Automation and Machine Learning (2023)
Clausius Scientific Press, Canada
DOI: 10.23977/autml.2023.040312
ISSN 2516-5003 Vol. 4 Num. 3

# *A Study of Strength Prediction of Multifiber Concrete Based on Improved Stacking*

**Ke Chao***

*School of Urban Construction, Yangtze University, Jingzhou, 434000, China*
*1614467197@qq.com*
**Corresponding author*

***Keywords:*** fiber reinforced concrete; Machine learning; Mechanical properties; stacking

***Abstract:*** In order to solve the expected problem of concrete strength in actual engineering, a mixed concrete database was constructed. Machine learning method was used to use multiple single learners and stacking integrated model. The results showed that the effect of stacking integrated model was far better than that of single learner, and the stacking model was improved. Obtain excellent prediction model of mixed concrete. The performance indexes of MAE, RMES and R2 of this model were significantly better than that of any single learner, and were significantly improved compared with the ordinary stacking model. It provides a reference for the prediction model of civil engineering industry in the future.

## 1. Introduction

Concrete has been widely used in buildings, civil engineering, and transportation infrastructure worldwide. Traditional concrete construction involves the mixing of water, cement, fine aggregate, and coarse aggregate. Concrete is the most commonly used construction material in buildings, bridges, and other infrastructure, with its performance parameters (such as compressive strength, Young's modulus, and slump) being crucial in reinforced concrete (RC) structures. Accurately predicting the compressive strength of concrete has become a key challenge in the current construction industry.

Concrete performance is influenced by various factors, and the relationships between these influencing factors and concrete properties exhibit highly complex non-linear patterns. Traditional linear or non-linear regression models are less ideal for predicting concrete performance. Machine learning, a branch of artificial intelligence (AI), provides computer systems with the ability to self-learn and improve without explicit programming. In recent years, machine learning has been widely applied in various civil engineering applications.

Hou Huiwei et al. [1] employed support vector machines, relevance vector machines, extreme learning machines, and traditional multiple linear regression to analyze and predict radial displacement of a certain arch dam. The results indicated that machine learning predictions significantly outperformed traditional multiple regression methods, although there is still room for improvement in accuracy.

Stacking, an advanced ensemble algorithm, was utilized by Han Z et al. [2] who applied bagging, boosting, stacking, and other ensemble algorithms to predict data sets. The prediction results showed that stacking had higher accuracy compared to other algorithms. To address the issues of long computation time and limited sample data associated with the Stacking algorithm, etc. [3] proposed an improved stacking algorithm based on new vector representation and accuracy-weighted cross-validation. The proposed model demonstrated a significant improvement in predictive performance compared to random forests and other stacking models based on MAE, MSE, and R2 indicators. Stacking research originates from other disciplines, and as an excellent ensemble algorithm, machine learning in the field of civil engineering is still in the expanding phase.

Zhou Hao et al. [4-5] combined support vector machines (SVM) and artificial bee colony algorithm (ABC) to establish a model for optimizing concrete mix proportions. Li Tiejun et al. [6] combined SVM with random forests (RF) to develop a model for predicting the impermeability of concrete. Machine learning demonstrates high accuracy in predicting concrete performance, but current research is focused on traditional concrete or single-fiber concrete. Therefore, this study designs the use of stacking ensemble algorithm for predicting various fiber-reinforced concretes. The stacking primary layer's multiple base learners are dynamically recombined, automatically selecting the most suitable combination of base learners for the data set, thereby improving the accuracy of stacking.

## 2. Problem Description for Strength Prediction of Multi-Fiber Reinforced Concrete

Concrete data is limited, and there is a lack of standardized data exchange. The databases used for training and validation typically consist of only a few hundred sets. The training time cost is negligible, and due to the limited data volume, there is a higher likelihood of low precision in the training results. For machine learning models predicting concrete properties, it is reasonable to increase the training time cost to enhance the final predictive accuracy.

