buildings

MDPI

Review

Research Status and Prospect of Machine Learning in Material Performance Prediction: A Case Study of Conventional and Solid Waste Concrete

Qingbiao Wang, Ming Fang, Zimo Wang *, Junjie Zhang and Rui Wang

School of Transportation Engineering, Shandong Jianzhu University, Jinan 250101, China
* Correspondence: 2025155133@stu.sdjzu.edu.cn

Abstract

Accelerated global urbanization places multiple pressures on the cement concrete industry, including resource consumption, carbon emissions, and solid waste disposal. Converting solid waste into construction materials offers a viable path towards its large-scale, high-value utilization, driving the green, low-carbon transition of the construction sector. However, solid waste concrete exhibits high-dimensional, multivariable, and strongly nonlinear characteristics owing to coupled factors such as waste type, source, substitution rate, mix proportions, and curing conditions. Consequently, traditional empirical methods fail to meet engineering needs, particularly in efficient mix design, performance prediction, and quality control. In conventional concrete research, predictive modeling has evolved from classical algorithms to ensemble and deep learning frameworks, often optimized via Bayesian optimization or genetic algorithms. Post-hoc interpretability methods such as SHAP and PDP are widely applied to accurately predict compressive strength, identify critical factors, reveal micro-mechanisms, and reconstruct empirical equations. In contrast, research on solid waste concrete has focused heavily on data challenges stemming from varied waste sources, including data heterogeneity, small sample sizes, and noise. In this field, swarm intelligence algorithms (such as particle swarm, grey wolf, and firefly algorithms) are commonly combined with ensemble models to improve prediction accuracy and engineering usability. However, limitations remain in data availability, feature consistency, cross-material transferability, and practical engineering application. Future research should leverage transfer learning, physics-informed neural networks (PINNs), and in-situ monitoring to advance models from standalone prediction tools into transferable, interpretable decision-support systems.

Keywords: concrete; solid waste concrete; machine learning; data-driven; performance prediction; intelligent design

Check for updates

Academic Editor: Salvatore Verre

Received: 15 September 2026
Revised: 4 October 2026
Accepted: 22 October 2026
Published: 26 October 2026

Copyright: © 2026 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.

1. Introduction

The expansion of modern engineering construction is inseparable from the support of cement concrete [1]. With the continuous development of population and cities, the demand for production is still huge in the future. This trend has also made the problem of resource consumption and carbon emissions more prominent [2–4]. International Energy Agency (IEA) data show that the carbon dioxide emissions of the cement industry have accounted for about 8% of the total anthropogenic carbon emissions worldwide, becoming the third largest industrial emission source after the energy and transportation fields [5]. Under

the goal of "carbon peak-carbon neutrality," the cement industry is considered to be a key industry with both emission reduction potential and difficulty because of its high emission intensity, difficult substitution, and strong demand rigidity [6,7]. Concrete Zero Global Policy Priorities 2026 released in February 2026 for the concrete initiative governments and standard-setting agencies must abolish the old compositional restrictions and fully shift to performance-based standards. It also calls for high-quality embodied-carbon data and a clear upper limit of carbon emissions in large-scale projects such as transportation [8].

With the deepening of industrialization, the production of solid waste has been at a high level for a long time. Its storage and disposal have become a difficult problem restricting the utilization efficiency of resources and the safety of ecological environment [9,10]. In order to promote resource recycling and green low-carbon transformation, the State Council of China has pointed out the route of solid waste treatment in the next few years in the latest policy documents such as the "Comprehensive solid waste management action plan." and realized the annual utilization of bulk solid waste by 4.5 billion tons by 2030 [11]. From the actual situation, China's industrial solid waste resource utilization still faces structural and technical constraints. How to break through the bottleneck of low utilization value, limited utilization, and insufficient added value of solid waste has become a key issue to realize the green transformation of industry.

Among many disposal pathways, engineering construction materials are considered one of the most practical options. They offer a realistic way to achieve large-scale and high-value utilization of solid waste. Concrete boasts huge consumption, diverse components, and high adaptability to raw materials. These characteristics create vast potential for the synergistic utilization of various solid wastes [12]. From 2024 to the present, the standardization of 'solid waste-based building materials' has been accelerated by global standards bodies. Global standards bodies such as the International Organization for Standardization (ISO) and the American Society for Testing and Materials (ASTM) are also accelerating the development of uniform performance benchmarks for solid waste-based building materials. The preparation of low-carbon concrete with solid waste instead of clinker has become a research hotspot [13,14]. The reasonable design of solid waste concrete can consider multiple performance and emission reduction targets [15–18]. However, the source of solid waste materials is complex, and their physical and chemical properties fluctuate greatly, resulting in a high degree of nonlinear characteristics in the reaction mechanism and performance evolution of solid waste concrete systems. Different types of industrial solid waste have obvious differences in chemical composition, mineral phase structure, and reactivity, and their mechanisms of action in concrete systems are also different [19–23]. With the exponential growth of the mixture ratio variable space, traditional test methods that rely on experience or single-factor control have exposed the problems of high costs, long cycles, the ease with which to fall into local optimum, and the inability to achieve simultaneous optimization of multi-objective performance. It is difficult to support its rational design simply from the material mechanism or empirical parameters [24]. Leveraging machine learning and big data technologies, researchers are building intelligent models that bridge "solid waste composition, microstructure, and macro-performance." This approach is emerging as a frontier in breaking through traditional research frameworks [25,26]. Although big data-driven algorithms show promise in materials research, their application to solid waste concrete remains fragmented. Existing models suffer from limited generalization across diverse material systems and engineering scenarios, while their predictions are often difficult to reconcile with underlying physical mechanisms.

From the perspective of engineering application and decision support, this study systematically reviews the progress of machine learning in traditional and multi-component solid waste concrete. We analyze the scope and limitations of various data-driven models

across performance prediction, mix proportion optimization, key variable identification, and mechanism interpretation. Furthermore, we compare their data requirements, generalization capabilities, and interpretability across different solid waste concrete systems. Based on this, the main bottlenecks in translating these models into practical applications are identified, spanning data foundations, modeling methodologies, and evaluation frameworks. This review provides a theoretical and methodological benchmark for the intelligent design of low-carbon concrete incorporating multi-solid waste, laying the foundation for a decision-support framework featuring interpretability, transferability, and engineering applicability.

Literature Search Strategy: To ensure a systematic and transparent literature collection, relevant studies were searched in the Web of Science, ScienceDirect, and CNKI databases using combinations of keywords related to "machine learning," "artificial intelligence," "concrete," "solid waste," "recycled materials," "prediction," and "mix proportion optimization." The search covered publications available up to September 2026. Studies were included if they focused on the application of machine learning or data-driven methods to concrete performance prediction, mix proportion optimization, durability assessment, or related research. Studies unrelated to concrete or machine learning, review articles, conference papers, and studies lacking sufficient information on the investigated materials or modeling methods were excluded. The retrieved literature was further screened based on titles, abstracts, and full texts, followed by manual screening of the reference lists of relevant studies. The selected studies were subsequently classified according to concrete type, the role of solid waste in concrete, key performance parameters, machine learning models, and optimization strategies. Detailed literature search and screening records can be provided upon request. In this study, CiteSpace (version 6.4.R1) was used to analyze relevant publications from 2016 to 2022 retrieved from the Web of Science (WoS). Figures 1, 7, and 10 correspond to three search queries: (1) "Machine learning"\"Property prediction"\"Materials design"; (2) "Machine learning"\"Concrete"\"Prediction"\"Mechanical properties"; and (3) "Solid waste"\"Concrete"\"Mechanical properties". The parameters were set as follows: the time slice was set to 1 year, and "Keywords" was chosen as the node type. Nodes were selected using the g-index criterion with k = 25. For network pruning, both Pathfinder and Pruning sliced networks were applied, and cluster labels were generated using the Log-Likelihood Ratio (LLR) algorithm.

2. Theoretical Basis and Model Comparison of Machine Learning

In recent years, the application of artificial intelligence (AI) in the field of engineering materials has been increasing. According to the keyword co-occurrence network diagram (Figure 1) of the Web of Science literature in the past five years, it can be found that it has been widely used in the research of machine learning material performance prediction.

By training the existing data, rules can be extracted using machine learning models, and these rules can be used to predict new unknown data. Traditional methods based on empirical formulas or single-factor analysis are inefficient and are no longer suitable for addressing rapidly developing technical needs. Machine learning provides an effective way to solve such complex nonlinear problems in materials science because of its ability to automatically extract potential rules from data [27]. Machine learning methods can be divided into three categories according to the training data forms and feedback mechanisms: supervised learning, unsupervised learning, and reinforcement learning (Figure 2). Supervised learning has a mature theoretical framework and high engineering applicability and is regarded as the most widely used mainstream method in current material performance prediction and optimization designs. This section systematically lists several representative

classical, ensemble learning, neural network, and deep learning algorithms in supervised learning and their internal operating mechanisms.

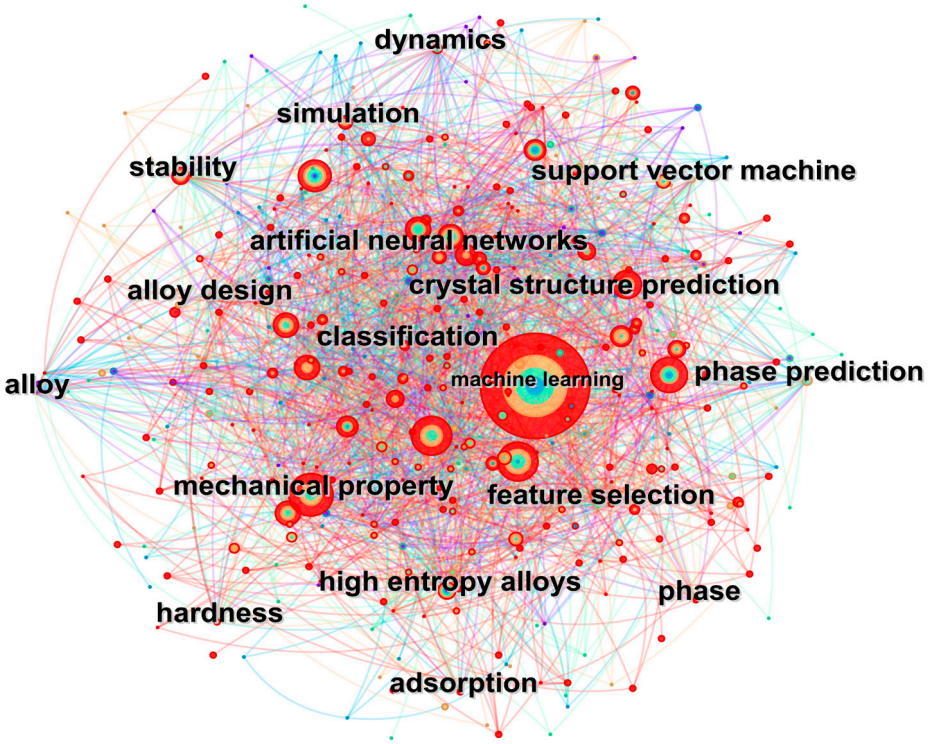


Figure 1. Citespace keyword co-occurrence diagram.

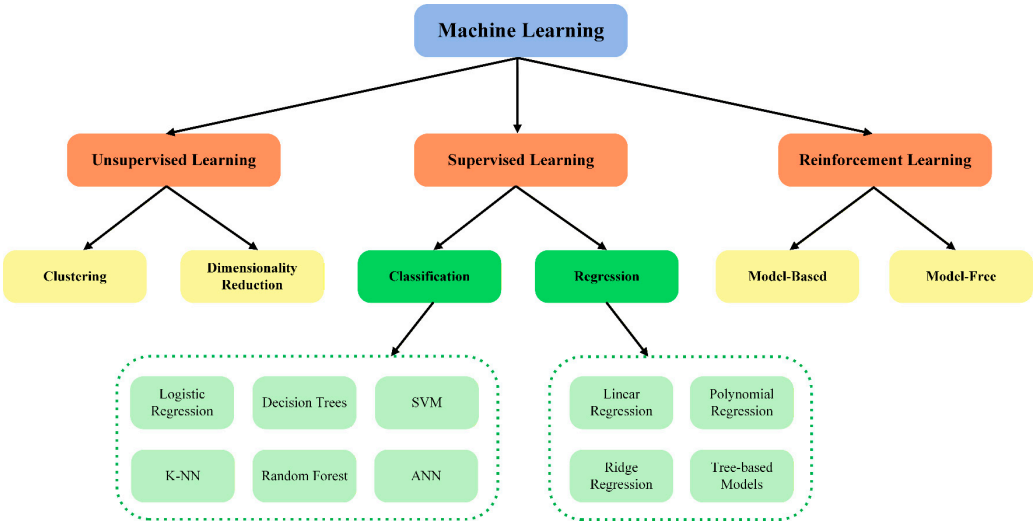


Figure 2. Machine learning classification.

2.1. Classical Machine Learning Model

Classical machine learning is one of the most mature branches in the field of artificial intelligence and has the most comprehensive theoretical system. With complete statistical theory and clear mathematical rules, it can efficiently extract key patterns from limited material feature data and establish a strong explanatory-mapping relationship. This type of model has significant advantages, such as fewer parameters and low computational overhead, which are more in line with the actual computing power and application requirements in the current engineering scenario.

(1) As one of the most representative classical algorithms in supervised learning, linear regression (LR) performs numerical prediction by constructing a linear mapping relationship between independent variables and dependent variables [28,29]. Linear regression is often used to explore the linear correlation between a single factor and the macroscopic properties of a material. However, in the face of significant nonlinear coupling effects in complex material systems, it is difficult to describe high-order interactions using LR. Therefore, it is usually used as a benchmark model to promote the evolution of research on complex nonlinear models [30].

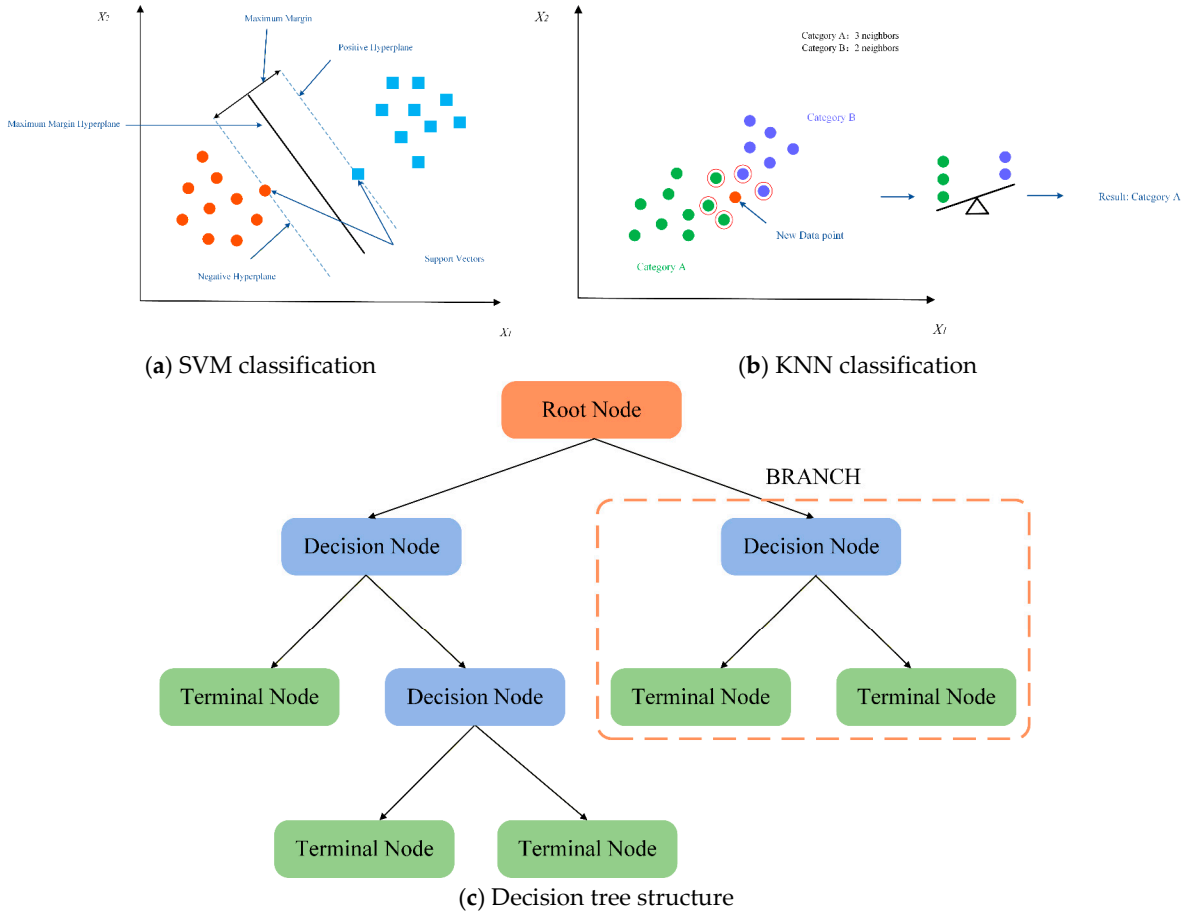
(2) Support vector machine (SVM) is based on statistical learning theory and structural risk minimization (SRM) principle. It maps the nonlinear problem in a low-dimensional space into a high-dimensional feature space through a kernel function and finds an optimal hyperplane that can maximize the distance between categories, so that different categories of data can be effectively distinguished. Figure 3a shows the geometric distribution of SVM for finding the maximum margin hyperplane and its core elements [31–33]. In the classification task, the algorithm is embodied as support vector classification (SVC), and the hyperplane is used to realize the accurate division of different types of samples. In the regression prediction task, the algorithm evolves into support vector regression (SVR), which brings fitting data points by establishing fault tolerance. The complexity of the model is determined by a few support vectors, and the generalization ability and robustness are excellent when dealing with small samples, high dimensions, and noisy data; however, in the face of large-scale samples, the training time cost of the underlying solution is higher.

(3) K-nearest neighbors (KNN) is a non-parametric instance learning method that does not require explicit pre-training. In the face of new data points, the algorithm filters out the nearest K known samples by calculating the spatial distance and then attributes them to the largest category of these K neighbors using 'majority voting', as shown in Figure 3b [33,34]. In the regression task, KNN generates predictions by averaging the output of the neighboring samples. It is logically transparent, does not require distribution assumptions, and is good at nonlinear problems, but the prediction accuracy strongly depends on the K value selection and data standardization. Each prediction requires a global distance calculation, resulting in low efficiency of large-scale data and vulnerability to dimensional disasters in high-dimensional scenarios.

(4) Decision trees (DTs) construct a tree-like structure using root nodes, decision nodes, and branches. Through this framework, the model effectively captures nonlinear relationships between features and targets, as shown in Figure 3c. Although a single decision tree is highly interpretable and computationally efficient, it is easy to overfit when dealing with complex material data; therefore, it is often used as a base learner. It is introduced into ensemble learning frameworks such as random forest (bagging) or gradient boosting tree (Boosting) to suppress prediction fluctuations through hyperparameter tuning and multi-model collaboration, thus exhibiting more robust generalization ability in complex material systems.

The classical machine learning model relies on clear underlying logic and strong physical interpretability. It provides a solid foundation for the numerical prediction of material properties. Real material test data are frequently constrained by complex compositions, high-dimensional features, and nonlinear interactions. Under such conditions, a satisfactory balance between prediction accuracy, computational cost, and generalization cannot be easily achieved by any single model. To break through the inherent performance bottleneck of a single algorithm, the current research path of material prediction has gradually evolved from a single model to multi-model collaboration. In this context, ensemble

learning has been gradually developed in the research on material performance prediction and reverse design.



(**a**) SVM classification (**b**) KNN classification

(**c**) Decision tree structure

Figure 3. Schematic illustrations of SVM, KNN, and decision tree [33,35].

2.2. Ensemble Learning Model

Ensemble learning reconstructs multiple basic 'weak learners' into 'strong learners" with excellent generalization ability through specific collaborative strategies. The prediction bias and variance of a single model are effectively suppressed, showing stronger robustness in the prediction of complex material systems. According to the different ways of generating individual learners, the ensemble framework is mainly divided into two categories: the parallelization method (bagging) represented by random forest (RF); and the serialization method (boosting) represented by AdaBoost, XGBoost, and LightGBM.

(1) Random Forest (RF) is a non-parametric supervised learning algorithm based on ensemble learning, as shown in Figure 4. It is primarily a forest system constructed using multiple decision trees. Its working mechanism is mainly based on Bootstrap Aggregating (Bagging) and random feature selection. The self-help method improves the stability and predictability of the model by constructing multiple independent weak learners in parallel and aggregating the prediction results of each learner into strong learners using the mean value or voting to improve the model stability and predictability [36]. RF can implicitly mine the complex coupling relationship between multiple variables without a preset data distribution and use the built-in feature importance mechanism to quantify the contribution of each material component to performance, providing an intuitive reference for the material design.

(2) AdaBoost (adaptive boosting) is a typical and widely used boosting method. It integrates multiple weak classifiers with slightly better performance than random guessing and constantly adjusts their weights to construct a strong classifier with higher prediction accuracy [37,38]. Unlike the parallel mechanism of random forest, AdaBoost uses a serial iterative framework to dynamically update the dual weights of samples and weak learners and outputs the final prediction results by weighted combination. The model has excellent performance in dealing with nonlinear data, but it is sensitive to outliers and high noise, and over-fitting is easy to cause. At the material application level, AdaBoost can indirectly evaluate the importance of features by weight and quantitatively analyze the contribution of each component to the macroscopic performance, thereby enhancing physical interpretability and guiding mix proportion optimization.

(3) Gradient Boosting Decision Tree (GBDT) is a classical ensemble learning algorithm based on the gradient boosting framework. It constructs a high-precision strong learner by combining multiple weak decision trees [39]. Unlike AdaBoost, which dynamically updates sample weights, GBDT is built upon a forward stagewise algorithm. During serial iterations, each newly constructed tree is trained to fit the negative gradient (i.e., the pseudo-residual) of the loss function from the previous model. The model shows high prediction accuracy when dealing with complex nonlinear mappings and can flexibly handle multidimensional heterogeneous features. By selecting different robust loss functions, its ability to resist outliers and high noise is better than that of AdaBoost. At the material application level, GBDT accurately models nonlinear mappings between complex mix proportions and macroscopic properties, while simultaneously evaluating feature importance. This dual capability deepens the understanding of co-evolution mechanisms among multiphase components, providing reliable guidance for formulating new materials.

