

Article

Machine Learning-Based Method for Predicting Compressive Strength of Concrete

Daihong Li ^{1,2}, Zhili Tang ³, Qian Kang ², Xiaoyu Zhang ^{1,*} and Youhua Li ¹

¹ China Gezhouba Group Three Gorges Construction Engineering Co., Ltd., Yichang 443000, China

² School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia

³ Beijing Jingtou Urban Utility Tunnel Investment Co., Ltd., Beijing 100027, China

* Correspondence: zhangxiaoyu16@mails.ucas.ac.cn

Abstract: Accurate prediction of the compressive strength of concrete is of great significance to construction quality and progress. In order to understand the current research status in the concrete compressive strength prediction field, a bibliometric analysis of the relevant literature published in this field in the last decade was conducted first. The 3135 journal articles published from 2012 to 2021 in the Web of Science core database were used as the database, and the knowledge map was drawn with the help of the visualisation software CiteSpace 6.1R2 to analyse the field at the macro level in terms of spatial and temporal distribution, hotspot distribution and evolutionary trends, respectively. Afterwards, we go into the detail and divide concrete compressive strength prediction methods into two categories: traditional and machine-learning methods, and introduce the typical methods of each. In addition, a boosting-based ensemble machine-learning algorithm, namely the gradient boosting regression tree (GBRT) algorithm, is proposed for predicting the compressive strength of concrete. 1030 sets of concrete compressive strength test data were collected as the dataset, of which 60% were used to train the model, 20% to validate the model and 20% to test the trained model. The coefficient of determination (R^2) of the GBRT model was 0.92, the mean square error (MSE) was 22.09 MPa, and the root mean square error (RMSE) was 4.7 MPa, which is an excellent prediction accuracy compared to prediction models constructed by other machine-learning algorithms. In addition, a five-fold cross-validation analysis was carried out, and the eight input variables were analyzed for their characteristic importance.



Citation: Li, D.; Tang, Z.; Kang, Q.; Zhang, X.; Li, Y. Machine Learning-Based Method for Predicting Compressive Strength of Concrete. *Processes* **2023**, *11*, 390.
<https://doi.org/10.3390/pr11020390>

Academic Editors: Shaohua Wu,
Han Xu and Xiang Sun

Received: 27 December 2022
Revised: 23 January 2023
Accepted: 26 January 2023
Published: 27 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Concrete is one of the essential building materials of our time. Due to its integrity, good durability and economy, it is used in a wide range of applications, not only in civil engineering but also in the mechanical industry, shipbuilding and other projects. In order to ensure that engineering structures serve safely and stably during their designed service life, it is vital to study the mechanical properties of concrete, of which compressive strength is the most important indicator, as it is directly related to the safety of engineering structures, so how quickly and accurately the strength of concrete can be judged is of great importance to the quality and progress of construction [1]. Generally speaking, concrete is a composite material consisting of cement as the cementitious material, sand and stone as coarse and fine aggregates, respectively, plus a certain proportion of water and admixtures. Accurately predicting the compressive strength of concrete is challenging due to its complex composition and the fact that there is not a simple linear relationship between the components and the strength of the concrete [2–4].

The most traditional method of measuring the compressive strength of concrete is by physical testing, by making a cubic or cylindrical concrete test block according to

the design specification, which can be measured very easily using a compression-testing machine after a period of standard curing [5–7]. This method is simple, but the time and economic costs are high. Some researchers have since proposed several empirical regression methods to predict the compressive strength of concrete for a given mix ratio [8–10]. However, the components of concrete do not show a simple linear relationship with concrete strength but rather a strongly non-linear relationship, which makes it extremely difficult to summarise the exact regression expressions directly [11,12]. In recent years, with the rapid development of artificial intelligence, the method of machine learning has been widely used in major research areas in structural engineering, such as structural system identification [13], structural element design [14–18] and concrete compressive strength prediction [19–24]. Compared to traditional regression methods, machine learning can build prediction models with the help of some algorithms to handle regression problems well [25], so predicting concrete compressive strength by machine learning has become a fashionable research trend [25–27]. To name a few, Chithra et al. [28] successfully used the artificial neural network (ANN) algorithm applied to predict the compressive strength of high-performance concrete (HPC) containing silica nanoparticles and copper slag. Ayat et al. [29] used ANN to predict the compressive strength of limestone-filled concrete with a correlation coefficient value as high as 0.976. Nguyen et al. [30] proposed four machine-learning algorithms for predicting the compressive and tensile strength of HPC, with the prediction models based on gradient boosting regressor (GBR) and extreme gradient boosting (XGBoost) having good output accuracy. Kumar et al. [31] collected 120 sets of data and developed several machine-learning algorithms for predicting the compressive strength of lightweight concrete (LWC), of which the support vector machine (SVM) model was the best. Ashrafian [32] used heuristic regression methods to predict fibrous concrete's strength and ultrasonic pulse velocity successfully. Zhang et al. [33] used random forest (RF) to predict the uniaxial compressive strength of lightweight self-compacting concrete and performed characteristic importance analysis on eight input variables.

However, most of the algorithms currently used for concrete compressive strength prediction are individual learning algorithms. In contrast, integrated learning algorithms have better prediction accuracy and robustness [13]. This is because ensemble learning algorithms use training data to train multiple weak learners, which are individual learning algorithms, simultaneously and then integrate them to build a strong learner to output results [34]. In addition, previous studies have only about two to three hundred datasets in the database, which is very small and can seriously affect the performance of the final prediction model. In this paper, the gradient boosting regression tree (GBRT) algorithm, a high-performing but rarely studied algorithm, is used and combined with 1030 sets of concrete experimental data to develop a model for concrete compressive strength prediction. Model performance evaluation metrics, including the coefficient of determination (R^2), the mean square error (MSE), and the root mean square error (RMSE) values, were calculated to determine the prediction accuracy of the GBRT model and compared with the performance of models using individual learning algorithms as well as other ensemble learning algorithms. In addition, a five-fold cross-validation analysis and a characteristic importance analysis of the input variables were carried out.

The structure of this paper is as follows. Chapter 1 introduces the importance of predicting the strength of concrete and briefly mentions several methods for predicting the compressive strength of concrete. Chapter 2 provides a literature review of the current state of research in the field of concrete compressive strength prediction, starting with a bibliometric analysis of the literature published in the field over the last decade, based on CiteSpace software, followed by a detailed description of the research methods for concrete strength prediction, which are divided into two categories: traditional approaches and machine-learning methods. In addition, gaps in current research are identified. Chapter 3 discusses the development of a concrete compressive strength prediction model using the GBRT algorithm and the highly accurate prediction results obtained, and a series of analyses are performed. Chapter 4 summarises the work carried out throughout the text

and draws several conclusions from the analysis of the results obtained, and concludes with an outlook for the future. Figure 1 illustrates the structure of this paper and the main steps in conducting the analysis.

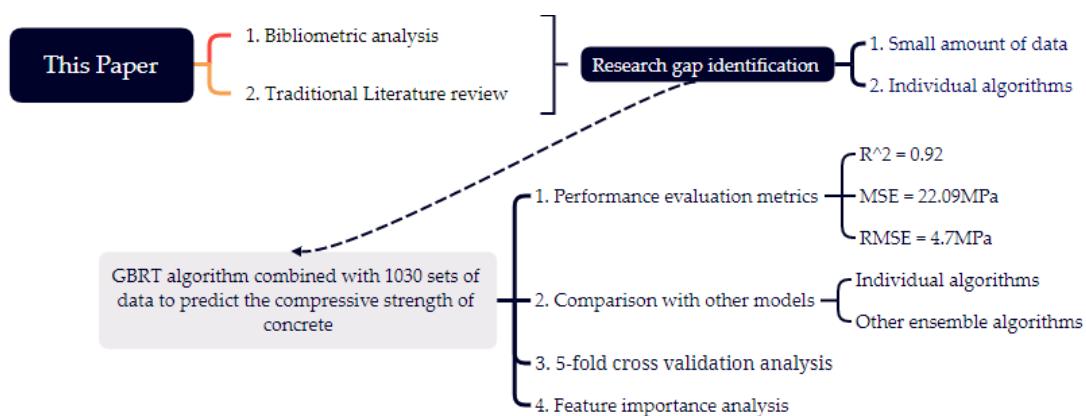


Figure 1. Structure of this paper.

2. Current Status of Research

In order to gain a more comprehensive understanding of the current state of research in the concrete compressive strength prediction field, this paper reviews the literature in the field from both macro and micro aspects. In the macro aspect, the 3135 publications in the Web of Science core database from 2012–2021 were used as a database to create a knowledge map with the help of the visualisation software CiteSpace, including the temporal distribution of the number of publications, the collaborative network of authors and research institutions, the keyword co-occurrence network, keyword clustering and a time-line map of hot research areas to illustrate the hotspots and the evolution of frontiers in re-search. At the micro level, the methods developed to rapidly and accurately predict the compressive strength of concrete have been divided into two main categories: traditional approaches and machine-learning methods. Traditional techniques include the preparation of concrete test blocks through physical experiments and summarising some empirical regression formulas. Machine-learning methods allow for high accuracy in predicting the compressive strength of concrete through a combination of stochasticity and non-linearity, such as artificial neural networks and support vector machines. It is worth noting that the significant advantage of machine-learning methods over traditional methods is that they can take into account the effects of multi-factor variables on the compressive strength of concrete, which has made machine-learning methods increasingly popular with research scholars.

2.1. Prediction of Concrete Compressive Strength Based on Bibliometric Analysis

Research into the strength of concrete can be traced back to the 19th century. After more than 100 years of development, a large amount of relevant literature has emerged. We are inevitably limited in reading them by traditional reading and statistical methods, making it difficult to classify and summarise the enormous amount of literature accurately. This problem can be effectively solved using CreateSpace software, a Java-based visualisation software developed by Professor Chaomei Chen of Drexel University, USA. It can analyse collaborative networks, keyword co-occurrence and literature co-citation among the literature [35,36], helping research scholars to quickly familiarise themselves with the current state of research in a research area. Therefore, this paper uses CiteSpace software to visualise and analyse the literature related to the subject term concrete compressive strength prediction in the Web of Science core database from 2012–2021. Combined with the knowledge graph output from the software, the basic background, development overview and frontier evolution of the research field of concrete compressive strength prediction on

a global scale are analysed and illustrated, providing references for further research in the field of concrete compressive strength prediction.

2.1.1. Research Methodology and Data Sources

- Research methodology

The scientific knowledge mapping software CiteSpace 6.1R2 was selected to visualise and analyse the concrete compressive strength prediction field. Using this software, we can demonstrate the development process of the knowledge field through knowledge mapping and identify the research frontiers expressed by citation node literature and co-citation clusters [37]. By importing the collected database into CiteSpace, an econometric analysis of the literature in terms of authors and their collaborations, research countries and their collaborations and research institutions and their collaborations can be performed, which in turn provides an understanding of the spatial and temporal distribution characteristics of the field of concrete compressive strength prediction. In addition, a keyword clustering analysis and a literature co-citation analysis can be performed to show the basic background, the distribution of research hotspots and future evolution trends in the concrete compressive strength prediction field. Figure 2 illustrates the five parts of the analysis that will be performed in the bibliometric section.

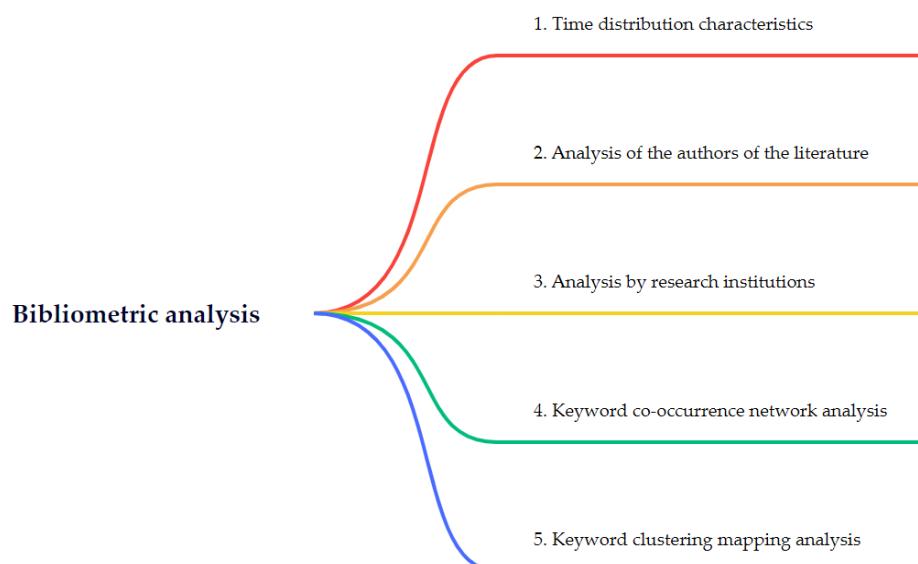


Figure 2. Five parts of the analysis in the bibliometric section. Data Source.

This paper selected the Web of Science core database as the literature database source. The subject term was set as a prediction of the compressive strength of concrete in the literature search. In order to obtain a more comprehensive and macroscopic research perspective, the literature search was set for the last decade (2012–2021), and the search results were screened and checked against the basic knowledge of concrete strength prediction to eliminate irrelevant content, resulting in a total of 3135 publications as the data source. The specific screening method of the database is shown in Table 1.

Table 1. Data sources.

Database	Web of Science Core Collection
Search method	Subject
Search vocabulary	Prediction of compressive strength of concrete
Time span	2012–2021
Search results	3135 articles

2.1.2. Overview of Research into the Prediction of Compressive Strength of Concrete

- Time distribution characteristics

Using Microsoft Excel software, we can count the number of publications by year and plot the time distribution of the number of publications (Figure 3). It can be found that the number of relevant literature publications in the last decade generally showed an increasing trend and can be roughly divided into three periods, namely the stable period (2012–2014), the slowly increasing period (2015–2018) and the period of rapid increase (2019–2021).

