BILINGUAL PUBLISHING GROUP

Journal of Building Material Science

https://journals.bilpubgroup.com/index.php/jbms

ARTICLE

Interpretable Machine Learning for Wear Depth Prediction in Sustainable Fly Ash Concrete

Payam Ismael Abdulrahman [1] **, *Hemin Said Mohammed Amin*** [2] **, *Ahmed Salih Mohammed*** [3*]

[1] *Civil Engineering Department, Tishk International University, Erbil 44001, Iraq; payam.ismael@tiu.edu.iq*

[2] *Geomatics Department, College of Engineering, Salahaddin University-Erbil, Erbil 44001, Iraq; hemin.mohammedamin@su.edu.krd*

[3] *Civil Engineering Department, College of Engineering, University of Sulaimani, Sulaymaniyah 46001, Iraq*

ABSTRACT

Wear depth is a key indicator of concrete durability under abrasive conditions, but traditional testing methods are labor-intensive, expensive, and slow. This research introduces a materials informatics framework that uses statistical models and machine learning to predict wear depth in fly ash concrete. A database with 439 concrete mixes was assembled from published sources. Input variables included cement, water, fly ash, aggregates, superplasticizer, air-entraining agent, curing age, and testing duration, with wear depth as the output. Four models were created and evaluated: nonlinear regression (NLR), multilinear regression (MLR), full quadratic (FQ) regression, and an artificial neural network (ANN). The models' effectiveness was measured by R^2, RMSE (root mean squared error), and MAE (mean absolute error). The ANN outperformed others, with R^2 around 0.99, RMSE of 0.097 mm, and MAE of 0.071 mm. Sensitivity analysis showed testing duration and curing age as the most influential factors, emphasizing the need to consider both when estimating wear depth. The results suggest the framework can quickly estimate abrasion resistance and aid in evaluating fly ash concrete mixes. Instead of replacing traditional testing, it provides engineers with a practical tool for mixture assessment and durability analysis based on the existing database.

*CORRESPONDING AUTHOR:

Ahmed Salih Mohammed, Civil Engineering Department, College of Engineering, University of Sulaimani, Sulaymaniyah 46001, Iraq; Email: ahmed.mohammed@univsul.edu.iq

ARTICLE INFO

Received: 18 July 2026 | Revised: 6 August 2026 | Accepted: 26 August 2026 | Published Online: 23 October 2026
DOI: https://doi.org/10.30564/jbms.v8i4.117236

CITATION

Abdulrahman, P.I., Amin, H.S.M., Mohammed, A.S., 2026. Interpretable Machine Learning for Wear Depth Prediction in Sustainable Fly Ash Concrete. Journal of Building Material Science. 8(4): 57–81. DOI: https://doi.org/10.30564/jbms.v8i4.117236

COPYRIGHT

Copyright © 2026 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (https://creativecommons.org/licenses/by-nc/4.0/).

Keywords: Materials Informatics; Sustainable Fly Ash Concrete; Wear Depth; Machine Learning; Artificial Neural Network; Durability Prediction

1. Introduction

Durability has always been a key concern with regard to concrete-based works. Wear depth in concrete is a fundamental factor when overall concrete durability is in question [1]. Mahmood et al. [2] developed a data-driven framework to predict the wear depth of fly ash concrete using a literature-based database comprising 391 experimental observations. The study evaluated four predictive models, including linear regression (LR), pure quadratic (PQ), interaction (IN), and M5P-tree, using concrete mixture proportions, curing age, and abrasion testing duration as input variables. Model performance was assessed using statistical indicators such as R^2, RMSE, MAE, scatter index (SI), objective function (OBJ), and a-20 index. Among the developed models, the M5P-tree demonstrated the highest prediction accuracy, indicating to capture the nonlinear relationships governing concrete wear behavior. Sensitivity analysis further revealed that abrasion testing duration and curing age were the most influential factors affecting wear depth. The study concluded that machine learning techniques can accurately estimate concrete wear depth and serve as efficient decision-support tools, reducing the need for extensive laboratory testing while facilitating the design of durable and sustainable fly ash concrete [2]. Abrasive properties include wear depth and subsequent mass loss. It relies mainly on the paste. Destructive and costly laboratory tests are often employed to determine the wear depth. Machine learning techniques have been widely adopted recently to reduce the cost and time required for testing [1,3]. Construction has been one of the most rapidly growing global industries, which in turn increases the demand for building materials, the most common of which are cement and concrete. Cement is the basic constituent that acts as a binding medium through hydration reactions, producing calcium-silicate-hydrate phases that act as a gel and bind all constituents (aggregates + water + others) into a single cohesive mass. The strength and durability of concrete are directly related to proportions of constituent materials. An increased water-cementitious material ratio (w/cm) can adversely affect the strength and bond strength of structural concrete and, hence, durability [4,5]. For both fresh and hardened concrete, a stable and ample air-void system is essential for several characteristics, including workability, strength, and durability under fire resistance, freezing and thawing cycles [6]. To improve workability, superplasticizers have been used to offset the strength reduction associated with increased water content. In addition to optimized concrete constituent proportions, incorporating superplasticizers in certain quantities plays a pivotal role in achieving enhanced mechanical and durability properties [7]. For cementitious materials, conventional research and practical methods are often time-consuming, labor-intensive, and environmentally toxic. For instance, the carbon footprint of cement, the main ingredient in cementitious composites, accounts for around 8% of global carbon emissions, primarily due to the calcination process used to produce clinker. Furthermore, many experimental trial-and-error procedures used in contemporary research and design processes lead to inefficiencies and increased material waste. These shortcomings underscore the pressing need for intelligent approaches that could increase research productivity while reducing environmental impacts [8,9]. Because cement production generates large volumes of carbon emissions and raises sustainability concerns, researchers have long focused on the “greenification” of concrete as part of global efforts to conserve the environment [10,11]. Fly ash has been proven to enhance strength and durability when integrated into a cementitious medium in certain quantities. Fly ash is a supplementary cementitious material commonly produced as a by-product of coal-fired power plants and is finer than cement. ASTM C618 [12] categorizes fly ash into two groups based on chemical composition (mainly differing in CaO content): Class C and Class F. Class C fly ash yields greater compressive strength than Class F fly ash when the cement dose and fly ash content are kept unchanged. Numerous experimental studies support this trend. The reaction rate is often higher in fly ash with higher CaO content, which is linked to increased concrete strength, especially at early ages. It also reduces environmental pressure, as using fly ash diminishes the amount that must be disposed of in landfills [13–16]. Furthermore,

