A Comprehensive Investigation of the Application of Machine Learning Models to Predict Concrete Strength

Wenhao Wang
Computer Science, Coventry University, Singapore

Keywords: Machine Learning, Concrete Strength Prediction, Deep Learning.

Abstract: Concrete is one of the mainstream building materials, so the strength of concrete has an important impact on the development of the world. This paper mainly shows some mainstream methods of machine learning models for predicting concrete strength. On the basis of various predicting concrete strength by machine learning models are analyzed, and the problems of overfitting and possible future development directions of a single machine learning model are discussed. Then, supervised learning modes, such as using Artificial Neural Network (ANN) models to supervise a single machine learning model, can effectively reduce the incidence of overfitting. However, machine learning models face many challenges in predicting concrete strength because the raw data of concrete is difficult to integrate and standardize. With the development and widespread use of expert systems, it may be possible to solve these problems in the future by allowing a variety of related professionals to work together to complete studies using machine learning to predict the strength of concrete.

1 INTRODUCTION

Concrete, as a material that emerged in 1940s (Ecosmartconcrete, 2024), still remains one of the most widely used construction materials today. It is extensively employed in the construction of buildings, roads, bridges, and other structures. According to a report by (Fazaeli, 2021), the use of concrete in the global construction industry is twice that of any other building material. Compared to traditional materials like wood or stone, concrete has advantages such as superior physical strength and chemical strength. Therefore, concrete will continue to be the preferred building material for most construction projects in the present and foreseeable future. However, the strength of concrete is affected by various factors e.g. temperature, humidity and material composition (Fazaeli, 2021). Thus, accurately and effectively measuring the impact of these objective factors on the strength of concrete has become a critical task.

A kind of effective approach currently is to input the raw parameters of the concrete into a machine learning model, and then predict the strength of concrete based on the results predicted by the machine learning model. By building machine learning models, features such as temperature, humidity, and material strength are input as parameters, allowing machine learning models to quantitatively analyze the influence of one or more factors on concrete strength. Computing the strength of concrete by machine learning models sometimes does not require a dedicated site, but only needs to load the computer to train the model and collect the data and will not be too affected by factors such as temperature or site size. The advantage is that it allows for the prediction of concrete strength under these influencing factors anywhere. At present, machine learning models that can be used to compute the strength of concrete include random forests, support vector machines.

There have been some research results on computing the strength of concrete by machine learning. This study points out the results of various machine learning models e.g. Linear Regression (LR), Support Vector Regression (SVR) and Artificial Neural Network (ANN) for predicting concrete strength, and briefly discusses the algorithms that can be used for computing concrete strength by machine learning and the preprocessing of data (Wan, 2021). Another article shows the discrete and large amount of data used to compute the strength of concrete by machine learning, which clearly points out the current challenges for this

Wang, W.
A Comprehensive Investigation of the Application of Machine Learning Models to Predict Concrete Strength.
DOI: 10.5220/0013331200004558

In *Proceedings of the 1st International Conference on Modern Logistics and Supply Chain Management (MLSCM 2024)*, pages 358-361
ISBN: 978-989-758-738-2

study and provides suggestions for subsequent experiments (Li, 2022). There are also articles that explain the problem that machine learning models have predictions that the results will be affected by the grade of concrete (Li, 2022).

The aim of this paper is to provide a review related to the application of machine learning models in concrete strength prediction. The remainder of the paper is organized as follows. First, in Section 2, this review will briefly describe the process Then in Section 3, the advantages and disadvantages in various methods, limitations and challenges in this field, and possible solutions are explained. Finally, this article is summarized in Section 4 to give a comprehensive result of prediction of concrete strength by machine learning models.

2 METHOD

2.1 ANN-Based Concrete Strength Prediction

Now, one of the main methods for computing the strength of concrete by machine learning is the ANN model. The advantages of ANN for calculating concrete strength are that it is good at processing large amounts of data with a variety of complex features, and it is very flexible to build a model of an ANN model, the ANN algorithm can learn to produce what it thinks is the best result, and can use nearly infinite outputs and outputs to calculate nonlinear problems such as concrete strength (Abdolrasol, 2021).

Wu et al. collects data on the ratio of concrete materials, sets them to 7 parameters, and then uses a linear transformation method to normalize the 7 parameters (Wu, 2021). Then they used the ANN model and included a backpropagation network with hidden layers (BPNN) on top of the ANN model and calculated the optimal number of neurons needed for the model. The input to the model is the parameters about the ratio of concrete, such as fly ash and water. In training this model, the study used an ANN model independent of the model topic to check the results of the BPNN model in computing the strength of concrete, and ANN model used a dataset that did not participate in the training of the BPNN model, which can more objectively demonstrate the true accuracy of the BPNN model. On this basis, the error between the predicted value of the model and the real experimental data was calculated to prevent overfitting (Lin,2021).

2.2 Random Forest-Based Concrete Strength Prediction

The random forest algorithm randomly extracts different sample training sets and selects different attribute combinations from them for training. This algorithm is characterized by being good at processing large data sets and has excellent noise immunity (Kim, 2023).