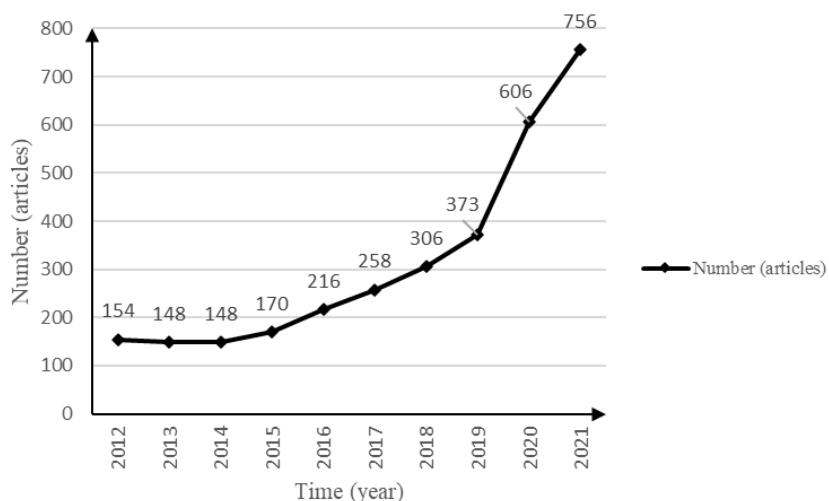


Figure 3. Number of publications in the literature on concrete compressive strength prediction 2012–2021.

The number of publications in the literature from 2012–2014 (stable period) was low, remaining at around 150 per year, a period that relied mainly on traditional methods such as linear regression and non-linear regression for concrete strength prediction, while the number of papers from 2015–2018 (slow growth period) increased year by year, mainly due to a 10–100-fold increase in the computational requirements of companies to train machine-learning models at the end of 2015, with large machine-learning models being developed one after another [38], and researchers in the concrete direction were able to apply these large machine-learning systems with better performance to the field of concrete strength prediction, and there is a spurt in the number of literature publications in 2019–2021 (rapid growth period), which is closely related to the fact that supervised-learning machine-learning models tend to be improved and deep learning starts to be applied to the field of concrete strength prediction.

- Analysis of the authors of the literature

Selecting the node type as the author in the CiteSpace software gives a co-occurring knowledge graph of the literature authors (Figure 4), where each dot indicates each author, and if there is a collaboration between the authors, lines of the same colour connect them. The collaboration network shows a large number of research scholars worldwide in the direction of concrete compressive strength prediction. However, most scholars collaborate less with each other, and only small-scale research collaborations exist. Of these, Professor Ali Nazari from Islamic Azad University has published the most papers as first author and co-author with 30 articles, occupying the most prominent node position in the knowledge graph. Under his leadership, Islamic Azad University has achieved many research results in the concrete compressive strength prediction field. The second most published article was by Professor To-gay Ozbakkaloglu with 19 papers, laying a solid research foundation for developing the field as one of the first scholars to work in the field (2012).

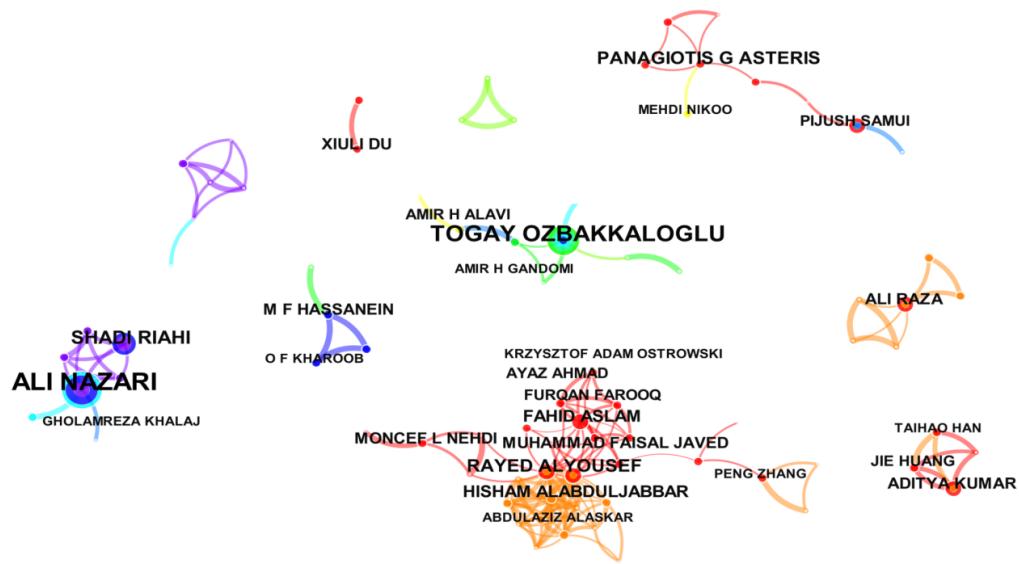


Figure 4. Research Author Collaboration Network.

- Analysis by research institutions

By setting the node type to the institution in the CiteSpace software settings, it was possible to obtain a co-occurring knowledge map of the institutions to which the literature was published between 2012 and 2021 (Figure 5), where each dot represents an institution. If there is cooperation between institutions, there will be the same colour link to connect them.

The institutional co-occurrence network diagram has 346 nodes and 542 links, indicating a relatively concentrated network of research institutions working on concrete compressive strength prediction and mainly concentrated in the Middle East and the Asia Pacific regions. In the Middle East region, Islamic Azad University published 90 articles and obtained the highest intermediary centrality of 0.22. In the Asia Pacific region, Tongji University, Southeast University and Harbin Institute of Technology in mainland China achieved good results in this area, with each institution publishing approximately 35 articles.

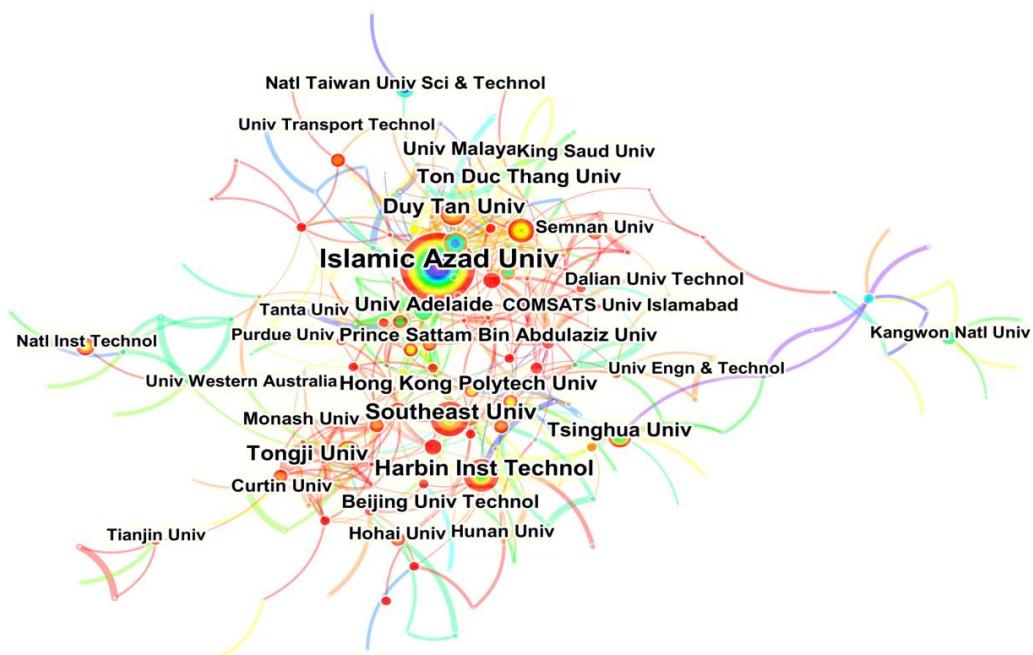


Figure 5. Collaborative network of research institutions.

- Keyword co-occurrence network analysis

By setting the node type to keywords in the CiteSpace software, the software can output a knowledge map of hot keyword co-occurrence networks based on the frequency and mediated centrality of keyword occurrences in the published literature, as shown in Figure 6. The knowledge map is tree-like, without many freely scattered points, with low-frequency keywords subordinated to high-frequency keywords and with high-frequency keywords interconnected, indicating that the theoretical system in the concrete compressive strength prediction field is relatively mature and that the internal sub-disciplines are closely linked. The most frequent keyword is compressive strength (737 times), followed by prediction (529 times), behaviour (435 times) and mechanical properties (364 times), which indicates that the study of compressive strength is most closely related to other aspects. This suggests that the study of compressive strength is closely linked to the other keywords, with compressive strength being the source of research leading to studies of prediction, behaviour and mechanical properties, which is also verified by the linkage of the keywords.

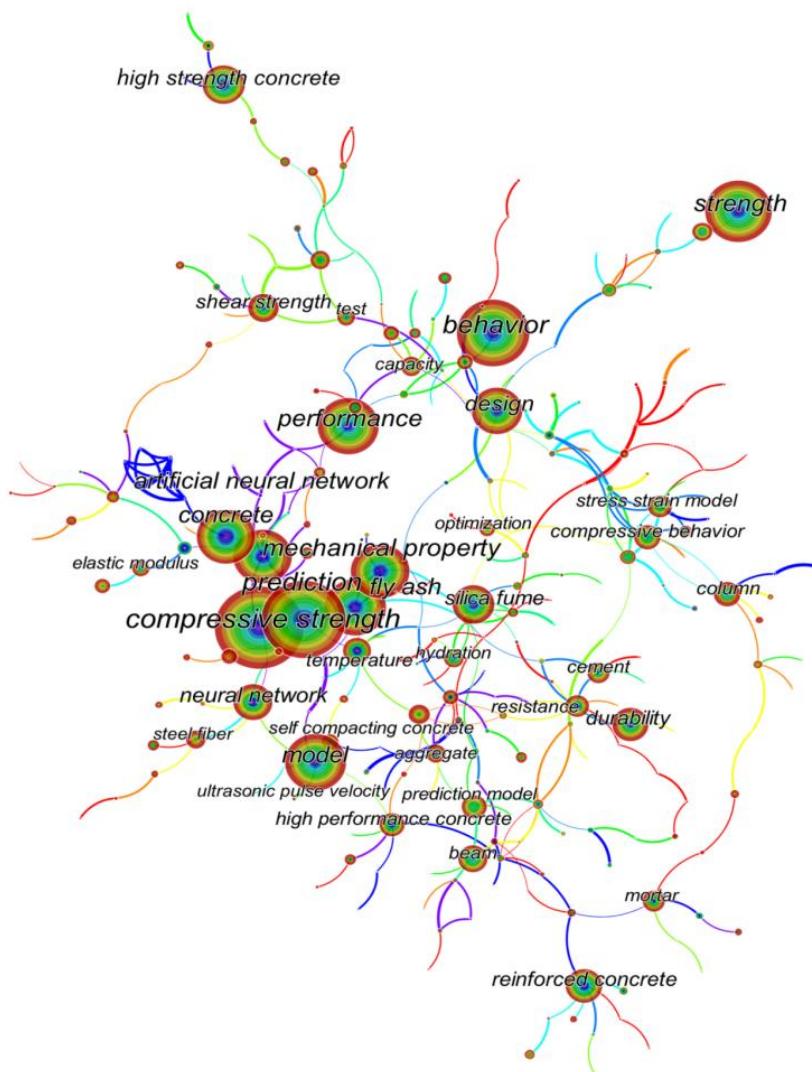


Figure 6. Co-occurrence network map of hot keywords.

- Keyword clustering mapping analysis

The CiteSpace software can use pathfinder clustering to generate a knowledge graph by clustering keyword tag words, as shown in Figure 7. Table 2 summarises the top ten clusters and their associated parameters, where the size value indicates the number of nodes

in each cluster, and the silhouette value is the average profile value of the clusters, which is an essential measure of cluster homogeneity. It is generally accepted that clusters with a silhouette value greater than 0.5 are reasonable, and if the value is more significant than 0.7, the clusters are convincing. The silhouette values for the first ten clusters in this study were all greater than 0.7, which indicates that all clustering results were convincing [39].

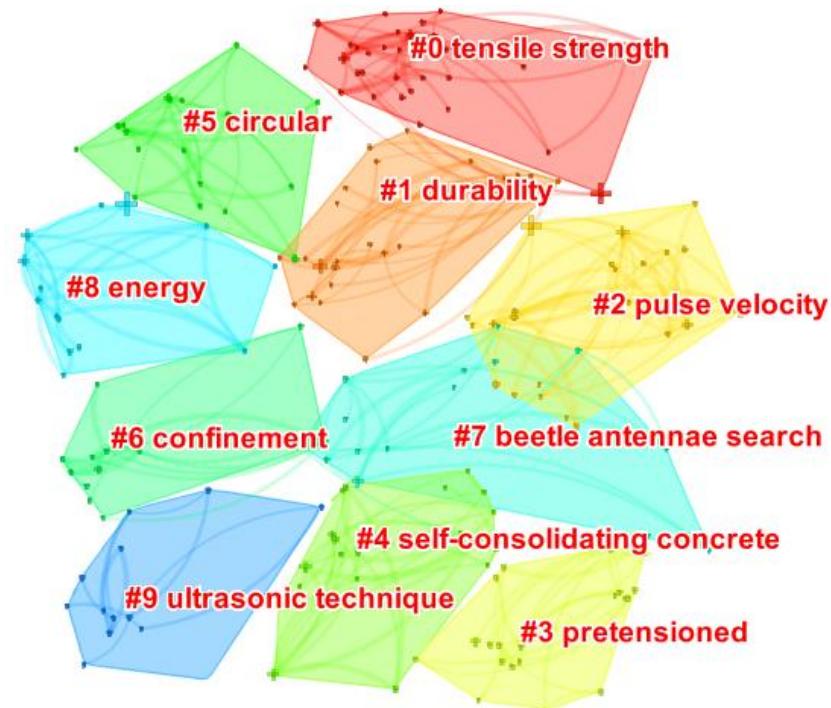


Figure 7. Co-occurrence network map of hot research areas.