(4) XGBoost, LightGBM, and CatBoost represent the three modern engineering variants of GBDT. By introducing targeted innovations to their underlying algorithms, these frameworks further break through the limits of prediction accuracy and computational efficiency. Designed as a deep extension of GBDT, XGBoost optimizes its objective function through two key mechanisms. A second-order Taylor expansion is utilized to accelerate convergence, while a regularization term is explicitly included to rigorously limit tree complexity [40,41]. Through this mechanism, the generalization capability and robustness of the model are significantly enhanced. Consequently, small-sample and highly nonlinear problems in engineering materials can be addressed more effectively. By accurately quantifying the contribution of features, XGBoost successfully overcomes the inherent 'black box' limitation of pure data fitting. Therefore, the prediction process is transformed into a material mechanism analysis with important physical reference values. LightGBM (Lightweight Gradient Boosting) introduces a histogram discretization algorithm and a leaf-wise mechanism, which makes up for the shortcomings of computing power and memory caused by large-scale or multidimensional data sets [42]. This local deep exploration strategy gives the model a stronger fitting depth and extreme computational efficiency under the same iteration scale. However, for the common small-sample characteristics in material tests, this mechanism is easy to cause over-fitting. In actual modeling, strict tree depth constraints and regularization penalties must be applied to achieve a balance between efficiency and accuracy. The main breakthrough of CatBoost lies in the original 'Ordered Boosting' strategy and Oblivious Trees structure. The problem of target leakage and prediction offset often encountered in traditional algorithms is fundamentally solved. In addition, it can also directly analyze those category features

quickly, eliminating the need to deal with cumbersome one-hot coding preprocessing steps. In the field of engineering materials, this mechanism is extremely suitable for processing modeling tasks containing multi-source heterogeneous information and shows excellent anti-overfitting ability while ensuring stability.

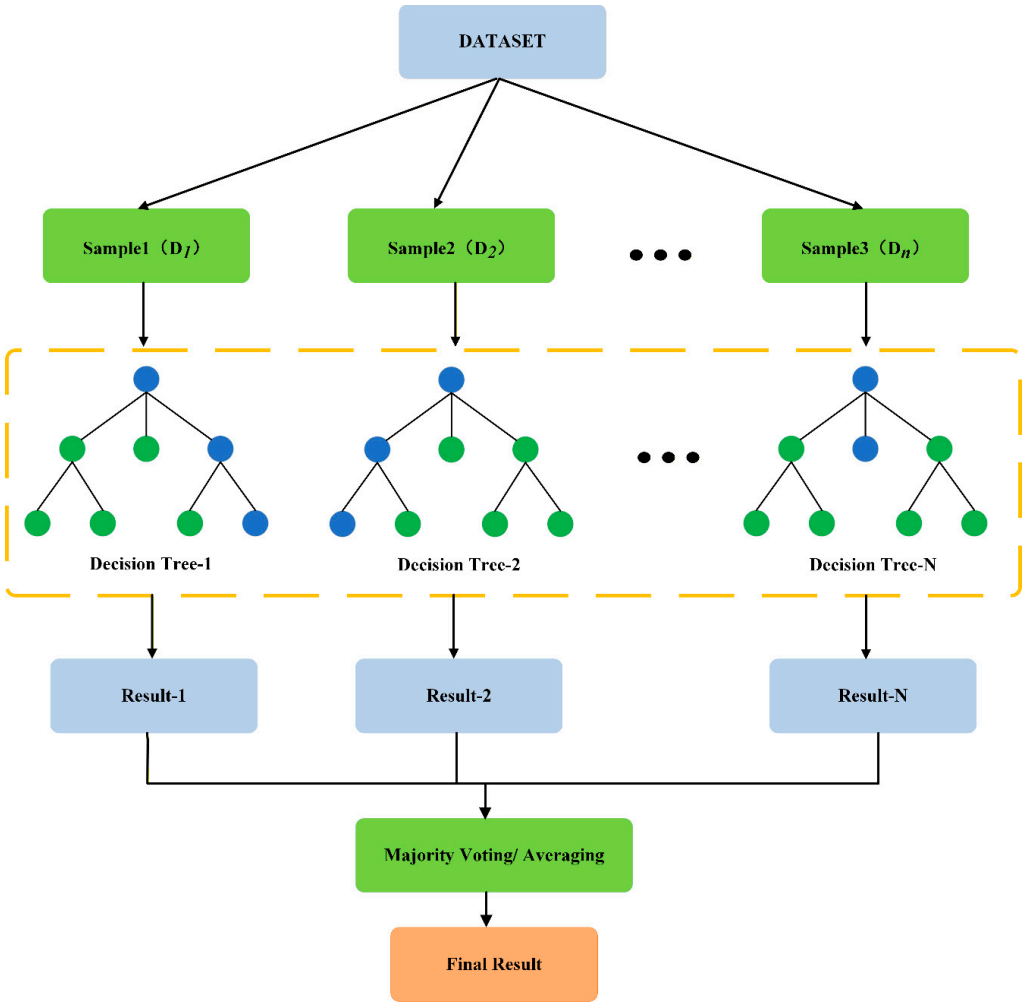


Figure 4. Schematic illustration of random forest construction and prediction.

As the most widely used and stable algorithm for material performance prediction, the ensemble learning algorithm system also has some shortcomings. In the ultra-high-dimensional data scenario, its representation ability is insufficient, and it is particularly difficult to effectively capture the internal complex topology and mapping relationship. Artificial neural networks (ANNs) and deep learning models have multi-layer nonlinear transformation structures and stronger representation abilities in high-dimensional spaces; therefore, they have gradually gained attention in research.

2.3. Neural Network and Deep Learning Model

Artificial neural networks (ANNs) and deep learning systems (including deep architectures such as DNNs and CNNs) have built one of the core frameworks for modern data-driven computing methods. This series of network models starts from the most basic neuron topology, realizes the efficient optimization of parameters by introducing the backpropagation mechanism (BPNN), and gradually evolves into a deep and complex architecture with a strong feature extraction ability. Although they have the same relationship

at the bottom of the calculation logic, there are still different emphases on the depth of the structure, specific connection mode, and respective application scenarios.

(1) Artificial neural networks (ANNs) are mathematical models that simulate the connection mechanism of brain neurons. Their theoretical prototype is largely inspired by the biological nervous system [43]. Its typical architecture is constructed based on highly interconnected neuron nodes in the input, hidden, and output layers, as shown in Figure 5 [44]. Through the role of multi-layer nonlinear mapping, an ANN can infinitely approximate complex continuous functions. As the most classic fully connected feedforward network, the multi-layer perceptron constitutes the most basic structural form of this type of model.

(2) BPNN, the most popular representative of classical feedforward neural networks, approximates the complex mapping relationship between the input and output through a multi-layer feedforward neural network structure [45]. Its core operational logic consists of two stages: forward propagation, which computes predicted values using weighted node connections and nonlinear activation functions, and backpropagation, which dynamically updates the network weights and biases via gradient descent based on the prediction error, thereby progressively minimizing the loss function. Although the prediction accuracy of BPNN is often affected by the hidden layer structure and hyperparameters, such as the learning rate, it is widely used in the field of engineering materials to construct the complex mapping relationship between multi-component parameters and macroscopic mechanical properties owing to its strong deep self-learning ability, which supports the performance prediction and mix proportion optimization of the materials.

(3) The Deep Neural Network (DNN) is also categorized as a nonlinear supervised learning algorithm based on artificial neural networks. Through the construction of a deep feedforward architecture with multiple hidden layers, complex mapping relationships between input variables and outputs are effectively modeled [46]. Compared with the previously mentioned BP neural network, DNN can automatically extract higher-order and more abstract features in complex data by increasing the number of hidden layers. However, an extremely deep network topology can easily lead to the risk of gradient disappearance and overfitting. To maintain the stability of the architecture during actual training, mechanisms such as L2 regularization, dropout, or early stopping are often introduced. Although DNN has high requirements for data volume and computing power and is often used as a ‘ black box ‘ model, it often needs to use post-attribution tools such as SHAP to enhance interpretability. However, when establishing this deep nonlinear mapping between material composition and macro performance, it still has very obvious advantages, which provide strong technical support for the accurate prediction and directional optimization of complex material systems.

(4) The CNN represents a specialized deep learning framework engineered for grid-like data topologies (e.g., images and time-series sequences). Through local connections and convolutional operations, spatial local features within the input data are adaptively extracted and hierarchically represented [47]. As shown in Figure 6, compared with traditional fully connected networks, CNN innovatively introduces convolutional and pooling layers. Relying on local connectivity and parameter sharing, it accurately captures spatial local features while significantly reducing the computational overhead. In the field of engineering materials, CNN has opened up the processing path of unstructured data, such that the microscopic physical and chemical characterization of materials (e.g., microscopic pore and fracture images) can be directly analyzed, thus effectively establishing the deep correlation between microscopic morphology and macroscopic mechanical properties.

From classical machine learning models to ensemble learning models and then to complex deep learning models, various algorithm models show different characteristics and are suitable for diverse application scenarios. Classical machine learning has laid the foundation for data pattern extraction and mapping relationship construction by virtue of complete statistical theory and clear underlying rules; however, it is often limited in dealing with complex nonlinear interactions. Ensemble learning effectively suppresses the bias and variance generated by a single model in prediction by combining multiple weak learners together for collaboration. The inherent performance bottleneck of the monomer algorithm is broken, so it shows strong generalization ability and robustness when dealing with small samples, high noise, and nonlinear data; deep learning relies on multi-layer architectures and the backpropagation mechanism, and the limitations of shallow models in representation are further broken through. It has the ability to adaptively analyze deep abstract features directly from original high-dimensional data, and grid-like topological data, such as images or sequences, can also be directly processed. To more intuitively show the differences in the characteristics of various underlying algorithms, Table 1 compares the core advantages and disadvantages of the above machine learning models and typical applicable scenarios.

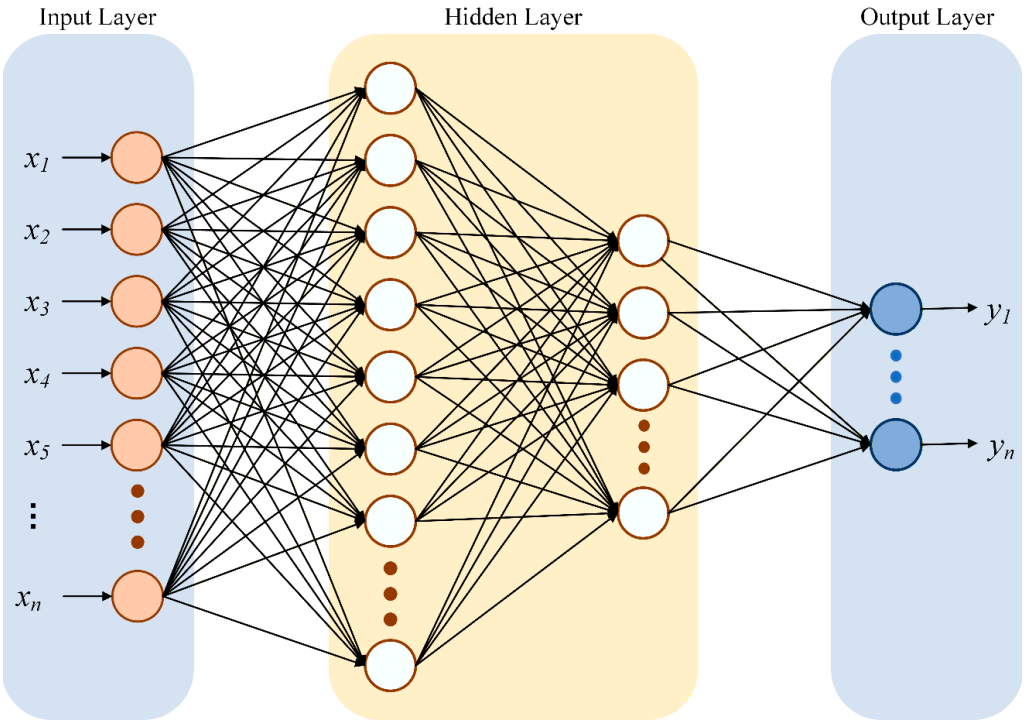


Figure 5. Working principle diagram of artificial neural network.

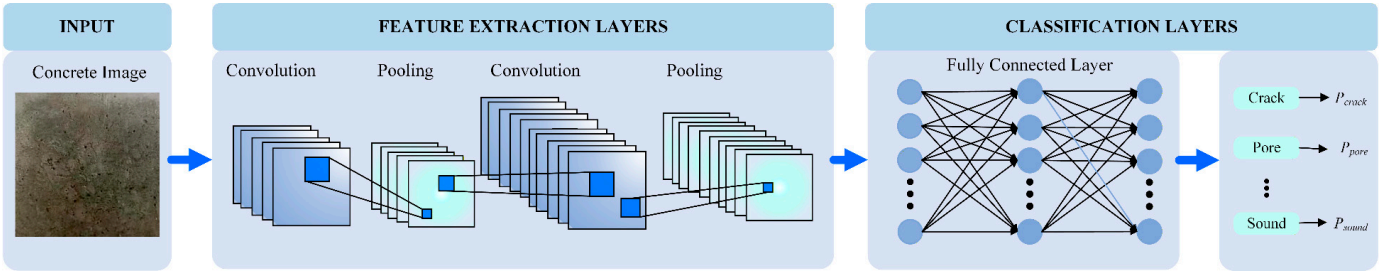


Figure 6. Convolutional neural network working principle diagram.

Table 1. The advantages and disadvantages of each machine learning model and the comparison of applicable scenarios.

Algorithm Category	Algorithm Name	Structural Complexity and Feature Extraction Capability	Advantages and Applicable Scenarios	Main Disadvantages
Classical Machine Learning	Linear Regression (LR) [28,29]	Extremely simple structure; relies on raw input data and manually engineered features.	Simple, interpretable, and computationally efficient; suitable for problems with approximately linear relationships.	Can only model linear relationships; sensitive to outliers; prone to underfitting.
	Support Vector Machine (SVM) [31–33]	Relatively simple structure; uses kernel functions to map low-dimensional features into a high-dimensional space for nonlinear modeling.	Performs well in high-dimensional spaces and has strong generalization; suitable for small- to medium-sized datasets and high-dimensional nonlinear classification tasks.	Slow to train on large datasets; sensitive to missing values; kernel selection is complex.
	K-Nearest Neighbor (KNN) [33,34]	No explicit training structure; directly matches features and performs classification based on spatial distance.	Intuitive concept and no separate training stage; has some tolerance to outliers. Suitable for simple classification tasks and baseline comparisons.	Computationally intensive; sensitive to the dimensionality of the feature space.
	Decision Tree [35]	Tree-structured branching; provides basic partitioning of the feature space.	Intuitive and highly interpretable; does not require data normalization and can handle missing values. Suitable when the decision logic must be explained.	Highly prone to overfitting; sensitive to small changes in the data.
Ensemble Learning	Random Forest (RF) [36]	Bagging ensemble of multiple decision trees; provides strong parallel feature extraction and robustness to interference.	High accuracy and noise robustness; can handle high-dimensional features and evaluate feature importance. Suitable for high-dimensional classification and regression.	Training becomes time-consuming when many trees are used; less interpretable than a single decision tree.
	AdaBoost [37,38]	Sequentially ensembles weak classifiers and dynamically adjusts their weights; can evaluate feature importance.	Strong nonlinear modeling capability; can quantify the contributions of material components and guide mix-proportion optimization.	Sensitive to outliers and high noise, which can readily lead to overfitting.
	GBDT [39]	Sequentially combines weak decision trees, fits residuals iteratively, and outputs feature importance.	Effective for multidimensional nonlinear mapping; suitable for accurate mix-proportion–performance prediction and analysis of component synergy.	Its core mechanism relies on sequential iterative fitting of residuals.
	Boosting Family (XGB/LGBM/Cat) [40–42]	Sequentially deepened ensemble structure; progressively strengthens feature learning and residual fitting.	High accuracy; can handle missing values and support parallel computation. Suitable for structured-data prediction.	Many parameters and complex tuning; prone to overfitting on data dominated by noise.
Neural Networks & Deep Learning	ANN/BPNN [43–45]	Usually has a single hidden layer or a shallow hidden-layer structure; performs shallow nonlinear mapping and relies on the original input features.	Strong nonlinear fitting and adaptive capabilities; suitable for small- to medium-sized, multidimensionally coupled, or poorly bounded problems.	Requires substantial computational resources and data; has a black-box nature and limited interpretability.
	DNN [46]	Deep architecture with multiple hidden layers and substantially increased network depth; automatically extracts and represents high-level latent features.	Can extract deep abstract features; suitable for large datasets, high-dimensional problems, and complex systems with unclear mechanisms.	Prone to vanishing or exploding gradients; has a very large number of parameters and is difficult to train.
	CNN [47]	Stacks convolutional and pooling layers to extract spatial features through local connectivity and parameter sharing.	Provides local perception, parameter sharing, and translation invariance; suitable for images and spatially structured data.	Complex architecture; generally less effective for non-spatial data.

3. The Application of Machine Learning Model in Conventional Concrete

The above machine learning models have been applied to material performance prediction and reverse design. From classical models to ensemble learning models and neural networks, they have been widely used in the performance modeling of various metal materials [48–50] and multi-objective reverse design of complex material systems [51–54], effectively revealing the macro- and micro-mechanisms of hydrogen storage materials [55–57]. Owing to its powerful high-dimensional characterization ability, the artificial neural network not only realizes the accurate prediction of polymer properties [58] and cross-scale physical mapping [59], but also overcomes the problem of small sample modeling by combining physical equations [60]. Based on this, frontier generation models (such as ETR and WGAN) break through the exploration boundary of existing data and successfully reverse-generate the mix ratio of new photocatalysts and multi-principal element alloys under a very small sample [61,62], which promotes material research to achieve a leap from passive prediction to active creation.

In the field of concrete, the material system has always been multiphase and multi-component, and there is a strong nonlinear mapping relationship between its macroscopic properties and mix ratio parameters [63–71]. In the past, the method of iteration back and forth by empirical formulas and laboratory tests was not only long in cycle and high in cost, but also difficult to completely describe the complex coupling effect between multiple variables, such as the water-binder ratio, admixture type and dosage, and aggregate characteristics. Machine learning, with its ability to autonomously mine complex patterns from data, has been widely used in the fields of concrete mechanical performance prediction, durability evaluation, and mix proportion optimization design [72–75]. By analyzing the keywords of Web of Science concrete and machine learning in the past five years, the time series evolution diagram of related keywords is sorted out, as shown in Figure 7. As illustrated in Figure 7, machine learning applications in concrete research have evolved significantly over the past five years. The paradigm has progressively shifted from classical machine learning and basic neural networks to ensemble learning and deep learning.

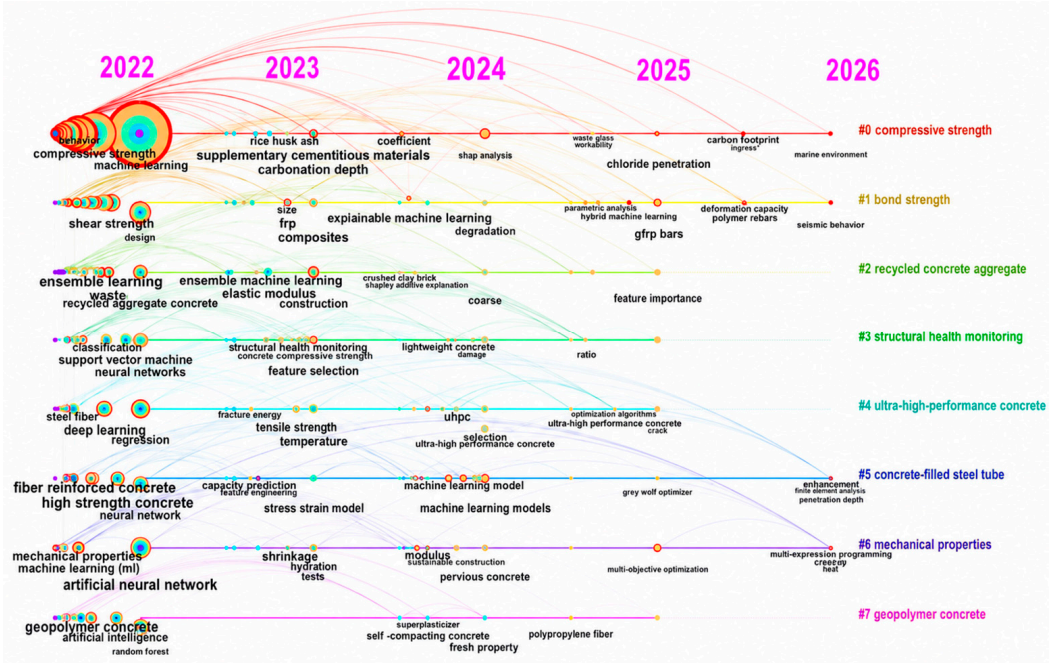


Figure 7. Citespace keyword timing evolution diagram.

3.1. The Application of Classical Machine Learning Model in Conventional Concrete

Because of its simple structure and strong interpretability, the classical machine learning algorithm is often used as a benchmark model or to solve specific small-sample nonlinear problems in the early stage of concrete research; however, its application boundary is also gradually emerging in more complex engineering scenarios.

(1) Basic modeling and data processing

At the basic modeling level, linear regression (LR), with its clear parameter weights, is regarded as an intuitive tool for analyzing the linear relationship between concrete compressive strength and the mix proportion. Concrete is a multiphase composite system in which aggregates, paste, and the interfacial transition zone (ITZ) work synergistically. Its strength development is governed by highly nonlinear physicochemical processes, including hydration reactions, microstructural evolution, and the propagation of defects. Linear regression is based on the assumption of independent linear superposition between variables, which makes it difficult to capture the above-mentioned complex mechanism. Therefore, it has been gradually replaced by nonlinear models in recent years [76]. In the data preprocessing stage, the k-nearest neighbor (kNN) algorithm is widely employed for missing value imputation owing to its strong capability for local similarity learning. For concrete datasets with complex compositions and high rates of missing data, data integrity is effectively restored using 10-nearest neighbor (10-NN) imputation, laying a reliable foundation for subsequent refined modeling [77]. However, the kNN algorithm exhibits clear limitations under sparse data distributions or when new sample features deviate significantly from the existing training patterns. In such cases, effective neighbors can hardly be identified using the "local similarity" strategy, and the prediction confidence is markedly reduced.

(2) Performance evaluation in specific scenarios

When the nonlinearity between the target parameters and features is controllable, algorithms such as support vector machine (SVM) and decision tree (DT) show high accuracy and efficiency in specific scenarios. Taking the durability evaluation as an example, in the classification and prediction of chloride ion penetration resistance of concrete, both SVM and DT can achieve rapid determination of chloride corrosion resistance grade with high precision, so as to better replace the time-consuming traditional experiment and bring an efficient method to the durability evaluation in coastal high salt environment [78]. This kind of application shows that the classical model can effectively replace the time-consuming traditional experiments such as electric flux method or RCM method in specific classification tasks, and provides a fast and efficient technical means for the durability evaluation of concrete structures in coastal high salt environment. Subsequent particle swarm optimization (PSO) [79]. This kind of application shows that the classical model can effectively replace the time-consuming traditional experiments such as electric flux method or RCM method in specific classification tasks, and provides a fast and efficient technical means for the durability evaluation of concrete structures in coastal high salt environment [80].

(3) Applicable boundary of classical model

In the face of extreme service environments or highly complex material systems, the classical model shows its limitations. In the study of the high-temperature compressive behavior of C120 high-strength concrete and mortar, in the face of complex physical and chemical evolution at high temperatures, the prediction accuracy of LR and support vector regression (SVR) is not as good as that of traditional standard tabulation data, and the shortcomings of kNN's serious lack of generalization performance are also exposed [81]. Similarly, in the strength prediction (compressive and flexural) of steel fiber-reinforced

concrete (SFRC), severe underfitting is exhibited by both linear regression (LR) and support vector regression (SVR) due to their inability to capture deep nonlinear interactions among mix proportion variables. Meanwhile, the "local similarity" strategy relied upon by kNN is rendered ineffective when confronted with new feature combinations, owing to the sparsity of experimental data [82]. In such complex problems with strong nonlinearity and multi-factor close coupling, the classical model has reached the upper limit of its own ability in terms of fitting accuracy and generalization performance.

