STRENGTH PREDICTING OF FLY ASH-BASED CONCRETE VIA MACHINE LEARNING ALGORITHMS

[a, b] ***Ayaz Ahmad *, [a, b] Yadong Jiang, [a, b] William Finnegan***

a: Civil Engineering, School of Engineering, College of Science & Engineering, National University of Ireland, Galway, University Road, Galway, Ireland, a.ahmad8@nuigalway.ie; yadong.jiang@nuigalway.ie; william.finnegan@nuigalway.ie;
b: MaREI Centre, Ryan Institute, National University of Ireland, Galway, University Road, Galway, Ireland.

Abstract- The application of the support vector machine (SVM) and random forest (RF) algorithms to anticipate the compressive strength of fly ash-based concrete has been investigated in the study. A predictive performance comparison was performed for these two algorithms, where statistical metrics, such as coefficient of determination (R^2), mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE), were used for the evaluation. The results reveal that the RF algorithm outperforms the support vector machine model in terms of predicting the compressive strength of concrete containing fly ash. The statistical checks and the investigation of k-fold cross validation also give confirmation of higher precision for the random forest model. The employed machine learning approaches successfully predicted the strength property of selected concrete and gives the indication that it can reduce the time, experimental efforts, and cost of the construction projects.

Keywords- concrete; fly ash; machine learning; modelling; validation.

1. Introduction

Portland cement is one of the most widely utilized ingredients in the concrete industry [1]. It is essential for developing the infrastructure required for social and economic growth but has a harmful influence on the environment [1, 2]. Cement manufacture results in substantial use of natural resources and substantial carbon dioxide (CO_2) emissions [3, 4]. In addition, it is widely known that CO_2 contributes to global warming. A few products or residues generated by various industrial sectors can react with calcium hydroxide to form calcium silicate hydrates similar to those formed during cement hydration [5-7]. These calcium silicate hydrates are referred to as pozzolans because they possess similar properties to Portland cement [8, 9]. The application of pozzolana in concrete can minimize the quantity of cement in it, which is one of the major causes of CO2 emission [10, 11]. Therefore, its use as a cement replacement appears to be a potential way to mitigate these issues [12] . The cement industry is responsible for around 8–10 percent of the world's anthropogenic CO_2 emissions [13]. As the demand for construction materials in the developing world continues to rise, this level of output is becoming increasingly troublesome, and future developments are required for the construction sector to become ecologically sustainable [14, 15].

Fly ash (FA), a byproduct of burning coal, is now frequently used as a cement component in concrete, along with other pozzolans [16-18]. FA is also acknowledged as an environmentally benign substance because the use of this material helps to reduce the cement industry's overall carbon footprint [19]. Application of FA in concrete material as a partial replacement of binding material plays a positive role in reducing the emission of CO_2 gasses [20, 21]. This approach not only helps to produce an environmentally friendly concrete but also fulfill the strength and durability requirements of the concrete [22].

Machine learning (ML) algorithms are widely using to forecast the strength properties of different types of concrete [23-25]. The connection of artificial inelegance with the field of engineering is of great interest for the researchers [26]. This approach not only reduces the effort of experiments but also minimizes the cost of the project [27]. A variety of ML techniques such as bagging, decision tree, random forest (RF), multilayer perceptron, and gene expression programming are widely using for predicting the required outputs.

Investigating how well machine learning algorithms can predict outcomes is the purpose of this study. In order to forecast the compressive strength of selected concrete, both the support vector machine (SVM) and the RF approaches have been investigated. In order to validate the validity of the model, statistical checks and a method known as k-fold cross validation were incorporated. To estimate the compressive strength of fly ash-based concrete, this study makes use of two distinct kinds of algorithms—a single-label machine learning approach (SVM) and a multi-label machine learning ensemble technique (RF)—in order to break new ground. This approach is the primary innovation of the study. While individual techniques execute the model in the conventional fashion, an ensemble machine learning approach divides the actual models into a number of sub-models.