partially replacing cement with fly ash in cement-based concrete reduces the demand for environmentally intensive cement production. Lower demand for cement production means fewer CO_2 emissions, thus promoting eco-friendly construction practices and aligning with global sustainable development goals[17]. Machine learning presents valuable tools for forecasting properties of cementitious materials. Jiang et al.[18] reviewed its use in performance prediction and sustainable mixture formulation, while Siddique et al.[19] and Aiyer et al.[20] explored neural network and vector machine methods for estimating the compressive strength of self-compacting concrete. Sathiparan and Jeyananthan[21] assessed soft-computing techniques for concrete blended with groundnut shell ash, and Ziyad Sami et al.[22] focused on predicting compressive and tensile strengths. Overall, these investigations form a solid methodological basis for exploring wear-depth prediction in fly ash concrete. Nonetheless, accurately predicting strength does not automatically guarantee reliable abrasion resistance or wear-depth estimates, emphasizing the importance of developing and validating models specifically against measured wear-depth data. Research on cementitious composites can advance sustainably and effectively thanks to the rapid expansion of artificial intelligence (AI), especially machine learning. Data-drivenmachine learning algorithms enable computational models to identify patterns in data, make judgments, and predict results. Several machine learning methods are used to estimate different properties of concrete. To mention a few, Artificial Neural Network (ANN) and Support Vector Machine (SVM) are among the most preferred models, along with linear regression, full quadratic, and tree regression models. A study by Moghaddas and Bao[23] describes a technique for estimating the depth of concrete wear due to abrasion based on complex machine learning models. A number of machine learning algorithms, such as CatBoost, XGBoost, LightGBM, Random Forest, and Extremely Randomized Trees, were employed and compared using standard machine learning performance measures, including root mean squared error, coefficient of determination, mean prediction. The mean absolute deviation, mean percentage deviation, and scatter index for regression analysis. Results indicated excellent performance of the machine learning models. The original dataset used for concrete wear depth (abrasion) prediction included cement, fly ash, silica fume, steel fiber, water, fine aggregate, coarse aggregate, water-reducing agent, age, testing time, abrasion condition, and velocity. SHAP analysis evaluated the effect of parameters on abrasion-induced wear depth, and the time-to-velocity ratio and the water-to-binder ratio (which showed a significant relationship between abrasion extent and water content) emerged as the top two parameters affecting abrasion-induced concrete wear depth. This study demonstrates the effectiveness of machine learning in predicting concrete durability issues related to abrasion, reassuring researchers of its potential to save time and effort by minimizing the need for arduous laboratory sessions. Malazdrewicz and Sadowski[24] explained how artificial intelligence using an artificial neural network may be used to estimate the depth of wear in a cementitious composite treated with high-calcium fly ash. The authors developed an intelligent model to estimate the depth of wear in cementitious composites. The mix proportions served as the basis for the model. Six C30/37 mixtures were involved in this research, including three control ones and three mixes with high calcium fly ash ranging from 18% to 34% of the total mass of binder. The air content varied from 6% to 1%, the water/cement ratio from 0.2 to 0.35, and the water/binder ratio from 0.31 to 0.34. Results indicated that the model's accuracy in estimating wear depth was high compared with laboratory tests such as the rotating cutter method. The linear correlation coefficient R was 0.998 for learning, 0.997 for testing, and 0.995 for validation, confirming the Artificial Neural Network's high reliability. It was claimed that, with a sufficiently large sample size of experimentally tested data, artificial intelligence can estimate the depth of wear in concrete with exceptional accuracy. The researchers partially replaced cement with calcined sludge to reduce cement use and recycle sludge in research on the compressive strength of cement-sludge composites. The authors intended to examine how several parameters, such as ball milling time, calcination temperature, sludge replacement rate, curing age, and the amounts of calcium chloride ($CaCl_2$) and calcium sulfate ($CaSO_4$) admixtures, affected the compressive strength of calcined sludge-cement composites. In addition, they used six regression-based machine learning models to predict compressive strength, including Linear Regression, Support Vector Machine Regression, Random Forest Regression, Multi-layer Perceptron-Artificial Neural Network Regression, Ensemble Regression, and Convolutional Neural Network Regression (CNN). The results

