Prediction of Concrete's Compressive Strength via Artificial Neural Network Trained on Synthetic Data

Saleh J. Alghamdi
Department of Civil Engineering, College of Engineering, Taif University, Saudi Arabia
sjalghamdi@tu.edu.sa (corresponding author)

Received: 25 October 2023 | Revised: 8 November 2023 | Accepted: 11 November 2023

 DOI: https://doi.org/10.48084/etasr.6560

ABSTRACT

Predicting concrete compressive strength using machine learning techniques has attracted the focus of many studies in recent years. Typically, given concrete mix ingredients, a machine learning model is trained on experimental data to predict properties of hardened concrete, such as compressive strength at 28 days. This study used computer-generated mix design data that contained mixed ingredients along with the corresponding theoretical strength of each mix to train a neural network and then test them on real-world experimental data. The developed model was able to predict the compressive strength of concrete specimens at 28 days with an R-value of 0.80. Furthermore, increasing the synthetic dataset increased the performance of the model to a point beyond which it started to decrease. The proposed sustainability-promoting method emphasizes the effectiveness of using synthetic data to train machine learning models that yield insightful predictions with acceptable accuracy.

Keywords-concrete; compressive strength; machine learning; AI; synthetic data; computer-generated data

I. INTRODUCTION

The application of Machine Learning (ML) algorithms in the field of construction materials and concrete technology has been trending for the past few years. Such models can learn from the data and establish relationships between inputs and outputs to solve complex engineering problems. The success of such models has been widely celebrated, increasing their prevalence in many fields of applied and theoretical sciences. In the fields of civil, architectural, construction engineering, and concrete technology, many studies attempted to increase the prediction accuracy of ML models to solve relevant problems. For example, several studies attempted to create models that accurately predict the properties of construction materials by training on data available in the literature [1-4]. In addition, many properties of fresh and hardened concrete of many different types have been studied and predicted using ML and deep learning models over the past two decades [5-21]. Meanwhile, many studies attempted to increase the predictive power of ML models by introducing more complexities or training them with additional experimental data. However, while heterogeneous and large data are often better for training artificial models [22], in the case of concrete compressive strength prediction, it can be very exhaustive and expensive to expand such a dataset. In particular, to add more instances to a dataset, materials for the mix design must be procured and studied, then mixed and cured for a standard period of 28 days, after which the samples should be crushed to determine their compressive strength.

In addition to the time and effort needed to expand an existing dataset, the experimental data may differ because they are reported by multiple educational laboratories, research centers, or concrete manufacturers. Differences also arise from the time of production of such a dataset and the design codes used. In other words, due to mix design variability, material differences, and mixing and testing procedures, it can be argued that these accumulated data are not homogeneous, hence affecting the performance of ML models. Even if the models that are trained on such inhomogeneous data achieve high accuracies, their application may be limited. Therefore, to overcome the difficulty of increasing the size and homogeneity of the concrete compressive strength dataset, this study proposes and showes that computer-generated data (synthetic data) can be used to train ML models to make accurate predictions of concrete compressive strength after 28 days. In addition, synthetic data are being increasingly used in many fields of applied and theoretical sciences. Synthetic data can be generated from existing analytical models or physical simulations and/or through artificial intelligence generative models [24]. A case in point is in the field of computer vision, where scenarios of tough driving terrains and environments, as well as scenarios of road accidents are being emulated to train artificial intelligence algorithms for self-driving cars.

II. MATERIALS AND METHODS

This study trained an ML model on synthetic mix design data to predict the compressive strength of concrete. The dataset contained mix designs of varying strengths and controlled variables, which aid in the training of the ML model, aiming to study how the synthetic inputs affect the model's predictions of concrete strength. The model was tested using actual laboratory-produced compressive strength data. Figure 1 shows the workflow of the proposed method.

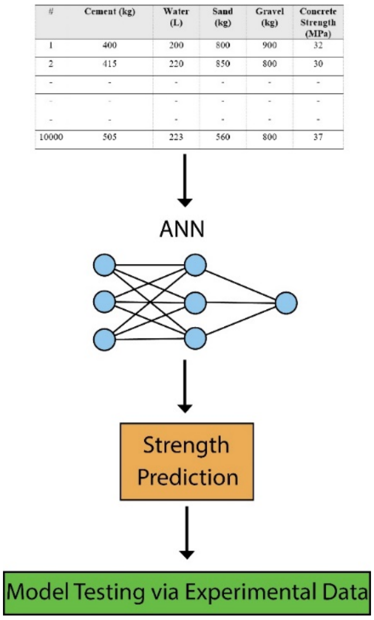

Fig. 1. An illustration of the workflow of the approach used in this study.