Table 2. Hot keyword clustering and tagged word information.

Cluster ID	Size	Silhouette	Representative Label (LLR)	Year Ave.
0	33	0.923	Tensile strength	2016
1	30	0.794	Durability	2016
2	24	0.96	Pulse velocity	2014
3	21	0.876	Pretensioned	2016
4	21	0.856	Self-consolidating concrete	2015
5	19	0.937	Circular	2018
6	16	0.875	Confinement	2016
7	15	0.966	Beetle antennae search	2016
8	14	0.903	Energy	2013
9	12	0.917	Ultrasonic technique	2019

CreateSpace can also plot a timeline view of clusters based on the clustered tag words extracted by the log-likelihood ratio test (LLR) method, as shown in Figure 8, where the horizontal line represents the timeline of the cluster, the length represents only the years in which high-frequency words appear in the popular keyword set for that cluster, and the size of the circle indicates the frequency of the keyword. The line diagram arranges the nodes in each cluster in chronological order, presenting the development of the cluster in the temporal dimension, and can help us to understand the hot representative words in each cluster at different stages.

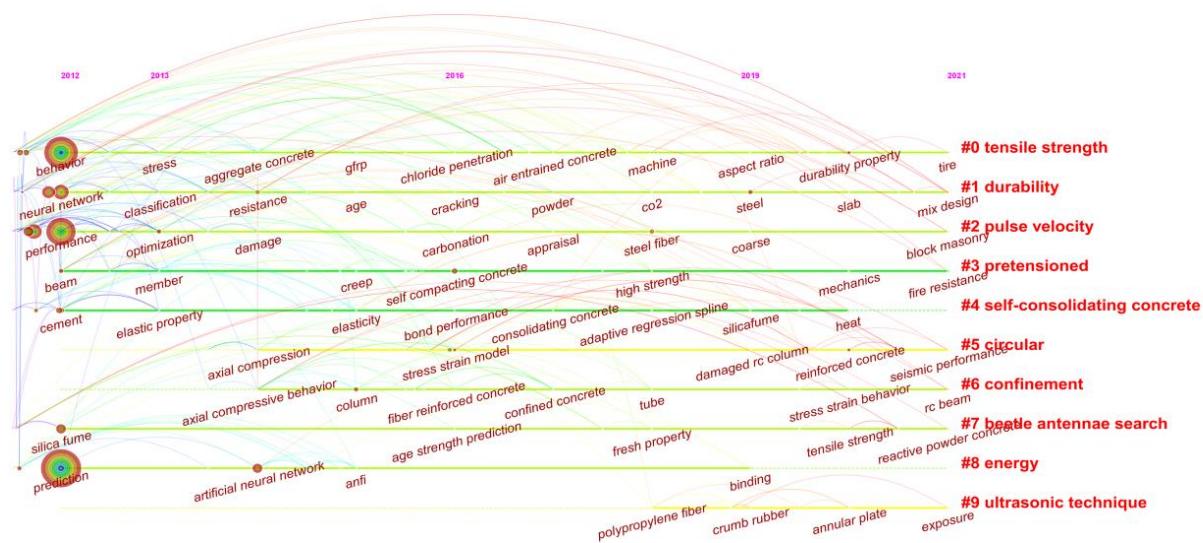


Figure 8. Timeline view of hot research areas.

2.2. Status of Research on Prediction Methods

2.2.1. Traditional Methods

The traditional methods of concrete strength prediction mainly include experimental methods and methods of summarizing empirical regression equations [40]. In 1971, scholars from the United States carried out a large number of tests related to the prediction of concrete strength [41]. They have summarised the three most representative experimental methods for predicting the strength of concrete: the self-heating method, the 35 °C warm-water method and the boiling-water method, which are popular among engineers [42–44]. Although the prediction accuracy of these three methods is good, the experimental procedure is complicated, and there are limitations in the scope of the application [45]. Therefore, more and more researchers have started to study how to summarise the empirical formulae based on the existing concrete experimental data to achieve the purpose of accurate concrete strength prediction. To name a few, Liu et al. [46] applied the FCT101 fresh concrete tester from Colebrang, a company in the UK, to detect the FCT (Fresh Concrete Test) values of concrete to predict the compressive strength of concrete with a relative error within 10% compared to the strength at 28 d of conventional standard curing. Soh and Bhalla [47] proposed a non-destructive testing method based on concrete impedance, using the electro-mechanical impedance (EMI) technique for the non-destructive determination of in-situ concrete strength. Zheng [48] derived an empirical equation based on the equivalent age theory of concrete, which summarised the variation of concrete strength with age, and the difference between the expected concrete strength values obtained, based on the empirical equation, and the real values of concrete strength measured by experiment was only 10%. Elaty [49] presented a mathematical model and three empirical formulas based on two constants defined by themselves, which successfully obtained the strength development pattern of portland cement concrete mixtures containing silica fume with age at room temperature and also successfully predicted the compressive strength of portland cement concrete containing nano-silica fume cured with water at room temperature at any age. Nambiar and Ramamurthy [50] extended Balshin's strength-porosity model and successfully predicted the compressive strength of foamed concrete with a final R^2 of 0.893. Table 3 presents information on conventional concrete strength prediction methods.

Table 3. Traditional concrete-strength prediction methods.

Author	Research Methods	Data Volume	Results
Liu et al. [46]	FCT prediction method	100	Relative error less than 10%
Soh and Bhalla [47]	EMI non-destructive testing	15	$R^2 = 0.955$
Zheng [48]	Equivalent age theory	54	Maximum error rate less than 10%
Elaty [49]	Summarising mathematical formulas	6	Not quantified
Nambiar and Ramamurthy [50]	Balshin's generalised model	11	$R^2 = 0.893$

In summary, traditional concrete strength prediction methods include classical experimental methods and regression methods based on mathematical statistics, which can be more accurate in predicting the compressive strength of concrete in simpler situations, but when more factors need to be considered, and the concrete is subject to more complex external effects, accurate prediction of the compressive strength of concrete will become very difficult [8]. In addition, the amount of data used to test the accuracy of the research method in previous work is generally only a few dozen groups, and the test data are generally not fully repeatable and reproducible, which leads to a small range of application and poor generalisation of the traditional concrete-strength prediction method.

2.2.2. Machine-Learning Methods

Compared to traditional prediction methods, machine-learning methods to predict the compressive strength of concrete are more favoured by scholars because machine-learning algorithms can tap into the deeper patterns of the input data and generate reliable prediction models through training, to output highly accurate results [43,51]. Many machine-learning algorithms have been applied to the field of concrete compressive strength prediction. To name a few, Lai and Serra [52] found that the performance of neural networks was independent of the number of neurons in the hidden layer (range 4–8), and the accuracy was the same (5%). Kewalramani and Gupta [53] used ANN to predict the compressive strength of concrete specimens and obtained results almost identical to the compressive strength obtained through physical experimental tests. Naderpour et al. [54] used ANN combined with 139 sets to predict recycled concrete's compressive strength (RAC) with a model MSE of 0.004447. Asteris and Kolovos [55] proposed a new heuristic algorithm to find the best heuristic for a multilayer feedforward backpropagation neural network based on the value of Pearson's correlation coefficient. The output values of the ANN prediction model trained using 205 sets of parameters were very close to the experimental results of the compressive strength of self-compacting concrete with R equal to 0.9828. Zhu et al. [56] proposed a genetic algorithm-optimised support vector machine model (GA-SVM) to establish the relationship between seven parameters and concrete strength and compared the prediction results with those of BP neural networks and found that the GA-SVM model performed better. Aiyer et al. [57] first applied the least square SVM, an advanced SVM concept, to the field of concrete strength prediction. Pham et al. [58] further optimised the least square SVM with the help of metaheuristic optimisation and successfully predicted the compressive strength of high-performance concrete. Li and Peng [59] used the neural network toolbox provided by Matlab to build a back propagation (BP), radial basis function (RBF) neural network model with a 4-dimensional input vector and a 1-dimensional output vector for the prediction of concrete compressive strength with good results. Gao and Hao [60] developed a back propagation artificial neural network (BP-ANN) model combining 30 sets of data to achieve a non-linear mapping between concrete rebound values, ultrasonic velocity values and compressive strength of concrete with an absolute error of less than 5.0% between predicted and measured values, and the model performed well for predicting the compressive strength of self-compacting

concrete. Based on BP neural networks, Ma and Liu [61] established a neural network prediction model with a non-linear mapping relationship between each input parameter and combination form and the compressive strength of the restrained column based on 251 sets of experimental data of carbon-fibre-reinforced plastics (CFRP) restrained concrete columns and proposed a theoretical calculation formula and a simplified formula. The overall results are better than the traditional linear regression results.

The machine-learning algorithms mentioned above are all learned individually. The training process for prediction models built using them is relatively simple, requiring only a tiny amount of data to obtain good prediction results. However, integrated learning algorithms, which offer better robustness and prediction accuracy, are more popular among researchers. The basic idea of integrated learning algorithms is first to train several weak learners (such as ANN, SVM and other individual learning algorithms) using the training set data and then combine several weak learners to generate a strong learner for prediction work. The common integrated learning algorithms can be divided into two categories, bagging-based algorithms and boosting-based algorithms [13], where the representative algorithms based on bagging are RF, while the representative algorithms based on boosting are adaptive boosting (AdaBoost) [62], GBRT, etc. Wu et al. [63] found that feature screening of variable importance indicators was critical to improving prediction accuracy when using the random forest algorithm for concrete-strength prediction. Cui et al. [64] compared the capabilities of three machine-learning algorithms, random forest, support vector regression and multilayer perceptron, on the same dataset (1030 sets of concrete compressive strength test data) and found that the RF algorithm has the best effect among these three algorithms. Farooq et al. [65] used RF and gene expression programming (GEP) to predict the compressive strength of high-strength concrete and investigated the relationship between cement content, coarse and fine aggregate ratio, water and superplasticizer and compressive strength. The final model has a R^2 of 0.96. Feng et al. [66] used the AdaBoost algorithm in an integrated learning algorithm to predict the compressive strength of concrete materials. The predicted model had an average R^2 of 0.952, an average mean absolute percentage error (MAPE) of 11.39% and an average RMSE of 4.856 MPa after 10-fold cross-validation. Table 4 lists specific information on the concrete-strength prediction methods based on machine-learning algorithms, including the machine-learning algorithms used, the number of datasets and the performance results of the predictive models, etc.

Table 4. Concrete-strength prediction methods using machine-learning algorithms.

Author	Algorithm	Data Volume	Results
Lai and Serra [52]	ANN	240	Relative error less than 5%
Kewalramani and Gupta [53]	ANN	864	Maximum error rate 25.69%
Naderpour et al. [54]	ANN	139	$R = 0.8926$, $MSE = 0.004447$
Asteris and Kolovos [55]	ANN	205	$R^2 = 0.919$
Zhu et al. [56]	GA-SVM	24	Maximum relative error 2.42%
Aiyer et al. [57]	SVM	80	$R = 0.94$
Pham et al. [58]	SVM	239	$R^2 = 0.87$, $RMSE = 4.86$, $MAPE = 9.81\%$
Li and Peng [59]	BP, RBF	19	Relative error less than 6%
Gao and Hao [60]	BP-ANN	30	Absolute error less than 5.0%
Ma and Liu [61]	BP	251	Coefficient of variation = 0.112
Wu et al. [63]	RF	56	$R^2 = 0.969$, $RMSE = 0.0149$
Cui et al. [64]	RF	1030	$R^2 = 0.902$, $MAE = 3.761$, $MAPE = 12.807$, $RMSE = 5.342$
Farooq et al. [65]	RF, GEP	357	$R^2 = 0.96$ (RF), $R^2 = 0.9$ (GEP)
Feng et al. [66]	AdaBoost	1030	$R^2 = 0.952$, $MAPE = 11.39\%$, $RMSE = 4.856$

Although more and more research scholars have tried to apply machine-learning algorithms to the field of concrete strength prediction in recent years, in general, research in this direction is still relatively limited and suffers from two main problems. As seen in Table 4, most researchers have collected a small amount of data, at most 1030 datasets.

Another problem is that most of the algorithms used by scholars are learned individually, with less research on integrated learning algorithms with better robustness and prediction accuracy. Therefore, this paper uses the boosting-based GBRT algorithm, one of the ensemble learning algorithms rarely used by scholars, combined with 1030 sets of concrete compressive strength data to build a prediction model and predict the compressive strength of concrete.

3. Compressive Strength Prediction Model for Concrete Based on GBRT Algorithm

The section uses the GBRT algorithm, combined with 1030 sets of concrete compressive strength data, to build a prediction model for the compressive strength of concrete and analyses the results obtained after running the model. It finds that the prediction model can predict the compressive strength of concrete better.

3.1. Introduction to the GBRT Algorithm

The GBRT is a typical ensemble learning boosting algorithm that features a cart regression tree model that weak learners can only use GBRT uses a forward distribution algorithm, the basic principle of which is to select an appropriate decision tree based on the current model and the fitting function to minimise the loss function [67].