2. Materials and method

Python coding was introduced in the software for each employed ML technique to run the models. The data set used for the models was retrieved from the literature [24]. The comparison between the experimental result and the result obtained from the models was shown with the regression graphs, while the amount of difference between both the results was indicated with the error distribution graphs. The software automatically splits the data set into 80 percent for training the model and 20 percent for testing the model. However, k-fold cross validation approach was adopted for validation purpose. All the ingredients used to prepare the concrete materials were taken as the input parameters, and the result of the strength was taken as the output parameter. The descriptive statistics of the input parameters taken from the mentioned literature are listed in Table 1. However, the relative frequency distribution for the same parameters can be seen in Figure 1.

Table 1. Descriptive statistics of the variables

Parameters	**Cement (kg/m^3)**	**Fly ash (kg/m^3)**	**Water (kg/m^3)**	**SP (kg/m^3)**	**CA (kg/m^3)**	**FA (kg/m^3)**	**Age (days)**
Mean	282.13	77.29	180.95	5.45	1003.76	794.19	44.50
Standard Error	3.78	2.46	0.72	0.21	2.90	2.71	2.34
Median	252.00	100.40	185.70	5.70	1006.40	794.90	28.00
Mode	213.50	0.00	192.00	0.00	968.00	613.00	28.00
Standard Deviation	94.88	61.91	17.97	5.28	72.84	68.18	58.66
Sample Variance	9001.86	3832.72	322.86	27.89	5305.62	4648.69	3441.28
Range	405.30	200.10	88.00	28.20	324.00	351.00	364.00
Minimum	134.70	0.00	140.00	0.00	801.00	594.00	1.00
Maximum	540.00	200.10	228.00	28.20	1125.00	945.00	365.00
Sum	178026.90	48768.30	114176.60	3439.70	633370.90	501134.40	28082.00
Count	631.00	631.00	631.00	631.00	631.00	631.00	631.00