The American Concrete Institute (ACI) mix design method was used to create the dataset to train the ML model. Computer software was developed that closely implements all steps of the ACI method for the concrete mix design. The software takes into account the different properties of the mixed ingredients, such as the specific gravities for gravel, sand, and cement, as well as the moisture content and absorption information of the aggregates, along with the fineness modulus of the fine aggregates. In addition, the software also considers the maximum nominal size of the gravel and the required workability, which is provided as a slump value. Moreover, the software requires knowledge of exposure conditions, in addition to requiring input of whether previous test records exist and how many tests are available, if any, accompanied with the necessary test results, such as the standard deviation. Lastly, the software can design both air-entrained and non-air-entrained concrete mixes.

The ACI method-based mix design software was used to create a dataset of non-air-entrained concrete mixes that had a range of specified compressive strength of 20 to 40 MPa with negligible exposure conditions and typical material properties. The chosen slump for all mix designs was 80-100 mm and an MNS for coarse aggregates of 10, 12.5, and 20 mm was specified. Figure 2 shows the histograms of the ingredients of all the mix designs in the dataset. As can be seen from the histograms, the ingredients of the mix designs were well distributed, some of them forming a frequency distribution that resembles a normal distribution, e.g. the amount of sand. However, for other ingredients, the ingredient distribution seems to be clustered in certain areas only, in other words, only a few columns represent the entirety of the data for this ingredient, e.g. the amount of water. The latter phenomenon is due to the way the ACI mix design method works to estimate the water content in a mix design.

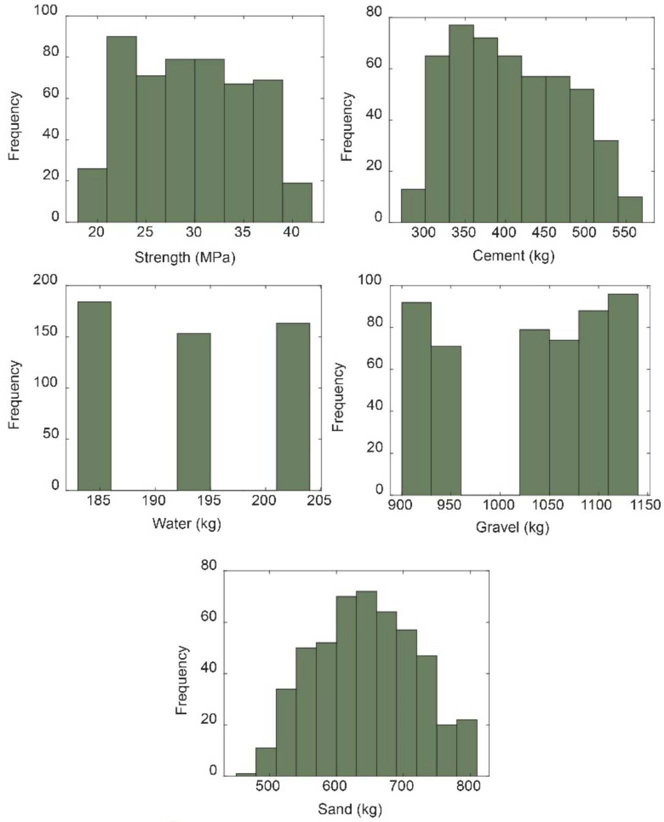

Fig. 2. Histograms of mix design constituents.

To train the model, 250, 500, 1,000 mixes, and 10,000 mixes were created. For each dataset, 80% was allocated to training and the remaining 20% was allocated to testing. In addition, the models were tested using external experimental compressive strength records. The external dataset in [25] was used to test the performance of the model on real-world data. The dataset contained the mix proportions and the resulting experimentally obtained compressive strengths. The dataset contained 1030 test records having 9 attributes, including the amount of cement, water, aggregates, blast furnace slag, fly ash, and superplasticizers. The compressive strengths in the dataset ranged from 2.33 to 82 MPa. To use this dataset as a test for the trained model, it was cleaned to have only records of samples that have strengths ranging from 20 to 40 MPa and made from only the basic four ingredients, i.e. cement, water, sand, and gravel. The cleaned dataset contained 42 records.