Since the GBRT algorithm uses the negative gradient value of the loss function in the current model as an approximation to the residuals in the boosted tree algorithm, a regression tree is fitted. This allows the performance of the GBRT model not to be significantly affected even if additional noise is present in the database.

As can be seen from the name GBRT, the algorithm consists mainly of gradient boosting (GB) and regression trees (RT), which are described separately below.

- GB

FreidMan originally proposed GB in 2000. The core idea of this algorithm is that each tree is learned from the residuals of all previous trees, and the negative gradient value of the loss function in the current model is used as an approximation to the residuals in the boosted tree algorithm as a way to adapt the regression or classification tree.

Based on the idea of GB above, M weak learner models need to be generated iteratively. The predictions of each weak learner model can then be summed up, where each later model $f_{m+1}(x)$ is generated based on the $f_m(x)$ of the previous learning model plus a new weak learner $h_m(x)$, as shown in Equation (1) [68].

$$f_{m+1}(x) = f_m(x) + h_m(x), \quad m \in [1, M] \quad (1)$$

where x is the vector with input variables, and m is the number of iterations.

If the objective function is the mean square error of the regression problem, it is easy to think that the ideal $h_m(x)$ should be able to fit $y - f_m(x)$ exactly, which is residual-based learning. The specific mathematical expression is shown in Equation (2).

$$h_m(x) = y - f_m(x) \quad (2)$$

where y is the target output or the test value of the output.

- RT

Decision trees can be divided into two main categories: regression trees, which are used to classify labelled values, and classification trees, which are used to predict actual values, i.e., classification and regression tree (CART) algorithms [69]. CART divides the feature space into cells. The test data is grouped into a cell according to its characteristics, resulting in the corresponding output [67]. Therefore, GBRT is essentially an iterative regression tree algorithm consisting of multiple regression trees, with the conclusions of all regression trees being aggregated as the final result.

3.1.1. GBRT Algorithm Steps

The steps of the GBRT algorithm can be summarised as follows.

1. Initialization of the weak learner function.

$$f_0(x) = \operatorname{argmin}_c \sum_{i=1}^N L(y_i, c) \quad (3)$$

where x is the vector with input variables, N is the total number of samples in the input training set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, L is the loss function, y_i is the sample's actual value, and c is the root node's class.

2. For the m -th iteration ($m \in [1, M]$):

For each sample (x_i, y_i) ($i \in [1, N]$) in the input dataset, the value of the negative gradient of the loss function in the current model is calculated as an estimate of the residuals.

$$r_{m,i} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)} \quad (4)$$

For $\{(x_1, r_{m,1}), (x_2, r_{m,2}), \dots, (x_i, r_{m,i})\}$, a regression tree is fitted to obtain the i -th leaf node region $R_{m,j}$ of the m -th tree, with j denoting the number of leaf nodes in each tree ($j \in [1, J]$).

For each leaf node of the regression tree, the value of the leaf node region is estimated using a linear search to minimise the loss function and calculate the best-fit value $c_{m,j}$.

$$c_{m,j} = \operatorname{argmin}_c \sum_{x_i \in R_{m,j}} L[y_i, f_{m-1}(x_i) + c], j \in [1, J] \quad (5)$$

Update the learner based on the following formula.

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J c_{m,j} I(x_i \in R_{m,j}) \quad (6)$$

3. After M iterations, the strong learner is finally obtained.

$$f_M(x) = \sum_{m=1}^M \sum_{j=1}^J c_{m,j} I(x_i \in R_{m,j}) \quad (7)$$

where m is the number of iterations ($m \in [1, M]$), j is the number of leaf nodes in each tree ($j \in [1, J]$), $c_{m,j}$ is the optimal fitting value, x_i is the sample in the input dataset ($i \in [1, N]$), and $R_{m,j}$ is the leaf node region of the m -th tree.

3.1.2. Implementation Process of GBRT

The implementation of GBRT can be summarised in the following four steps. Figure 9 clearly illustrates this process [67].

1. Collection and processing of experimental data, including data normalisation, setting up input/output variables and training/test dataset partitioning.
2. The GBRT algorithm combines data from the training set to obtain a preliminary model, and data from the validation set are used to validate the preliminary model and adjust the hyperparameters to improve the algorithm's learning performance and obtain the final predictive model.
3. Test the performance of the trained prediction model with the test dataset.
4. Apply the predictive model to a real problem.

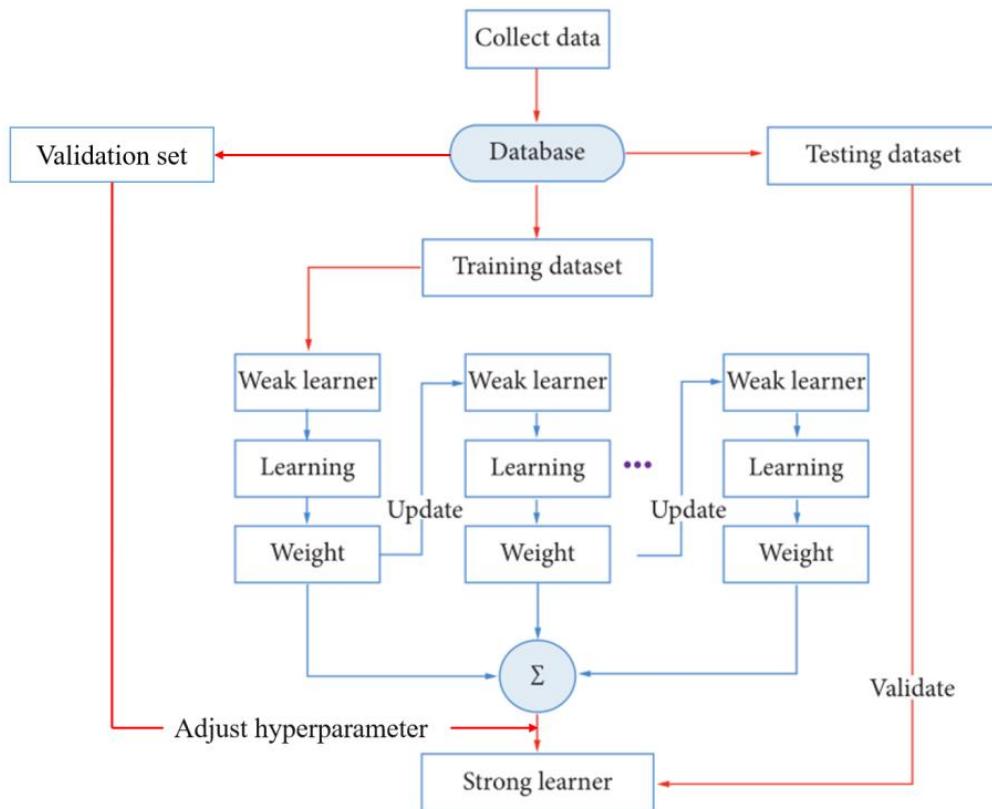


Figure 9. Flowchart of the implementation process of GBRT [67].

3.1.3. Advantages and Disadvantages of the GBRT Algorithm

According to the researcher's summary, the advantages and disadvantages of the GBRT algorithm are as follows.

Advantages

1. Ability to handle mixed data types, including continuous and discrete values, flexibly;
2. High predictive power;
3. Good robustness benefits from a strong loss function, including least squares, least absolute deviation function, Huber and quantile in the case of outliers in the output space.

Disadvantages

4. Poor scalability. Parallel training data is challenging due to dependencies between weak learners.

3.2. Datasets

In order to ensure the accuracy of the prediction model, a large amount of experimental data on the compressive strength of concrete is required to train and test the model [66,70]. In this paper, a dataset containing 1030 sets of concrete compressive strength test data was used as the data source, which was obtained in an experiment by a group led by Professor Yi-Zheng Yeh [71,72] at the Chinese University of Taiwan and later donated free of charge to the Machine Learning Laboratory at the University of California, Irvine. Professor Yi-Zheng Yeh and his team fabricated cylindrical concrete specimens with a height of 150 mm and subjected them to classical compressive tests after a period of standard curing, during which the following nine parameters were collected: cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, age, and compressive strength of concrete. In order to understand these nine parameters more intuitively, this paper uses descriptive statistics for the data and draws box plots corresponding to these nine parameters, as shown in Figure 10, where the top horizontal line in each graph

is the maximum value, the bottom horizontal line represents the minimum value, the middle horizontal line is the median, and then the hollow square in the middle represents the mean value. In addition, with the distribution fitting function in Origin software, histograms of each parameter were plotted, which not only reflected the distribution of the involved parameters but also fitted the corresponding normal distribution curves, as shown in Figure 11.

To ensure the generalisation ability of the prediction model, eight parameters (cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate and age) were taken as the input parameters of the model and the compressive strength of concrete was taken as the output parameter of the model, which means the relationship between the eight variables and the dependent variable of concrete compressive strength was considered simultaneously, as shown in Table 5.

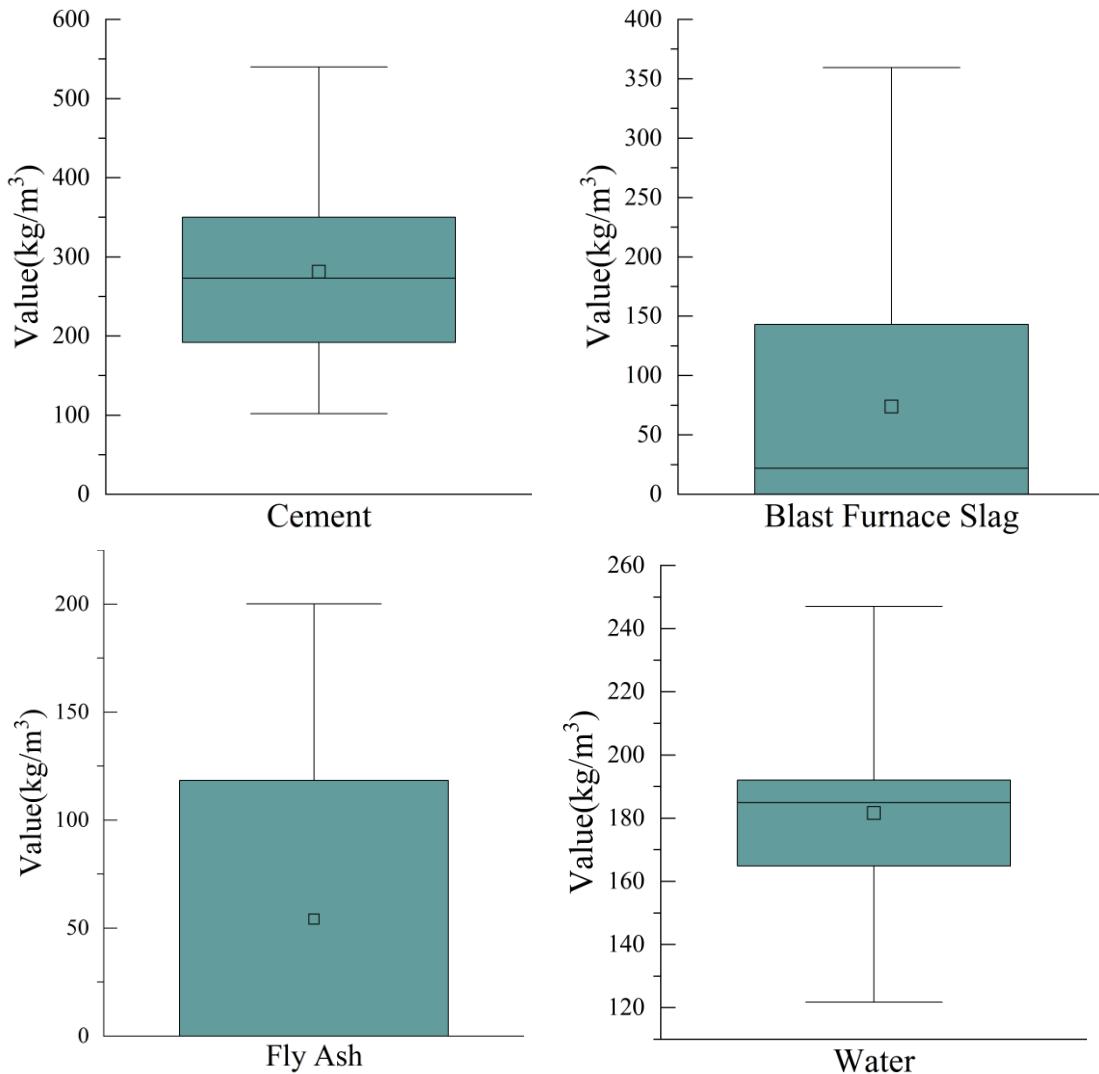


Figure 10. Cont.