From the above analysis, it can be seen that the classical machine learning method does set up an early analysis framework for the prediction of concrete performance. These methods still have certain application value for tasks with relatively simple data characteristics and controllable nonlinearity. However, once a complex system with strong nonlinearity and multi-factor deep interweaving, such as high temperature, high strength, or fiber reinforcement, is entered, the classical model encounters a bottleneck in both fitting accuracy and generalization performance. Driven by these requirements, ensemble learning techniques with enhanced interaction modeling and generalization performance have gained prominence. Consequently, they are now widely implemented as standard solutions for such problems.

3.2. Application of Ensemble Machine Learning Model in Conventional Concrete

The limitations of these classical models have prompted the ensemble learning method to become mainstream, as it can independently capture the nonlinear interaction between variables quickly. Ensemble algorithms, such as gradient boosting regression trees, XGBoost, and random forest, abandon traditional linear assumptions and show extremely high accuracy and engineering value in concrete performance prediction [83].

(1) A high-precision prediction model that breaks through the traditional linear hypothesis

Ensemble algorithms have obvious advantages in dealing with nonlinear mapping and can accurately capture the importance of complex variables. In the compressive strength study, random forest (RF) and (boosted tree) models were applied. The researchers found that the important properties of inconsistent variables in these captured data were effectively captured by nonlinear models, resulting in the best performance [76]. For nano-silica fiber-reinforced concrete (NS-FRC) with a complex composition, the traditional SVM and artificial neural network (ANN) models were greatly exceeded by the XGBoost model in terms of prediction accuracy, robustness, and computational efficiency [84,85]. In an acidic erosion environment, for the fracture toughness of ordinary, glass fiber, and micro-silicon reinforced concrete, by comparing the 12 machine learning models, the limit tree regression (ETR) finally obtained the optimal prediction results [76]. In terms of interfacial bonding and fracture properties, the high-precision prediction of the initial, ultimate, and residual bonding stresses was greatly improved by using four mainstream integration algorithms: RF, AdaBoost, GBRT, and XGBoost [86,87]. A novel CatBoost-based strength prediction framework was developed by researchers by considering the intrinsic properties of superabsorbent polymers (SAP). This study not only overcame the limitations of traditional empirical regression methods and models based only on macroscopic porosity, but also provided a screening process for mix-design recommendations for SAP concrete based on strength requirements. This framework provides an important tool for the intelligent design of concrete materials with high durability and low autogenous shrinkage [88].

(2) Performance jump and fine design combined with hyperparameter optimization

To further explore the potential of the i ensemble algorithm, Bayesian and other hyperparameter optimization techniques were introduced, which not only improved the prediction upper limit of the model but also provided a new path for the reverse design of materials. In the horizontal comparison of various ensemble algorithms, a Bayesian

optimization model was introduced to significantly improve the prediction performance of the combined stress [86]. The Tree-structured Parzen Estimator (TPE) within the Bayesian optimization framework was integrated with XGBoost. Through this integration, highly reliable and accurate optimal hyperparameter configurations can be identified at a relatively low computational cost. On three different datasets, the prediction accuracy of TPE-XGBoost comprehensively and significantly exceeded that of other mainstream, unoptimized machine learning models [89]. Similarly, the Bayesian hyper-parameter-optimized XGBoost model successfully achieved high-precision prediction of complex properties, such as flexural strength, compressive strength, and compactness, in ultra-high-performance concrete (UHPC). Based on this, the researchers proposed two types of UHPC fine design ideas based on the XGBoost model, which were combined with a physical model and genetic algorithm [90].

(3) From Model Interpretability to Explicit Equations

Machine learning has always been criticized for the 'black box' problem. Combining the ensemble algorithm with interpretable technology, such as SHAP, this situation can be effectively improved. This method clearly displays the complex physical coupling mechanism hidden in the underlying layer. In contrast, the gap between the data-driven model and the traditional engineering experience specification is eliminated, making the model output an explicit physical formula that is easy to use. Using interpretability techniques, researchers can analyze the internal mechanism of the model. This method not only improves the prediction accuracy to a new level but also greatly enhances the interpretability of the model on the physical layer, which clearly explains how different characteristic variables couple to affect the mechanical behavior of concrete [89]. Similarly, in a corrosive environment, a convolution-based ensemble deep learning algorithm combined with SHAP was used to quantify the relative contributions of corrosion level, transverse reinforcement ratio, and other factors to the predicted bonding performance of corroded reinforced concrete, providing a more transparent basis for interpreting the predicted residual bearing capacity of corroded structures [91]. With further extension of interpretability techniques, innovative applications have been spawned to directly convert machine learning models into explicit physical equations. Previously, in predicting the bond strength between FRP and UHPC, traditional empirical equations were often constrained by insufficient accuracy. Meanwhile, high-precision machine learning models could hardly be directly converted into practical engineering formulas. In response to this pain point, a breakthrough ' machine learning-guided canonical form equation ' hybrid framework was proposed. After the ensemble model is trained, SHAP is used to break down its complex nonlinear prediction into independent additive contributions from each parameter. Then, a nonlinear function that follows physical and mechanical laws is fitted and seamlessly embedded into the traditional ACI additive equation structure. This new explicit equation greatly improves the prediction accuracy and successfully captures the hidden physical trends. At the same time, it also retains the core advantages of the traditional standard formula, which is transparent, simple, and can be directly calculated without running the algorithm [92].

However, SHAP-based post-hoc interpretation has several limitations. Highly correlated input features may make it difficult to distinguish their individual contributions, potentially leading to biased attribution among correlated variables. Although SHAP can characterize feature interactions, complex nonlinear and higher-order interactions may be difficult to interpret in concrete mixtures with multiple interacting parameters. In addition, global SHAP importance reflects the overall contribution across the dataset, whereas local explanations may vary considerably among individual samples, limiting the direct generalization of local patterns. SHAP analysis also introduces additional computational costs, particularly for complex models and interaction analysis. Moreover, SHAP results are

sensitive to the underlying model structure, hyperparameters, and training data, and different models may yield different feature contributions. Therefore, SHAP results should be interpreted as model-based associations rather than direct physical or causal mechanisms.

3.3. Application of Neural Network and Deep Machine Learning Model in Conventional Concrete

Research on concrete materials is continuously expanding towards micro-morphology analysis. Consequently, concrete intelligent analysis methods with stronger nonlinear capabilities are urgently required when massive high-dimensional data or complex feature extraction tasks are involved. Artificial neural networks and their derived deep learning models have gradually become another research focus and have evolved rapidly along the path from basic networks to deep architectures.

(1) Construction and optimization of basic neural network

In the prediction of the mechanical properties of concrete, the BPNN is one of the earlier models. Through the error backpropagation mechanism, the ability to approximate any complex nonlinear function is theoretically provided by this algorithm. Therefore, natural advantages are shown when the mapping relationship between the mix proportions and macroscopic strength is established. In early research, scholars realized the prediction of the compressive strength of specific types of concrete by constructing a BPNN model, which preliminarily verified the feasibility of neural networks in this field [93]. To address the problem that traditional BP networks easily fall into local minima, genetic algorithms (GA) were introduced in subsequent research to optimize the initial weights and thresholds of the network. Consequently, the strength prediction accuracy for ultra-high-performance concrete (UHPC) was greatly improved. Furthermore, a new idea was proposed for reverse mix design based on algorithm models [94]. Based on this, the GA-BP model was further applied to explore the internal relationships between compressive strength and multiple parameters, including sample side length, aspect ratio, and air content. According to the fitting results, the prediction accuracy of the model was significantly improved after GA optimization [95]. To further improve global optimization efficiency, the Logistic chaotic mapping and a nonlinear inertia weight were introduced into the sparrow search algorithm (SSA) [96], developing the Logistic chaotic sparrow search algorithm (LCSSA). Subsequently, an RF-LCSSA-BPNN collaborative model was constructed by coupling LCSSA with the feature screening mechanism of random forest (RF). Compared with the traditional model, this model shows a better convergence speed and prediction accuracy in dealing with the multi-component complex nonlinear problems of high-performance concrete (HPC) [97]. Similarly, when dealing with extreme environments such as high temperatures, the BP-ANN calculation framework can capture the bond strength degradation law of FRP bars and concrete interfaces with high precision, which reflects the good generalization ability of the neural network [98]. A machine learning prediction system was developed for the multi-performance mechanical properties of polyethylene fiber-reinforced engineered cementitious composites (PE-ECC). In this study, three algorithms were used to predict the mechanical properties of PE-ECC. Among them, the BPANN model showed the best overall performance and was able to accurately describe the nonlinear relationship between mix proportions and the ultra-high ductile mechanical response. When high-strength and high-ductility PE-ECC is designed, strength should not be improved only by reducing the water-to-binder ratio, as the ultimate tensile strain may be reduced. In engineering practice, a high-performance water-reducing agent should also be used to ensure sufficient fiber dispersion, so that a balance between ultra-high strength and ultra-high ductility can be achieved [99].

(2) Convolutional neural network and image-driven analysis

The integration of deep learning technology opens a new way for intelligent detection of concrete based on image analysis and makes the application of machine learning in this field more diversified. Researchers have proposed an automatic microstructure analysis method for SEM images based on the U-Net convolutional neural network, which can accurately segment different phases, such as aggregate, anhydrous cement, hydrated cement, and pores, and provide a reliable basis for quantitative characterization of key microscopic parameters, such as hydration degree and porosity [100]. Similarly, researchers have proposed a physics-guided artificial intelligence framework that tightly integrates image segmentation with key physical metrics in materials engineering (phase evolution, pore refinement, and particle geometric evolution) to automatically characterize cement hydration microstructures from X-ray μCT images, as shown in Figure 8 [101].

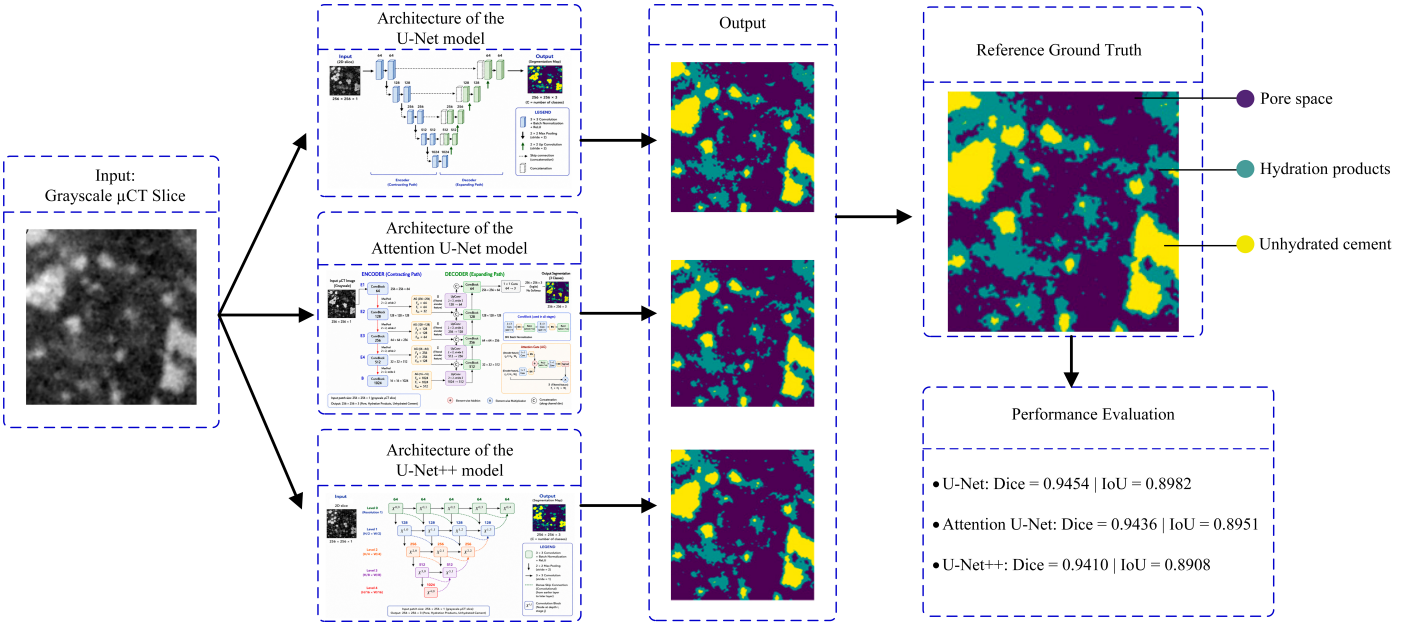


Figure 8. Physics-guided artificial intelligence framework [101].

In terms of macroscopic non-destructive testing, the MPaConvNet-SVR model was constructed by combining CNN and support vector regression (SVR) (Figure 9). Through this model, compressive strength can be quickly and non-destructively predicted using only concrete surface images, while a certain level of interpretability is provided for the decision-making process [102].

(3) The integration trend of deep learning and other methods

The research paradigm of deep learning (DL) in the field of concrete materials is evolving into two significant trends. First, heterogeneous fusion is achieved between deep networks and classical machine learning, such as through cascaded CNN and SVR architectures. In this approach, the automatic feature extraction capability of deep learning is combined with the stable convergence of statistical learning models. Second, new topologies, such as the attention mechanism and gated recurrent unit (GRU), have been gradually introduced. Through these structures, the limited focus of traditional networks on multi-source heterogeneous information is resolved, and long-range temporal dependencies can be effectively captured. The application of this type of composite strategy greatly expands the representation dimension and adaptability of the model in analyzing the nonlinear behavior of multiphase complex concrete systems.

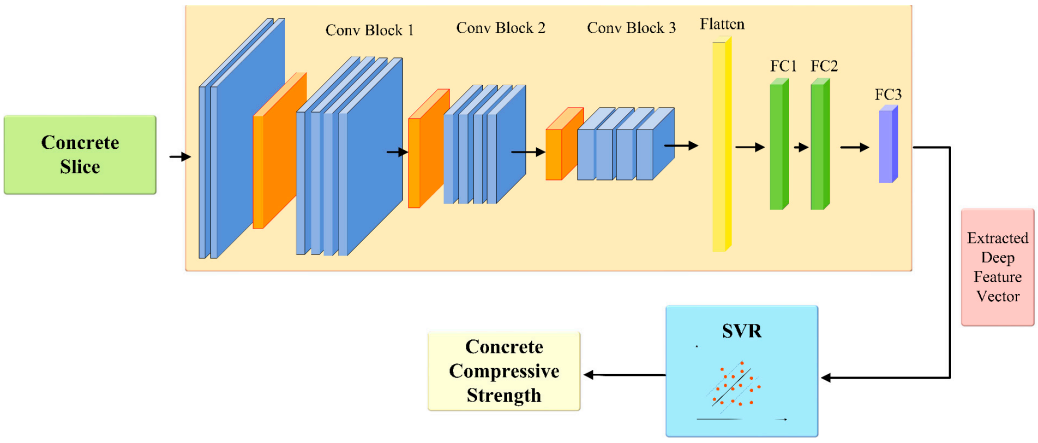


Figure 9. Optimization of CNN Models via the MPaConvNet-SVR Architecture.

However, the deep learning architecture requires extremely high data requirements owing to its large parameter space. In the field of material testing, especially in the evaluation of special mix designs or extreme service conditions, it is often necessary to incur high time and economic costs to obtain high-quality test data. Such datasets have the characteristics of small samples and high dispersion. It is often difficult to reach the sample threshold that drives the convergence of deep networks, and the risk of model overfitting is easily caused. Therefore, the current application of deep learning in concrete engineering is mostly focused on scenes with relatively abundant unstructured data, such as microscopic image analysis and structural health monitoring. In data-constrained conventional mechanical performance prediction and mixture ratio reverse optimization tasks, the ensemble learning framework still occupies the mainstream position with excellent generalization ability and anti-noise robustness under small sample conditions.

In summary, in the performance prediction of conventional concrete, research often focuses on the deep mining of relatively clear and standardized material feature space. Therefore, in the selection of numerical optimization algorithms, researchers have been inclined to adopt mature algorithms with deep mathematical foundations and rigorous optimization paths, such as Bayesian optimization and genetic algorithms (GA) (as shown in Table 2). Such algorithms show high efficiency in the face of small samples or datasets with stable feature distributions. They not only help the model avoid local optimal traps but also gradually reduce the prediction error to ensure the fitting accuracy and stability of the model. With the gradual expansion of research on micro-image analysis and dynamic service-state evaluation, convolutional neural networks (CNNs) and their improved architectures have also been widely used. Specifically, the feature extraction capability of CNNs is combined with classical regression models, such as SVR, in some studies. Through this combination, the end-to-end evaluation of non-destructive testing was successfully achieved. Some studies have also introduced gated recurrent units (GRU) and attention mechanisms (Attention) to correct multi-source sequence deviations and alleviate gradient anomaly problems. This ensures that the model can still make high-intensity and accurate judgments on the whole life cycle durability and multi-objective prediction under the condition of limited data size.

Table 2. Comparison of the advantages of the optimized model in concrete research.

Model Category	Specific Model	Introduced Algorithm	Optimized Model	Core Advantages in Solid Waste Concrete Research
Classical Machine Learning	SVM	Particle Swarm Optimization (PSO)	PSO-SVM [80]	Automatically optimizes the SVR penalty factor, improving fitting and prediction accuracy; suitable for predicting the mechanical properties of concrete.
Ensemble Machine Learning	XGBoost	Bayesian Optimization (e.g., TPE algorithm)	Bayesian Optimization Ensemble Model [87–90]	Efficient search with high accuracy and strong generalization; suitable for accurate multi-objective performance prediction with missing data.
Neural Networks & Deep Learning	Neural Network (BPNN)	Genetic Algorithm (GA)	GA-BP [94,95]	Avoids local optima in the BP network and improves stability and accuracy with small samples; suitable for mix-proportion–performance modeling and inverse design.
		Sparrow Search Algorithm (LCSSA)\Random Forest (RF) Feature Selection	RF-LCSSA-BPNN [97]	Balances global and local optimization, providing high accuracy and low error; suitable for high-dimensional, noisy concrete-strength prediction and field assessment.
	Convolutional Neural Network (CNN)	Support Vector Regression (SVR)	MPaConvNet-SVR [102]	Combines deep feature extraction with regression for nondestructive, end-to-end assessment; suitable for concrete nondestructive testing and microstructural image analysis.

4. The Application of Machine Learning Model in Solid Waste Concrete

The utilization of solid waste concrete is of great significance in reducing the environmental burden and promoting the construction industry to move towards low-carbon transformation [103,104]. The keyword co-occurrence map of literature related to solid waste concrete from 2022 to the present is shown in Figure 10. From this map, it is revealed that the main research efforts in academia are still focused on evaluating mechanical properties and durability through traditional methods [103–113]. The application of machine learning in this field is still in its infancy, and it is mostly aimed at a single solid waste system, and the modeling research of multi-solid waste collaboration is even more scarce. Although the solid waste concrete can learn from the conventional concrete model method, the data characteristics are completely different. In view of this, this section will sort out the modeling application status of different solid waste systems according to the three types of industrial solid waste concrete, construction solid waste concrete, and other solid waste concrete, to clarify the different needs of various solid waste materials in formulating modeling strategies.

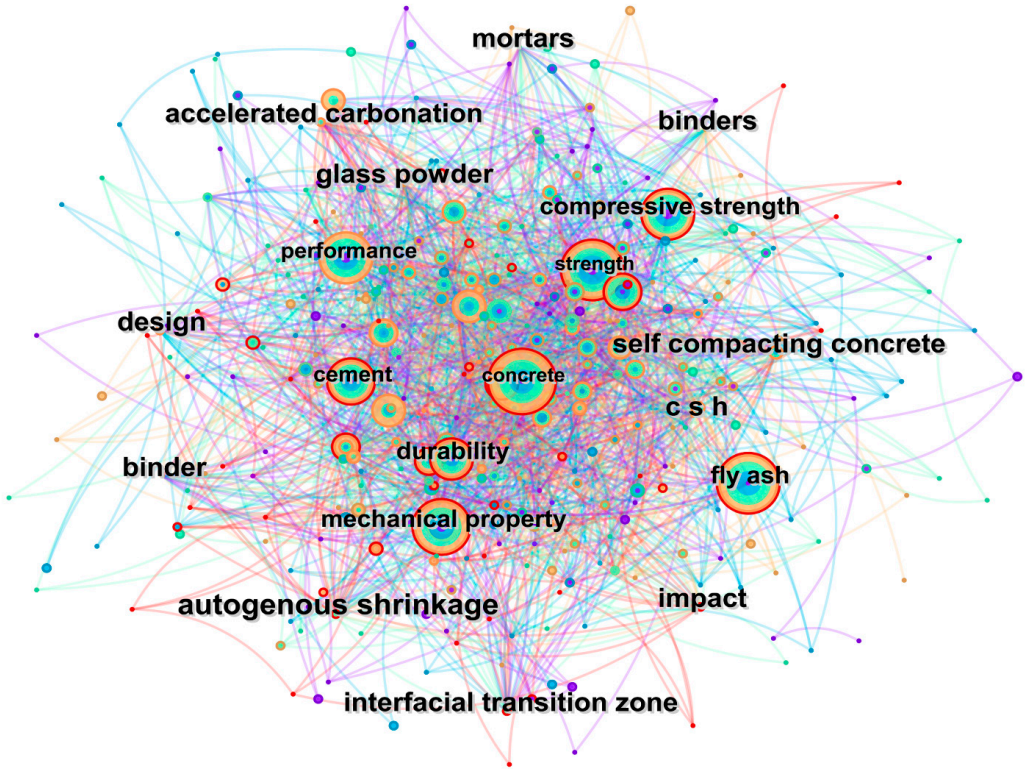


Figure 10. Cite space keyword co-occurrence graph.

4.1. Classification and Data Characteristics of Solid Waste Concrete

From the perspective of source properties and material characteristics, solid waste concrete can be divided into three categories: industrial, construction, and other solid waste concrete (Figure 11). There are fundamental differences in the role and mechanism of action of the three types of solid waste in concrete, resulting in completely different data characteristics. To more intuitively sort out the differences between them, Table 3 compares the characteristics of these three types of solid waste concrete from multiple key dimensions.