Within the primary layer of stacking, individual base learners, when subjected to cross-validation, yield differing learning models for each fold. Even when utilizing the same database and the same combination of base learners, results may exhibit variations. The machine learning models with the best and worst training outcomes may originate from the same combination of base learners. In the absence of considering time costs, it is possible to compare the training results for each combination and identify the most optimal learning model for the current iteration.

## 3. Modeling the K-MEANS-GSSA Model based on the Standard Stacking Algorithm

### 3.1 Stacking Model

Stacking algorithm, as a typical meta-learning approach, generally employs a two-layer structure: the first layer consists of primary learners (base learners) used for training and predicting on the original samples, while the second layer involves a secondary learner (meta-learner) used to combine the predictions of the first layer and perform further learning and prediction. This ensemble algorithm can be either a homogeneous integration of a single learning algorithm or, more commonly, a heterogeneous integration of multiple algorithms [19], leveraging the advantages of various algorithms to achieve superior predictive performance compared to individual base learners.

Stacking adopts a learning strategy where the primary learners train on the original dataset, and the secondary learner performs a second-level learning on the new dataset generated by the first layer. The input features for the secondary layer are the output results of the first layer, and its dimension depends on the number of individual learners in the primary layer. In this way, the use of the secondary learner replaces the voting strategy, reducing bias and variance.

### 3.2 KMSA Model

The main issues with the standard stacking algorithm lie in data selection and base learner selection.

The learning data for the second-layer stacking meta-learner is generated from the results of the first-layer base learners. The effectiveness of using the data from base learners with excellent training results as meta-learner data is unknown, and the choice of base learners significantly influences the performance of the meta-learner. Before generating stacking model results, the optimal combination of base learners for the database is unknown.

Within the same database, it can be partitioned into multiple data groups, and for each data group, the optimal combination may not be the same. The KMSA model utilizes k-means to partition the training set into multiple clusters, with each cluster representing a new dataset. Utilizing a grid algorithm, KMSA exhaustively determines the optimal combination of base learners for each cluster type.

The construction of the KMSA model involves four stages:

(1)Utilizing the k-means clustering algorithm to partition the original dataset into several clusters.

(2)For each stacking model training, selecting only one cluster as the validation and test set. Additionally, to enhance data diversity and prevent overfitting, weighting the remaining data of this cluster type and combining it with data from other cluster types to generate the training set.

(3)Using grid training, adding a filtering layer to stacking, employing mainstream machine learning models as base learners. Training these models with the same training set and validation set, the results are combined in vector form for three types of base learners, four types, five types, and six types, resulting in 42 matrices (20 for three types of base learners, 15 for four types, 6 for five types, and 1 for six types). These matrices are further separated using linear regression.

(4)Repeating the stacking combinations multiple times, each time fixing the test set from one cluster type to ensure an equal selection frequency for each cluster type. Combining the predicted results of the obtained data samples vertically and taking the average generates a comprehensive dataset.

(5)Identifying the optimal model combination for each cluster type.

(6)When predicting on a new dataset, the machine model first classifies the new data and existing data combinations. The new data is classified into different clusters.

## 4. Model-Based Testing

### 4.1 Data Processing

The database is sourced from literature [6-20], comprising a total of 304 sets. Among them, there are 107 sets of steel fiber-reinforced concrete, 148 sets of polypropylene fiber-reinforced concrete, and 49 sets of steel fiber-polypropylene hybrid fiber-reinforced concrete. In accordance with the research by [15] and others, features with relatively low correlation, such as water reducer, were excluded. Finally, fifteen features, including steel fiber, polypropylene, cement, water, etc., along with their corresponding compressive strength, were selected as the dataset for this study. The 306 data points in the dataset were divided into two parts using the k-means algorithm, with 80% of the data serving as the training set and 20% as the test set. Each time the test set data is fixed, it is partitioned from one cluster. The remaining data is weighted and assigned to the training set. The k-fold cross-validation algorithm is employed to divide the training data into five folds for the learning of base learners.