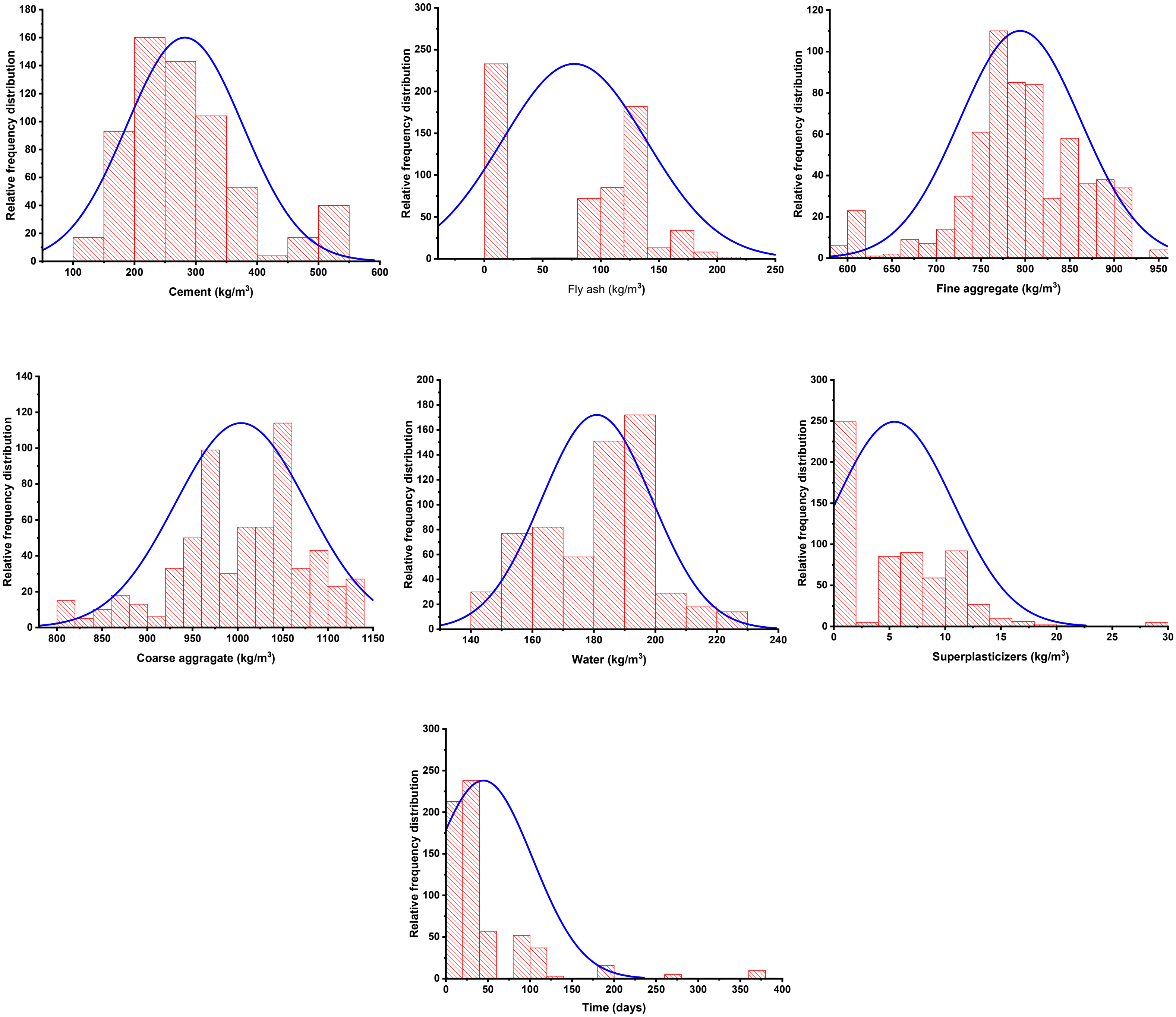


Figure 1. Relative frequency dispersal of the variables used for running the models

3. Employed Algorithms

A linear model that can be used for classification and regression issues is known as an SVM [28]. It is effective for a wide variety of real-world challenges and can solve both linear and non-linear issues. The basic notion of SVM is as follows: The algorithm draws a line or a hyperplane that divides the data into the appropriate categories. In recent years, one of the most effective machine learning approaches has been SVMs. SVMs have been successfully applied to numerous engineering related applications, including those of the petroleum and mining industries.

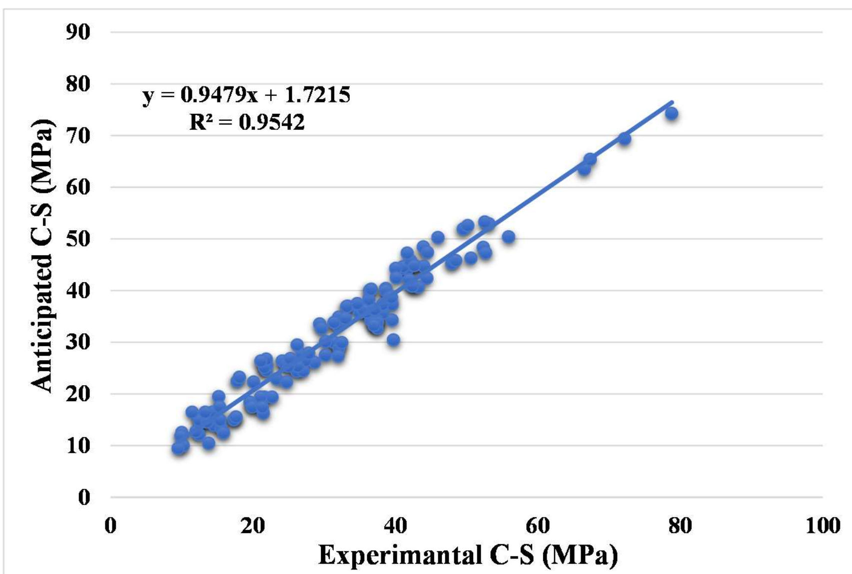
RF is a type of classification technique that is made up of several different decision trees [29]. When building each individual tree, it employs bagging and feature randomness to try to produce an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. This forest of trees would then be used to make a prediction. This is among the most important aspects of the RF Algorithm, as it can handle a data set containing both continuous variables, which is used in the case of regression and categorical variables for classification. This capability is one of the most significant characteristics of the Random Forest Algorithm. The outcomes of categorization issues are improved, as a result of its use.

4. Result and discussion

4.1. SVM algorithm's outcome.

The relationship between the experimental result for compressive strength (C-S) and the result obtained from the SVM model is depicted in the upper part of Figure 2 (a). The fact that the R^2 value is equal to 0.95 demonstrates that the relationship is both reasonable and strong. Figure 2 (b), on the other hand, illustrates the disparity between the observed C-S data and the predicted data. The difference between these two values demonstrates that the highest value is equal to 9.16 MPa, while the lowest value that was reported was equal to 0 MPa.