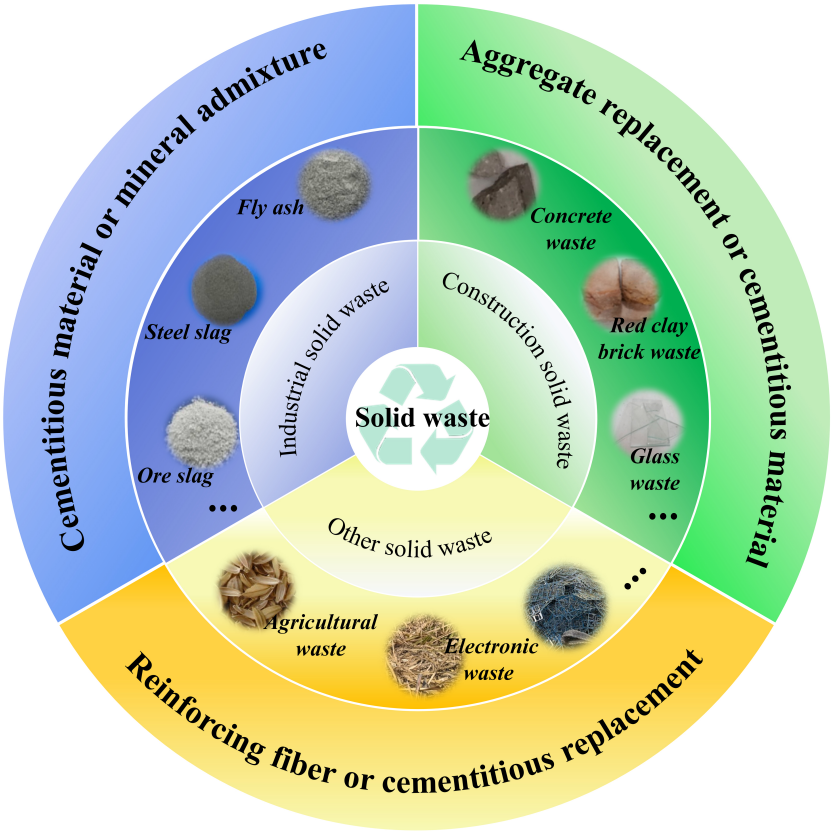


Figure 11. Categorization of Solid Wastes and Their Applications in Concrete Materials.

Table 3. The core characteristics of the application of different solid waste categories in concrete.

Solid Waste Category	Typical Materials	The Core Mechanism of Action	Core Data Characteristics	Modeling Difficulties
Industrial Solid Waste	Cementitious materials/mineral admixtures: fly ash, steel slag, slag, coal gangue, silica fume, etc.	Chemical activation and micro-aggregate filling effect	High non-linear strength; relatively controllable dispersion.	Synergistic and interaction mechanisms of multi-component active materials.
Construction solid waste	Aggregate replacement: recycled concrete aggregate (RCA), crushed brick waste, waste glass sand, etc.; Cementitious materials: finely ground glass powder, recycled concrete fines, etc.	Skeleton support and pozzolanic activity	Extremely high dispersion; significant noise level.	Robust modeling and uncertainty quantification under aggregate quality fluctuations.
Other Solid Wastes	Reinforcing fibers: electronic waste fibers, waste rubber, agricultural natural fibers, etc.; Cementitious replacement: agricultural biomass ash (e.g., rice husk ash, bagasse ash)	Fiber toughening, localized pozzolanic reaction, and pore structure densification	Insufficient data volume; large variations in physical properties.	Reliable modeling under extremely small sample conditions.

(1) Data characteristics and modeling challenges of industrial solid waste concrete

Industrial solid waste is mainly used as cementitious materials or active admixtures, and its chemical composition and mineral structure can be relatively controlled. The core feature of this type of dataset is that nonlinearity is strong, but the data dispersion is relatively low [89]. When building the model, the main challenge is to accurately grasp the synergistic effect of various active materials, such as fly ash, slag, and steel slag, at different curing ages. At the same time, physical and chemical laws, such as hydration kinetics, should be integrated into data-driven algorithms. Through this integration, the interpretability and extrapolation generalization ability of the model are significantly improved.

(2) Data characteristics and modeling challenges of building solid waste concrete

Construction solid waste is mainly used as an aggregate in concrete to provide a physical skeleton support for concrete. The most prominent feature of this kind of data is marked by the high variability in raw material quality. For instance, waste concrete from different sources is often characterized by large differences in attached mortar content and water absorption, by which a large amount of noise is introduced into the dataset. Consequently, the internal relationships between mix proportions and final performance can easily be concealed. In addition, the use of recycled aggregates also introduces multiple interfacial transition zones, resulting in a significant increase in the complexity of the entire system [114]. Therefore, when carrying out modeling, the main challenge is to extract a stable mapping relationship from such highly discrete and noisy data and to make a reasonable quantitative assessment of the prediction uncertainty caused by the fluctuation of aggregate quality.

(3) Data characteristics and modeling challenges of other solid waste concrete

In addition to the aforementioned solid waste types, agricultural and electronic waste have different physical and chemical properties from those of conventional concrete components. Consequently, the existing theoretical framework is challenged to some extent [115]. From the perspective of datasets, such research generally faces a serious shortage of samples and scattered distributions. Simultaneously, there is still a broad extension space for the related discussion of multi-type solid waste co-mixing. When building the prediction model, the main challenges are overcoming the limitations of very small sample data to establish a stable and reliable model and accurately characterizing the huge physical property differences between special solid waste and the concrete matrix.

4.2. Application of Classical Machine Learning Model in Industrial Solid Waste Concrete

Classical machine learning models are mostly used as benchmark methods or preliminary explorations for small-sample scenarios in the study of solid waste concrete. Compared with the application of conventional concrete mentioned above, the shortcomings of the classical model in solid waste systems are often exposed earlier. This issue is primarily caused by the strong heterogeneity and compositional fluctuations of solid waste materials. Consequently, linear hypotheses and local similarity inferences are ineffective when dealing with datasets with high noise and strong variability.

(1) Application of classical machine learning model in industrial solid waste concrete

Industrial solid waste is the most concentrated type of solid waste in the application of classical machine learning models, among which support vector machines and decision trees are widely used. Support vector machines (SVM) demonstrated excellent generalization ability and high stability in predicting the performance of fly ash machines (SVM). When compared with other algorithms, such as BPNN, the optimal cross-dataset generalization advantage is shown by SVM. Furthermore, feature importance analysis was combined to quantify and reveal how key components, such as SiO_2, Al_2O_3, and alkaline activators, drive compressive strength [116,117]. However, with the deepening of the discussion, the

modeling work for such complex materials is no longer limited to the direct application of a single algorithm but is more focused on the customization of the basic model and the integration of multiple algorithms. A composite design framework was constructed by combining an ensemble support vector machine (SVM) with the modified Andreasen and Andersen (MAA) particle packing model. Through this framework, a green UHPC with a 20% steel slag powder substitution rate was successfully developed. In dealing with this kind of small and medium-sized dataset with clear characteristic variables after preliminary screening by MAA physics, SVM is more robust than a complex ANN neural network, and it is not easy to over-fit [118]. In intelligent prediction, excellent nonlinear relationship capturing capabilities are demonstrated by decision trees (DT). Through this method, powerful data-driven support is provided for the optimized mix design and application of waste iron slag in green concrete [119]. In the study of the resource utilization of copper mine tailings (CMT) as an auxiliary cementitious material in concrete, researchers constructed a new hybrid model based on support vector regression (SVR), combined with three metaheuristic optimization algorithms, namely the firefly algorithm (FFA) [120], grey wolf optimizer (GWO) [121], and particle swarm optimization (PSO) (Figure 12), and compared its performance with decision tree (DT) and random forest (RF) models. The SVR-FFA hybrid model showed the highest prediction accuracy [122].

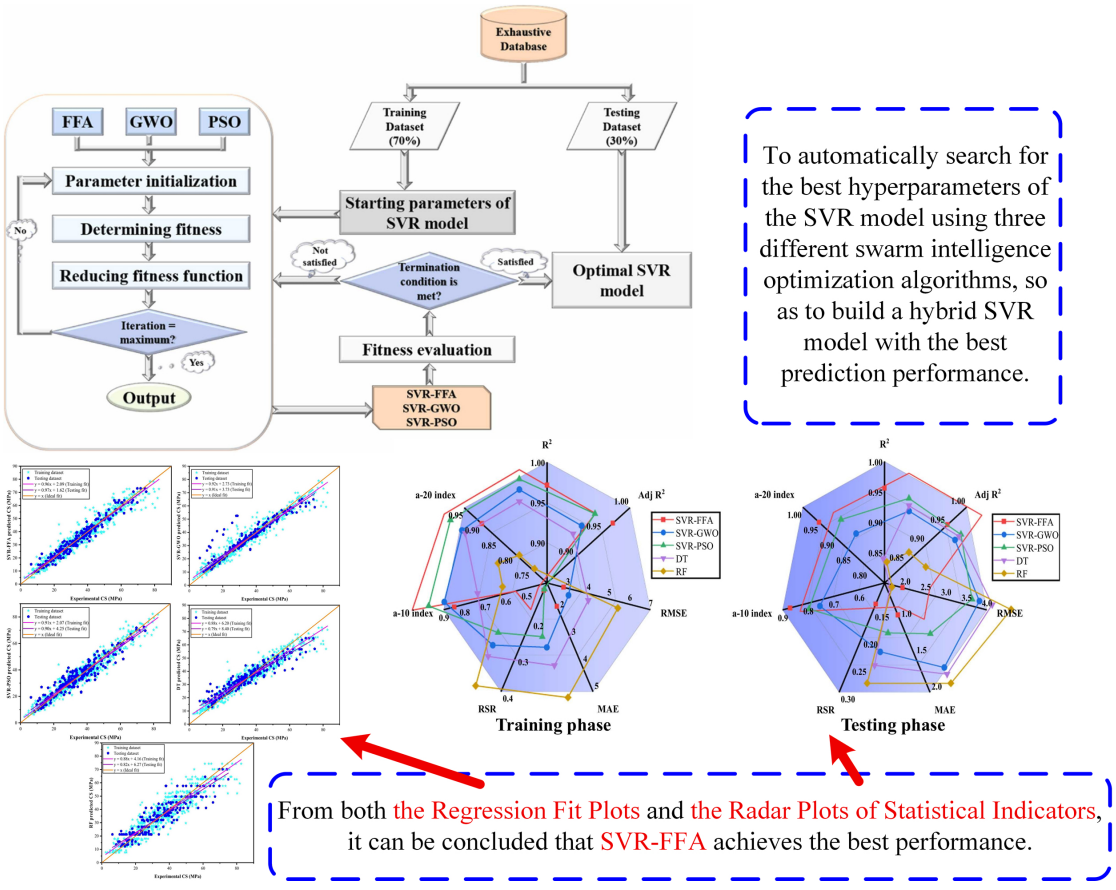


Figure 12. SVM Algorithm Framework Based on Metaheuristic Optimization [122].

(2) Application of classical machine learning models in construction solid waste and other solid waste concrete

The value of the independent application of classical machine learning models in building solid waste concrete is limited, which is mainly limited by the inherent defect of insufficient generalization ability when dealing with high dispersion and high noise data. In the study of partial or total replacement of natural aggregate by recycled concrete

aggregate (RCA), some studies have used decision tree, KNN, and XGBoost to predict the compressive strength (Fc) and splitting tensile strength (STS) of RCA concrete. The results show that the decision tree is flat, the prediction ability of KNN is limited, and the ensemble learning model XGBoost has more advantages [123]. Classical machine learning models, such as KNN, SVM, and DT, show limitations such as parameter sensitivity and limited generalization ability in complex nonlinear relationship modeling. However, the parameters of a single model can be optimized by introducing intelligent optimization algorithms to construct a hybrid machine learning model. Through this combination, the prediction accuracy, stability, and generalization ability were significantly improved. In developing new sustainable ultra-high-performance concrete (S-UHPC), local and solid waste substitute materials, including waste glass powder (Re-WGP), micronized quartz powder, and crushed desert sand, were utilized. Based on DT and SVM models, hybrid machine learning models (such as DT-PSO and SVR-GB) were constructed to predict material durability. Furthermore, SHAP sensitivity analysis identified the water-to-binder ratio, cement content, and aggregate ratio as the most influential features in the model predictions of durability indicators [124].

In general, the application frequency of single shallow machine learning models such as SVM and KNN for evaluating the performance of construction solid waste and other types of solid waste concrete is not high. This is because strong heterogeneity, composition fluctuations, and surface defects are commonly found in solid waste materials. After these materials are added to concrete, the structures of the micro-pores and the interfacial transition zone (ITZ) become complicated. Consequently, single models are often proven insufficient when dealing with such high-noise and high-variability datasets [125,126]. Therefore, ensemble learning frameworks with better anti-noise performance or neural networks with stronger feature extraction capabilities have been adopted in most subsequent studies, together with various optimization algorithms for parameter tuning. Through these approaches, the obstacles caused by data uncertainty were overcome, and both the accuracy and stability of prediction results were further improved.

4.3. Application of Ensemble Machine Learning Model in Solid Waste Concrete

In the current study, the ensemble learning method exhibited outstanding performance in capturing the complex nonlinear correlation between multiple variables and was robust to high noise data. Therefore, it has gradually become the mainstream application method in the field of solid waste concrete performance prediction. Compared with the modeling applications for conventional concrete in Chapter 2, hyperparameter optimization algorithms and interpretability analysis tools are more required by ensemble models when applied to solid waste systems. Through this combination, the extra uncertainty caused by compositional fluctuations in raw solid waste materials is better suppressed. Consequently, the reliability of the prediction results was guaranteed.

(1) Application of ensemble machine learning model in industrial solid waste concrete

To address the modeling problem of multi-component synergy in industrial solid waste systems, the boosting ensemble model represented by XGBoost and gradient boosting decision tree shows better prediction accuracy than the classical model. Some researchers have used the gradient boosting decision tree (GBDT) model to achieve a high-precision performance prediction. Through sensitivity analysis, the physical laws that silica fume enhances early strength and fly ash contributes to late strength were verified. Additionally, the different effects of polypropylene fibers on tensile and compressive strengths were revealed, providing data-driven theoretical support for optimizing complex cementitious systems [127]. A forensic-based investigation optimization algorithm was developed to predict the working performance and multiple mechanical strength indices of sustainable

environmental protection concrete (CCWFS) containing foundry waste sand. A large amount of literature data were collected for model training and validation. Using this data, the hyperparameters of AdaBoost were effectively optimized by the Forensic-Based Investigation Optimization (FBIO) algorithm [128], which significantly improved the prediction accuracy. Compared with traditional or single machine learning methods, such as SVM and RF, the AdaBoost-FBIO model demonstrated superior generalization ability and reliability [128]. For durability prediction, three hybrid models, including SCSO-GBDT, SCSO-LGBM, and SCSO-CatBoost [129], were developed by combining the sand cat swarm optimization (SCSO) algorithm (Figure 13). These models were specifically developed to predict the carbonation depth of fly ash concrete. By comparing with random forest, BP neural network, k nearest neighbor, and other models, it was proven that SCSO-optimized ensemble learning achieves the best prediction performance and the highest efficiency [129]. Further research has introduced the Artificial Bee Colony (ABC) algorithm [130], a metaheuristic method, to perform global optimization of hyperparameters. Studies show that although the ABC-optimized Random Forest model achieves high predictive performance, it is affected by the initial parameter boundaries and assumptions inherent to the ABC algorithm [131]. For solid waste concrete with multidimensional mixed component data (including 16 features such as curing temperature, sodium silicate, fly ash, and various oxides), the researchers systematically compared the performance of XGBoost, CatBoost, LightGBM, and support vector regression (SVR). The evaluation results show that the ensemble algorithms based on the boosting framework exhibit excellent ability to capture nonlinear relationships and high stability, with the XGBoost model performing the best. This study successfully confirmed the reliability of advanced ensemble models represented by XGBoost and CatBoost in accurately evaluating the mechanical properties of sustainable building materials [132].

(2) Application of Ensemble Machine Learning Model in Construction Solid Waste

One of the main problems with concrete made from C&D waste is the high variability in the quality of the recycled aggregates. The ensemble algorithm model has become the main method and research direction for solving this problem because of its anti-overfitting ability and robustness to data-related noise. Previous studies have explored the long-term service performance of recycled concrete in detail and systematically evaluated the applicability of five common machine learning algorithms for predicting the creep behavior of recycled concrete (RAC). Traditional empirical models struggle to accurately predict the drying shrinkage of recycled aggregate concrete (RAC). To solve this problem, a study evaluated eight machine learning algorithms, and XGBoost achieved the best baseline. Furthermore, key features, including drying age, elastic modulus, water-to-binder ratio, and recycled aggregate replacement rate, were identified using the SHAP method based on their contributions to the model predictions. Accordingly, the whale optimization algorithm (WOA) was introduced to optimize the hyperparameters of the XGBoost model. The constructed WOA-XGBoost model provided better prediction accuracy than traditional empirical formulas [133]. In addition to drying shrinkage as a deformation property, traditional empirical models also face challenges of insufficient accuracy and poor generalization when evaluating the multidimensional durability of recycled aggregate concrete (RAC) under harsh environments, which promotes the adoption of tree-based ensemble learning algorithms: for chloride penetration resistance in marine or deicing salt environments, multidimensional features such as mixture proportions and aggregate water absorption were screened by dimensionality reduction, and several algorithms were integrated; it was found that the gradient boosting decision tree (GBDT) model achieved the best accuracy in predicting the passed electric charge, water content was identified as the most critical factor governing ion permeability, and a performance-based mixture design method alongside a service life prediction framework was proposed [134]; regarding carbonation damage induced by atmospheric exposure, where traditional empiri-

cal formulas based on diffusion laws suffer from limited generalization, a gradient boosting regression tree (GBRT) model was developed based on a large dataset of accelerated and natural carbonation tests, and its predictive performance was found to significantly outperform classic semi-empirical models, with carbonation depth shown to be mainly dominated by exposure time, water-to-binder ratio, and concrete compressive strength [135]; furthermore, focusing on freeze–thaw spalling damage in cold regions, the extreme gradient boosting (XGBoost) model was confirmed to provide the highest generalization capability in predicting the mass loss rate among various nonlinear models, optimal ranges of aggregate proportion, water-to-cement ratio, and air-entraining agent dosage were suggested through model interpretability analysis, and a graphical user interface was developed to enable convenient interactive prediction of freeze–thaw durability [136]. From the above studies, it can be found that ensemble learning algorithms represented by gradient boosting can effectively overcome the nonlinear effects caused by defects such as high porosity and high water absorption of recycled aggregates. While the complex internal structure–property relationships of recycled aggregate concrete (RAC) are accurately captured, high-precision predictions are achieved for deformation and multidimensional durability properties, including drying shrinkage, chloride penetration resistance, carbonation resistance, and freeze–thaw resistance.

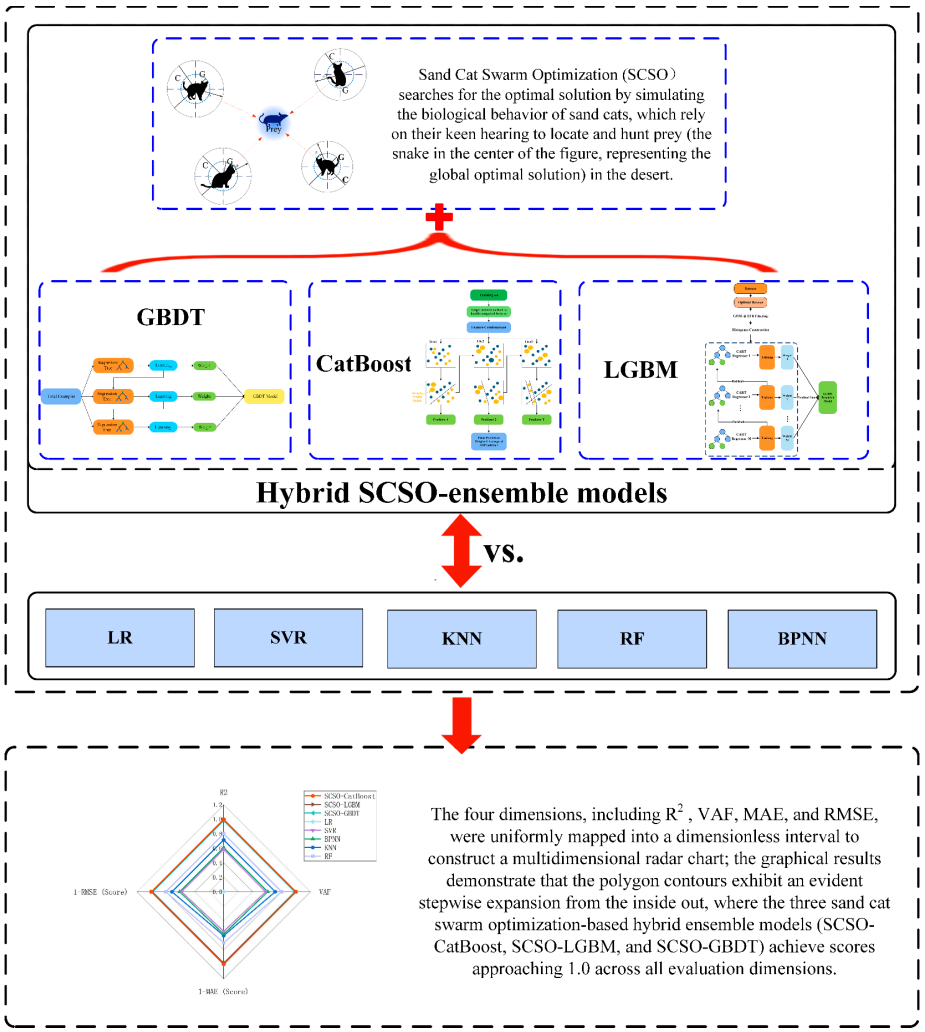


Figure 13. Ensemble Algorithm Model Constructed Based on the Sand Cat Swarm Optimization (SCSO) Algorithm.

Large fluctuations are often observed in the mechanical properties of steel fiber-recycled aggregate concrete (SFR-RAC), whereas traditional empirical models show low

applicability and reliability. To solve this problem, machine learning technologies have been widely introduced in current research. Consequently, an accurate dual-performance prediction framework for both compressive and splitting tensile strengths was successfully constructed. The study covered SVM, BP-ANN, AdaBoost, GBRT, and RF models. Through a comprehensive comparison, it was found that GBRT performs well in predicting the splitting tensile strength and overall stability, whereas RF and AdaBoost perform well in predicting the compressive strength and overall accuracy control, and their prediction errors can be controlled within a lower range. By integrating SHAP values with partial dependence plots (PDPs), the nonlinear response patterns of concrete performance predictions with respect to key parameters, such as the water-to-binder ratio, recycled aggregate replacement rate, steel fiber content, and fiber length, can be further characterized. According to the quantitative analysis of the deep mechanism, the study not only overcomes the modeling problem caused by the heterogeneity of materials but also provides a scientific and reasonable optimization ratio interval and a very important design critical value for practical engineering applications [137–139].

Extreme gradient boosting (XGBoost) and random forest were combined with SHAP and PDP interpretability analyses for the resource utilization of waste bricks and tiles. From a data-driven perspective, the mechanism by which the water-to-binder ratio affects performance was explored. Furthermore, the optimal admixture ratio was identified and verified [140]. In the study of recycled brick and tile (RBT) aggregates, the XGBoost model was optimized using meta-heuristic algorithms such as SFO, HGS, and HBA, which effectively alleviated the overfitting phenomenon and achieved high accuracy and stability. Using SHAP analysis, it was found that test age, cement content, and total coarse aggregate made relatively large contributions to the model predictions of concrete compressive strength [141]. In a study on replacing coarse aggregates with waste glass aggregates (WGA) to prepare environmentally friendly self-compacting concrete (SCC), the random forest (RF) model exhibited higher regression accuracy and reliability than other machine learning algorithms. Consequently, scientific guidance was provided for the efficient utilization of waste glass in high-performance concrete. At the same time, in the study of waste glass and recycled concrete aggregate (RCA) to jointly optimize the performance of concrete, the XGBoost model shows excellent accuracy in predicting compressive strength, and SHAP analysis revealed that cement content, water-to-cement ratio, and natural aggregate content were the most important features in the model predictions of concrete strength [142].