### 4.2 Evaluation Criteria

MSE (Mean Squared Error), RMSE (Root Mean Squared Error), R2 (Coefficient of Determination), and MAE (Mean Absolute Error) are selected as evaluation criteria. In these formulas, N represents the sample size, Y represents the true values, and $\hat{Y}$ represents the model's predicted values.

Mean Squared Error (MSE) is commonly used as a loss function for linear regression. Additionally, it magnifies deviations with larger prediction errors, enhancing detection sensitivity.

Root Mean Squared Error (RMSE) is more intuitive for data errors and directly reflects the deviation between predicted values and error values.

Coefficient of Determination (R2) is used to assess whether the model fits the experimental data curve. A higher R2 indicates a better fit to the data, approaching 1, while a lower R2 suggests a poorer fit of the regression line.

Mean Absolute Error (MAE), also known as average absolute deviation, is used to assess whether the average of predicted values aligns with the true values.

## 5. Analysis and discussion

### 5.1 Efficiency of Base Learners

To ensure the objectivity of the final results, it was chosen to train each base learner simultaneously using the same training set. The training set provided for each training session is a new combination of k-fold, and the final results for each combination undergo multiple rounds of 5-fold k-fold training, as shown in Table 1.

Table 1: Efficiency of base learner (mean)

| | LGB | RF | SVR | XGB | GBR | MLP |
|---|---|---|---|---|---|---|
| RMSE | 5.728 | 4.619 | 6.29 | 4.04 | 4.44 | 5.67 |
| R2 | 0.818 | 0.877 | 0.77 | 0.905 | 0.88 | 0.82 |

### 5.2 KMSA Model

The original dataset is partitioned into kmeans-1 and kmeans-2, followed by grid search training. The optimal base learner combination for the kmeans-1 cluster is [ 'RF-', 'SVR-', 'GBR-', 'MLP-'], while the optimal machine learning combination for kmeans-2 is ['LGB-', 'SVR-', 'GBR-', 'XGB-'].

Using the improved stacking model as a reference for the KMSA model, the MSE, RMSE, R2, and MAE average results for both models are employed as evaluation criteria to assess the performance of the improved stacking model.

Table 2: Original optimal stacking model and improved stacking data

| | MSE | RMSE | $R^2$ | MAE |
|---|---|---|---|---|
| Stacking(mean) | 12.02 | 3.431 | 0.9378 | 2.36 |
| stacking(min) | 19.62 | 4.427 | 0.9162 | 2.75 |
| stacking(max) | 6.31 | 2.511 | 0.9579 | 1.83 |
| KMSA(mean) | 10.97 | 3.287 | 0.9432 | 2.33 |
| KMSA(min) | 17.91 | 4.232 | 0.9234 | 2.70 |
| KMSA(max) | 2.33 | 2.379 | 0.9671 | 1.81 |

The improved stacking model demonstrates a significantly superior performance compared to the best-performing original stacking model. The optimal combination model of the original stacking deviates from the origin with more data points than the improved stacking model. Moreover, the data points of the improved stacking model are more concentrated around the origin. In Table 2, the average R2 value of the KMSA model is 0.7% higher than that of the original stacking model, while RMSE and MAE have increased by 4.2% and 1.3%, respectively. The KMSA model exhibits more stability on points where conventional stacking models with larger errors perform poorly. Whether in the best or worst training results, the improved stacking consistently outperforms the best combination in the original stacking model.

## 6. Reliability analysis of KMSA model

Yongjian Li et al.[21] constructed a machine learning model for predicting the compressive strength of steel fiber-reinforced concrete using supervised learning. This model was employed as a comparative group for the KMSA (K-means Stacking Algorithm) in predicting the strength of steel fiber, as shown in Table 3.