(a)



(b)

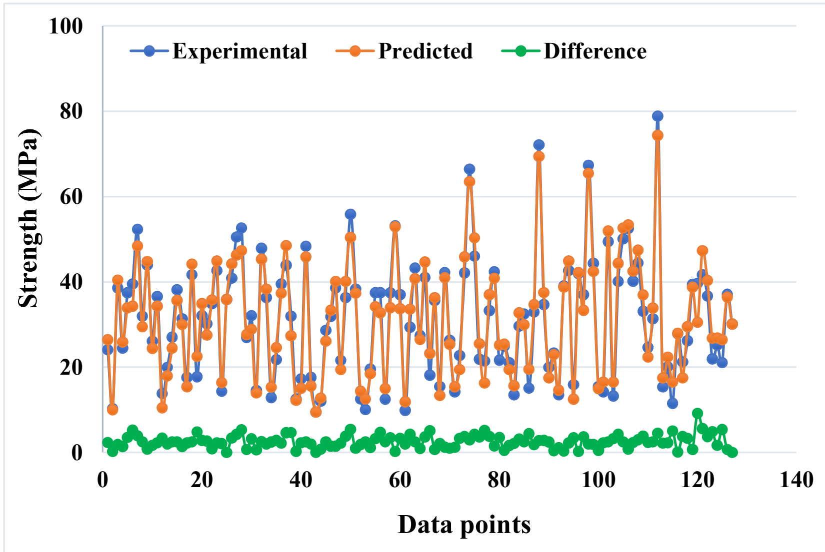
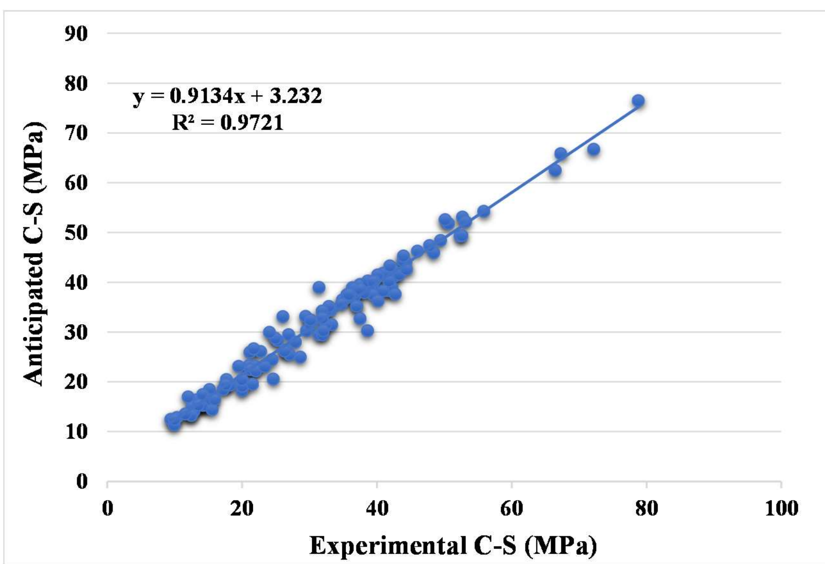


Figure 2. Regression and error's graphs: (2a) Experimental and predictive result's relationship of the C-S from SVM model; (2b) Difference of the predictive results and experimental C-S results from SVM model

4.2. RF algorithm's outcome.

Figure 3 illustrates the relationship that exists between the actual C-S and the output that the model produces for the C-S. (a). When compared to the output of the SVM model, this relationship reveals a predictive result that is both more effective and accurate, as indicated by the R2 value, which is equal to 0.97. Figure b presents the error distribution that pertains to the RF model, which is another point of interest (b). This dispersal has a range of values, the lowest of which is 0.0092 MPa and the highest of which is 8.32 MPa.