(3) The application of ensemble machine learning models in other solid wastes

Ensemble learning has become a primary means of predicting the performance of waste concrete. Although research on different solid wastes is scattered, it is generally found that the ensemble model has a good effect on dealing with small sample data and discovering complex nonlinear relationships. The introduction of bagasse ash (SCBA) and rice husk ash (RHA) into concrete systems has become a feasible way to improve the sustainable use of materials in research and development of agricultural waste recycling and new environmentally friendly building materials. Ensemble machine learning (EML) technology has been introduced into such multi-component complex systems. Through this approach, an effective method was provided for accurately predicting material properties and optimizing mix design. When exploring the application potential of SCBA lightweight concrete, the random forest (RF) model based on the bagging ensemble method has obvious advantages. It is superior to the traditional artificial neural network in predicting the accuracy of new mixing and mechanical properties. It was successfully and efficiently verified that the 10% SCBA substitution rate was the best solution for both microstructure densification and carbon emission reduction benefits [143]. For the prediction of the compressive strength of RHA environmentally friendly concrete with a larger data size, the

state-of-the-art LightGBM based on Boosting integration strategy further demonstrated the advantages over traditional regression models. Owing to the high architecture compatibility of ensemble tree models, SHAP analysis was successfully integrated into the model interpretation. The results showed that the water-to-cement ratio and rice husk ash (RHA) content made relatively large contributions to the model predictions of concrete strength [144]. A study was conducted to predict the slump, compressive strength, and splitting tensile strength of betel nut shell fiber-reinforced concrete (BNHF). Five ensemble machine learning models were compared. The results show that the CatBoost model performs best in predicting slump and compressive strength, with the highest prediction accuracy and the lowest uncertainty. The XGBoost model stands out when predicting the splitting tensile strength, showing the best generalization ability and stability [145]. In the latest research, amorphous silicon-rich waste (SMW) produced by the decommissioning of photovoltaic panels was also used to optimize the mix ratio of concrete. Combined with the XGBoost prediction model and SHAP-based interpretation, the framework simultaneously achieved maximum compressive strength, maximum waste utilization, and minimum cement content [146].

4.4. Application of Neural Network and Deep Learning Model in Solid Waste Concrete

Compared with ensemble algorithm models, the application of neural networks and deep learning in solid waste concrete is still in its infancy, and the related research results are fewer than those of conventional concrete research. However, existing studies have preliminarily verified its potential for capturing deep nonlinear features, which is more obvious when dealing with a large amount of data or the need to fuse various types of information.

(1) Application of neural network and deep learning model in industrial solid waste

In the field of industrial solid waste, the ANN is the earliest basic deep learning model. Some researchers used the ANN model to predict the compressive strength of fly ash concrete under different fly ash compositions, water quality days, and sulfate exposure periods. Through sensitivity analysis, the 'applicable boundary' of the model is clarified; that is, when the fly ash replacement rate is within 50%, the prediction is the most reliable. At the same time, it can be found that the model prediction results obtained after training are in good agreement with the actual data obtained from 162 experiments, which proves that the neural network has good accuracy for small sample and multivariable datasets [147]. For the durability design of solid waste concrete, researchers have introduced metaheuristic algorithms, such as the Bat Algorithm (BA), Ant Colony Optimization (ACO) [148], and Particle Swarm Optimization (PSO), into BPNN (Figure 14). The dual predictive model optimized by PSO can provide forward guidance for rebar corrosion and concrete damage failure time under saline-alkali environments, and its prediction accuracy is significantly better than that of the unoptimized base model [148]. In the study of predicting the compressive strength of molybdenum tailings concrete, the multi-layer perceptron (MLP) model stands out with its excellent nonlinear relationship capture ability and achieves the best comprehensive prediction performance. SHAP analysis further identified cement content as an important feature in the model predictions of compressive strength. The applicability of MLP in small sample complex mix proportion data is verified [149]. This provides a reference for the high-value utilization of molybdenum tailings. In the study of high-performance fiber-reinforced concrete (HPFRC) prepared by industrial cable recycling stainless steel fiber (SFC), a deep neural network based on an improved gray wolf optimizer (DNN-IGWO) was constructed to demonstrate excellent generalization ability and prediction accuracy. IGWO retains the basic framework of GWO, while the search strategy is improved. Not only the effectiveness of SFC in HPFRC is verified; a data-driven framework is also provided for the future large-scale design of environmentally friendly,

high-performance, and low-cost green building materials [150]. Furthermore, researchers proposed a hybrid deep learning regression model based on a one-dimensional convolutional neural network and long short-term memory (1D-CNN-LSTM), which utilizes mix proportion parameters and curing duration to predict the compressive strength and carbonation depth of fly ash (FA) concrete; this hybrid framework achieved outstanding predictive results, as illustrated in Figure 15 [151].

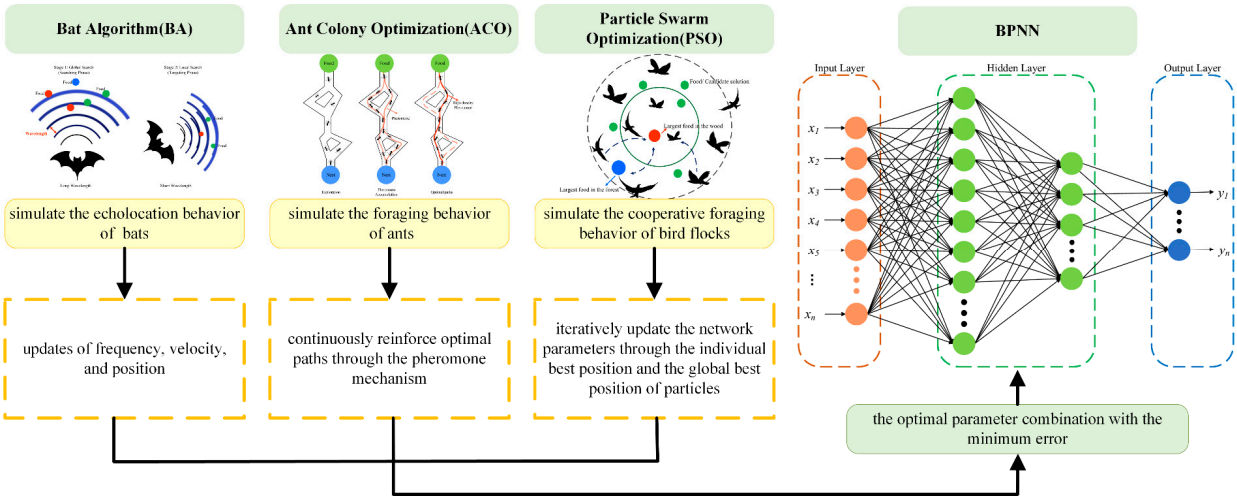


Figure 14. BPNN Model Framework Based on Metaheuristic Algorithms.

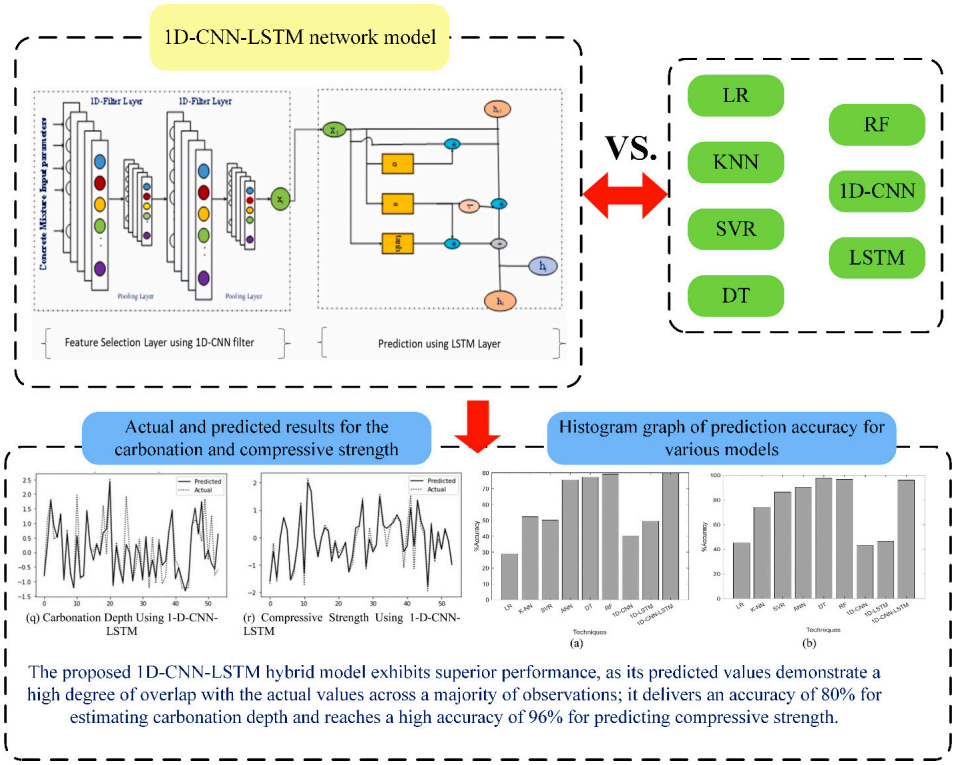


Figure 15. 1D-CNN-LSTM Network Model and Its Predictive Performance [151].

(2) Application of neural network and deep learning model in construction solid waste

In solid waste recycling, the complex composition and high variability of solid waste replacement aggregate concrete pose challenges. To address these challenges, key input features, such as physical and chemical properties, were integrated, and better robustness was demonstrated by artificial neural networks (ANN) compared with the traditional

response surface method (RSM). The performance evaluation system for recycled materials was significantly optimized [152–154]. In the study of construction and demolition waste (CDW) as an auxiliary cementitious material, researchers used recycled concrete powder (RCP) instead of cement and combined it with a polypropylene fiber to prepare sustainable concrete. The dynamic impact strength was accurately predicted using static measurement data obtained using an artificial neural network (ANN) [155]. Simultaneously, the incorporation of solid waste particles affects the rheological properties of concrete, and a quantitative analysis of this microscopic mechanism is often an important step to improve the accuracy of model prediction.

In previous studies, key indicators, including water film thickness (WFT), paste film thickness (PFT), and interstitial fluid viscosity, were measured. Using these parameters, the shear thickening or shear thinning behavior in solid waste systems can be accurately evaluated. Specifically, fine particles, such as fly ash, can play a filling effect to increase the WFT, which significantly reduces the yield stress and slurry viscosity. With the introduction of such multidimensional key features, including space, rheology, and image attributes, the research focus has gradually shifted from relying solely on a single numerical input in the early stage to deep mining of multidimensional features. In this process, convolutional neural networks (CNN) are often used to deeply mine the deep coupling characteristics between the aggregate replacement rate and strength, which is significantly better than traditional neural networks in dealing with complex feature mining. In this process, convolutional neural networks (CNN) are often used for deep mining of aggregate replacement. Some researchers have further proposed a CNN-LSTM-ATT hybrid model based on the LSTM (Long Short-Term Memory) model, with ATT referring to the attention mechanism. Compared with CNN-LSTM, this framework dynamically weights input features by introducing an attention mechanism, achieving high-accuracy prediction for both compressive strength and fatigue life. At the same time, SHAP analysis showed opposite response patterns of recycled aggregate content in the model predictions of compressive strength and fatigue life [156]. However, as far as the current situation is concerned, the application of deep learning in the field of construction solid waste concrete is still insufficient. In particular, the use of image recognition technology to detect the quality of recycled aggregates remains to be explored.

(3) The application of neural network and deep learning model in other solid wastes

For other solid waste, neural network technology is mainly used in biochar and rubber asphalt concrete. From traditional artificial neural networks (ANN) to deep learning models represented by deep neural networks (DNNs) and convolutional neural networks (CNNs), related technologies have been widely used in the field of biomass industrial waste resource utilization, and show good performance prediction and optimization potential. Regarding the effect of palm oil fuel ash (POFA) as a cement substitute on concrete performance, machine learning techniques such as artificial neural networks (ANN), combined input neural networks (ANNX), and random forest (RF) have been applied in existing studies. Through these technologies, multiple mechanical properties of eco-friendly concrete mixed with POFA were predicted and optimized. Consequently, the influence of POFA on concrete properties was comprehensively evaluated [157,158]. The ANN model also shows excellent prediction accuracy in predicting the effectiveness of oil palm shell (OPS) lightweight aggregate concrete uniaxial compressive strength (UCS). In the study, the optimal ANN model with a “6-13-3-1” double hidden layer structure was adopted, demonstrating extremely high reliability and generalization ability. Specifically, the “6–13–3–1” architecture indicates the number of neurons in each layer of the ANN, with 6, 13, 3, and 1 representing the input layer, the first hidden layer, the second hidden layer, and the output layer, respectively. Through this architecture, a good balance between network depth and overfitting was effectively achieved, and the prediction error was

reduced to the lowest level [159]. To predict the compressive strength of rubber concrete, a new method based on a convolutional neural network and long short-term memory network combined with an attention mechanism (CNN-LSTM-ATT) was proposed. In this approach, an enhanced firefly algorithm (EFA) with chaotic initialization was used for hyperparameter optimization. The resulting hybrid model achieved higher accuracy and better stability compared to traditional machine learning methods. Further, SHAP analysis showed that coarse aggregate content, rubber content, and concrete age made relatively large contributions to the model predictions of compressive strength. This study shows that deep learning and bio-inspired optimization are effective for solving complex nonlinear material problems. The above work preliminarily verifies the high efficiency of deep learning and bionic optimization in capturing the complex nonlinear behavior of solid waste concrete. However, due to the small data set, the deep learning method has not been widely expanded to more types of solid waste concrete.

Based on the above studies, the roles of different solid wastes in concrete were associated with differences in their machine learning applications. Several representative studies were selected and are summarized in Table 4. It can be seen that the dataset sizes varied greatly among different studies, indicating that the current machine learning research on solid waste concrete still has an imbalance in data size. In terms of model applications, relatively high prediction accuracy was obtained in most studies, with R^2 values above 0.90 in the test sets or the reported prediction results. However, most existing studies have focused on mechanical properties, especially compressive strength, while relatively few machine learning studies have been conducted on key durability-related properties such as pore structure, chloride penetration, and carbonation. Therefore, future studies should further consider the different roles of solid waste in concrete, extend machine learning applications to multiple performance indicators, and develop more systematic methods for the performance prediction and optimization of solid waste concrete.

Simultaneously, it was found that, compared with conventional concrete, solid waste concrete datasets usually show high dimensionality, strong noise, and high nonlinearity owing to the different sources of waste materials and the variations in their composition. In the complex feature space formed by multi-component synergies, the optimization routes of models were altered, and heuristic swarm intelligence algorithms with global search capabilities were widely introduced to escape local minima traps. As shown in Table 5, the firefly algorithm (FFA), gray wolf optimizer (GWO), bat algorithm (BA), and sandcat swarm algorithm (SCSO) are widely used in SVR, RF, GBDT, and other basic models. Owing to the good positive feedback mechanism and anti-interference capability of these algorithms, high-noise experimental data were processed smoothly. Meanwhile, clear applicability was demonstrated in the global optimization of new material mix proportion spaces and the mapping of macroscopic parameters. In the framework of deep learning, in the face of more complex micro-morphology and long-term evolution of solid waste materials, the improved network based on CNN reflects different emphasis directions. In conventional concrete research, the GRU structure is commonly used. In contrast, the integration of CNN with long short-term memory (LSTM) and attention mechanisms is more frequently preferred in the field of solid waste concrete. In this CNN-LSTM-ATT architecture, key information was focused on through dynamic weighting of the attention mechanism, and complex long-term time-series dependencies were effectively handled by the fine gating mechanism of LSTM. Through this process, the deep fusion of multiple microstructural features was successfully achieved. Consequently, a dynamic joint high-precision prediction for multiple service indicators was realized. The advantages of each optimized algorithm for multi-component solid waste concrete are summarized in Table 5.

Table 4. Machine Learning Applications of Solid Wastes with Different Roles in Concrete.

The Role of Solid Waste	Typical Solid Waste Materials	Optimal Dosage/Substitution Rate	Dataset Size	Input Features	Output Targets	Optimal Model	Predicted Performance
Aggregate	RCA [137]	≈30%	465 (CS); 339 (STS)	Cement, W/B, Sand, CA, r, Vf, Df, Lf	CS; STS	AdaBoost; GBRT	CS: R^2 = 0.924, RMSE = 3.872 MPa, MAE = 2.624 MPa, MAPE = 5.277%; STS: R^2 = 0.946, RMSE = 0.523 MPa
	RCA [138]	<70%	324	Cement, W/B, Sand, CA, r, Vf, Df, Lf	CS; STS	RF; GBRT	CS: R^2 = 0.915, MAPE = 6.1%; STS: R^2 = 0.877, MAPE = 10.8%
	Recycled brick and tile [141]	Coarse aggregate ≤25%; fine aggregate ≤50%	393	C, W, TCA, B-CA, T-CA, O-CA, TFA, B-FiA, T-FiA, O-FiA, Age	CS	XGB	R^2 = 0.985, RMSE = 1.539 MPa, MAE = 0.849 MPa, MAPE = 4%
Binder	Steel slag powder [118]	<20%	168	Cement, SF, FA, Sand, SFiber, Water, SP	CS	SVM	R^2 = 0.88, RMSE = 12.7 MPa, MAE = 9.13 MPa;
	Copper mine tailings [122]	≈15%	1061	C, FA, BFS, W, FiA, CoA, CMT, SP, Age	CS	SVR-FFA	R^2 = 0.96, RMSE = 1.94 MPa, MAE = 0.94 MPa
	SCBA [143]	≈10%	NR	SCBA replacement, Curing age	CS	RF	Overall: R^2 = 0.989; CS: R^2 = 0.9797, RMSE = 0.926 MPa, MAE = 0.672 MPa
	RHA [144]	0–66.67%	348	FiA, SP, W/C, RHA, CA	CS	LightGBM	R = 0.998, RMSE = 1.711 MPa, MAE = 1.188 MPa, RRMSE = 0.029
Fiber	JF [140]	0.5%	NR	W/B, Cement, Waste brick powder, Jute fiber, CA, FiA, Water, Age	CS; STS	RF	R^2 = 0.957
	BNHF [145]	1%	119	BNHF, Fiber content, Fiber length, W/C, Cement, FiA, CA, W, Age	Slump; CS; STS	CatBoost; XGBoost	Slump: CatBoost, R^2 = 0.989, RMSE = 2.116 mm; CS: CatBoost, R^2 = 0.918, RMSE = 1.986 MPa; STS: XGBoost, R^2 = 0.862, RMSE = 0.406 MPa

Abbreviations: C, cement; W, water; CA, coarse aggregate; FiA, fine aggregate; TCA, total coarse aggregate; TFA, total fine aggregate; B-CA, crushed brick in coarse aggregate; T-CA, crushed tile in coarse aggregate; O-CA, other materials in coarse aggregate; B-FiA, crushed brick in fine aggregate; T-FiA, crushed tile in fine aggregate; O-FiA, other materials in fine aggregate; W/B, water-to-binder ratio; W/C, water-to-cement ratio; RCA, recycled coarse aggregate; CMT, copper mine tailings; SCBA, sugarcane bagasse ash; RHA, rice husk ash; BNHF, betel nut husk fiber; FA, fly ash; SF, silica fume; BFS, blast furnace slag; SP, superplasticizer; Vf, fiber volume fraction; Df, fiber diameter; Lf, fiber length; Age, test age; CS, compressive strength; STS, splitting tensile strength; NR, not reported.

Table 5. Comparison of the advantages of the optimized model in the study of solid waste concrete.

Model Category	Specific Model	Introduced Algorithm	Optimized Model	Core Advantages in Solid Waste Concrete Research
Classical Machine Learning	SVR	Firefly Algorithm (FFA)	SVR-FFA [122]	Automatically searches for optimal parameters and enhances nonlinear mapping; suitable for strength prediction with multicomponent solid waste mix proportions.
	SVR	Grey Wolf Optimizer (GWO)	SVR-GWO [122]	Robust with strong local-search capability; suitable for durability assessment and fine-tuning of mix proportions for noisy solid waste concrete data.
	DT/SVR	Particle Swarm Optimization (PSO)/Gradient Boosting (GB)	DT-PSO/SVR-GB [122,124]	Improves parameter stability and generalization when base models are used to model complex nonlinear relationships.
Ensemble Machine Learning	RF	Artificial Bee Colony (ABC)	ABC-RF [131,160]	Simple structure, fast convergence, and high efficiency; suitable for predicting durability and other properties of solid waste concrete.
	AdaBoost	Forensic-Based Investigation Optimization (FBIO)	AdaBoost-FBIO [128]	Optimizes hyperparameters and improves generalization and prediction reliability; suitable for evaluating the mechanical properties of solid waste materials from new sources.
	GBDT/LGBM/CatBoost	Sand Cat Swarm Optimization (SCSO)	SCSO-GBDT/SCSO-LGBM/SCSO-CatBoost [128]	Effective for complex nonlinear degradation prediction; suitable for long-term durability assessment and mechanism analysis of solid waste concrete.
	XGBoost	Meta-heuristic algorithms such as SFO/HGS/HBA/WOA	Heuristic-Optimized XGBoost [133,141]	Reduces overfitting, improves accuracy, and identifies key variables; suitable for nonlinear modeling of multiphase solid waste concrete.
Neural Networks &Deep Learning	BPNN	Bat Algorithm (BA)	BA-BPNN [148,161]	Enhances global search through frequency adjustment; suitable for optimizing multidimensional solid waste concrete mix proportions.
	BPNN	Ant Colony Optimization (ACO)	ACO-BPNN [148,162]	Uses positive pheromone feedback for robust search; suitable for multi-objective mix-proportion and preparation-route optimization of solid waste concrete.
	DNN	Improved Grey Wolf Optimizer (IGWO)	DNN-IGWO [150]	Efficiently searches hyperparameters and network architectures, with high stability and strong generalization; suitable for modeling large solid waste concrete databases.
	CNN	Long Short-Term Memory (LSTM)/Attention Mechanism	NN-LSTM-ATT [156]	Dynamically focuses on key features through the attention mechanism, enabling microstructural-feature fusion and joint prediction of multiple service-related indicators.

In conventional concrete research, a comprehensive data system and testing standards have been accumulated. Consequently, model training was rapidly carried out in a unified and standardized data environment. Although various solid waste materials (such as industrial, construction, and other solid wastes) were introduced, the data organization method of solid waste concrete databases still largely followed the traditional database structure. In this structure, the selection of core variables (such as water-binder ratio, composition of cementitious materials, aggregate particle size distribution, and curing regime) was kept consistent. Based on this, characteristic parameters of solid waste materials were further added. Through this integration, comprehensive data coverage and the optimization of concrete performance evaluation were successfully achieved. The application of machine learning in the field of solid waste concrete is slower than that of conventional concrete, but its model development is also the same as the former. From the initial simple machine learning technology to the current ensemble machine learning, deep learning, and integration of professional knowledge-oriented model, showing a deeper and more complex trend.