The model used was an artificial neural network with four inputs, i.e. the ingredients of the mix design, which were fed to two internal layers and one output layer that represents the predicted compressive strength of concrete. The Levenberg-Marquardt algorithm was used for training and the Mean Square Error (MSE) was used for performance monitoring.

Training stopped when the model was trained for 1000 epochs or the stopping criterion was met. Data are the backbone of ML. Generally, for models that depend on data such as artificial neural networks, the more training data, the better the performance of the model. To see if this is the case with synthetic data, the model was trained on data having from 250 to 10,000 mix designs. Table I shows a sample of the training data. MATLAB R2021b environment was used to create and test the models, particularly the Regression Learner and Neural Network toolbox.

TABLE I. A SAMPLE OF THE TRAINING DATA.

#	Cement (kg)	Water (L)	Sand (kg)	Gravel (kg)	Concrete Strength (MPa)
1	400	200	800	900	32
2	415	220	850	800	30
…	…	…	…	…	…
10000	505	223	560	800	37

III. RESULTS AND DISCUSSION

After training the neural network model on 80% of the mix design dataset, it produced satisfactory training and validation (test) results with very low MSE and high correlation value R. Training required less than 800 epochs, due to meeting the predefined stopping criterion, which was a constantly very low MSE, as shown in Figure 3.

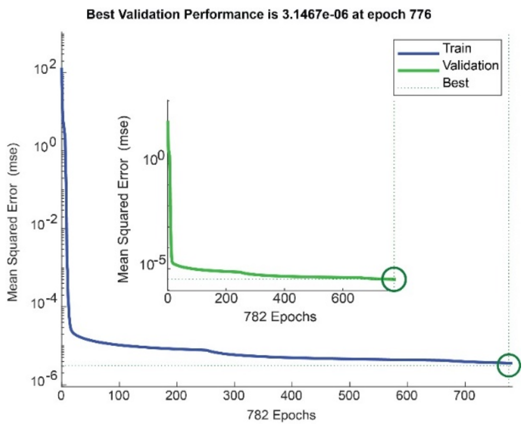

Fig. 3. Training performance of the neural network model.

The resulting trained neural network model achieved an MSE of 3.1467e-6 and a correlation coefficient of 1.0. Figure 4 shows the regression plots of the training performance, indicating that the model achieved a very high correlation, as evidenced by the training R-value, which was 1.0. However, at the same time, the model did not overfit, which was evidenced by the validation R-value, which was 1.0. However, there was a very small error associated with the prediction of the model, as shown in the error distribution in Figure 5. It can be said that the error distribution is well distributed across a center of zero error, with a maximum error not exceeding ±0.008 MPa.

The trained neural network model was tested using external experimental data. The external dataset in [25] was used to test the performance of the model in real-world examples. The dataset contained mix proportions and their experimentally obtained compressive strengths. Similarly to the training dataset, the testing dataset contained four features, i.e. the four basic mix ingredients, and one output, which was the compressive strength of each mix. The number of mixes in the testing dataset was 42. The model was tested using this experimental dataset and achieved a correlation coefficient of 0.80. Figure 6 shows the regression of the neural network model when tested using experimental data with 42 instances, indicating that the data were well distributed around the fit line, with some outliers above and below it. It can be argued that achieving a 0.80 correlation between real-world data and the output of an exclusively synthetic data-trained model is considered a very good result.

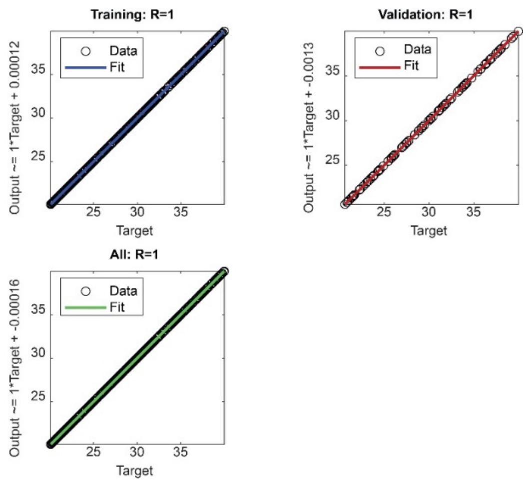

Fig. 4. Regression results for training and validation of the neural network model; *x*-axes refer to actual compressive strength while *y*-axes refer to compressive strength predicted by the ANN model.