(a)



(b)

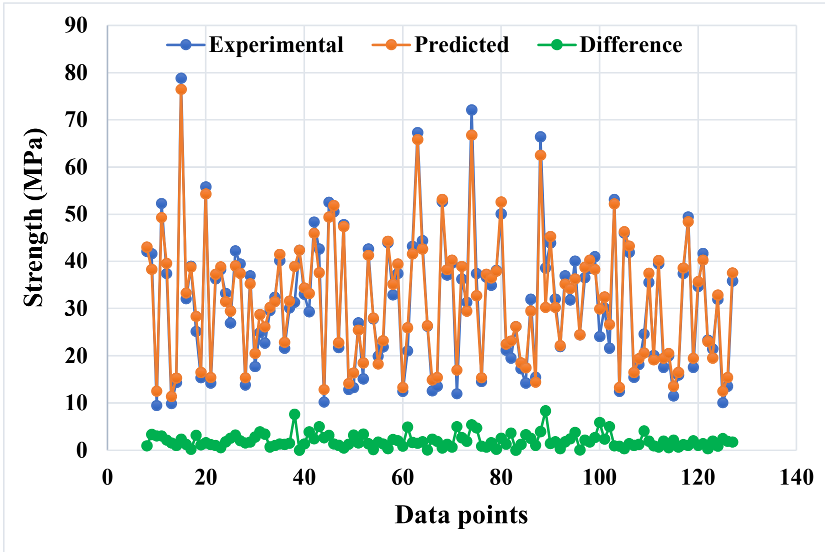
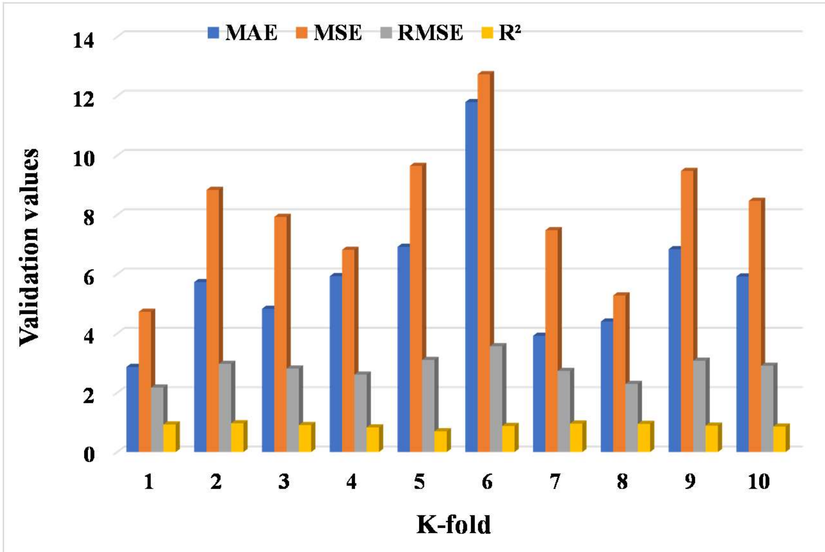


Figure 3. Regression and error's graphs: (3a) Experimental and predictive result's relationship of the C-S from RF model; (3b) Difference of the predictive results and experimental C-S results from RF model

5. K-fold cross validation (CV)

This method was used to validate the employed models. The legitimacy of the models was confirmed by the adoption of the CV approach. It uses nine out of ten subsets, with the exception of one subset that is utilized for the purpose of model validation. The results of the C-V analysis are evaluated using the root mean square error (RMSE), the mean absolute error (MAE), the mean square error (MSE), and the correlation coefficient (R^2). In comparison to the SVM model, the RF model demonstrates a significantly lower result of the specified errors and a significantly higher result of R^2. The average values of the MAE, MSE, RMSE, and R^2 from the validation process for SVM models were 5.92 MPa, 8.14 MPa, 2.83 MPa, and 0.89, respectively, as shown in Figure 3 (a). Similarly, the average results for the same parameters from the RF model during the validation process were noted as 5.17 MPa, 6.24 MPa, 2.46 MPa, and 0.90, respectively, as depicted in Figure 3 (b).