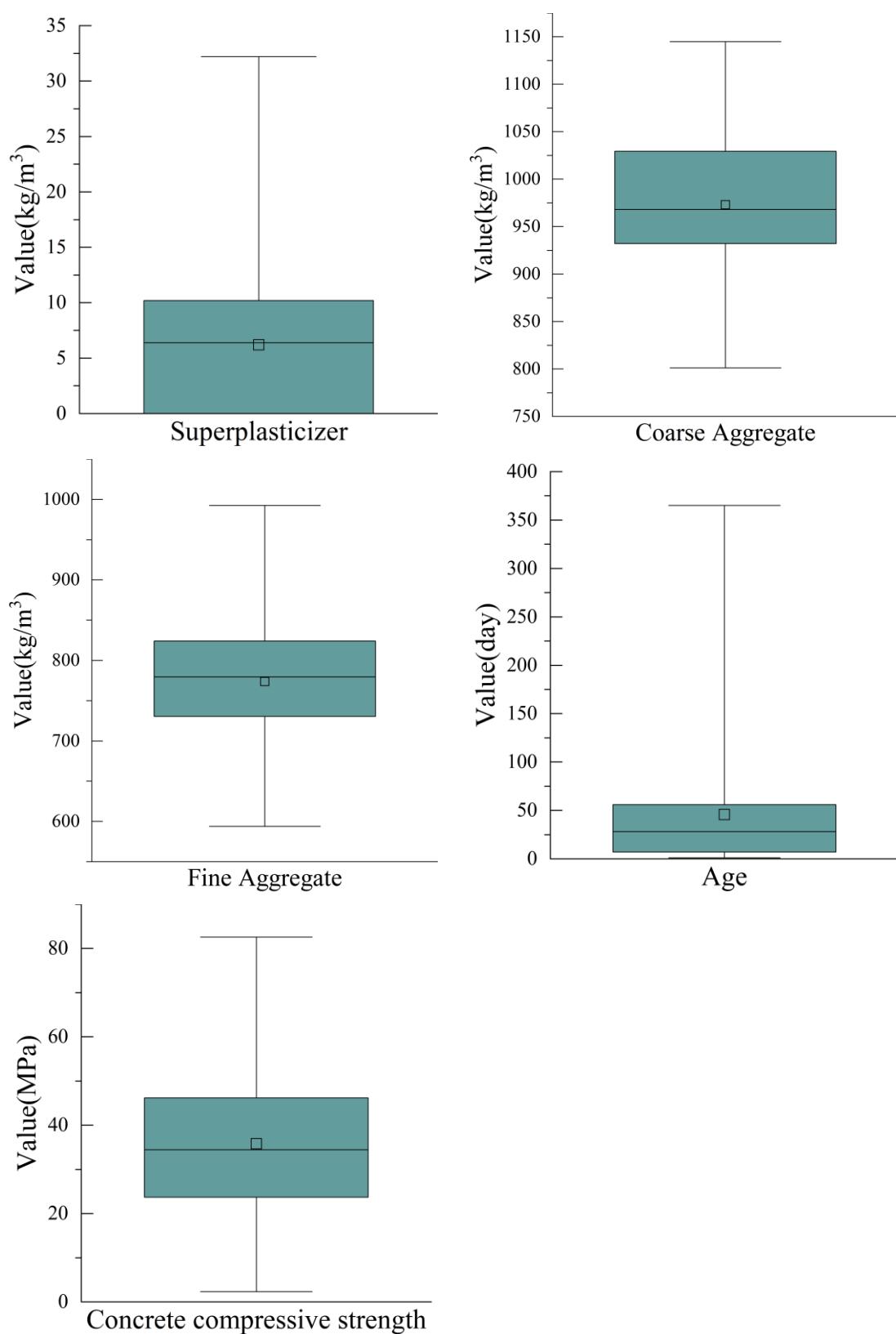
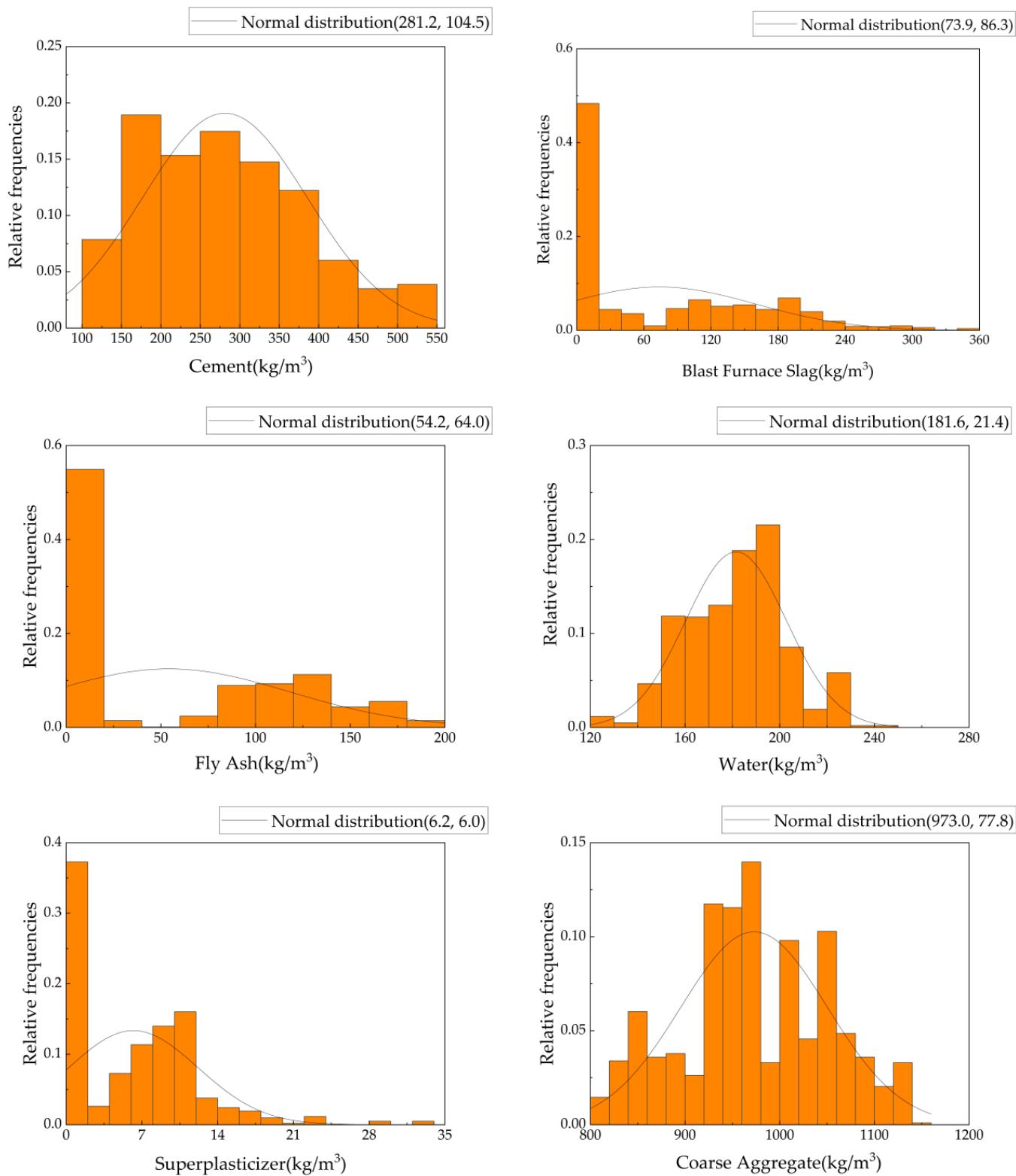


Figure 10. Range of values for nine parameters.

**Figure 11. Cont.**

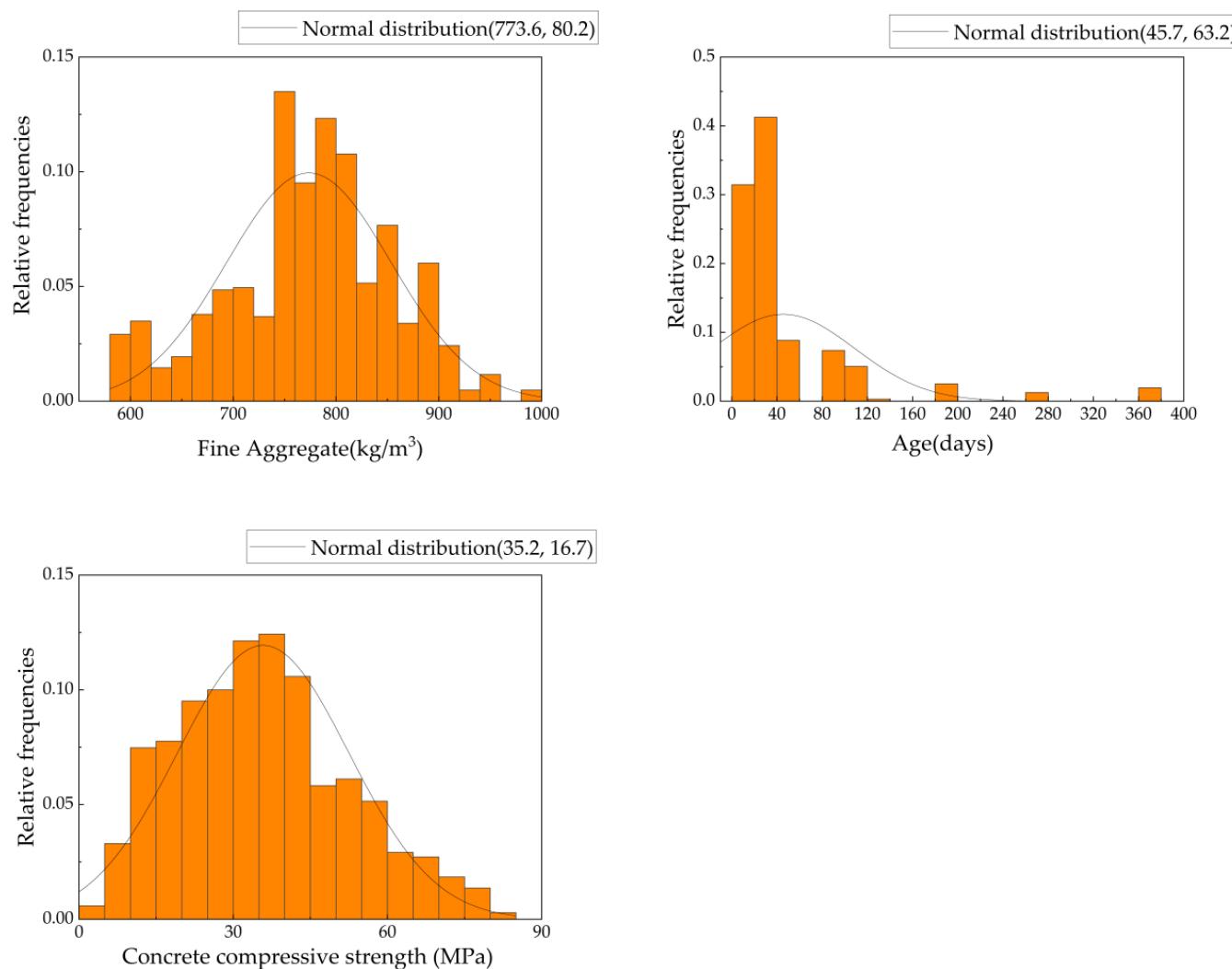


Figure 11. Histogram of nine parameters.

Table 5. Numerical characteristics of the parameters.

Parameter	Range	Mean	Variance	Standard Deviation	Type
Cement (kg/m^3)	102.0–540.0	281.2	10911.1	104.5	Input
Blast Furnace Slag (kg/m^3)	0.0–359.4	73.9	7436.9	86.2	Input
Fly Ash (kg/m^3)	0.0–200.1	54.2	4091.6	64.0	Input
Water (kg/m^3)	121.8–247.0	181.6	455.6	21.3	Input
Superplasticizer (kg/m^3)	0.0–32.2	6.2	35.6	6.0	Input
Coarse Aggregate (kg/m^3)	801.0–1145.0	972.9	6039.8	77.7	Input
Fine Aggregate (kg/m^3)	594.0–992.6	773.6	6421.9	80.1	Input
Age (days)	1–365	45.7	3986.6	63.1	Input
Concrete compressive strength (MPa)	2.3–82.6	35.8	278.8	16.7	Output

3.3. Model Building

There are eight input parameters, each with a different physical meaning. In order to avoid small absolute errors in the minor part of the parameters and significant absolute

errors in the larger part, the input and output parameters need to be normalised, a process whose mathematical formula is shown in Equation (8).

$$X_1 = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (8)$$

where X is the data to be processed, X_{max} is the maximum value in the data sequence, X_{min} is the minimum value in the data sequence, X_1 is the normalised data. To build the GBRT prediction model, this manuscript uses the most classical way of dividing the database, which means 60% of the entire experimental dataset is used as the training set, 20% of the dataset is used as the validation set, and the remaining 20% of the data is used as the test set. Figure 12 shows in detail how the database partitioning was done and how it was used. As the three concepts of the training set, validation set and test set are essential and easily confused, they will be sorted out and explained in detail below [66]. Table 6 summarises the role of these three concepts and the specific amount of data used in the text.

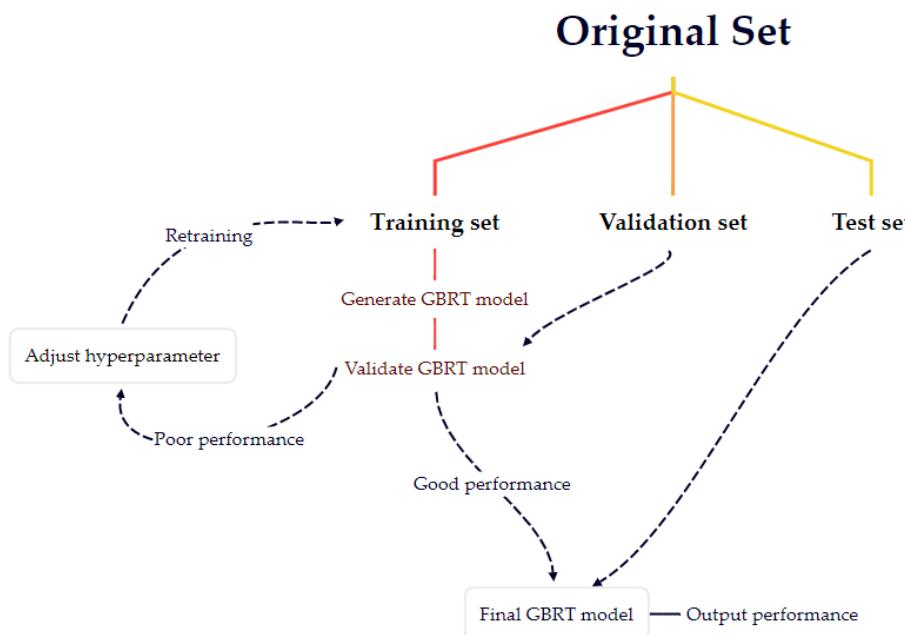


Figure 12. Database division.