RF models and gene expression models were used in this study to predict concrete strength, replacing the original training data with new training data from bootstrap samples, and finally calculating the error of each random tree to demonstrate the efficiency of the random forest model. The study also uses gene expression programming (GEP), the GEP model includes a set of functions, control parameters, etc, and after completing the prediction, the GEP model compares the results with the predicted results and calculates the fitness for each data point, then selects the best chromosome to send to the next generation and repeats the process, which ultimately yields an optimal result (Farooq, 2020).

2.3 Vector Machines-Based Concrete Strength Prediction

The SVM model has excellent performance in analyzing many types of data, so it can be used to analyze concrete strength, while SVM can analyze a wide range of results with only a small sample of data (Khan, 2021).

In this study, 98 sets of data were used in this study, each of which included six raw parameters on concrete, and the interaction between these data was very complex, and the data were highly discrete, which was difficult to accurately express using traditional equations. By processing into specific six parameters, it can be convenient to build and train the model in the future. Next, the study used a SVR algorithm to predict concrete strength, which can find the most suitable fitting equation for the sample points while minimizing the total variance of the sample points. In order to reduce the influence of errors caused by overfitting on the prediction results, relaxation variables and penalty parameter C are introduced to enhance the resistance of the model prediction results to overfitting, which can reduce the overfitting situation and improve the accuracy of the prediction results. In the part of testing the accuracy of the SVR model, the correlation coefficient (R), MSE, MRSE, MAPE and MAPE were introduced in this study, and GS was introduced to select hyperparameters for the SVR model and perform tenfold cross-validation training. The final results show that the GS-SVR model with GS selection hyperparameters has higher accuracy than the SVR model alone, the data points of the GS-SVR model are closer along the diagonal, and the relationship coefficient R of the GS-SVR model has more than 0.98 in both the training set and the test set (Tang, 2022).

2.4 Clustering Model-Based Concrete Strength Prediction

It is also a good choice to clustering model to compute the strength of concrete by clustering model, which can improve the accuracy of the model's prediction by decomposing the problem domain into subdomains in a systematic and structured way. The difference between a clustering model and ordinary hierarchical modeling is that a clustering model can show hidden causal relationships in clustered data.

For data collection and processing, this study collected 1030 cylindrical samples made of Portland cement, and each sample recorded nine properties such as cement and fly ash, in order to improve the uniformity of the test data, In this study, the data were processed using a linear mapping function. After that, the 981 data samples obtained after removing outliers were divided into two groups, of which 70% were used to train the model and 30% were retained as the test data set. After that, for building the model, by using clustering techniques (UPGMA, HC) to identify the classification of data in the feature space, the next step is to train and test the classifier based on the dataset derived from UPGMA and HC, and then use the same training method to build a classifier that is used to distinguish HC and UPGMA from other sub-clusters. This study trains and optimizes a regression model for each subcluster that has passed the test, and ultimately creates an optimal HCR model from these filtered and optimized subclusters (Demetriou, 2024).

3 DISCUSSIONS

Machine learning models cannot explain the relationship between predictors and outcomes to a certain extent, and will form a "black box", where operators cannot obtain the process from input parameters to results, and thus cannot judge the rationality of the causal relationship of the output results (Aria,2021). For the support vector regression model projected concrete strength, The quality of the prediction depends on the choice of hyperparameters. Because the raw data on concrete is very discrete and the amount of data is huge, inaccurate test results are often obtained if only a single machine learning model is used. the process of prediction by the machine learning model cannot be well demonstrated, which means that this kind of process is not well interpretable, as mentioned above, the person who observes the parameters and results cannot directly derive a complete causal relationship, which leads to a defect in the credibility of the data, if it is necessary to truly check the credibility of computing the strength of concrete by a machine learning model, it can only be tested by setting the same parameters for real concrete. However, doing so will increase the cost of time and manpower.

At present, a method to improve the strength of concrete tested by machine learning includes the use of supervised learning algorithms, which can display and correct some errors of the underlying algorithm under the influence of supervised algorithms or filter out the optimal test results and test parameters. Studies have shown that a single machine learning model is often affected by too much noise or overfitting when predicting the strength of concrete, a task with discrete and large data volume in the original data, and many studies have used ANN algorithms to supervise mutual optimization (Ahmad, 2021). Because the ANN algorithm is good at handling multivariate analysis, this makes the ANN algorithm very suitable as a supervised part (Wu, 2018).

In addition, since the research on using machine learning models to compute concrete strength involves many different disciplines, and professionals on one side who do not know other related disciplines (engineering project managers, builders, algorithm designers) may be prevented from computing compressive strength of concrete by machine learning, so if it is easy to operate, it is possible to set the parameters freely, and the client that uploads the raw data to the database to predict the concrete strength of concrete will be very convenient for the actual operation in this field.

It is also a good practice to optimize a machine learning model using an expert system, which can randomly generate a network with user-specified parameters, and then the expert system determines a random set of inputs that will be used to test the machine learning model (Straub,2021). The parameters mentioned above for computing the compressive strength of concrete are very discrete and numerous, and the machine learning model optimized by expert systems (Liao, 2005) is more resistant to noise from the original parameters of computing the compressive strength of concrete. The versatility of the expert system means that for the task of predicting the strength of concrete, which may be subject to other categories, the tester can add parameters to obtain more accurate and realistic data.