(a)



(b)

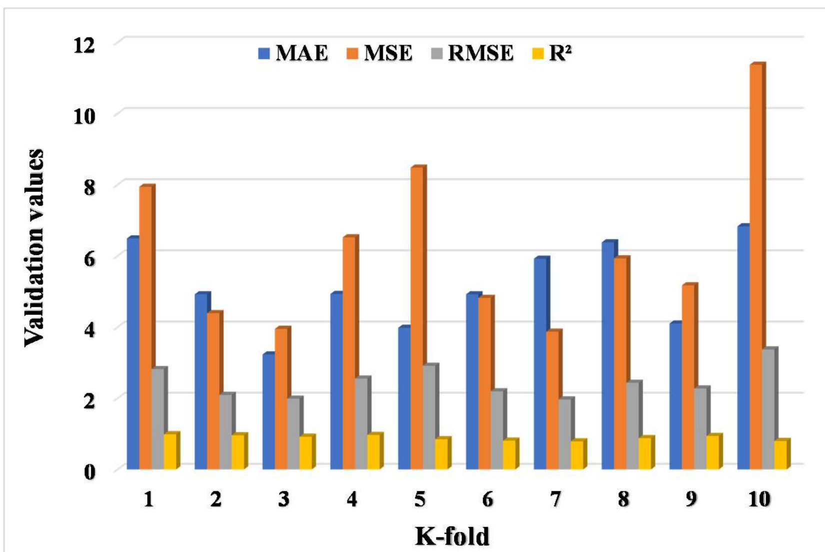


Figure 4. Validation results of the employed models (4a) K-fold CV for SVM algorithm; (4b) K-fold CV for SVM algorithm

Table 2. Statistical metrics for SVM and RF models

Algorithms	**MAE (MPa)**	**MSE (MPa)**	**RMSE (MPa)**
SVM	2.54	8.76	2.96
RF	1.97	6.25	2.50

6. Conclusions

This study describes the comparative study of two ML algorithms towards the prediction of C-S of concrete containing fly ash. The SVM and RF approaches have been introduced for the required outcome. However, the following conclusion can be drawn from the study.

- The RF model showed the affective predictive performance by giving an R^2 value of 0.97 as compared to the SVM model's R^2 value of 0.95.
- The lower values of the errors (MAE, MSE) also confirm the high accuracy of the RF model.
- The K-fold cross validation method was adopted to validate the employed models and confirm their legitimacy towards the predicted C-S of the concrete.
- The ML algorithms can be successfully applied to the selected data set to anticipate the required outcome.

This research has the potential to benefit the construction industries by reducing the overall cost of the projects that are being undertaken. The utilization of such techniques to predict the performance of composite materials in a limited time period is beneficial. This not only reduces the physical effort on the tests in the laboratory, but it also saves the valuable time of the project, which is a significant advantage.

Acknowledgment

The authors would like to acknowledge the support of Science Foundation Ireland (SFI), through the MaREI Research Centre for Energy, Climate and Marine (Grant no. 12/RC/2302_2).