Table 6. Summary of the three concepts.

Data Set Type	Role	Data Volume
Training set	Training and generating models	618 (60%)
Validation set	Adjusting hyperparameters & preventing overfitting	206 (20%)
Test set	Evaluating model performance	206 (20%)

- Training set

A set of examples is used to train a model, which is a collection of data samples that the model fits, for example, by training to fit some parameters to build a regressor. For this paper, the data from the training set was first used to train a weak learner, which was eventually integrated to produce a strong learner as the GBRT prediction model.

- Validation set

A separate set of data set aside during model training to determine the network structure or parameters that control the complexity of the model, the validation set provides

an initial assessment of the model's capabilities. During iterative model training, it is often used to tune the parameters of the classifier (regressor) to improve the final model's performance and prevent overfitting of the model.

In this paper, the data from the validation set is used during model training to tune the hyperparameters of the GBRT prediction model while ensuring that the model is not over or under-fitted. The validation set further improves the model's prediction performance and generalisation capability.

- Test set

A dataset that can only be used to evaluate how good a model is, i.e., to assess how well the final model performs. The test set is not involved in training and is primarily used to test the accuracy capability of the trained model, etc., but cannot be used as a basis for algorithm-related choices such as tuning parameters, selecting features, etc.

For this paper, the data from the test set was used to test the accuracy of the trained GBRT model in predicting the compressive strength of concrete.

In order to accurately and objectively evaluate the performance of the GBRT model when predicting the compressive strength of concrete, three commonly used metrics for assessing the performance of machine-learning models were introduced: R^2 , MSE and RMSE. R^2 reflects the degree of linear correlation between a model's predicted and actual data values. Generally, a model is considered valid when its R^2 value is more significant than 0.8 and accurate. The closer the R^2 value is to 1, the higher the model's prediction accuracy [73]. At the same time, MSE and RMSE demonstrate the deviation between the predicted and tested values. The smaller the MSE and RMSE values, the higher the model's prediction accuracy. The mathematical formulae for these three evaluation metrics are shown in Equations (9) to (11).

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - X_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (9)$$

$$MSE = \frac{\sum_{i=1}^n (Y_i - X_i)^2}{n} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - X_i)^2}{n}} \quad (11)$$

where Y_i is the predicted concrete compressive strength model value, X_i is the actual concrete compressive strength value, \bar{X} is the mean of all the true concrete compressive strength values, and n is the total number of samples in the dataset.

3.4. Results and Analysis

Figures 13–15 show the relationship between the predicted values of concrete compressive strength output by the GBRT model and the true values of concrete compressive strength in the validation and test groups. Each of the blue dots in Figures 13 and 14 represents a coordinate where the horizontal coordinate is the true value in the validation and test sets, and the vertical coordinate is the corresponding predicted value output by the GBRT model. It can be noted that the predicted and true values show a highly linear relationship. The orange dashed lines plotted through these circles are the fitted lines with fitted equations of $y = 0.943x + 2.391$ (for validation set) and $y = 0.934x + 2.565$ (for test set) respectively, both very close to the ideal $y = x$ with very little dispersion. This demonstrates that the predicted values obtained when predicting the compressive strength of concrete using the GBRT model are very close to the true values of the compressive strength of concrete, which is also verified in Figure 15.

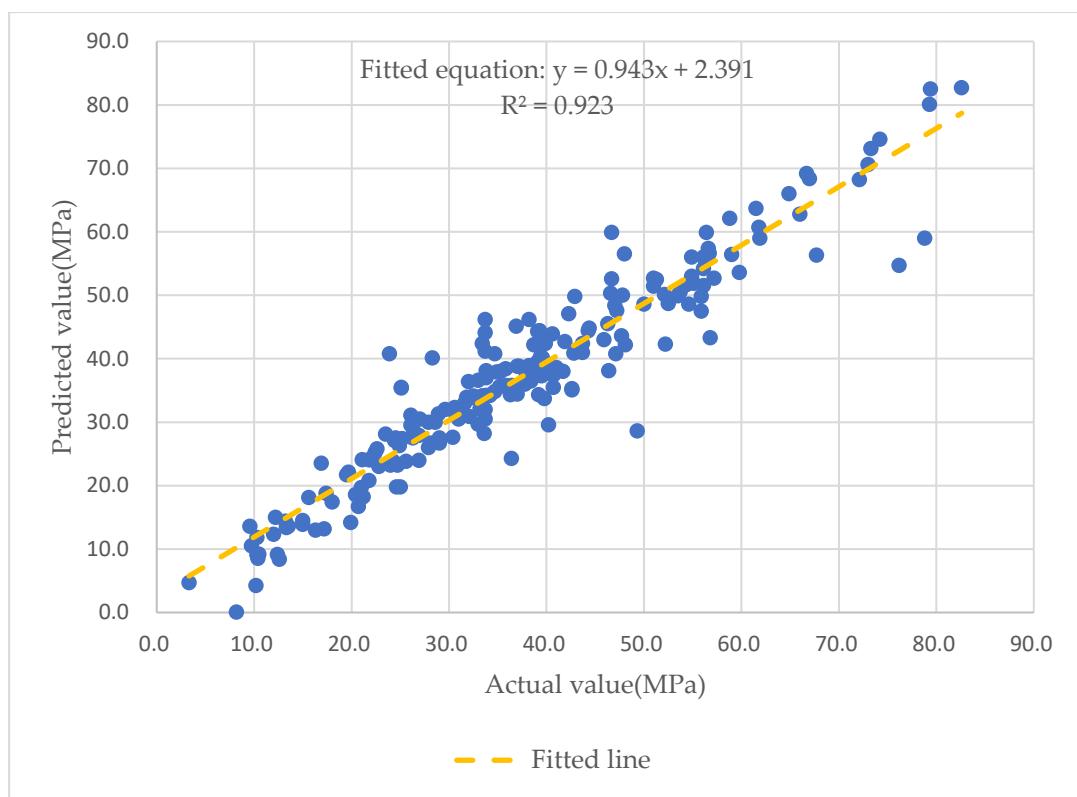


Figure 13. Relationship between tested and predicted compressive strength in the validation set.

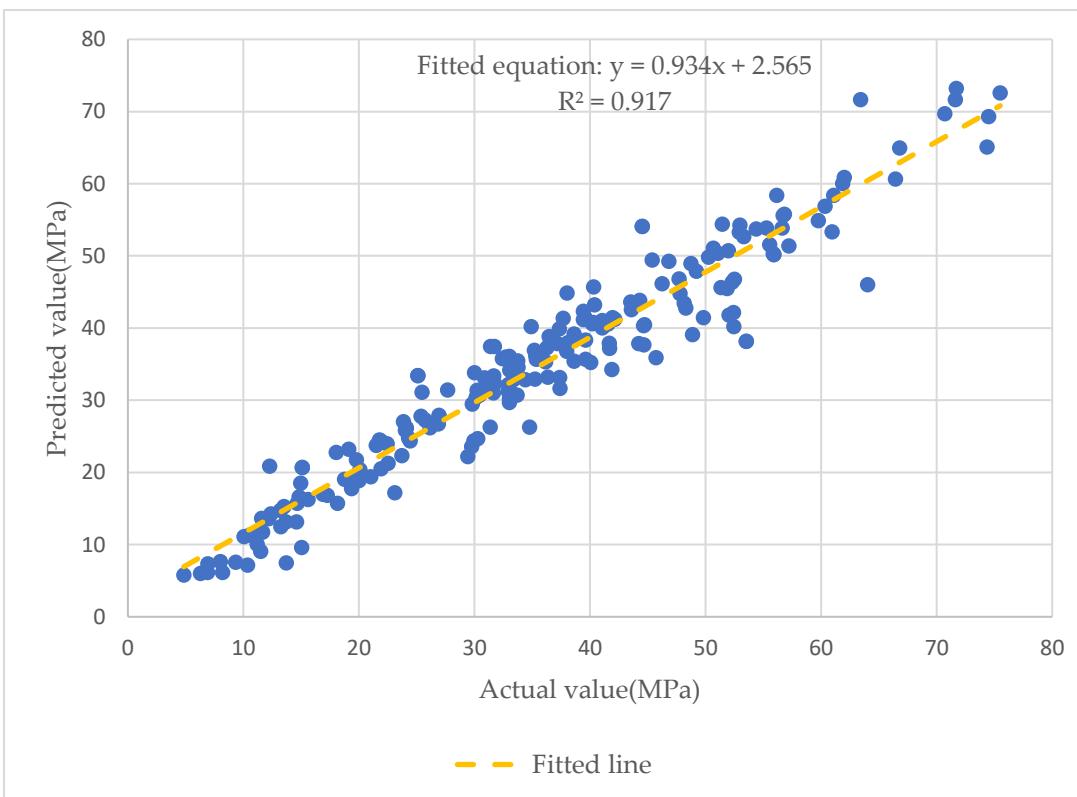


Figure 14. Relationship between tested and predicted compressive strength in the test set.

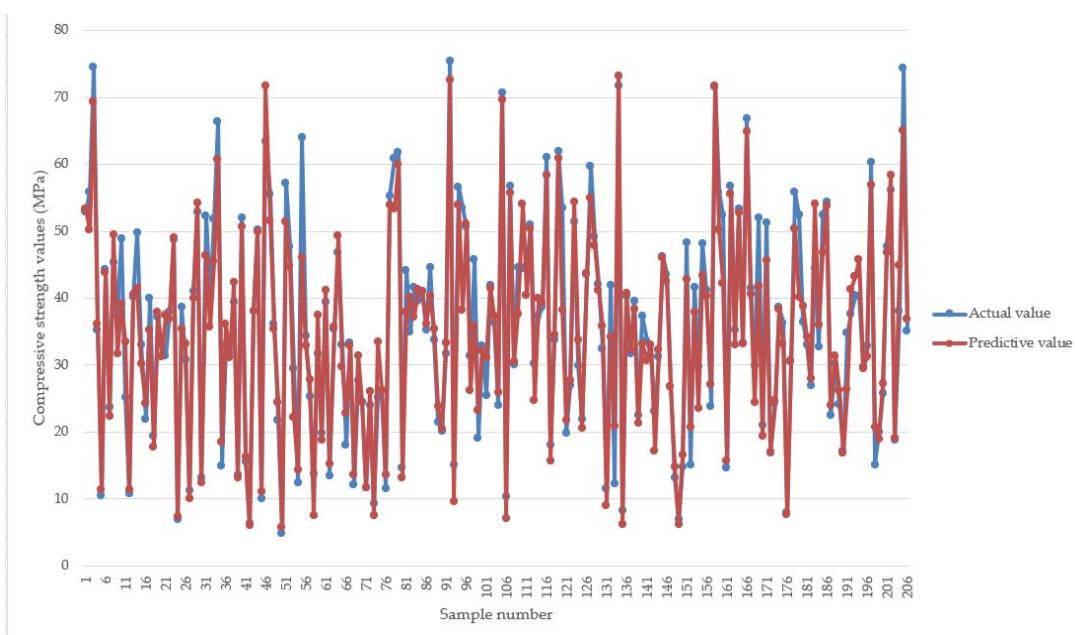


Figure 15. Scatter plot analysis of predicted and tested values in the test set.

For the testing data, the performance evaluation metrics for the GBRT model were R^2 of 0.92, MSE of 22.09 MPa and RMSE of 4.7 MPa. To more accurately and objectively evaluate the performance of the GBRT model, it was necessary to compare the results of the GBRT model with the same database used, which ensured that the data volume variable would not impact the comparison results. The evaluation metrics output by the model were compared with the evaluation metrics output by machine-learning models constructed by previous researchers for predicting the compressive strength of concrete.

3.4.1. Comparison with Individual Machine-Learning Algorithms

To more accurately demonstrate the superiority of GBRT, an ensemble learning method, over individual learning methods in predicting the compressive strength of concrete, the GBRT algorithm was compared with the widely adopted, well-known individual learning methods ANN and SVM. Many research scholars in previous studies used the ANN and SVM algorithms to predict the compressive strength for the same 1030 sets of data as the database in this paper. In Table 7, it is evident that the GBRT model significantly outperforms the individual learning methods, with an increase in R^2 from 0.86 to 0.92 and a decrease in RMSE from 6.28 MPa to 4.7 MPa, a significant improvement in prediction accuracy. The reason for this may be that the GBRT ensemble learning algorithm integrates several weak learners generated by the individual learning algorithm, in which weak learners that perform well will receive higher weights, and weak learners that perform poorly will receive lower weights.

Table 7. Comparison with individual machine-learning algorithms with the same dataset.

Algorithm	Data Volume	Evaluation Indicators		Refs.
		R^2	RMSE(MPa)	
GBRT	1030	0.92	4.70	This paper
ANN	1030	0.90	5.14	[66]
ANN	1030	0.91	5.03	[74]
ANN	1030	0.91	5.57	[75]
SVM	1030	0.89	5.62	[74]
SVM	1030	0.86	6.28	[66]

3.4.2. Comparison with Other Ensemble Machine-Learning Algorithms

In Table 8, the performance of the GBRT model is compared with the performance of models presented in other papers using the same dataset and different ensemble machine-learning algorithms. The GBRT model proposed in this paper outperforms the RF model proposed by Cui et al. [64]. Compared to the AdaBoost model constructed by Feng et al. [66]. Although the AdaBoost model performs better in terms of R^2 values, the GBRT model can be found to perform better when the model performance is evaluated based on MSE and RMSE values. This indicates that the GBRT integrated algorithm can learn well from the data of the training set and can predict the compressive strength of concrete with higher accuracy at the end of the training compared with other ensemble algorithms.