5. Research Status Summary, Challenges and Prospects

After years of development, a relatively complete theoretical framework has been established for applying machine learning to conventional concrete. In this framework, ensemble models, hyperparameter optimization techniques, and SHAP value analyses were integrated. Consequently, a smooth transition from classical algorithms to deep learning was achieved and further improved. However, owing to the high noise levels and scarcity of high-quality data in solid waste systems, the application boundaries of classical models and deep learning in solid waste concrete are significantly narrowed. Instead, ensemble models represented by XGBoost and Random Forest, combined with various meta-heuristic algorithms, have been widely recognized as mainstream in this field. Across both domains, most current studies have focused on evaluating the compressive strength of materials, while long-term service performance has been largely neglected. Most current waste management systems for intelligent analysis are designed only for single-component treatment. Collaborative modeling among different types of waste has rarely been studied, leaving a significant gap in practical engineering needs. A comparison of the research depth regarding machine learning applications in traditional and solid waste concrete is presented in Table 6.

As machine learning models are increasingly applied to solid waste concrete, several issues also require further consideration. Due to the high data noise and limited high-quality data in solid waste systems, the direct transfer of these models to new solid waste concrete systems should be treated with caution. Prediction uncertainty, extrapolation risk, and model applicability should be carefully considered. Confidence intervals, Bayesian uncertainty analysis, and ensemble model variance can be used to assess prediction reliability. When new material characteristics or experimental conditions fall outside the range of the training data, model extrapolation may lead to unreliable predictions. Therefore, the applicability domain should be evaluated before model transfer to avoid the blind application of models beyond their validated range.

Table 6. Comparison of research status of machine learning in the field of traditional and solid waste concrete.

	Conventional Concrete Field	Solid Waste Concrete Field
Mainstream Model Paradigm	Experienced three stages: classical foundation, ensemble mainstreaming, and deep extension. The collaborative framework of ensemble models, hyperparameter optimization, and SHAP analysis was most mature; Bayesian optimization and genetic algorithms were mostly preferred for optimization algorithms.	Ensemble models served as the recognized mainstream paradigm. Hyperparameter optimization strategies were more heavily relied on, and meta-heuristic algorithms were widely introduced for global optimization.
Application Positioning of Classical Models	Provided introductory frameworks and baseline methods. In strong non-linear systems such as high-temperature environments, high strength, and fiber reinforcement, fitting capacity and generalization performance were noticeably insufficient.	Application boundaries were narrower than in traditional domains. Independent value remained in industrial solid waste; in construction and other solid wastes, coping with high noise and high variability was difficult.
Focus of Interpretability (SHAP)	SHAP revealed the coupled influence mechanisms of water-binder ratio, curing age, and component compositions.	SHAP emphasized chemical-driven mechanisms in industrial solid waste, while physical-driven mechanisms were prioritized in construction waste.
Application Intensity of Deep Learning	Development was in-depth, extending automatic feature extraction and end-to-end learning. Architectures covered BP networks, CNN (and deep architectures integrating attention mechanisms and GRU).	Constrained by the scarcity of high-quality experimental data, conditions for large-scale applications were temporarily lacking. BP networks and MLP were dominant, while CNN and LSTM were rarely explored.
System Coverage and Data Foundation	Mainly built on data foundations of conventional concrete and single- or multi-component modified concrete, leaving room for expansion into solid waste systems.	Single solid wastes were predominant. Binary blending of identical solid wastes, cross-category solid waste synergy, and systems with three or more components were relatively rare.
Performance Indicators and Research Focus	Highly concentrated on mechanical properties. Shrinkage, creep, and other long-term performance and service life predictions were lacking.	Compressive strength was the mainstream. Predictions for workability and durability were insufficient.

5.1. Potential Challenges

In conventional concrete performance evaluation and mix proportion optimization design, machine learning technology has obvious advantages and has formed a relatively mature design scheme. For recycled concrete from solid waste, there are increasing methods of machine learning assistance; however, the research is still relatively small and has a limited scope. At present, there are two main challenges in the research on waste concrete.

(1) Data insufficiency and heterogeneity

The total amount of experimental data for solid waste concrete was much smaller than that for conventional concrete. Currently, these data are highly concentrated in two systems: fly ash concrete and recycled coarse aggregate concrete. In contrast, data scarcity was clearly observed for solid wastes, such as steel slag, coal gangue, waste glass, and biomass ash. In terms of performance indicators, the compressive strength data are far more than the durability and long-term performance data, and the laboratory data are far more than the engineering measured data, resulting in a serious narrowing of the performance dimension covered by the model. Significant differences were found among various solid wastes in chemical composition, mineral structure, and physical properties. Through this diversity, the material reactivity in cement-based systems was directly affected. Consequently, the variation laws of macroscopic mechanical properties and durability became difficult to grasp. Therefore, universal performance prediction models or unified mix proportion design methods were difficult to establish across different solid waste materials. Even for the same type of solid waste, performance fluctuations can easily occur owing to differences in sources, processing methods, and mixed components. Through these factors, the discreteness and uncertainty of research results were further increased. At the same time, because raw solid waste materials come from diverse sources, consistent standards are often lacking among research teams in material classification, experimental testing

procedures, and data statistical methods. As a result, strong heterogeneity was observed in the collected datasets. This difference is not only reflected in the specific numerical distribution but also in the definition of variables. For example, the representation method of dosage has various forms, such as mass fraction, volume ratio, and substitution rate, which brings practical difficulties to data integration and model training.

In addition, the high heterogeneity of data sources places higher requirements on data splitting and validation in machine learning models. Data from the same literature source, material batch, or experimental study may have similar raw material characteristics, testing conditions, and testing methods. If these data are randomly divided into training and test sets, data from the same source may be included in both sets, resulting in potential data leakage and an overestimation of model performance. Therefore, in future machine learning studies of solid waste concrete, the grouping characteristics of data sources should be considered. Group-based data splitting strategies based on literature source, material batch, or experimental study should be adopted, and cross-validation methods such as leave-one-study-out validation should be considered when appropriate. In this way, the generalization performance of the models can be evaluated more realistically.

(2) The feature system is not unified

In recent years, transfer learning and domain adaptation have gradually been used to address the differences in data distributions and the lack of target data in material property prediction. Previous studies have shown that feature-based, instance-based, and parameter-based domain adaptation methods can improve the prediction performance for out-of-distribution materials to some extent, but their performance is affected by the differences between the source and target domains [163]. Few-shot transfer based on transfer component analysis (TCA) has further shown that knowledge can be transferred between different material systems by reducing the distribution difference between the source and target domains and using a small amount of target-domain data and material-related features [164]. In the field of concrete, parameter transfer combined with data augmentation has been used to transfer knowledge from conventional concrete to recycled aggregate concrete, and improved prediction performance has been achieved under some small-data conditions [165]. Further cross-domain transfer studies have shown that knowledge learned from conventional concrete can be transferred to different recycled aggregate concrete systems. However, the transfer performance is affected by the characteristics of the target system and the transfer strategy, and inappropriate transfer may lead to negative transfer [166]. Overall, transfer learning provides a new approach to address the limited data problem in solid waste concrete, but the large differences between different solid waste systems also increase the difficulty of knowledge transfer.

These material differences are reflected not only in the sources and compositions of solid waste, but also in the data and feature definitions used in different studies. Due to the lack of unified data standards and guidelines, the input variables selected in solid waste concrete studies are also quite different. One type of research tends to use mix proportion parameters such as the water-binder ratio and solid waste content, while the other is more concerned with the microstructure or curing conditions. This inconsistency in feature selection makes the existing models often rely on specific datasets to build, and it is difficult to directly migrate applications in different research or engineering scenarios. Owing to strong data heterogeneity and inconsistent feature definitions, strong generalization performance is difficult to achieve using traditional models and single-source methods. Consequently, the systematic development and engineering applications of solid waste concrete research were restricted to a certain extent.

In view of the above problems of strong data heterogeneity and inconsistent feature systems in the study of solid waste concrete, machine learning can be used not only for

performance prediction but also play an important role in the classification of solid waste materials. Owing to significant differences in chemical composition, physical structure, and reactivity, different solid wastes are often simply categorized based on experience (e.g., fly ash and slag). Through this empirical categorization, essential differences in micro-physicochemical properties among materials are difficult to reveal scientifically. A new technical path for the accurate classification of solid waste materials was provided through the introduction of classification algorithms in machine learning. For example, high-precision boosting algorithms are widely used in practical modeling. Among them, the CatBoost algorithm has integrated an efficient processing mechanism for Categorical Features within itself; therefore, it shows good applicability in the face of complex multiple solid waste classification tasks. Figure 16 shows the basic logic of the algorithm. which takes the chemical composition of solid waste (such as the specific contents of SiO_2 and CaO), the physical characteristics of particles and the activity index as multi-dimensional feature inputs. Driven by these data, the model automatically clusters and reclassifies materials. This allows samples with similar physical and chemical properties to be grouped together for unified processing. Compared with the past practice of rough classification based on sources, this data-driven classification method performs better in objectivity and rationality. For solid waste concrete, the uncertainty caused by material diversity is a core issue. Machine learning provides a more systematic response by starting from the data source, first optimizing the solid waste classification results, and then using the classified results for subsequent performance prediction. The strategy of 'classifying first and then predicting' has practical application prospects in the intelligent construction of solid waste concrete.

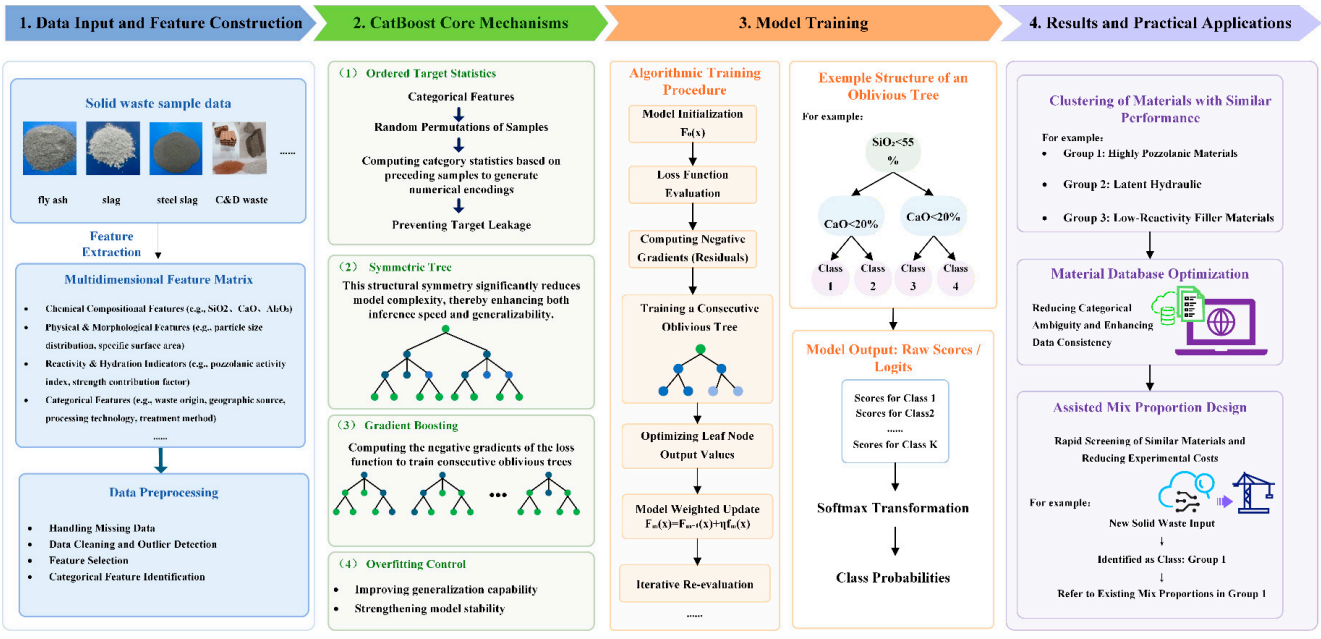


Figure 16. Schematic Diagram of CatBoost Model Application in Solid Waste Concrete [167,168].

5.2. Future Perspectives

Machine learning is reshaping the research model of performance prediction and ratio optimization in solid waste concrete research, and Its advantages in modeling accuracy and computational efficiency have also been shown in existing studies. There are still some key problems to be solved if these algorithm models are to be truly applied to engineering applications and promote the extension of basic theories. Future work can be conducted in the following four directions.

(1) Data processing and model construction of solid waste concrete

The main characteristics of solid waste concrete systems include diverse sources, complex components, and large performance fluctuations. In addition, reaching uniform standards for testing conditions and data formatting across different tests was difficult. Consequently, the generalization performance of the established models was significantly constrained. To address this limitation, future research should focus on cross-dataset knowledge transfer and generalization improvement. Transfer learning can be introduced to transplant modeling experience from mature material systems into new solid waste systems, thereby improving prediction accuracy under small-sample conditions. Furthermore, generative models, such as generative adversarial networks (GAN) and variational autoencoders (VAE), can also be employed. Existing datasets can be reasonably expanded by generating virtual samples that conform to physical statistical distributions, fulfilling the training demands of the models.

Owing to the relatively small size of existing datasets for solid waste concrete, direct application of deep learning architectures with massive parameters can easily cause overfitting. Therefore, a two-step training strategy can be adopted. First, a basic model can be trained on a conventional concrete performance database to capture fundamental physical mechanisms, such as cement hydration and aggregate gradation. Subsequently, the trained model parameters can be transferred to the solid waste concrete database for fine-tuning. For specific solid waste materials, prediction experience from well-researched solid wastes (such as fly ash) can be transferred to other materials with similar physical properties (such as stone powder or waste glass powder). Through this cross-category knowledge sharing, reliance on experimental data for new materials can be significantly reduced, while the generalization adaptability of the model under small-sample scenarios is enhanced. In the future, machine learning predictions can be combined with performance-based design standards for solid waste concrete. According to different service environments and performance requirements, the relationships between solid waste type, replacement level, mix proportions, and target properties can be established. This can help shift concrete mix design from performance prediction to performance-based intelligent design.

(2) Development of Physical Information Neural Network (PINN)

Most existing models are still mainly data-driven, with limited consideration of material physical mechanisms. This limitation may reduce the reliability of model predictions when dealing with complex operating conditions. Because solid waste materials show complex multiscale and multiphysical behavior, existing data-driven algorithms are still not well integrated with these material physical mechanisms. Future research should further embed classical mechanics, chemical reaction kinetics, and durability transport mechanisms into machine learning algorithms, particularly through physics-informed neural networks (PINNs). Using this approach, governing physical laws can be introduced into the model training process and formulated as penalty residuals within a composite loss function to constrain model outputs based on physical laws (Figure 17). For example, hydration kinetics can be embedded into the model to constrain deviations in the degree of hydration and reaction rate from reasonable ranges, thereby describing early-age heat release and hydration development in mass concrete with high-volume fly ash or slag. Under variable curing temperatures, maturity theory can be used to constrain the growth of equivalent age, enabling in situ strength tracking of precast solid-waste members. In addition, for durability prediction, mass conservation and Fickian diffusion equations can be incorporated into the model, and the flux divergence residual can be constrained to ensure continuous chloride transport and strict mass balance in marine recycled aggregate concrete. By incorporating these constitutive relationships, reaction kinetics, and durability transport mechanisms into multi-objective loss objectives, high prediction accuracy can

be maintained while the physical consistency, engineering credibility, and cross-domain extrapolation of the models can be further improved. This approach can provide a more reliable theoretical basis for the intelligent prediction and design of complex solid waste concrete systems.

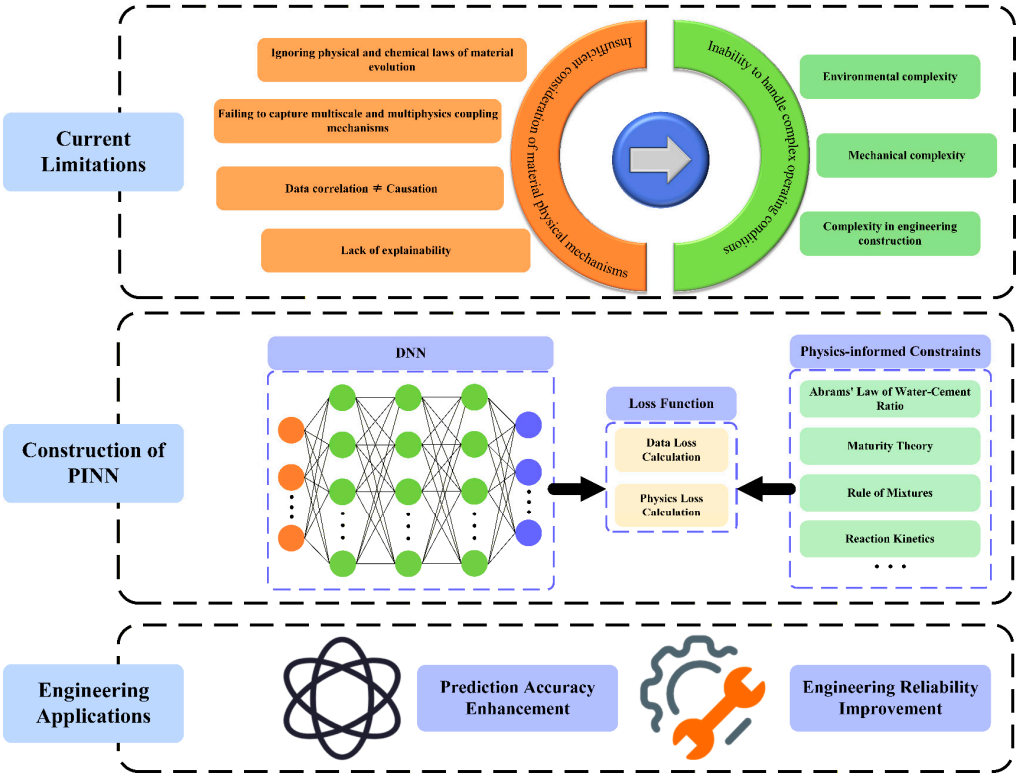


Figure 17. PINN Application Framework Diagram.

(3) Intelligent modeling and design of multi-solid waste collaboration

The synergistic use of multiple solid wastes is an important pathway for the large-scale engineering application of solid waste concrete. Traditional research on multi-solid-waste concrete has mainly used multi-component mix proportion tests. By changing the types and contents of different solid wastes, their effects on the macroscopic properties of concrete have been studied, and microscopic tests have been used to explain the synergistic effects between different solid wastes. Previous studies combined recycled aggregate with fly ash, slag, and silica fume, and found that the proper combination of different solid waste binders could partly compensate for the negative effects of recycled aggregate on strength, shrinkage, and durability [169]. Similarly, the combined use of fly ash, GGBS, and silica fume in recycled aggregate geopolymer concrete was used to improve resistance to acid and sulfate attack. XRD, SEM, and MIP tests were also used to further reveal the densification of the matrix and the formation of gel products [170]. These traditional studies on multi-solid-waste concrete have shown that physical, chemical, and multiscale synergistic effects can occur between different solid wastes. However, traditional experiments and empirical models have limited ability to systematically describe complex relationships among multiple components. Therefore, machine learning is needed to provide a new approach for performance prediction and mix proportion optimization of multi-solid-waste concrete. Faced with the dual challenges of mix design space explosion and unclear cross-category waste interactions, a unified feature representation system needs to be constructed for cross-category solid waste. To bridge different feature spaces between industrial and construction wastes, graph neural networks (GNNs) or transformer architectures can be

employed to build this multi-dimensional system, laying the groundwork for capturing complex coupling relationships. To handle non-additive physicochemical interaction effects among multiple solid wastes, Bayesian optimization and symbolic regression algorithms can be utilized to extract potential mathematical relationships from limited experimental data. Under data scarcity, the learning directions of the models can be constrained by embedding physical prior knowledge. Finally, multi-component waste proportioning can be formulated as a high-dimensional optimization problem. Based on generative models, inverse intelligent design across multiple components can be realized under multi-objective constraints, including strength, durability, cost, and carbon emissions. Future studies should move from single-property optimization toward multi-objective design that considers mechanical performance, durability, carbon emissions, life-cycle environmental impacts, and cost. Such models could provide a more comprehensive basis for the design and application of multi-solid-waste concrete.

(4) Post-hoc interpretation analysis and intelligent monitoring methods

The "black-box" nature of machine learning severely limits its reliability and large-scale adoption in demanding engineering environments. To overcome this limitation, post-hoc explanation techniques should be widely integrated into future solid waste-concrete research. Currently, SHAP is widely applied based on game theory, enabling the precise quantification of positive or negative contributions from each input feature to individual predictions. Through SHAP summary plots, nonlinear relationships between solid waste components and macroscopic properties can be intuitively revealed. Similarly, interpretability tools such as LIME (Local Interpretable Model-agnostic Explanations) and partial dependence plots (PDP) can also be utilized to analyze interactions among solid waste components and macroscopic performance, as illustrated in Figure 18.

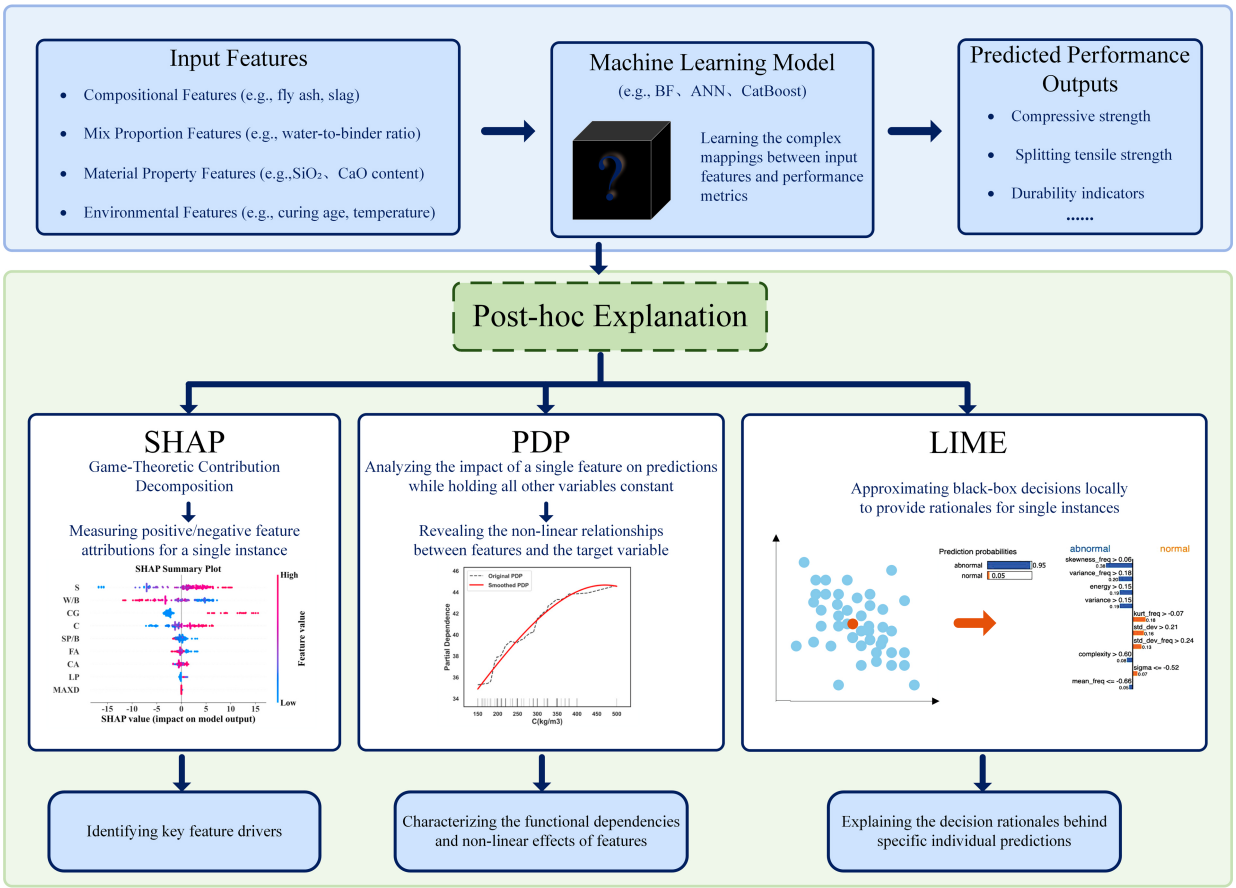


Figure 18. Post-hoc interpretation technical work diagram [171,172].