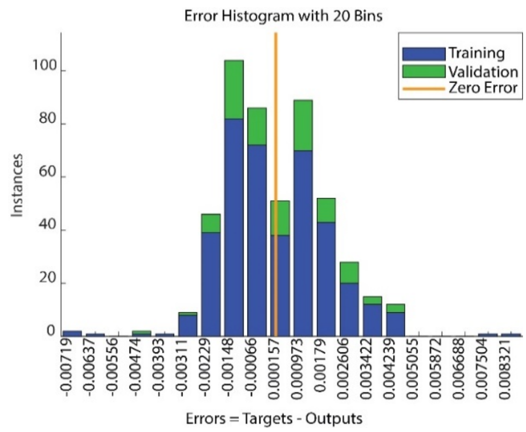

Fig. 5. Error histogram of the model's output in training and validation.

The error histograms of the model when tested on experimental data show that the model achieved a very good correlation, evidenced by the testing R-value, which was 0.8. However, there are some errors associated with the prediction of the model, as shown in the error distribution in Figure 7. The error distribution is skewed, with more errors present above the zero-error line. In other words, the neural network model tends to overshoot.

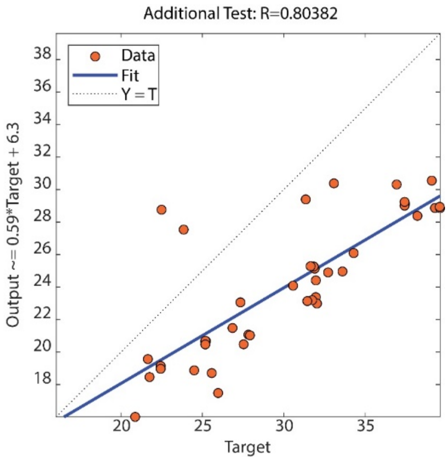

Fig. 6. Performance of the neural network model when tested on experimental data with 42 instances; The *x*-axis refers to the actual compressive strength while *y*-axis refers to the predicted compressive strength.

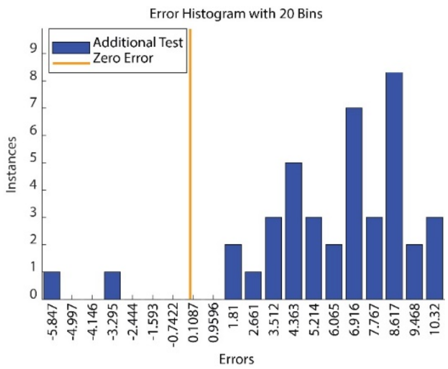

Fig. 7. Error histogram of the model's output in the testing stage using external experimental data.

The prediction of concrete compressive strength using ML methods has been studied using many methods, including Linear Regression (LR), Decision Trees (DT), Support Vector Machine (SVM) and Artificial Neural Networks (ANN). The previously reported findings regarding the prediction of concrete strength using ML methods used experimental data to train the models, while this study used computer-generated data for training. Table II compares the findings of this study with recent relevant studies. It can be seen that although synthetic data were used for training, the proposed model outperformed some experimental data-trained models.

Generally, for models that depend on data, such as the one used in this study, the more the training data, the better the results. To see if this is the case with synthetic data, the model was trained on data ranging from 250 up to 10,000 mix designs. The results showed that for 250 mix designs, the model had an R-value of 0.46. Doubling the amount of training data resulted in an R-value of 0.80. Increasing the dataset to 750, then 1,000, and lastly to 10,000 records only decreased the performance of the model, to 0.78, 0.78, and 0.75, respectively. This decline in performance can be explained by the fact that as synthetic data increased, meaningful data within them were limited, leading to an overfit of the model and a decrease in performance.

TABLE II. COMPARISON OF THIS STUDY'S RESULTS WITH OTHER ML METHODS

Reference	ML Algorithm	Training data	R
[26]	ANN	Experiments	0.9
[26]	LR	Experiments	0.81
[27]	DT	Experiments	0.87
[27]	ANN	Experiments	0.9
[28]	ANN	Experiments	0.76
[28]	LR	Experiments	0.63
[28]	SVM	Experiments	0.67
This study	ANN	Synthetic	0.80