Table 3: Comparison of prediction models for coagulation strength of different steel fibers

| | Yongjian Li model | | | KMSA model | | |
|---|---|---|---|---|---|---|
| | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| 1 | 8.29 | 0.91 | 3.62 | 3.49 | 0.93 | 2.48 |
| 2 | 4.29 | 0.94 | 3.43 | 3.00 | 0.96 | 2.10 |
| 3 | 9.65 | 0.70 | 8.24 | 3.33 | 0.94 | 2.34 |
| 4 | 4.29 | 0.94 | 3.49 | 3.30 | 0.94 | 2.34 |
| 5 | 7.89 | 0.87 | 4.58 | 3.85 | 0.93 | 2.51 |
| 6 | 8.76 | 0.79 | 7.82 | 3.71 | 0.93 | 2.69 |
| 7 | 10.7 | 0.69 | 9.40 | 3.50 | 0.95 | 2.57 |
| 8 | 5.83 | 0.81 | 4.79 | 3.37 | 0.95 | 2.12 |
| 9 | 4.90 | 0.94 | 3.23 | 2.49 | 0.96 | 1.86 |
| 10 | 3.49 | 0.96 | 2.46 | 3.42 | 0.93 | 2.31 |

R. Tuğrul Eredm [22] constructed an artificial neural network (ANN) model for predicting the compressive strength of polypropylene fiber-reinforced concrete. This model was utilized as a comparative group for the KMSA (K-means Stacking Algorithm) in predicting the strength of polypropylene fiber-reinforced concrete.

Table 4: Comparison of coagulation strength prediction models of different polypropylene fibers

| | KMSA model | Eredm model |
|---|---|---|
| $R^2$(max) | 0.97 | 0.91 |

KMSA exhibits the highest R2 value, which is close to that of other models, and demonstrates significantly higher stability compared to other models. In ten model tests, only the R2 of KMSA remains consistently above 0.90, with an average R2 that is noticeably superior to other models. The MSE and MAE of KMSA are 1-5 units lower than those of other models. Additionally, in the efficiency scatter plot, data points approaching the origin are predominantly occupied by KMSA, highlighting the outstanding performance of KMSA within the database, as shown in Table 4.

The absolute error of the KMSA model is significantly smaller than that of other models. Therefore, the KMSA model exhibits higher accuracy compared to existing fiber-reinforced concrete models. In terms of precision, the practicality of the model proposed in this study far exceeds existing concrete strength prediction models. In terms of breadth, the model can predict various types of fibers as well as fiber-mixed concrete, making its predictive scope more extensive

## 7. Conclusion

(1)A database for hybrid fiber-reinforced concrete was established, and six different machine learning models were trained, with XGB exhibiting the best performance.

(2)The stacking model was applied to train the hybrid fiber-reinforced concrete database, revealing superior performance compared to any individual base learner. Additionally, as the number of base learners increased, the stability of the prediction model improved, albeit with a reduced probability of achieving the optimal solution. For the specific database in this study, three/four types of base learners simultaneously balanced stability and the probability of reaching the optimal solution.

(3)An improved version of the stacking ensemble learning model was developed, demonstrating superior performance compared to any original combination of base learners in the stacking model.

(4)A machine prediction model for hybrid concrete was constructed. In comparison with existing models for predicting compressive strength in steel fiber-reinforced concrete and polypropylene fiber-reinforced concrete, the proposed model outperformed significantly and offered a broader range of predictions, providing a reference solution for future concrete predictions.

## References


*[1] Hou Huiwei, Zheng Dongjian, Huang Hanbing, Fan Bowei. Deformation Prediction Model of Hyperbolic Arch Dam in Concrete Considering the Correlation of Adjacent Points[J]. Journal of Hydroelectric Engineering, 2022, 40(02): 105-109.*

*[2] Zhou Hao, Chen Bin, Meng Meili. Concrete Mix Ratio Optimization Based on SVM-ABC Model[J]. Journal of Hydroelectric Engineering, 2021, 39(06): 127-130.*

*[3] Li Tiejun, Hu Yi, Liu Xi, Wu Xianguo, Wang Lei. Research on Concrete Permeability Prediction Model Based on Support Vector Machine and Random Forest [J]. Journal of Hydroelectric Engineering, 2021, 39(02): 119-122+118.*

