol. 18, no. 6, pp. 461-468, 2004.

[8] F. de Larrard and A. Belloc, "The influence of aggregate on the compressive strength of normal and high-strength concrete," Materials Journal, vol. 94, no. 5, pp. 417-426, 1997.

[9] Z. M. Yaseen et al., "Predicting compressive strength of lightweight foamed concrete using extreme learning machine model," Advances in Engineering Software, vol. 115, pp. 112-125, 2018.

[10] M.-Y. Cheng, P. M. Firdausi, and D. Prayogo, "High-performance concrete compressive strength prediction using Genetic Weighted Pyramid Operation Tree (GWPOT)," Engineering Applications of Artificial Intelligence, vol. 29, pp. 104-113, 2014.

[11] A.-D. Pham, N.-T. Ngo, Q.-T. Nguyen, and N.-S. Truong, "Hybrid machine learning for predicting strength of sustainable concrete," Soft Computing, pp. 1-16, 2020.

[12] M. DeRousseau, E. Laftchiev, J. Kasprzyk, B. Rajagopalan, W. J. C. Srubar III, and B. Materials, "A comparison of machine learning methods for predicting the compressive strength of field-placed concrete," vol. 228, p. 116661, 2019.

[13] Y. Jang, Y. Ahn, and H. Y. Kim, "Estimating compressive strength of concrete using deep convolutional neural networks with digital microscope images," Journal of Computing in Civil Engineering, vol. 33, no. 3, p. 04019018, 2019.

[14] Z. Keshavarz and D. Mostofinejad, "Porcelain and red ceramic wastes used as replacements for coarse aggregate in concrete," Construction and Building Materials, vol. 195, pp. 218-230, 2019.

[15] Y. Cheng, S. Liu, B. Zhu, R. Liu, and Y. Wang, "Preparation of preplaced aggregate concrete and experimental study on its strength," Construction and Building Materials, vol. 229, p. 116847, 2019.

[16] (06 Jan. 2020) Data for Predicting Compressive Strength of Concrete. Available: kaggle.com/pavanraj159/concrete-compressive-strength-data-set

[17] L. Lepetit, F. Strobel, and T. H. Tran, "An alternative Z-score measure for downside bank insolvency risk," Applied Economics Letters, pp. 1-6, 2020.

[18] G. Engel, "Machine Learning to Approximate Solutions of Ordinary Differential Equations: Neural Networks vs. Linear Regressors," in International Conference on Computational Science, 2019, pp. 169-177: Springer.

[19] A. Tanaka and H. Imai, "A Fast Cross-Validation Algorithm for Kernel Ridge Regression by Eigenvalue Decomposition," IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, vol. 102, no. 9, pp. 1317-1320, 2019.

[20] N. Zhang and S. Zeng, "A gradient descending solution to the LASSO criteria," in Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., 2005, vol. 5, pp. 2942-2947: IEEE.

[21] M.-T. Tran-Nguyen, L.-D. Bui, Y.-G. Kim, and T.-N. Do, "Decision tree using local support vector regression for large datasets," in Asian Conference on Intelligent Information and Database Systems, 2018, pp. 255-265: Springer.

[22] T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer, "Robust and accurate shape model fitting using random forest regression voting," in European Conference on Computer Vision, 2012, pp. 278-291: Springer.

[23] D. P. Solomatine and D. L. Shrestha, "AdaBoost. RT: a boosting algorithm for regression problems," in 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), 2004, vol. 2, pp. 1163-1168: IEEE.

[24] A. Keprate and R. C. Ratnayake, "Using gradient boosting regressor to predict stress intensity factor of a crack propagating in small bore piping," in 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2017, pp. 1331-1336: IEEE.

[25] R. Sawhney, R. R. Shah, V. Bhatia, C.-T. Lin, S. Aggarwal, and M. Prasad, "Exploring the Impact of Evolutionary Computing based Feature Selection in Suicidal Ideation Detection," in 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2019, pp. 1-6: IEEE.

[26] M. De Bois, M. A. El Yacoubi, and M. Ammi, "Study of short-term personalized glucose predictive models on type-1 diabetic children," in 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1-8: IEEE.

[27] M. Tanveer et al., "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," ACM Transactions on Multimedia Computing, Communications, and Applications, 2019.

[28] G. Tripathi, K. Singh, and D. K. Vishwakarma, "Convolutional neural networks for crowd behaviour analysis: a survey," The Visual Computer, vol. 35, no. 5, pp. 753-776, 2019.