References

[1] C. Meyer, "The greening of the concrete industry," *Cement and concrete composites,* vol. 31, no. 8, pp. 601-605, 2009.
[2] A. Torres, L. Bartlett, and C. Pilgrim, "Effect of foundry waste on the mechanical properties of Portland Cement Concrete," *Construction and building materials,* vol. 135, pp. 674-681, 2017.
[3] J. Hussain, A. Khan, and K. Zhou, "The impact of natural resource depletion on energy use and CO2 emission in Belt & Road Initiative countries: a cross-country analysis," *Energy,* vol. 199, p. 117409, 2020.
[4] G. Habert *et al.*, "Environmental impacts and decarbonization strategies in the cement and concrete industries," *Nature Reviews Earth & Environment,* vol. 1, no. 11, pp. 559-573, 2020.
[5] C. Hepburn *et al.*, "The technological and economic prospects for CO2 utilization and removal," *Nature,* vol. 575, no. 7781, pp. 87-97, 2019.
[6] R. Heede, "Tracing anthropogenic carbon dioxide and methane emissions to fossil fuel and cement producers, 1854–2010," *Climatic change,* vol. 122, no. 1, pp. 229-241, 2014.
[7] J. S. Gregg, R. J. Andres, and G. Marland, "China: Emissions pattern of the world leader in CO2 emissions from fossil fuel consumption and cement production," *Geophysical Research Letters,* vol. 35, no. 8, 2008.
[8] A. H. Buchanan and B. G. Honey, "Energy and carbon dioxide implications of building construction," *Energy and Buildings,* vol. 20, no. 3, pp. 205-217, 1994.
[9] S. D. Casler and A. Rose, "Carbon dioxide emissions in the US economy: a structural decomposition analysis," *Environmental and resource economics,* vol. 11, no. 3, pp. 349-363, 1998.
[10] S. Ahmad, O. S. B. Al-Amoudi, S. M. S. Khan, and M. Maslehuddin, "Effect of silica fume inclusion on the strength, shrinkage and durability characteristics of natural pozzolan-based cement concrete," *Case Studies in Construction Materials,* vol. 17, p. e01255, 2022.
[11] Y. Zhang, Y. Hua, and X. Zhu, "Investigation of the durability of eco-friendly concrete material incorporating artificial lightweight fine aggregate and pozzolanic minerals under dual sulfate attack," *Journal of Cleaner Production,* vol. 331, p. 130022, 2022.
[12] A. A. Ramezanianpour, "Cement replacement materials," *Springer Geochemistry/Mineralogy, DOI,* vol. 10, pp. 978-3, 2014.
[13] L. Barcelo, J. Kline, G. Walenta, and E. Gartner, "Cement and carbon emissions," *Materials and structures,* vol. 47, no. 6, pp. 1055-1065, 2014.
[14] H. Jiang and C. Liu, "Forecasting construction demand: A vector error correction model with dummy variables," *Construction Management and Economics,* vol. 29, no. 9, pp. 969-979, 2011.
[15] J. C. S. Tang, P. Karasudhi, and P. Tachopiyagoon, "Thai construction industry: demand and projection," *Construction Management and Economics,* vol. 8, no. 3, pp. 249-257, 1990.
[16] M. D. A. Thomas, *Optimizing the use of fly ash in concrete*. Portland Cement Association Skokie, IL, USA, 2007.
[17] K. Wesche, *Fly ash in concrete: properties and performance*. CRC Press, 1991.
[18] R. C. Joshi and R. P. Lohita, *Fly ash in concrete: production, properties and uses*. CRC Press, 1997.

[19] K. G. Babu and G. S. N. Rao, "Efficiency of fly ash in concrete with age," *Cement and concrete research,* vol. 26, no. 3, pp. 465-474, 1996.

[20] U. N. Environment, K. L. Scrivener, V. M. John, and E. M. Gartner, "Eco-efficient cements: Potential economically viable solutions for a low-CO2 cement-based materials industry," *Cement and Concrete Research,* vol. 114, pp. 2-26, 2018.

[21] S. A. Miller, A. Horvath, P. J. M. Monteiro, and C. P. Ostertag, "Greenhouse gas emissions from concrete can be reduced by using mix proportions, geometric aspects, and age as design factors," *Environmental Research Letters,* vol. 10, no. 11, p. 114017, 2015.

[22] A. I. Nicoara *et al.*, "End-of-life materials used as supplementary cementitious materials in the concrete industry," *Materials,* vol. 13, no. 8, p. 1954, 2020.

[23] D.-C. Feng *et al.*, "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach," *Construction and Building Materials,* vol. 230, p. 117000, 2020.

[24] K. Khan, A. Ahmad, M. N. Amin, W. Ahmad, S. Nazar, and A. M. Arab, "Comparative Study of Experimental and Modeling of Fly Ash-Based Concrete," *Materials,* vol. 15, no. 11, 2022.

[25] A. Ahmad *et al.*, "Prediction of Geopolymer Concrete Compressive Strength Using Novel Machine Learning Algorithms," *Polymers,* vol. 13, no. 19, 2021.

[26] W. Z. Taffese and E. Sistonen, "Machine learning for durability and service-life assessment of reinforced concrete structures: Recent advances and future directions," *Automation in Construction,* vol. 77, pp. 1-14, 2017.

[27] M. Shang *et al.*, "Predicting the Mechanical Properties of RCA-Based Concrete Using Supervised Machine Learning Algorithms," *Materials,* vol. 15, no. 2, 2022.

[28] S. Suthaharan, "Support vector machine," in *Machine learning models and algorithms for big data classification*: Springer, 2016, pp. 207-235.

[29] G. Biau and E. Scornet, "A random forest guided tour," *Test,* vol. 25, no. 2, pp. 197-227, 2016.