*[4] Han Z. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan [J]. Landslides, 2020, 17(3):641-658.*

*[5] Li Pengqin, Zhang Changsheng, Li Yingna, Li Chuan. Photovoltaic Power Prediction with Improved Stacking Algorithm [J]. Journal of Applied Sciences, 2022, 40(02): 288-301.*

*[6] Yoo D Y, Yoon Y S, Banthia N. Flexural response of steel-fiber-reinforced concrete beams: Effects of strength, fiber content, and strain-rate [J]. Cement & Concrete Composites, 2015, 64(2):84-92.*

*[7] Atis C D, Karahan O. Properties of steel fiber reinforced fly ash concrete [J]. Construction & Building Materials, 2009, 23(1):392-399.*

*[8] Yoon E S, Park S B. An Experimental Study on the Mechanical Properties and Long-Term Deformations of High-Strength Steel Fiber Reinforced Concrete [J]. Journal of the Korean Society of Civil Engineers A, 2006, 26(2A):401-409.*

*[9] Lee J H, Cho B, Choi E. Flexural capacity of fiber reinforced concrete with a consideration of concrete strength and fiber content [J]. Construction and Building Materials, 2017, 138(MAY1):222-231.*

*[10] Soulioti D V, Barkoula N M, Paipetis A, et al. Effects of Fibre Geometry and Volume Fraction on the Flexural Behaviour of Steel-Fibre Reinforced Concrete[J]. Strain, 2009, 47.*

*[11] Dvorkin L, Bordiuzhenko O, Tekle B H, et al. A Method for the Design of Concrete with Combined Steel and Basalt Fiber [J]. Applied Sciences, 2021, 11(19): 8850.*

*[12] Hasan A, Maroof N, Ibrahim Y. Effects of Polypropylene Fiber Content on Strength and Workability Properties of Concrete[J]. Polytechnic Journal, 2019, 9(1):7-12*

*[13] Fard A, Soheili H, Movafagh S R, et al. Combined effect of glass fiber and polypropylene fiber on mechanical properties of self-compacting concrete. 2016.*

*[14] Xu L, Li B, Ding X, et al. experimental investigation on damage behavior of polypropylene fiber reinforced concrete under compression international journal of concrete structures and materials [J]. 2019.*

*[15] Wang Zhizhao. Experimental Study on the Comprehensive Performance of Polypropylene Fiber Reinforced Concrete [D]. Zhejiang University, 2004.*

*[16] Shan Fengzhi. Research on the Mechanical Properties of Polypropylene Fiber Reinforced Concrete and Its Application in Bridge Deck Pavement [D]. Southwest Jiaotong University, 2005.*

*[17] Huang Gongxue. Experimental Study on the Physical and Mechanical Properties of Polypropylene Fiber Reinforced Concrete (Mortar)[D]. Zhengzhou University, 2005.*

*[18] Li Huaming. Study on the Performance of Polypropylene Fiber Reinforced Concrete and Its Application in Tunnel Engineering [D]. Southwest Jiaotong University, 2005.*

*[19] Hao Zhuan. Research and Application of Mechanical Characteristics of Polypropylene Fiber Reinforced Concrete Panels [D]. Northwest A&F University, 2007.*

*[20] Yu Chun. Study on the Performance of High-Performance Polypropylene Fiber Reinforced Concrete[D]. Chongqing Jiaotong University, 2009.*

*[21] Li Y, Zhang Q, Kamiński P, et al. Compressive Strength of Steel Fiber-Reinforced Concrete Employing Supervised Machine Learning Techniques[J]. Materials, 2022, 15(12): 4209.*

*[22] Eredm RT, Kantar E, Gücüyen E, Anil O. Estimation of compression strength of polypropylene fibre reinforced concrete using artificial neural networks. Comput Concr 2013; 12: 613–625.*