From an engineering perspective, the performance of existing machine learning models in scenarios such as real-time prediction and in situ monitoring needs to be further improved. Future work can combine real-time data collected by online sensors with dynamically updated algorithms. By constructing dynamic computing architectures capable of handling time-series data—such as stream learning or Long Short-Term Memory networks (LSTMs)—dynamic physical indicators measured in situ (e.g., strain, temperature, and impedance) can be mapped to internal micro-damage evolution and residual load-bearing capacity of concrete. Through this approach, continuous tracking and early failure warnings for the service performance evolution of solid waste concrete can be achieved. This provides reliable, data-driven support for structural health monitoring and digital maintenance decision-making.

Looking ahead, machine learning for solid waste concrete will focus on better generalization, physical constraints, multiscale modeling, and intelligent design. As computing capacity, data sources, and cross-disciplinary studies continue to grow, these methods will offer broader value in the practical design of green-building materials.

Author Contributions: Q.W.: Writing—original draft, Conceptualization, Supervision, Methodology, Data curation, Funding acquisition. M.F.: Writing—original draft, Methodology, Investigation, Visualization. Z.W.: Writing—review & editing, Visualization, Data curation. J.Z.: Writing—review & editing, Supervision, R.W.: Writing—review & editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Shandong Province, China (grant number ZR2026LJT013), the National Natural Science Foundation of China (grant number 52009074), and the Science and Technology Support Plan for Youth Innovation of Colleges and Universities of Shandong Province, China (grant number 2024KJH088).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Perez, O.F.A.; Arrieta, V.S.; Ospina, J.H.G.; Herrera, S.H.; Rojas, C.F.R.; Navarro, A.M.S. Carbon dioxide emissions from traditional and modified concrete. A Review. *Environ. Dev.* **2024**, *52*, 101036. [CrossRef]
2. Zhang, P.; Zhang, L.; Wang, K.; Zhang, T.; Hu, S. A state-of-the-art review on abrasion resistance of concrete. *J. Build. Eng.* **2025**, *111*, 113277. [CrossRef]
3. Boobalan, S.C.; Shereef, M.S.; Saravanaboopathi, P.; Siranjeevi, K. Studies on green concrete-a review. *Int. Conf. Adv. Constr. Mater. Struct. (ICCMS)* **2021**, *65*, 1404–1409. [CrossRef]
4. Nilimaa, J. Smart materials and technologies for sustainable concrete construction. *Dev. Built Environ.* **2023**, *15*, 100177. [CrossRef]
5. IEA. *Technology Roadmap—Low-Carbon Transition in the Cement Industry*; International Energy Agency: Paris, France, 2018.
6. Obrist, M.D.; Kannan, R.; Schmidt, T.J.; Kober, T. Decarbonization pathways of the swiss cement industry towards net zero emissions. *J. Clean. Prod.* **2021**, *288*, 125413. [CrossRef]
7. Nie, S.; Zhou, J.; Yang, F.; Lan, M.; Li, J.; Zhang, Z.; Chen, Z.; Xu, M.; Li, H.; Sanjayan, J.G. Analysis of theoretical carbon dioxide emissions from cement production: Methodology and application. *J. Clean. Prod.* **2022**, *334*, 130270. [CrossRef]
8. Climate Group. *Concretezero Global Policy Priorities 2026*; Climate Group: London, UK, 2026.
9. Ferreira, R.L.S.; Pereira, J.E.S.; Andrade, S.R.S.; de Azevedo, A.R.G.; Carneiro, A.M.P. Utilization of a diatomaceous earth by-product as a supplementary cementitious material: Hydration behavior, strength development and environmental efficiency. *J. Build. Eng.* **2026**, *120*, 115471. [CrossRef]
10. Zhan, H.; Wen, Z.; Hu, Y.; He, K.; Hao, J. Construction paths and development suggestions for china's waste recycling system. *Strateg. Study CAE* **2024**, *26*, 210–222. [CrossRef]
11. Ministry of Ecology and Environment of the People's Republic of China, Comprehensive Solid Waste Management Action Plan. 2026. Available online: https://www.mee.gov.cn/zcwj/gwywj/202601/t20260105_1139840.shtml (accessed on 21 October 2026).

12. Gupta, N.; Siddique, R.; Belarbi, R. Sustainable and greener self-compacting concrete incorporating industrial by-products: A review. *J. Clean. Prod.* **2021**, *284*, 124803. [CrossRef]
13. Ma, S.; Liu, X.; Zhang, Z.; Shao, Y.; Sun, Y.; Li, S.; Du, W.; Zhu, L.; Wang, J. Synergistic utilization of bauxite residue (red mud) and multiple solid wastes for low carbon precast concrete materials: Mechanical performance, mechanism and sustainability assessment. *Green Energy Resour.* **2026**, *4*, 100172. [CrossRef]
14. Chen, H.; Chow, C.L.; Lau, D. Developing green and sustainable concrete in integrating with different urban wastes. *J. Clean. Prod.* **2022**, *368*, 133057. [CrossRef]
15. Jiang, Z.; Gao, W.; Yang, Q.; Li, C.; Ren, Q. Technical principles and approaches for low carbon concrete. *J. Build. Mater.* **2023**, *26*, 1143–1150. [CrossRef]
16. He, X.; Qiao, X.; Yu, P.; Zhou, J.; Niu, J. Effect of industrial solid waste powder on the properties of self-compacting high performance concrete. *Silic. Bull.* **2023**, *42*, 4017–4026. [CrossRef]
17. Ma, Z.; Jiang, H.; Wang, X.; Ming, Y.; Liu, Y. Advances in research and application of shrinkage in industrial solid waste concrete. *Acta Mater. Compos. Sin.* **2026**, *43*, 1343–1367. [CrossRef]
18. Cai, M.; Zhu, H.; Wan, Y.; Zhu, H.; Zhuang, X. Using ultrafine fly ash to achieve low-carbon, high strength and high toughness engineered cementitious composites LC-HSTC. *Case Stud. Constr. Mater.* **2024**, *52*, 1000–1008+1988. [CrossRef]
19. Jiang, J.; Jin, H.; Qiu, W.; Wang, X.; Tian, S.; Huang, R. Research status and prospects of resource utilization of coal fly ash. *Mater. Rep.* **2026**, *40*, 25070212-15. Available online: https://www.mater-rep.com/EN/10.11896/cldb.25070212 (accessed on 21 October 2026).
20. Wang, J.; Qin, Z.; Xu, X.; Wang, W.; Dang, Y.; Guo, S.; Liu, Y. Progress and trends in research on fly ash application in road concrete engineering: A citespace bibliometric review. *J. Inn. Mong. Univ. (Nat. Sci. Ed.)* **2026**, *57*, 102–112. [CrossRef]
21. Xia, Z.; Luo, Y.; Zhang, Y.; Yang, Y.; Chu, F. Performance evaluation for alkali-activated fly ash-slag concrete with recycled coarse aggregate: Compressive properties, co2 emission impacts and cost-benefits. *Constr. Build. Mater.* **2026**, *506*, 144888. [CrossRef]
22. Xue, G.; Qiu, Y.; Yao, W.; Liu, J. Fatigue performance and damage evolution of steel slag fine aggregate concrete with various slag contents under uniaxial compression. *J. Build. Eng.* **2026**, *122*, 115781. [CrossRef]
23. Zhang, Y.; Li, Z.; Gu, X.; Nehdi, M.L.; Marani, A.; Zhang, L. Utilization of iron ore tailings with high volume in green concrete. *J. Build. Eng.* **2023**, *72*, 106585. [CrossRef]
24. Leslie, M. Construction industry innovation takes aim at reducing carbon emissions. *Engineering* **2022**, *19*, 7–10. [CrossRef]
25. Van Damme, H. Concrete material science: Past, present, and future innovations. *Cem. Concr. Res.* **2018**, *112*, 5–24. [CrossRef]
26. Raabe, D.; Mianroodi, J.R.; Neugebauer, J. Accelerating the design of compositionally complex materials via physics-informed artificial intelligence. *Nat. Comput. Sci.* **2023**, *3*, 198–209. [CrossRef] [PubMed]
27. Zhou, T.; Song, Z.; Sundmacher, K. Big data creates new opportunities for materials research: A review on methods and applications of machine learning for materials design. *Engineering* **2019**, *5*, 1017–1026. [CrossRef]
28. Su, X.; Yan, X.; Tsai, C.L. Linear regression. *Wiley Interdiscip. Rev. Comput. Stat.* **2012**, *4*, 275–294. [CrossRef]
29. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 6th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2021.
30. Abdallah, M.; Talib, M.A.; Feroz, S.; Nasir, Q.; Abdalla, H.; Mahfood, B. Artificial intelligence applications in solid waste management: A systematic research review. *Waste Manag.* **2020**, *109*, 231–246. [CrossRef] [PubMed]
31. Chen, N.; Lu, W.; Yang, J.; Li, G. *Support Vector Machine in Chemistry*; World Scientific: Singapore, 2004.
32. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
33. Bansal, M.; Goyal, A.; Choudhary, A. A comparative analysis of k-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decis. Anal. J.* **2022**, *3*, 100071. [CrossRef]
34. Beretta, L.; Santaniello, A. Nearest neighbor imputation algorithms: A critical evaluation. *BMC Med. Inform. Decis. Mak.* **2016**, *16*, 74. [CrossRef] [PubMed]
35. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [CrossRef]
36. Yang, Y.; Wang, H. Random forest-based machine failure prediction: A performance comparison. *Appl. Sci.* **2025**, *15*, 8841. [CrossRef]
37. Mayr, A.; Binder, H.; Gefeller, O.; Schmid, M. The evolution of boosting algorithms. From machine learning to statistical modelling. *Methods Inf. Med.* **2014**, *53*, 419–427. [CrossRef] [PubMed]
38. Mannodi-Kanakkithodi, A.; Pilania, G.; Ramprasad, R. Critical assessment of regression-based machine learning methods for polymer dielectrics. *Comput. Mater. Sci.* **2016**, *125*, 123–135. [CrossRef]
39. Ye, J.; Chow, J.-H.; Chen, J.; Zheng, Z. Stochastic gradient boosted distributed decision trees. In Proceedings of the 18th ACM Conference on Information and Knowledge Management, Hong Kong, China, 2–6 November 2009; pp. 2061–2064. [CrossRef]

40. Xu, C.; Liu, X.; Wang, H.; Li, Y.; Jia, W.; Qian, W.; Quan, Q.; Zhang, H.; Xue, F. A study of predicting irradiation-induced transition temperature shift for rpv steels with xgboost modeling. *Nucl. Eng. Technol.* **2021**, *53*, 2610–2615. [CrossRef]
41. Duan, J.; Asteris, P.G.; Nguyen, H.; Bui, X.-N.; Moayedi, H. A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ica-xgboost model. *Eng. Comput.* **2021**, *37*, 3329–3346. [CrossRef]
42. Fan, J.; Ma, X.; Wu, L.; Zhang, F.; Yu, X.; Zeng, W. Light gradient boosting machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric. Water Manag.* **2019**, *225*, 105758. [CrossRef]
43. Wang, S.C. (Ed.) Artificial neural network. In *Interdisciplinary Computing in Java Programming*; Springer: Boston, MA, USA, 2003; pp. 81–100.
44. Zhao, Z.; Liu, Y.; Lu, Y.; Ji, C.; Lin, C.; Yao, L.; Pu, Z.; de Brito, J. Prediction of properties of recycled aggregate concrete using machine learning models: A critical review. *J. Build. Eng.* **2024**, *90*, 109516. [CrossRef]
45. Duan, J.; Zuo, H.; Bai, Y.; Duan, J.; Chang, M.; Chen, B. Short-term wind speed forecasting using recurrent neural networks with error correction. *Energy* **2021**, *217*, 119397. [CrossRef]
46. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]
47. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems 25, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105. [CrossRef]
48. Fang, S.; Wang, M.; Qi, W.; Zheng, F. Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials. *Comput. Mater. Sci.* **2008**, *44*, 647–655. [CrossRef]
49. Wu, R.; Zeng, L.; Fan, J.; Peng, Z.; Zhao, Y. Composition, heat treatment, microstructure and loading condition based machine learning prediction of creep life of superalloys. *Mech. Mater.* **2023**, *187*, 104819. [CrossRef]
50. Tan, Y.; Wang, X.; Kang, Z.; Ye, F.; Chen, Y.; Zhou, D.; Zhang, X.; Gong, J. Creep lifetime prediction of 9% Cr martensitic heat-resistant steel based on ensemble learning method. *J. Mater. Res. Technol.* **2022**, *21*, 4745–4760. [CrossRef]
51. Wang, J.; Lu, Y.; Wang, X.; Liang, S.; Xiong, J.; Zhen, L.; Liu, L. Target-driven design of high strength yet corrosion resistant medium Mn steel via interpretable machine-learning. *Mater. Des.* **2025**, *260*, 115217. [CrossRef]
52. Xing, J.; Qiao, A.; Zhou, H.; Wu, F.; Huang, Y.; Jia, J.; Tao, H. Inverse design of optical fiber core glasses with multi-objective properties using machine learning. *J. Non-Cryst. Solids* **2025**, *667*, 123732. [CrossRef]
53. Dai, M.; Zhang, F.; Dai, S.; Xing, C.; Xiao, S.; Lv, H.; Tan, Y. Optimizing asphalt mix design through predicting volumetric properties using interpretable machine learning. *Powder Technol.* **2024**, *444*, 119954. [CrossRef]
54. Xu, C.; Chen, X.; Zeng, Q.; Cai, M.; Zhang, W.; Yu, B. A framework of integrating machine learning model and pavement life cycle assessment to optimize asphalt mixture design. *Constr. Build. Mater.* **2025**, *469*, 140481. [CrossRef]
55. Salehi, K.; Rahmani, M.; Atashrouz, S. Machine learning assisted predictions for hydrogen storage in metal-organic frameworks. *Int. J. Hydrogen Energy* **2023**, *48*, 33260–33275. [CrossRef]
56. Thanh, H.V.; Taremsari, S.E.; Ranjbar, B.; Mashhadimoslem, H.; Rahimi, E.; Rahimi, M.; Elkamel, A. Hydrogen Storage on Porous Carbon Adsorbents: Rediscovery by Nature-Derived Algorithms in Random Forest Machine Learning Model. *Energies* **2023**, *16*, 2348. [CrossRef]
57. Zhou, P.; Xiao, X.; Zhu, X.; Chen, Y.; Lu, W.; Piao, M.; Cao, Z.; Lu, M.; Fang, F.; Li, Z.; et al. Machine learning enabled customization of performance-oriented hydrogen storage materials for fuel cell systems. *Energy Storage Mater.* **2023**, *63*, 102964. [CrossRef]
58. Xi, C.; Sztandera, L.; Cartwright, H.M. A neural network approach to prediction of glass transition temperature of polymers. *Int. J. Intell. Syst.* **2008**, *23*, 22–32. [CrossRef]
59. Bhutada, A.; Kumar, S.; Gunasegaram, D.; Alankar, A. Machine learning based methods for obtaining correlations between microstructures and thermal stresses. *Metals* **2021**, *11*, 1167. [CrossRef]
60. Kim, J.M.; Ha, T.; Lee, J.; Lee, Y.-S.; Shim, J.-H. Prediction of pressure-composition-temperature curves of ab2-type hydrogen storage alloys by machine learning. *Met. Mater. Int.* **2023**, *29*, 861–869. [CrossRef]
61. Zhou, H.; Liu, H.; Chen, Z.; Li, J.; Liu, B.; Xiao, L.; Liu, X.; Ling, Y.; Lai, M.; Chen, Z.; et al. Few-shot learning breaks the data bottlenecks in perovskite materials design. *Mater. Today Phys.* **2025**, *59*, 101904. [CrossRef]
62. Li, Z.; Li, S.; Birbilis, N. A machine learning-driven framework for the property prediction and generative design of multiple principal element alloys. *Mater. Today Commun.* **2024**, *38*, 107940. [CrossRef]
63. Zhang, C.; Wang, Y.; Zhang, X.; Ding, Y.; Xu, P. Mechanical properties and microstructure of basalt fiber-reinforced recycled concrete. *J. Clean. Prod.* **2021**, *278*, 123252. [CrossRef]
64. Zhang, J.; Yang, Z.; Li, S.; Peng, H. Investigation on mechanical property adjustment of multi-scale hybrid fiber-reinforced concrete. *Case Stud. Constr. Mater.* **2022**, *16*, e01076. [CrossRef]
65. Grinys, A.; Augonis, A.; Dauksys, M.; Pupeikis, D. Mechanical properties and durability of rubberized and sbr latex modified rubberized concrete. *Constr. Build. Mater.* **2020**, *248*, 118584. [CrossRef]

66. Abdelmonem, A.; El-Feky, M.S.; Nasr, E.-S.A.R.; Kohail, M. Performance of high strength concrete containing recycled rubber. *Constr. Build. Mater.* **2019**, *227*, 116660. [CrossRef]
67. Tawfik, M.; El-said, A.; Deifalla, A.; Awad, A. Mechanical properties of hybrid steel-polypropylene fiber reinforced high strength concrete exposed to various temperatures. *Fibers* **2022**, *10*, 53. [CrossRef]
68. Pham, T.M.; Elchalakani, M.; Hao, H.; Lai, J.; Ameduri, S.; Iran, T.M. Durability characteristics of lightweight rubberized concrete. *Constr. Build. Mater.* **2019**, *224*, 584–599. [CrossRef]
69. Jiang, X.; Zhu, G.; Ai, J.; Zhang, B.; Yan, Y. Mechanical properties of nano-sio2 modified ultra-high-performance concrete. *Vibroeng. Procedia* **2025**, *58*, 229–234. [CrossRef]
70. Zhang, Q.; Li, C.; Li, G.; Chen, D.; Wu, X.; Wang, Y.; Mao, Y.; Zhang, K. The effect of modification with nano-alumina, nano-silica, and polypropylene fiber on the frost resistance of concrete. *Buildings* **2025**, *15*, 4002. [CrossRef]
71. Tanimola, J.O.; Efe, S. Recent advances in nano-modified concrete: Enhancing durability, strength, and sustainability through nano silica (ns) and nano titanium (nt) incorporation. *Appl. Eng. Sci.* **2024**, *19*, 100189. [CrossRef]
72. Nguyen, T.A.; Trinh, S.H.; Ly, H.B. Enhanced bond strength prediction in corroded reinforced concrete using optimized ml models. *Structures* **2024**, *63*, 106461. [CrossRef]
73. Li, Y.; Liu, Q.; Guo, T. Prediction method of concrete compressive strength based on CART-SVR. *Concrete* **2022**, *8*, 40–44. [CrossRef]
74. Li, F.; Rana, M.S.; Qurashi, M.A. Advanced machine learning techniques for predicting concrete mechanical properties: A comprehensive review of models and methodologies, Multiscale and Multidisciplinary Modeling. *Exp. Des.* **2024**, *8*, 110. [CrossRef]
75. Jiang, J.; Lin, J.; Jin, L.; Wang, F.; Liu, Z.; Li, Y.; Lu, Z. Advancements in machine learning for the development of cementitious composites toward an intelligent and green lifecycle: A state-of-the-art review. *Engineering* **2026**, *in press*. [CrossRef]
76. DeRousseau, M.A.; Laftchiev, E.; Kasprzyk, J.R.; Rajagopalan, B.; Srubar, W.V., II. A comparison of machine learning methods for predicting the compressive strength of field-placed concrete. *Constr. Build. Mater.* **2019**, *228*, 116661. [CrossRef]
77. Lyngdoh, G.A.; Zaki, M.; Krishnan, N.M.A.; Das, S. Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning. *Cem. Concr. Compos.* **2022**, *128*, 104414. [CrossRef]
78. Taffese, W.Z.; Espinosa-Leal, L. Prediction of chloride resistance level of concrete using machine learning for durability and service life assessment of building structures. *J. Build. Eng.* **2022**, *60*, 105146. [CrossRef]
79. Gad, A.G. Particle swarm optimization algorithm and its applications: A systematic review. *Arch. Comput. Methods Eng.* **2022**, *29*, 2531–2561. [CrossRef]
80. Zhu, M.; Lin, J.; Cao, G.; Zhang, J.; Zhang, X.; Zhou, J.; Gao, Y. Prediction of constitutive model for basalt fiber reinforced concrete based on PSO-KNN. *Heliyon* **2024**, *10*, e32240. [CrossRef] [PubMed]
81. Li, S.; Liew, J.Y.R. Experimental and data-driven analysis on compressive strength of steel fibre reinforced high strength concrete and mortar at elevated temperature. *Constr. Build. Mater.* **2022**, *341*, 127845. [CrossRef]
82. Kang, M.C.; Yoo, D.Y.; Gupta, R. Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete. *Constr. Build. Mater.* **2021**, *266*, 121117. [CrossRef]
83. Zhang, H.; Wang, X.; Yan, C.; Dong, Y. Prediction of thermal insulation performance of concrete lining with multi-component solid waste aggregate under freeze-thaw action based on machine learning. *Constr. Build. Mater.* **2026**, *512*, 145396. [CrossRef]
84. Tuan, N.S.; Wakim, J.; Dong, T.Q.; Ngoc, V.M.; Duong, N.T.; Trung, N.T. Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. *Constr. Build. Mater.* **2020**, *260*, 119757. [CrossRef]
85. Amin, M.N.; Khan, K.; Sufian, M.; Al-Ahmad, Q.M.S.; Deifalla, A.F.; Alsharari, F. Predicting parameters and sensitivity assessment of nano-silica-based fiber-reinforced concrete: A sustainable construction material. *J. Mater. Res. Technol.* **2023**, *23*, 3943–3960. [CrossRef]
86. Albaijan, I.; Fakhri, D.; Mohammed, A.H.; Mahmoodzadeh, A.; Ibrahim, H.H.; Elhag, A.B.; Rashidi, S. Several machine learning models to estimate the effect of an acid environment on the effective fracture toughness of normal and reinforced concrete. *Theor. Appl. Fract. Mech.* **2023**, *126*, 103999. [CrossRef]
87. Zhang, S.; Xu, J.; Lai, T.; Yu, Y.; Xiong, W. Bond stress estimation of profiled steel-concrete in steel reinforced concrete composite structures using ensemble machine learning approaches. *Eng. Struct.* **2023**, *294*, 116725. [CrossRef]
88. Kim, D.; Oh, S.; Choi, S.; Hong, G. Machine-learning-based prediction of the compressive strength of superabsorbent polymer-incorporated cementitious materials: Modeling and experimental insights into polymer characteristics. *Dev. Built Environ.* **2026**, *26*, 100919. [CrossRef]
89. Akber, M.Z.; Anwar, G.A.; Chan, W.-K.; Lee, H.-H. Tpe-xgboost for explainable predictions of concrete compressive strength considering compositions, and mechanical and microstructure properties of testing samples. *Constr. Build. Mater.* **2024**, *457*, 139398. [CrossRef]
90. Yu, R.; Chen, F.; Fan, D.; Xu, W.; Zhang, L.; Li, W.; Ji, D. Machine learning-based performance prediction and precision design of ultra-high performance concrete. *J. Chin. Ceram. Soc.* **2025**, *53*, 1165–1176. [CrossRef]