4 CONCLUSIONS

In this work, this article mainly discusses some of the current achievements of Machine learning be a tool for calculating the compressive strength of concrete, and there are now a variety of machine learning models that can accomplish this task, such as ANN, RF, SVR, and clustering models. Many experimental findings show that the use of machine learning to predict concrete strength is a very promising field, but it also faces many challenges, such as the problem that data preprocessing is challenging to be perfect and the prediction accuracy of a single model is not high, but it may be solved by supervised learning and using expert system methods. In the future, there will be better models or data processing methods that can be applied in this field.

REFERENCES

Abdolrasol MGM, Hussain SMS, Ustun TS, Sarker MR, Hannan MA, Mohamed R, Ali JA, Mekhilef S, Milad A. 2021. Artificial Neural Networks Based Optimization Techniques: A Review. Electronics; 10(21):2689

Ahmad A, Chaiyasarn K, Farooq F, Ahmad W, Suparp S, Aslam F. 2021. Compressive Strength Prediction via Gene Expression Programming (GEP) and Artificial Neural Network (ANN) for Concrete Containing RCA. Buildings. 11(8):324.

Aria, M., Cuccurullo, C., Gnasso, A. 2021. A comparison among interpretative proposals for Random Forests, Machine Learning with Applications, Volume 6, 100094, ISSN 2666-8270

Cascardi A, Micelli F. 2021. ANN-Based Model for the Prediction of the Bond Strength between FRP and Concrete. Fibers. 9(7):46.

Demetriou D, Polydorou T, Nicolaides D, Petrou MF. 2024. A clustering machine learning approach for improving concrete compressive strength prediction. Engineering Reports. e12934.

Deshpande G, Batliner A, Schuller BW. 2022. AI-Based human audio processing for COVID-19: A comprehensive overview. Pattern recognition, 122, 108289.

Ecosmartconcrete. 2024. Statistics. https://ecosmart concrete.com/?page_id=208

Fazaeli, H., Seyed Javad Vaziri Kang Olyaei & Mohammad Ali Ziari. 2021. Evaluation of Effects of Temperature, Relative Humidity, and Wind Speed on Practical Characteristics of Plastic Shrinkage Cracking Distress in Concrete Pavement Using a Digital Monitoring Approach

Farooq F, Nasir Amin M, Khan K, Rehan Sadiq M, Faisal Javed M, Aslam F, Alyousef R. 2020. A Comparative Study of Random Forest and Genetic Engineering Programming for the Prediction of Compressive Strength of High Strength Concrete (HSC). Applied Sciences. ; 10(20):7330.

Kim M-C, Lee J-H, Wang D-H, Lee I-S. 2023. Induction Motor Fault Diagnosis Using Support Vector Machine, Neural Networks, and Boosting Methods. Sensors. ; 23(5):2585.

Kim, J., Lee, D., Ubysz, A. 2024. Comparative analysis of cement grade and cement strength as input features for machine learning-based concrete strength prediction, Case Studies in Construction Materials, Volume 21.e03557, ISSN 2214-5095,

Khan, M.A., Memon, S.A., Farooq, F., Javed, Muhammad F., Aslam, F., Alyousef, R. 2021. Compressive Strength of Fly-Ash-Based Geopolymer Concrete by Gene Expression Programming and Random Forest, Advances in Civil Engineering, 2021, 6618407, 17 pages.

Liao, S. H. 2005. Expert system methodologies and applications—a decade review from 1995 to 2004. Expert systems with applications, 28(1), 93-103.

Li, Z., Yoon, J., Zhang, R. et al. 2022. Machine learning in concrete science: applications, challenges, and best practices. npj Comput Mater 8, 127.

Lin C-J, Wu N-J. 2021. An ANN Model for Predicting the Compressive Strength of Concrete. Applied Sciences. ; 11(9):3798.

Straub, J. 2021. Machine learning performance validation and training using a ‘perfect’ expert system, MethodsX, Volume 8, 101477, ISSN 2215-0161

Tang, F., Wu, Y., Zhou, Y. 2022. Hybridizing Grid Search and Support Vector Regression to Predict the Compressive Strength of Fly Ash Concrete, Advances in Civil Engineering, 2022, 3601914, 12 pages.

Wan Z, Xu Y, Šavija B. 2021. On the Use of Machine Learning Models for Prediction of Compressive Strength of Concrete: Influence of Dimensionality Reduction on the Model Performance. Materials. 14(4):713.

Wang, H. 2022. AI-Based Music Recommendation Algorithm under Heterogeneous Network Platform. Mobile Information Systems, 2022(1), 7267012.

Wu N-J. 2021. Predicting the Compressive Strength of Concrete Using an RBF-ANN Model. Applied Sciences. ; 11(14):6382.

Wu, Y. C., & Feng, J. W. 2018. Development and application of artificial neural network. Wireless Personal Communications, 102, 1645-1656.