IV. CONCLUSIONS

Recently, intense interest has been placed on using ML techniques to predict the compressive strength of concrete. Such predictive models are typically trained using experimental data on concrete mix ingredients and hardened concrete properties. This study proposed that concrete mixes designed using the ACI mix design method can be used as a training dataset for ML models. The study then showed that synthetic data can be used to train ML models to make accurate predictions of concrete strength after 28 days, ranging from 20 to 40 MPa. To test the performance of the trained model, an external experimental dataset was used, which contained more than 40 instances. When tested in the external experimental dataset, the model was able to predict the 28-day compressive strength of concrete with an R-value greater than 0.8. Meanwhile, it is important to note that the size and quality of the synthetic dataset are important factors that can affect the performance of the ML model. This study aimed to shed some light on the efficacy of using computer-generated data to train artificial intelligence models to provide practical and beneficial predictions. This innovative technique promotes sustainability, as it can alleviate the reliance on experimental data to train ML models, which can be costly and time-consuming.

DATA AVAILABILITY STATEMENT

Data can be provided upon request.

ACKNOWLEDGMENT

The author would like to acknowledge the Deanship of Scientific Research, Taif University for funding this work.

REFERENCES

[1] I. C. Yeh, "Modeling of strength of high-performance concrete using artificial neural networks," *Cement and Concrete Research*, vol. 28, no. 12, pp. 1797–1808, Dec. 1998, https://doi.org/10.1016/S0008-8846(98)00165-3.

[2] H. G. Ni and J. Z. Wang, "Prediction of compressive strength of concrete by neural networks," *Cement and Concrete Research*, vol. 30, no. 8, pp. 1245–1250, Aug. 2000, https://doi.org/10.1016/S0008-8846(00)00345-8.

[3] S. C. Lee, "Prediction of concrete strength using artificial neural networks," *Engineering Structures*, vol. 25, no. 7, pp. 849–857, Jun. 2003, https://doi.org/10.1016/S0141-0296(03)00004-X.

[4] S. Akkurt, G. Tayfur, and S. Can, "Fuzzy logic model for the prediction of cement compressive strength," *Cement and Concrete Research*, vol.

34, no. 8, pp. 1429–1433, Aug. 2004, https://doi.org/10.1016/j.cemconres.2004.01.020.

[5] M. M. Alshihri, A. M. Azmy, and M. S. El-Bisy, "Neural networks for predicting compressive strength of structural light weight concrete," *Construction and Building Materials*, vol. 23, no. 6, pp. 2214–2219, Jun. 2009, https://doi.org/10.1016/j.conbuildmat.2008.12.003.

[6] S. J. Alghamdi, "Classifying High Strength Concrete Mix Design Methods Using Decision Trees," *Materials*, vol. 15, no. 5, Jan. 2022, Art. no. 1950, https://doi.org/10.3390/ma15051950.

[7] S. J. Alghamdi, "A Deep Learning Based Non-Destructive Method for Estimating Concrete Strength using Continuous Wavelet Transform of Vibration Signals Acquired using A Smartphone's Accelerometer," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 12, no. 2, pp. 47–53, Jul. 2023, https://doi.org/10.35940/ijrte.b7738.0712223.

[8] F. Deng, Y. He, S. Zhou, Y. Yu, H. Cheng, and X. Wu, "Compressive strength prediction of recycled concrete based on deep learning," *Construction and Building Materials*, vol. 175, pp. 562–569, Jun. 2018, https://doi.org/10.1016/j.conbuildmat.2018.04.169.

[9] H. Erdal, M. Erdal, O. Simsek, and H. I. Erdal, "Prediction of concrete compressive strength using non-destructive test results," *Computers and Concrete*, vol. 21, no. 4, pp. 407–417, Apr. 2018, https://doi.org/10.12989/cac.2018.21.4.407.

[10] K. C. Williams and P. Partheeban, "An experimental and numerical approach in strength prediction of reclaimed rubber concrete," *Advances in Concrete Construction*, vol. 6, no. 1, pp. 87–102, Feb. 2018, https://doi.org/10.12989/acc.2018.6.1.087.

[11] J. Kasperkiewicz, J. Racz, and A. Dubrawski, "HPC Strength Prediction Using Artificial Neural Network," *Journal of Computing in Civil Engineering*, vol. 9, no. 4, pp. 279–284, Oct. 1995, https://doi.org/10.1061/(ASCE)0887-3801(1995)9:4(279).

[12] W. P. S. Dias and S. P. Pooliyadda, "Neural networks for predicting properties of concretes with admixtures," *Construction and Building Materials*, vol. 15, no. 7, pp. 371–379, Oct. 2001, https://doi.org/10.1016/S0950-0618(01)00006-X.