91. Cavaleri, L.; Barkhordari, M.S.; Repapis, C.C.; Armaghani, D.J.; Ulrikh, D.V.; Asteris, P.G. Convolution-based ensemble learning algorithms to estimate the bond strength of the corroded reinforced concrete. *Constr. Build. Mater.* **2022**, *359*, 129504. [CrossRef]
92. Zhang, P.; Zhang, D.; Zhao, Q.; Tuerxunmaimaiti, Y.; Yang, X.; Zhao, X.; Iqbal, M. Machine learning-guided interpretable modelling of code-based equation for bond strength of composite bars in ultra-high-performance concrete. *Adv. Eng. Inform.* **2026**, *73*, 104542. [CrossRef]
93. Zou, P.; Ke,, G.; Song, B.; Fang, Y. Experiment and its prediction artificial neural networks model study on the compressive strength of coal gangue powder concrete. *Concrete* **2018**, *12*, 31–33. [CrossRef]
94. Chen, Q.; Ma, R.; Jiang, Z.; Wang, H. Compressive strength prediction and mix proportion design of uhpc based on ga-bp neural network. *J. Build. Mater.* **2020**, *23*, 176–183+191.
95. Zheng, Z.; Tian, C.; Wei, X.; Zeng, C. Numerical investigation and ann-based prediction on compressive strength and size effect using the concrete mesoscale concretization model. *Case Stud. Constr. Mater.* **2022**, *16*, e01056. [CrossRef]
96. Xue, J.; Shen, B. A novel swarm intelligence optimization approach: Sparrow search algorithm. *Syst. Sci. Control Eng.* **2020**, *8*, 22–34. [CrossRef]
97. Wang, K.; Ren, J.; Yan, J.; Wu, X.; Dang, F. Research on a concrete compressive strength prediction method based on the random forest and lcssa-improved bp neural network. *J. Build. Eng.* **2023**, *76*, 107150. [CrossRef]
98. Huang, L.; Chen, J.; Tan, X. Bp-ann based bond strength prediction for frp reinforced concrete at high temperature. *Eng. Struct.* **2022**, *257*, 114026. [CrossRef]
99. Chen, W.; Fediuk, R.; Yu, J.; Nikolayevich, K.; Vatin, N.; Bazarov, D.; Yu, K. Performance prediction and analysis of engineered cementitious composites based on machine learning. *Dev. Built Environ.* **2024**, *18*, 100459. [CrossRef]
100. Bangaru, S.S.; Wang, C.; Zhou, X.; Hassan, M. Scanning electron microscopy (sem) image segmentation for microstructure analysis of concrete using u-net convolutional neural network. *Autom. Constr.* **2022**, *144*, 104602. [CrossRef]
101. Tanimola, J.O.; Efe, S. Physics-informed artificial intelligence framework for automated characterization of cement hydration microstructures from X-ray μCT images. *Appl. Sci.* **2026**, *16*, 8632. [CrossRef]
102. Celik, G.; Ozdemir, M. Determination of concrete compressive strength from surface images with the integration of cnn and svr methods. *Measurement* **2024**, *238*, 115331. [CrossRef]
103. Andeobu, L.; Wibowo, S.; Grandhi, S. Artificial intelligence applications for sustainable solid waste management practices in australia: A systematic review. *Sci. Total Environ.* **2022**, *834*, 155389. [CrossRef]
104. He, X.; Guo, J.; Shen, D.; Lin, R. Application of machine learning in anaerobic digestion of organic solid waste: Recent advances and prospects. *J. Southeast Univ. (Nat. Sci. Ed.)* **2025**, *55*, 743–750.
105. Lu, J.; Chen, J.; Lee, C.-H.; Li, Y.; Ruan, H.D.; Jiang, S.Y. Feasibility of fly ash-based geopolymer as a sustainable alternative to cement in concrete production. *Constr. Build. Mater.* **2026**, *509*, 145143. [CrossRef]
106. Zhang, Q.; Zou, C.; Guo, X.; Song, Y.; Zheng, J.; Zhao, Y. Improvement of microstructure and coupling wind erosion with freeze-thaw durability of fly ash concrete by sodium hydroxide. *J. Southeast Univ. (Nat. Sci. Ed.)* **2024**, *54*, 131–140.
107. Tung, T.M.; Babalola, O.E.; Le, D.-H. Experimental investigation of the performance of ground granulated blast furnace slag blended recycled aggregate concrete exposed to elevated temperatures. *Clean. Waste Syst.* **2023**, *4*, 100069. [CrossRef]
108. Jiang, J.; Liu, C.; Liu, A.; Kong, D.; Wang, L. Experimental study on impermeability of steel slag fine aggregate red mud-based concrete. *Non-Met. Mines* **2022**, *45*, 27–29.
109. Bai, Q.; Deng, W.; Lu, F.; Zhai, J.; Lai, H.; Mao, N.; Shi, C.; Xiong, R. Research progress on effect of steel slag on durability of plateau concrete. *Bull. Chin. Ceram. Soc.* **2023**, *42*, 4332–4340+4350. [CrossRef]
110. Ye, C. Influence of steel slag activator on mechanical property and durability of offshore concrete. *J. Highw. Transp. Res. Dev.* **2024**, *41*, 106–112.
111. Zhao, J.; Ni, K.; Su, Y.; Shi, Y. An evaluation of iron ore tailings characteristics and iron ore tailings concrete properties. *Constr. Build. Mater.* **2021**, *286*, 122968. [CrossRef]
112. Zhang, T.; Wang, H.; Tang, J.; Gao, S. Mechanical and environmental performance of structural concrete with coal gangue fine aggregate. *J. Build. Eng.* **2024**, *84*, 108488. [CrossRef]
113. Li, J.; Wang, J. Comprehensive utilization and environmental risks of coal gangue: A review. *J. Clean. Prod.* **2019**, *239*, 117946. [CrossRef]
114. Silva, R.V.; de Brito, J.; Dhir, R.K. Properties and composition of recycled aggregates from construction and demolition waste suitable for concrete production. *Constr. Build. Mater.* **2014**, *65*, 201–217. [CrossRef]
115. Aprianti, E.; Shafigh, P.; Bahri, S.; Farahani, J.N. Supplementary cementitious materials origin from agricultural wastes—A review. *Constr. Build. Mater.* **2015**, *74*, 176–187. [CrossRef]
116. Azimi-Pour, M.; Eskandari-Naddaf, H.; Pakzad, A. Linear and non-linear svm prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete. *Constr. Build. Mater.* **2020**, *230*, 117021. [CrossRef]
117. Peng, Y.; Unluer, C. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques. *Constr. Build. Mater.* **2022**, *316*, 125785. [CrossRef]

118. Sun, G.; Du, M.; Shan, B.; Shi, J.; Qu, Y. Ultra-high performance concrete design method based on machine learning model and steel slag powder. *Case Stud. Constr. Mater.* **2022**, *17*, e01682. [CrossRef]
119. Raju, M.R.; Ahmad, S.I.; Hasan, M.M.; Hasan, N.M.S.; Islam, M.M.; Basit, M.A.; Hossain, I.T.; Santo, S.A.; Alam, M.S.; Rahman, M. Machine learning for predicting strength properties of waste iron slag concrete. *Heliyon* **2025**, *11*, e42133. [CrossRef] [PubMed]
120. Dong, J.F.; Xu, Y.; Guan, Z.W.; Wang, Q.Y. Freeze-thaw behaviour of basalt fibre reinforced recycled aggregate concrete filled CFRP tube specimens. *Eng. Struct.* **2022**, *273*, 115088. [CrossRef]
121. Kamboj, V.K.; Bath, S.K.; Dhillon, J.S. Solution of non-convex economic load dispatch problem using Grey Wolf Optimizer. *Neural Comput. Appl.* **2016**, *27*, 1301–1316. [CrossRef]
122. Alyami, M.; Onyelowe, K.; AlAteah, A.H.; Alahmari, T.S.; Alsubeai, A.; Ullah, I.; Javed, M.F. Innovative hybrid machine learning models for estimating the compressive strength of copper mine tailings concrete. *Case Stud. Constr. Mater.* **2024**, *21*, e03869. [CrossRef]
123. Khan, A.; Manan, A.; Umar, M.; Mehmood, M.; Onyelowe, K.C.; Arunachalam, K.P. Enhancing concrete strength for sustainability using a machine learning approach to improve mechanical performance. *Sci. Rep.* **2025**, *15*, 23067. [CrossRef] [PubMed]
124. Essam, A.; Alanazi, N.; Elrahman, M.A.; Mostafa, S.A.; Khan, M.; Tahwia, A.M. Waste-to-resource design of eco-friendly ultra-high-performance concrete using reprocessed glass powder: Life cycle sustainability assessment and durability prediction. *Case Stud. Constr. Mater.* **2025**, *23*, e05617. [CrossRef]
125. Bassi, A.; Manchanda, A.; Singh, R.; Patel, M. A comparative study of machine learning algorithms for the prediction of compressive strength of rice husk ash-based concrete. *Nat. Hazards* **2023**, *118*, 209–238. [CrossRef]
126. Amlashi, A.T.; Golafshani, E.M.; Ebrahimi, S.A.; Behnood, A. Estimation of the compressive strength of green concretes containing rice husk ash: A comparison of different machine learning approaches. *Eur. J. Environ. Civ. Eng.* **2023**, *27*, 961–983. [CrossRef]
127. Li, Q.; Song, Z. High-performance concrete strength prediction based on ensemble learning. *Constr. Build. Mater.* **2022**, *324*, 126694. [CrossRef]
128. Ghanizadeh, A.R.; Amlashi, A.T.; Dessouky, S. A novel hybrid adaptive boosting approach for evaluating properties of sustainable materials: A case of concrete containing waste foundry sand. *J. Build. Eng.* **2023**, *72*, 106595. [CrossRef]
129. Wang, M.; Mitri, H.S.; Zhao, G.; Wu, J.; Xu, Y.; Liang, W.; Wang, N. Performance comparison of several explainable hybrid ensemble models for predicting carbonation depth in fly ash concrete. *J. Build. Eng.* **2024**, *98*, 111246. [CrossRef]
130. Karaboga, D.; Basturk, B. On the performance of artificial bee colony (ABC) algorithm. *Appl. Soft Comput.* **2008**, *8*, 687–697. [CrossRef]
131. Bali, M.; Mishra, V.P.; Yenkikar, A. Artificial bee colony optimized random forest model for prediction of fly ash concrete compressive strength. *MethodsX* **2025**, *14*, 103412. [CrossRef] [PubMed]
132. Bahram, M.; Khosravi, H. Machine learning-based prediction of geopolymer concrete compressive strength using boosting and svr models. *Appl. Eng. Sci.* **2026**, *25*, 100307. [CrossRef]
133. Wang, Q.; Dai, R.; Zhang, H.; Zheng, H.; Liang, X. Machine learning-based prediction method for drying shrinkage of recycled aggregate concrete. *J. Build. Eng.* **2024**, *96*, 110493. [CrossRef]
134. Liu, K.-H.; Zheng, J.-K.; Pacheco-Torgal, F.; Zhao, X.-Y. Innovative modeling framework of chloride resistance of recycled aggregate concrete using ensemble-machine-learning methods. *Constr. Build. Mater.* **2022**, *337*, 127613. [CrossRef]
135. Nunez, I.; Nehdi, M.L. Machine learning prediction of carbonation depth in recycled aggregate concrete incorporating SCMs. *Constr. Build. Mater.* **2021**, *287*, 123027. [CrossRef]
136. Li, Y.; Jin, K.; Lin, H.; Shen, J.; Shi, J.; Fan, M. Analysis and prediction of freeze-thaw resistance of concrete based on machine learning. *Mater. Today Commun.* **2024**, *39*, 108946. [CrossRef]
137. Zhang, S.; Chen, W.; Xu, J.; Xie, T. Use of interpretable machine learning approaches for quantificationally understanding the performance of steel fiber-reinforced recycled aggregate concrete: From the perspective of compressive strength and splitting tensile strength. *Eng. Appl. Artif. Intell.* **2024**, *137*, 109170. [CrossRef]
138. Yousafzai, M.H.; Javed, M.F.; Rehan, M.; Jameel, M.; Alabduljabbar, H.; Ahmad, F. Application of ensemble learning, deep neural networks, and explainable ai in predicting compressive and split tensile strength of steel fiber-reinforced recycled aggregate concrete. *Case Stud. Constr. Mater.* **2025**, *23*, e05392. [CrossRef]
139. Manan, A.; Pu, Z.; Ahmad, J.; Umar, M. Multi-targeted strength properties of recycled aggregate concrete through a machine learning approach. *Eng. Comput.* **2025**, *42*, 388–430. [CrossRef]
140. Islam, M.H.; Datta, S.D.; Mondol, S.; Sobuz, M.H.R.; Ashraf, J.; Aayaz, R.; Akid, A.S.M.; Kabbo, M.K.I. Experimental and machine learning assessment of sustainable natural fiber reinforced concrete incorporating waste brick powder as an alternative binder materials. *Ain Shams Eng. J.* **2025**, *16*, 103755. [CrossRef]
141. Ly, H.B.; Nguyen, T.A. Enhancing compressive strength and sustainability: A machine learning approach to recycled brick and tile concrete mix design. *Mater. Today Commun.* **2024**, *40*, 109642. [CrossRef]
142. Yehia, S.A.; Shahin, R.I.; Fayed, S. Compressive behavior of eco-friendly concrete containing glass waste and recycled concrete aggregate using experimental investigation and machine learning techniques. *Constr. Build. Mater.* **2024**, *436*, 137002. [CrossRef]

143. Sobuz, M.H.R.; Al, I.; Datta, S.D.; Jabin, J.A.; Aditto, F.S.; Hasan, N.M.S.; Hasan, M.; Zaman, A.A.U. Assessing the influence of sugarcane bagasse ash for the production of eco-friendly concrete: Experimental and machine learning approaches. *Case Stud. Constr. Mater.* **2024**, *20*, e02839. [CrossRef]
144. Alyami, M.; Nassar, R.-U.-D.; Khan, M.; Hammad, A.W.A.; Alabduljabbar, H.; Nawaz, R.; Fawad, M.; Gamil, Y. Estimating compressive strength of concrete containing rice husk ash using interpretable machine learning-based models. *Case Stud. Constr. Mater.* **2024**, *20*, e02901. [CrossRef]
145. Hasan, M.; Hassan, M.S.; Hasan, K.; Tushar, F.H.; Khan, M.; Jaya, R.P. Interpretable machine learning approach for predicting the workability and mechanical properties of betel nut husk fiber-reinforced concrete. *Structures* **2025**, *81*, 110152. [CrossRef]
146. Khan, H.; Ali, M.M.; Ahmad, J.; Irfan, S. Multi-objective optimisation of sustainable concrete incorporating solar module waste using an integrated machine learning and nsga-ii framework. *Case Stud. Constr. Mater.* **2026**, *24*, e06107. [CrossRef]
147. Sahoo, S.; Mahapatra, T.R. Ann modeling to study strength loss of fly ash concrete against long term sulphate attack. *Int. Conf. Adv. Mater. Manuf. Appl.* **2017**, *5*, 24595–24604. [CrossRef]
148. Feng, Q.; Xie, X.; Wang, P.; Qiao, H.; Zhang, Y.; Ma, Y. Prediction of durability of reinforced concrete based on hybrid-bp neural network. *Constr. Build. Mater.* **2024**, *425*, 136091. [CrossRef]
149. Cao, W.; Ge, R.; Jiang, C.; Zhu, L.; Bai, Q. Application of multiple machine learning models integrated with shap analysis in predicting the compressive strength of molybdenum tailings concrete. *Mater. Today Commun.* **2026**, *51*, 114853. [CrossRef]
150. Laouissi, A.; Benkhelladi, A.; Boumaaza, M.; Karmi, Y.; Hani, M.; Belaadi, A.; Zaitri, R.; Alshaikh, I.M.H.; Ghernaout, D.; Chetbani, Y. Deep neural network modeling of the properties of sustainable high-performance concrete from industrial waste materials. *Results Eng.* **2025**, *27*, 106818. [CrossRef]
151. Kumar, M.; Singh, S.; Kim, S.; Anand, A.; Pandey, S.; Hasnain, S.M.M.; Ragab, A.E.; Deifalla, A.F. A hybrid model based on convolution neural network and long short-term memory for qualitative assessment of permeable and porous concrete. *Case Stud. Constr. Mater.* **2023**, *19*, e02254. [CrossRef]
152. 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]
153. Hammoudi, A.; Moussaceb, K.; Belebchouche, C.; Dahmoune, F. Comparison of artificial neural network (ANN) and response surface methodology (RSM) prediction in compressive strength of recycled concrete aggregates. *Constr. Build. Mater.* **2019**, *209*, 425–436. [CrossRef]
154. Cai, R.; Han, T.; Liao, W.; Huang, J.; Li, D.; Kumar, A.; Ma, H. Prediction of surface chloride concentration of marine concrete using ensemble machine learning. *Cem. Concr. Res.* **2020**, *136*, 106164. [CrossRef]
155. Moein, M.M.; Rahmati, K.; Moein, A.M.; Rigby, S.E.; Saradar, A.; Karakouzian, M. Utilizing construction and demolition waste in concrete as a sustainable cement substitute: A comprehensive study on behavior under short-term dynamic and static loads via laboratory and numerical analysis. *J. Build. Eng.* **2024**, *97*, 110778. [CrossRef]
156. Zhang, Y.; Jiang, Y.; Li, C.; Bai, C.; Zhang, F.; Li, J.; Guo, M. Prediction of cement-stabilized recycled concrete aggregate properties by cnn-lstm incorporating attention mechanism. *Mater. Today Commun.* **2025**, *42*, 111137. [CrossRef]
157. Hasan, N.M.S.; Sobuz, M.H.R.; Shaurdho, N.M.N.; Meraz, M.M.; Datta, S.D.; Aditto, F.S.; Kabbo, M.K.I.; Miah, M.J. Eco-friendly concrete incorporating palm oil fuel ash: Fresh and mechanical properties with machine learning prediction, and sustainability assessment. *Heliyon* **2023**, *9*, e22296. [CrossRef] [PubMed]
158. Kellouche, Y.; Tayeh, B.A.; Chetbani, Y.; Zeyad, A.M.; Mostafa, S.A. Comparative study of different machine learning approaches for predicting the compressive strength of palm fuel ash concrete. *J. Build. Eng.* **2024**, *88*, 109187. [CrossRef]
159. Soleymani, A.; Eidgahee, D.R.; Tamimi, A.; Jahangir, H.; Hasani, H.; Nehdi, M.L. Machine learning assessment of mechanical properties of oil palm shell concrete. *Mater. Today Commun.* **2025**, *49*, 114239. [CrossRef]
160. Tian, H.; Zhang, Y.; Feng, Q.; Qiao, H. Performance analysis and strength prediction of polymer-modified manufactured sand concrete using artificial bee colony optimization algorithm. *Mater. Today Commun.* **2025**, *49*, 113854. [CrossRef]
161. Zaini, N.; Malek, M.A.; Yusoff, M.; Osmi, S.F.C.; Mardi, N.H.; Norhisham, S. Bat algorithm and neural network for monthly streamflow prediction. In Proceedings of the 4th International Conference on Green Design and Manufacture (IConGDM), Ho Chi Minh, Vietnam, 29–30 April 2018; Volume 2030. [CrossRef]
162. Zhang, L.; Sun, Z.; Zhang, C.; Dong, F.; Wei, P. Numerical investigation of the dynamic responses of long-span bridges with consideration of the random traffic flow based on the intelligent aco-bpnn model. *IEEE Access* **2018**, *6*, 28520–28529. [CrossRef]
163. Hu, J.; Liu, D.; Fu, N.; Dong, R. Realistic material property prediction using domain adaptation based machine learning. *Digit. Discov.* **2024**, *3*, 300–312. [CrossRef]
164. Liu, Y.; Wei, S.-Z.; Jiang, T. Property prediction for high-chromium high-vanadium steel based on transfer component analysis with few-shot guided. *J. Mater. Res. Technol.* **2023**, *24*, 9754–9764. [CrossRef]
165. Ford, E.; Maneparambil, K.; Kumar, A.; Sant, G.; Neithalath, N. Transfer (machine) learning approaches coupled with target data augmentation to predict the mechanical properties of concrete. *Mach. Learn. Appl.* **2022**, *8*, 100271. [CrossRef]

166. Wang, C.; Chu, W.; Liu, J.; Zhao, Y.; Liu, F.; Zhang, Y.; Rashidi, M.; Yu, Y. Unified axial capacity prediction of FRP-confined concrete across natural and recycled aggregate systems using a Kolmogorov-Arnold networks based cross-domain transfer-learning framework. *Structures* **2026**, *90*, 112255. [CrossRef]
167. Wei, H.; Wan, H.; Yuan, S.; Liu, G.; Teng, J.; Shi, N.; Liu, Z. A new gelling material: Properties of recycled aggregate concrete under conditions of complete cement replacement using steel slag, ore slag, and fly ash. *Constr. Build. Mater.* **2025**, *464*, 140180. [CrossRef]
168. Akduman, S.; Aktepe, R.; Aldemir, A.; Ozcelikci, E.; Alam, B.; Sahmaran, M. Opportunities and challenges in constructing a demountable precast building using c&d waste-based geopolymer concrete: A case study in türkiye. *J. Clean. Prod.* **2024**, *434*, 139976. [CrossRef]
169. Guo, Z.; Jiang, T.; Zhang, J.; Kong, X.; Chen, C.; Lehman, D.E. Mechanical and durability properties of sustainable self-compacting concrete with recycled concrete aggregate and fly ash, slag and silica fume. *Constr. Build. Mater.* **2020**, *231*, 117115. [CrossRef]
170. Singh, R.P.; Vanapalli, K.R.; Jadda, K.; Mohanty, B. Durability assessment of fly ash, GGBS, and silica fume based geopolymer concrete with recycled aggregates against acid and sulfate attack. *J. Build. Eng.* **2024**, *82*, 108354. [CrossRef]
171. Zhang, J.; Wang, Z.; Shen, S.; Sheng, S.; He, H.; He, C. Hybrid explainable machine learning models with metaheuristic optimization for performance prediction of self-compacting concrete. *Buildings.* **2026**, *16*, 225. [CrossRef]
172. Zereen, A.N.; Das, A.; Uddin, J. Machine fault diagnosis using audio sensors data and explainable ai techniques-lime and shap. *Comput. Mater. Contin.* **2024**, *80*, 3463–3484. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.