Table 8. Comparison with other ensemble learning algorithm models with the same dataset.

Algorithm	Data Volume	Evaluation Indicators		Refs.
		R^2	RMSE(MPa)	
GBRT	1030	0.92	4.70	This paper
RF	1030	0.90	5.34	[64]
AdaBoost	1030	0.95	4.86	[66]

3.4.3. K-Fold Cross Validation Analysis

K-fold cross-validation is often used to minimise the bias associated with a random sampling of the training dataset [51]. This paper used a five-fold cross-validation approach to further validate the performance and generalisation ability of the GBRT prediction model. The experimental data samples were equally divided into five subsets, four of which were used to construct the strong learner to form the final prediction model, and the remaining subset was used to validate the model. Detailed statistical information on the results obtained from these five operations is given in Table 9. The mean value of R^2 is 0.906, and the mean RMSE value is 4.875 MPa which is relatively small compared to the mean compressive strength value of 35.8 MPa. These values indicate that the GBRT model has a small prediction error. In addition, Figure 16 shows the model performance evaluation metrics for each fold. It can be seen that although there are some fluctuations in the results for the five folds, they all maintain a high level of accuracy.

Table 9. Five-fold cross validation results.

Number of Folds	Evaluation Indicators	
	R^2	RMSE(MPa)
Fold 1	0.901	4.689
Fold 2	0.891	4.933
Fold 3	0.916	4.567
Fold 4	0.930	4.484
Fold 5	0.890	5.700
Average	0.906	4.875

3.4.4. Analysis of the Importance of the Characteristics of the Input Variables

The GBRT model constructed in this paper is highly accurate in predicting the compressive strength of concrete. However, it is a black box model with a complex detailed mechanism behind it, making it difficult for us to explain the relationship between each input variable and the dependent variable in detail [51]. Although the GBRT model is a black box, we can use the Gini index to calculate the importance of each feature variable on a single tree and then explore the contribution each feature variable makes on each tree, and then take the average and normalise it to calculate the global importance of each feature variable. In this paper, the importance factor for each input parameter was calculated for the case where the effect of eight input variables on the final concrete compressive

strength was considered simultaneously. The specific values are listed in Table 10. To more demonstrate the effect of these input parameters on the compressive strength of concrete in a more visual way, Figure 17 is plotted, from which it can be seen that the influence of age and cement content on compressive strength is dominant, with the two together occupying nearly 70% of importance, which is in line with engineering practice.

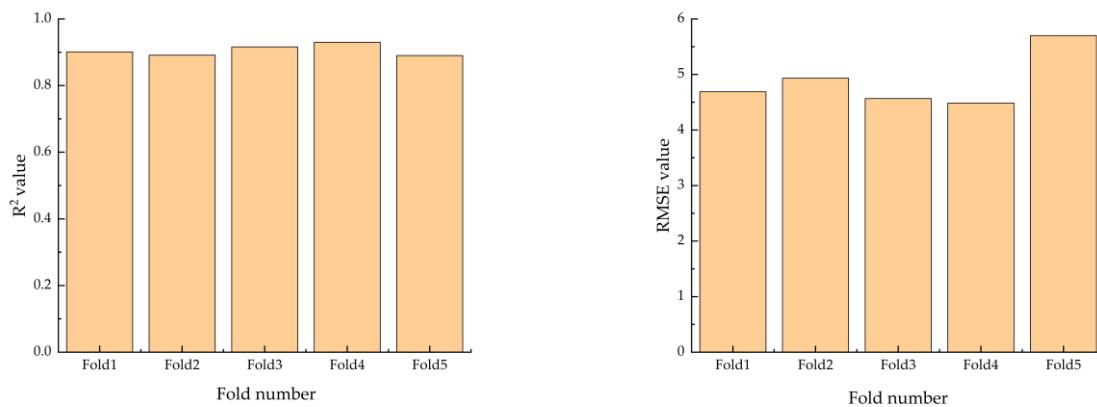


Figure 16. R^2 and RMSE value of five folds.

Table 10. Importance factors for the eight input parameters.

Parameter	Importance Factor
Cement (kg/m^3)	0.3154
Blast Furnace Slag (kg/m^3)	0.0802
Fly Ash (kg/m^3)	0.0121
Water (kg/m^3)	0.1218
Superplasticizer (kg/m^3)	0.0622
Coarse Aggregate (kg/m^3)	0.0157
Fine Aggregate (kg/m^3)	0.0366
Age (days)	0.3560

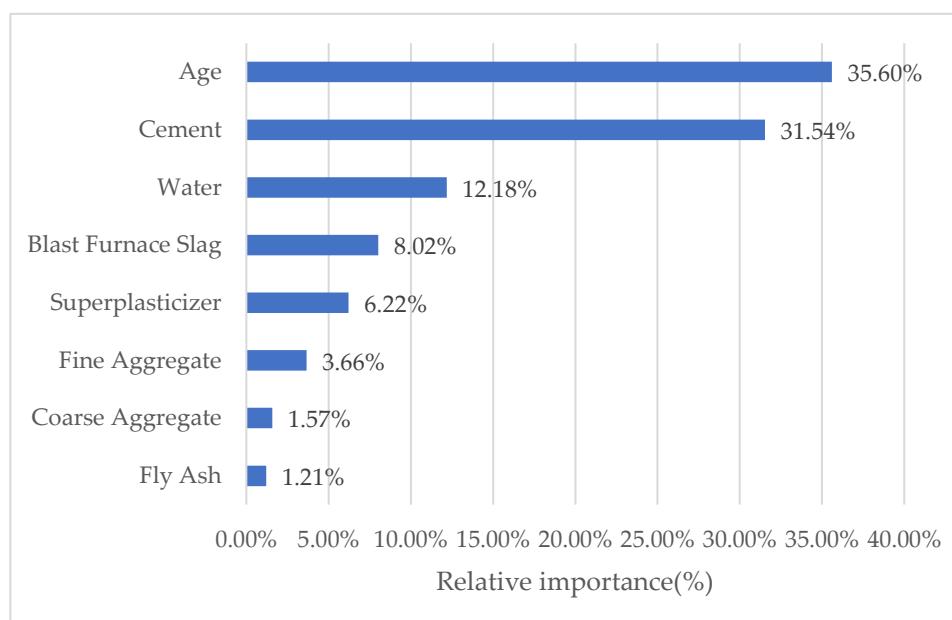


Figure 17. Relative importance of the input parameters.

4. Conclusions

Accurate and rapid prediction of the compressive strength of concrete is of great significance to engineering practice and has become a fashionable research area in recent years. In order to provide a complete understanding of the current state of research in this field, not only is the traditional literature research adopted and a review conducted at the micro level in this paper, but a bibliometric analysis of 3135 papers published in the last decade at the macro level is also conducted by means of CiteSpace software. The following conclusions were obtained.

1. At the macro level. Since 2015, the field has flourished with an increasingly mature theoretical system, driven by pioneers represented by Ali Nazari, a professor at Islamic Azad University, and has given rise to numerous hot research directions related to compressive strength.
2. At the microscopic level. Concrete compressive strength prediction methods are mainly divided into traditional approaches and machine learning. Traditional methods include using the FCT prediction method, summarising empirical mathematical formulas, and using equivalent age theory, etc. Machine-learning methods can be divided into individual learning algorithms, such as ANN and SVM, and ensemble learning algorithms, such as BP and RF.
3. The problems of the small amount of data and few studies of ensemble learning algorithms exist in the current research of using machine-learning algorithms to predict the compressive strength of concrete.

To fill the research gap, this paper uses the GBRT algorithm, an ensemble learning method based on lifting, to predict the compressive strength of concrete materials. 1030 sets of concrete compressive test data were collected while considering the relationship between eight input variables (cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate and age) and one output variable, the compressive strength of concrete. The total dataset was divided into a training set, a validation set and a test set in the ratio of 6:2:2. The prediction model was generated from the training set by the GBRT algorithm and combined with the validation set to adjust the hyperparameters as well as to avoid overfitting, and then evaluated by the testing set. Based on the results, the following conclusions can be drawn.

1. The R^2 of 0.92, MSE of 22.09 MPa and RMSE of 4.7 MPa for the GBRT model prove that the model has high prediction accuracy in predicting the compressive strength of concrete.
2. Using the same database, the GBRT model was compared with prediction models constructed using classical individual learning algorithms such as ANN and SVM from previous work species. The GBRT model performed significantly better than these models. Moreover, the GBRT model has an advantage even when compared with other ensemble learning algorithms such as RF and AdaBoost.
3. The R^2 and RMSE values were calculated for each fold through a five-fold cross-validation analysis, and the model performance was found to be accurate.
4. The importance coefficients of the eight input parameters were calculated by analysing the feature importance, and the effects of age and cement on concrete strength were found to be dominant.

The current study results in the construction of a black box, the GBRT model, where the user does not need to know the detailed mechanism behind the operation of the model but only needs to predict the compressive strength of concrete accurately and effectively given the input variables, which is very simple and convenient for engineers.

As a suggestion for future research, there is a need to enrich the dataset further and increase the amount of data, as well as to consider more factors relating to the compressive strength of concrete to provide data support for predicting the compressive strength of special concrete such as recycled concrete and steel-fibre concrete. In addition, although this study considers the relationship between multiple input factors and compressive strength

simultaneously through the GBRT model, how to extend the model to a multiple output model, such as the simultaneous output of compressive strength and slump, needs further study. Multiple input parameters can calculate the concrete mix ratio, so optimising the concrete mix ratio for a given compressive strength and slump will be an essential research direction in the future. Finally, it is worth mentioning that many previous papers [52–66] simply divided the database into a training set and a test set without setting up a validation set. Although the final model evaluation metrics obtained look good, this is likely to be the result obtained after overfitting the prediction model and is inaccurate. Future research should set a separate validation set as in this paper to further improve the generalisation ability of the prediction model.

Author Contributions: Conceptualisation, D.L. and Z.T.; methodology, D.L. and X.Z.; software, D.L. and Z.T.; validation, D.L., X.Z. and Q.K.; formal analysis, Z.T. and Y.L.; data curation, D.L. and Q.K.; writing—original draft preparation, D.L.; writing—review and editing, X.Z., Z.T., Y.L. and Q.K.; visualization, D.L., Y.L. and Q.K.; supervision, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gong, Z.; Pu, X.; Wu, H. Concrete compressive strength test based on vector machine optimized by particle swarm optimization algorithm. *Concrete* **2013**, *12*, 11–13.
2. Feng, D.; Ren, X.; Li, J. Stochastic damage hysteretic model for concrete based on micromechanical approach. *Int. J. Non-Linear Mech.* **2016**, *83*, 15–25. [[CrossRef](#)]
3. Feng, D.; Li, J. Stochastic nonlinear behavior of reinforced concrete frames. II: Numerical simulation. *J. Struct. Eng.* **2016**, *142*, 04015163. [[CrossRef](#)]
4. Chen, Z.; Feng, Y.; Zhang, X.; Guo, X.; Shao, L.; Cao, Y.; Li, S.; Gao, L. Similarity criterion for the nonlinear thermal analysis of the soil freezing process: Considering the dual effect of nonlinear thermal parameters and boundary conditions. *Acta Geotech.* **2022**, *17*, 5709–5719. [[CrossRef](#)]
5. Bischoff, P.; Perry, S. Compressive behaviour of concrete at high strain rates. *Mater. Struct.* **1991**, *24*, 425–450. [[CrossRef](#)]
6. Lessard, M.; Challal, O.; Aticin, P. Testing high-strength concrete compressive strength. *Mater. J.* **1993**, *90*, 303–307.
7. Shi, H.; Xu, B.; Zhou, X. Influence of mineral admixtures on compressive strength, gas permeability and carbonation of high performance concrete. *Constr. Build. Mater.* **2009**, *23*, 1980–1985. [[CrossRef](#)]
8. Bhanja, S.; Sengupta, B. Investigations on the compressive strength of silica fume concrete using statistical methods. *Cem. Concr. Res.* **2002**, *32*, 1391–1394. [[CrossRef](#)]
9. Bharatkumar, B.; Narayanan, R.; Raghuprasad, B.; Ramachandramurthy, D. Mix proportioning of high performance concrete. *Cem. Concr. Compos.* **2001**, *23*, 71–80. [[CrossRef](#)]
10. Zain, M.F.M.; Abd, S.M. Multiple regression model for compressive strength prediction of high performance concrete. *J. Appl. Sci.* **2009**, *9*, 155–160. [[CrossRef](#)]
11. Zhu, X. Strength prediction of high strength concrete using two nonlinear methods. *Concrete* **2011**, *12*, 28–30.
12. Chen, Z.; Guo, X.; Shao, L.; Li, S.; Tian, X. Design of a three-dimensional earth pressure device and its application in a tailings dam construction simulation experiment. *Acta Geotech.* **2021**, *16*, 2203–2216. [[CrossRef](#)]
13. Zhou, Z. *Ensemble Methods: Foundations and Algorithms*; CRC press: London, UK, 2012.
14. Ababneh, A.; Alhassan, M.; Abu-Haifa, M. Predicting the contribution of recycled aggregate concrete to the shear capacity of beams without transverse reinforcement using artificial neural networks. *Case. Stud. Constr. Mat.* **2020**, *13*, e00414. [[CrossRef](#)]
15. Alshboul, O.; Almasabha, G.; Shehadeh, A.; Mamlook, R.E.A.; Almuflah, A.S.; Almakayeel, N. Machine learning-based model for predicting the shear strength of slender reinforced concrete beams without stirrups. *Buildings* **2022**, *12*, 1166. [[CrossRef](#)]
16. Moein, M.M.; Saradar, A.; Rahmati, K.; Mousavinejad, S.H.G.; Bristow, J.; Aramali, V.; Karakouzian, M. Predictive models for concrete properties using machine learning and deep learning approaches: A review. *J. Build. Eng.* **2022**, *63*, 105444. [[CrossRef](#)]
17. Almasabha, G.; Alshboul, O.; Shehadeh, A.; Almuflah, A.S. Machine Learning Algorithm for Shear Strength Prediction of Short Links for Steel Buildings. *Buildings* **2022**, *12*, 775. [[CrossRef](#)]
18. Mukhtar, F.; Deifalla, A. Shear strength of FRP reinforced deep concrete beams without stirrups: Test database and a critical shear crack-based model. *Compos. Struct.* **2022**, *307*, 116636. [[CrossRef](#)]