[13] A. Öztaş, M. Pala, E. Özbay, E. Kanca, N. Çagˇlar, 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, Nov. 2006, https://doi.org/10.1016/j.conbuildmat.2005.01.054.

[14] E. Ghafari, M. Bandarabadi, H. Costa, and E. Júlio, "Prediction of Fresh and Hardened State Properties of UHPC: Comparative Study of Statistical Mixture Design and an Artificial Neural Network Model," *Journal of Materials in Civil Engineering*, vol. 27, no. 11, Nov. 2015, Art. no. 04015017, https://doi.org/10.1061/(ASCE)MT.1943-5533.0001270.

[15] F. Almohammed *et al.*, "Assessment of Soft Computing Techniques for the Prediction of Compressive Strength of Bacterial Concrete," *Materials*, vol. 15, no. 2, Jan. 2022, Art. no. 489, https://doi.org/10.3390/ma15020489.

[16] A. Nafees *et al.*, "Predictive Modeling of Mechanical Properties of Silica Fume-Based Green Concrete Using Artificial Intelligence Approaches: MLPNN, ANFIS, and GEP," *Materials*, vol. 14, no. 24, Jan. 2021, Art. no. 7531, https://doi.org/10.3390/ma14247531.

[17] M. M. Alshihri, A. M. Azmy, and M. S. El-Bisy, "Neural networks for predicting compressive strength of structural light weight concrete," *Construction and Building Materials*, vol. 23, no. 6, pp. 2214–2219, Jun. 2009, https://doi.org/10.1016/j.conbuildmat.2008.12.003.

[18] R. Siddique, P. Aggarwal, and Y. Aggarwal, "Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks," *Advances in Engineering Software*, vol. 42, no. 10, pp. 780–786, Oct. 2011, https://doi.org/10.1016/j.advengsoft.2011.05.016.

[19] L. N. Hussain, M. J. Hamood, and E. A. Al-Shaarbaf, "A Finite Element Approach to Evaluate and Predict the Shear Capacity of Steel Fiber-Reinforced Concrete Beams," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11497–11504, Aug. 2023, https://doi.org/10.48084/etasr.6054.

[20] K. P. Rusna and V. G. Kalpana, "Using Artificial Neural Networks for the Prediction of the Compressive Strength of Geopolymer Fly Ash," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9120–9125, Oct. 2022, https://doi.org/10.48084/etasr.5185.

[21] H. A. Maddah, M. Kheimi, and M. A. El-Wafa, "Predictive Regression Models for the Compressive Strength of Fly Ash-based Alkali-Activated Cementitious Materials via Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8241–8247, Apr. 2022, https://doi.org/10.48084/etasr.4713.

[22] A. D. Trister, "The Tipping Point for Deep Learning in Oncology," *JAMA Oncology*, vol. 5, no. 10, pp. 1429–1430, Oct. 2019, https://doi.org/10.1001/jamaoncol.2019.1799.

[23] X. Wang, R. Jiang, L. Li, Y. Lin, X. Zheng, and F.-Y. Wang, "Capturing Car-Following Behaviors by Deep Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 910–920, Mar. 2018, https://doi.org/10.1109/TITS.2017.2706963.

[24] R. J. Chen, M. Y. Lu, T. Y. Chen, D. F. K. Williamson, and F. Mahmood, "Synthetic data in machine learning for medicine and healthcare," *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 493–497, Jun. 2021, https://doi.org/10.1038/s41551-021-00751-8.

[25] I. C. Yeh, "Concrete Compressive Strength," https://doi.org/10.24432/C5PK67.

[26] B. A. Young, A. Hall, L. Pilon, P. Gupta, and G. Sant, "Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods," *Cement and Concrete Research*, vol. 115, pp. 379–388, Jan. 2019, https://doi.org/10.1016/j.cemconres.2018.09.006.

[27] H. Song *et al.*, "Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms," *Construction and Building Materials*, vol. 308, Nov. 2021, Art. no. 125021, https://doi.org/10.1016/j.conbuildmat.2021.125021.

[28] S. Wadhawan, A. Bassi, R. Singh, and M. Patel, "Prediction of Compressive Strength for Fly Ash-Based Concrete: Critical Comparison of Machine Learning Algorithms," *Journal of Soft Computing in Civil Engineering*, vol. 7, no. 3, pp. 68–110, Jul. 2023, https://doi.org/10.22115/scce.2023.353183.1493.