19. Deng, F.; He, Y.; Zhou, S.; Yu, Y.; Cheng, H.; Wu, X. Compressive strength prediction of recycled concrete based on deep learning. *Constr. Build. Mater.* **2018**, *175*, 562–569. [[CrossRef](#)]
20. Ozturan, M.; Kutlu, B.; Ozturan, T. Comparison of concrete strength prediction techniques with artificial neural network approach. *Build. Res. J.* **2008**, *56*, 23–36.
21. Tayfur, G.; Erdem, T.K.; Kirca, Ö. Strength prediction of high-strength concrete by fuzzy logic and artificial neural networks. *J. Mater. Civ. Eng.* **2014**, *26*, 04014079. [[CrossRef](#)]
22. Younis, K.H.; Pilakoutas, K. Strength prediction model and methods for improving recycled aggregate concrete. *Constr. Build. Mater.* **2013**, *49*, 688–701. [[CrossRef](#)]
23. Amini, K.; Jalalpour, M.; Delatte, N. Advancing concrete strength prediction using non-destructive testing: Development and verification of a generalizable model. *Constr. Build. Mater.* **2016**, *102*, 762–768. [[CrossRef](#)]
24. Chen, Z.; Guo, X.; Shao, L.; Li, S.; Gao, L. Sensitivity analysis of the frozen soil nonlinear latent heat and its precise transformation method. *Geophys. J. Int.* **2022**, *228*, 240–249. [[CrossRef](#)]
25. Salehi, H.; Burgueño, R. Emerging artificial intelligence methods in structural engineering. *Eng. Struct.* **2018**, *171*, 170–189. [[CrossRef](#)]
26. Feng, D.; Wang, Z.; Wu, G. Progressive collapse performance analysis of precast reinforced concrete structures. *Struct. Des. Tall Spec. Build.* **2019**, *28*, e1588. [[CrossRef](#)]
27. Feng, D.; Ren, X.; Li, J. Softened damage-plasticity model for analysis of cracked reinforced concrete structures. *J. Struct. Eng.* **2018**, *144*, 04018044. [[CrossRef](#)]
28. Chithra, S.; Kumar, S.S.; Chinnaraju, K.; Ashmita, F.A. A comparative study on the compressive strength prediction models for High Performance Concrete containing nano silica and copper slag using regression analysis and Artificial Neural Networks. *Constr. Build. Mater.* **2016**, *114*, 528–535. [[CrossRef](#)]
29. Ayat, H.; Kellouche, Y.; Ghrici, M.; Boukhatem, B. Compressive strength prediction of limestone filler concrete using artificial neural networks. *Adv. Comput. Des.* **2018**, *3*, 289–302.
30. Nguyen, H.; Vu, T.; Vo, T.P.; Thai, H.-T. Efficient machine learning models for prediction of concrete strengths. *Constr. Build. Mater.* **2021**, *266*, 120950. [[CrossRef](#)]
31. Kumar, A.; Arora, H.C.; Kapoor, N.R.; Mohammed, M.A.; Kumar, K.; Majumdar, A.; Thinnukool, O. Compressive strength prediction of lightweight concrete: Machine learning models. *Sustainability* **2022**, *14*, 2404. [[CrossRef](#)]
32. Ashrafian, A.; Amiri, M.J.T.; Rezaie-Balf, M.; Ozbakkaloglu, T.; Lotfi-Omran, O. Prediction of compressive strength and ultrasonic pulse velocity of fiber reinforced concrete incorporating nano silica using heuristic regression methods. *Constr. Build. Mater.* **2018**, *190*, 479–494. [[CrossRef](#)]
33. Zhang, J.; Ma, G.; Huang, Y.; Aslani, F.; Nener, B. Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. *Constr. Build. Mater.* **2019**, *210*, 713–719. [[CrossRef](#)]
34. Barkhordari, M.S.; Armaghani, D.J.; Mohammed, A.S.; Ulrich, D.V. Data-Driven Compressive Strength Prediction of Fly Ash Concrete Using Ensemble Learner Algorithms. *Buildings* **2022**, *12*, 132. [[CrossRef](#)]
35. Chen, C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57*, 359–377. [[CrossRef](#)]
36. Lee, Y.; Chen, C.; Tsai, X. Visualizing the Knowledge Domain of Nanoparticle Drug Delivery Technologies: A Scientometric Review. *Appl. Sci.* **2016**, *6*, 11. [[CrossRef](#)]
37. Zhou, J.; Dou, W.; Quan, D. CiteSpace-based analysis of domestic spatial governance research hotspots and frontiers. In Proceedings of the Annual National Planning Conference, Chengdu, China, 25–30 September 2021.
38. Sevilla, J.; Heim, L.; Ho, A.; Besiroglu, T.; Hobbhahn, M.; Villalobos, P. Compute trends across three eras of machine learning. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022.
39. Xiang, X.; Yang, C.; Chen, J. Bibliometric Analysis of Transit-oriented Development Research. *Urban Rapid Rail Transit* **2020**, *33*, 15–21.
40. Liu, Y.; Lebedev, M.; Zhang, Y.; Wang, E.; Li, W.; Liang, J.; Feng, R.; Ma, R. Micro-cleat and permeability evolution of anisotropic coal during directional CO₂ flooding: An in situ micro-CT study. *Nat. Resour. Res.* **2022**, *31*, 2805–2818. [[CrossRef](#)]
41. Safiuddin, M.; Hearn, N. Comparison of ASTM saturation techniques for measuring the permeable porosity of concrete. *Cem. Concr. Res.* **2005**, *35*, 1008–1013. [[CrossRef](#)]
42. Chang, C.; Ho, M.; Song, G.; Mo, Y.-L.; Li, H. A feasibility study of self-heating concrete utilizing carbon nanofiber heating elements. *Smart Mater. Struct.* **2009**, *18*, 127001. [[CrossRef](#)]
43. Ammari, M.; Belhadj, B.; Bederina, M.; Ferhat, A.; Quéneudec, M. Contribution of hybrid fibers on the improvement of sand concrete properties: Barley straws treated with hot water and steel fibers. *Constr. Build. Mater.* **2020**, *233*, 117374. [[CrossRef](#)]
44. Ozkul, M.H. Efficiency of accelerated curing in concrete. *Cem. Concr. Res.* **2001**, *31*, 1351–1357. [[CrossRef](#)]
45. Liu, Y.; Wang, E.; Jiang, C.; Zhang, D.; Li, M.; Yu, B.; Zhao, D. True Triaxial Experimental Study of Anisotropic Mechanical Behavior and Permeability Evolution of Initially Fractured Coal. *Nat. Resour. Res.* **2023**, *1*–19. [[CrossRef](#)]
46. Liu, J.; Zhou, B.; Qu, H. Application on Inspecting Technique of Fresh Concrete Quality. *J. Jinan Univ.* **2002**, *16*, 251–253.
47. Soh, C.K.; Bhalla, S. Calibration of piezo-impedance transducers for strength prediction and damage assessment of concrete. *Smart Mater. Struct.* **2005**, *14*, 671. [[CrossRef](#)]

48. Leidong, Z. Study on the Performance of Double-combined with Mineral Admixture Concrete and Prediction Model of Compressive Strength. Master's Thesis, Zhejiang University, Hangzhou, China, 13 March 2012.
49. Abd Elaty, M.A.A. Compressive strength prediction of Portland cement concrete with age using a new model. *HBRC J.* **2014**, *10*, 145–155. [[CrossRef](#)]
50. Nambiar, E.; Ramamurthy, K. Models for strength prediction of foam concrete. *Mater. Struct.* **2008**, *41*, 247–254. [[CrossRef](#)]
51. Chou, J.; Tsai, C.; Pham, A.D.; Lu, Y.H. Machine learning in concrete strength simulations: Multi-nation data analytics. *Constr. Build. Mater.* **2014**, *73*, 771–780. [[CrossRef](#)]
52. Lai, S.; Serra, M. Concrete strength prediction by means of neural network. *Constr. Build. Mater.* **1997**, *11*, 93–98. [[CrossRef](#)]
53. Kewalramani, M.A.; Gupta, R. Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks. *Autom. Constr.* **2006**, *15*, 374–379. [[CrossRef](#)]
54. Naderpour, H.; Rafiean, A.H.; Fakharian, P. Compressive strength prediction of environmentally friendly concrete using artificial neural networks. *J. Build. Eng.* **2018**, *16*, 213–219. [[CrossRef](#)]
55. Asteris, P.G.; Kolovos, K.G. Self-compacting concrete strength prediction using surrogate models. *Neural Comput. Appl.* **2019**, *31*, 409–424. [[CrossRef](#)]
56. Zhu, W.; Shi, C.; Li, N. Prediction model for compressive strength of recycled concrete based on genetic algorithm optimized support vector machine. *J. China Foreign Highw.* **2014**, *34*, 311–314.
57. Aiyer, B.G.; Kim, D.; Karingattikkal, N.; Samui, P.; Rao, P.R. Prediction of compressive strength of self-compacting concrete using least square support vector machine and relevance vector machine. *KSCE J. Civ. Eng.* **2014**, *18*, 1753–1758. [[CrossRef](#)]
58. Pham, A.; Hoang, N.; Nguyen, Q. Predicting compressive strength of high-performance concrete using metaheuristic-optimized least squares support vector regression. *J. Comput. Civ. Eng.* **2016**, *30*, 06015002. [[CrossRef](#)]
59. Li, H.; Peng, T. Prediction of Concrete Compression Strength Based on BP and RBF Neural Network Theories. *J. Wuhan Univ. Technol.* **2009**, *31*, 33–36.
60. Gao, F.; Hao, Q. Concrete Compression Strength Prediction based on Matlab7.2 Neural Network Toolbox. *J. Shanxi Datong Univ.* **2012**, *28*, 60–62, 96.
61. Ma, G.; Liu, K. Prediction of Compressive Strength of CFRP-confined Concrete Columns Based on BP Neural Network. *J. Hunan Univ.* **2021**, *48*, 88–97.
62. Wu, X.; Kumar, V.; Ross Quinlan, J.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [[CrossRef](#)]
63. Wu, X.; Liu, P.; Chen, H.; Zeng, T.; Xu, W. Characteristic screening and prediction of high-performance concrete compressive strength based on random forest method. *Concrete* **2022**, *01*, 17–20, 24.
64. Cui, X.; Wang, Q.; Zhang, R.; Dai, J.; Xie, C. Prediction of Compressive Strength of Concrete Based on Random Forests. *J. Lanzhou Jiaotong Univ.* **2021**, *40*, 1–6, 14.
65. Farooq, F.; Nasir Amin, M.; Khan, K.; Rehan Sadiq, M.; Javed, M.F.; Aslam, F.; Alyousef, R. A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC). *Appl. Sci.* **2020**, *10*, 7330. [[CrossRef](#)]
66. Feng, D.; Liu, Z.; Wang, X.; Chen, Y.; Chang, J.; Wei, D.; Jiang, Z. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Constr. Build. Mater.* **2020**, *230*, 117000. [[CrossRef](#)]
67. Feng, D.; Fu, B. Shear strength of internal reinforced concrete beam-column joints: Intelligent modeling approach and sensitivity analysis. *Adv. Civ. Eng.* **2020**, *2020*, 8850417. [[CrossRef](#)]
68. Guelman, L. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Syst. Appl.* **2012**, *39*, 3659–3667. [[CrossRef](#)]
69. Razi, M.A.; Athappilly, K. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Syst. Appl.* **2005**, *29*, 65–74. [[CrossRef](#)]
70. Sun, X.; Luo, T.; Wang, L.; Wang, H.; Song, Y.; Li, Y. Numerical simulation of gas recovery from a low-permeability hydrate reservoir by depressurization. *Appl. Energy* **2019**, *250*, 7–18. [[CrossRef](#)]
71. Yeh, I.C. Modeling of strength of high-performance concrete using artificial neural networks. *Cem. Concr. Res.* **1998**, *28*, 1797–1808. [[CrossRef](#)]
72. Yeh, I.C. Modeling slump of concrete with fly ash and superplasticizer. *Comput. Concr.* **2008**, *5*, 559–572. [[CrossRef](#)]
73. Gandomi, A.H.; Babanajad, S.K.; Alavi, A.H.; Farnam, Y. Novel approach to strength modeling of concrete under triaxial compression. *J. Mater. Civ. Eng.* **2012**, *24*, 1132–1143. [[CrossRef](#)]
74. Chou, J.; Chiu, C.; Farfoura, M.; Al-Taharwa, I. Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. *J. Comput. Civ. Eng.* **2011**, *25*, 242–253. [[CrossRef](#)]
75. Erdal, H.I.; Karakurt, O.; Namli, E. High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform. *Eng. Appl. Artif. Intel.* **2013**, *26*, 1246–1254. [[CrossRef](#)]