showed that CNN and Ensemble Regression models offer superior prediction accuracy. According to the robustness comparison, curing age has the greatest effect on compressive strength; however, ball milling time and $CaSO_4$ have little effect, which aligns with the experimental results [25,26].

This study evaluates several advanced machine learning models that use analytical techniques to estimate wear depth in fly-ash concrete based on a dataset of 439 samples incorporating parameters including Cement 'C', Water 'W', Fly ash 'F', Fine aggregate 'FA', Coarse aggregate 'CA', Superplasticizer 'P', Air entraining 'AE', Age 'A', Time of testing '*t*', and Depth of wear 'Dw'. Predictive models include full quadratic (FQ), nonlinear (NLR), multiple linear regression (MLR), and artificial neural network (ANN) models. The validity and reliability of the generated models are assessed using a range of statistical techniques, including the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), scatter index (SI), objective function (OBJ), and the a20-index. This study offers definitive recommendations on model selection for durability-related applications by thoroughly comparing models to estimate concrete wear depth. The 439 datasets used in this study were taken from Mahmood et al. [2,27].

Although many studies have used machine learning to predict concrete's abrasion resistance, most focus on general mixtures or a limited set of algorithms, with little attention to sustainable fly ash concrete or thorough comparisons. Additionally, previous research rarely employed a materials informatics framework to systematically analyze large datasets and identify key factors affecting wear depth. To address these gaps, this study introduces a comprehensive materials informatics approach based on 439 experimental mixtures and compares four predictive models (MLR, NLR, FQ, and ANN). This framework not only provides highly accurate wear depth predictions but also offers insights into the importance of mixture components and curing variables, enabling the sustainable design of durable fly ash concrete while greatly reducing experimental effort.

2. Methodology

Figure 1 shows a detailed flowchart of the steps followed in this study's methodology. The main steps are outlined as follows:

1. Collecting relevant data from literature on concrete wear depth ranging from 0.05 to over 3 mm,
2. Considering the parameters of the study to be used as target model predictors. The models, in turn, are considered to predict wear depth of tested cement concrete samples, including samples of fly ash-based cement concrete,
3. Random categorization of data into training and testing datasets and development of proposed models using models including Full Quadratic (FQ), Nonlinear Regression (NLR), Multiple Linear Regression (MLR), and Artificial Neural Networks (ANN) models,
4. Accuracy assessment of the proposed models on the basis of common indices such as R^2, RMSE, MAE, OBJ, SI, and the a-20 index,
5. Confirmation of the key parameter in the prediction of concrete wear depth through sensitivity analysis.

2.1. Data Collection

A machine learning framework was created to predict concrete wear depth, utilizing a database compiled from previously published experimental studies. The dataset includes 439 concrete samples sourced from existing literature, organized to support model development. These studies provided measured wear depths along with mixture proportions and testing parameters necessary for prediction. This database covers a variety of concrete mixtures and compositions from earlier research, offering a comprehensive foundation for developing and testing the prediction models [2,27–29]. The compiled dataset included key variables known to influence concrete wear performance, including fly ash content, cement content, fine aggregate content, coarse aggregate content, curing age, and different admixtures. These parameters served as model inputs, while the experimentally measured concrete wear depth was the target output. After collecting and organizing the 439 samples, the database was used to train and evaluate the ML models to establish relationships between concrete mixture characteristics, curing conditions, and wear depth. This approach enabled systematic use of experimental information from the published literature to develop predictive models of concrete wear behavior [29–32]. To ensure model applicability, the input variables spanned a wide range of material parameters and testing conditions.

Specifically, cement content ranged from 107 to 470 kg/m^3, water content from 123 to 208 kg/m^3, fly ash content from 0 to 316 kg/m^3, fine aggregate content from 607 to 715 kg/m^3, and coarse aggregate content from 916 to 1,264 kg/m^3. In addition, the superplasticizer dosage ranged from 2.60 to 2.90 kg/m^3, whereas the air-entraining agent dosage ranged from 0 to 1.38 L/m^3. The concrete specimens were tested at ages ranging from 7 to 365 days, with testing durations varying between 5 and 60 min. The measured concrete wear depth, the output parameter, ranged from 0.05 to 3.68 mm. These broad parameter ranges provide a diverse dataset for creating reliable concrete wear depth prediction models. Based on the literature data, we divided the dataset into training and testing subsets.

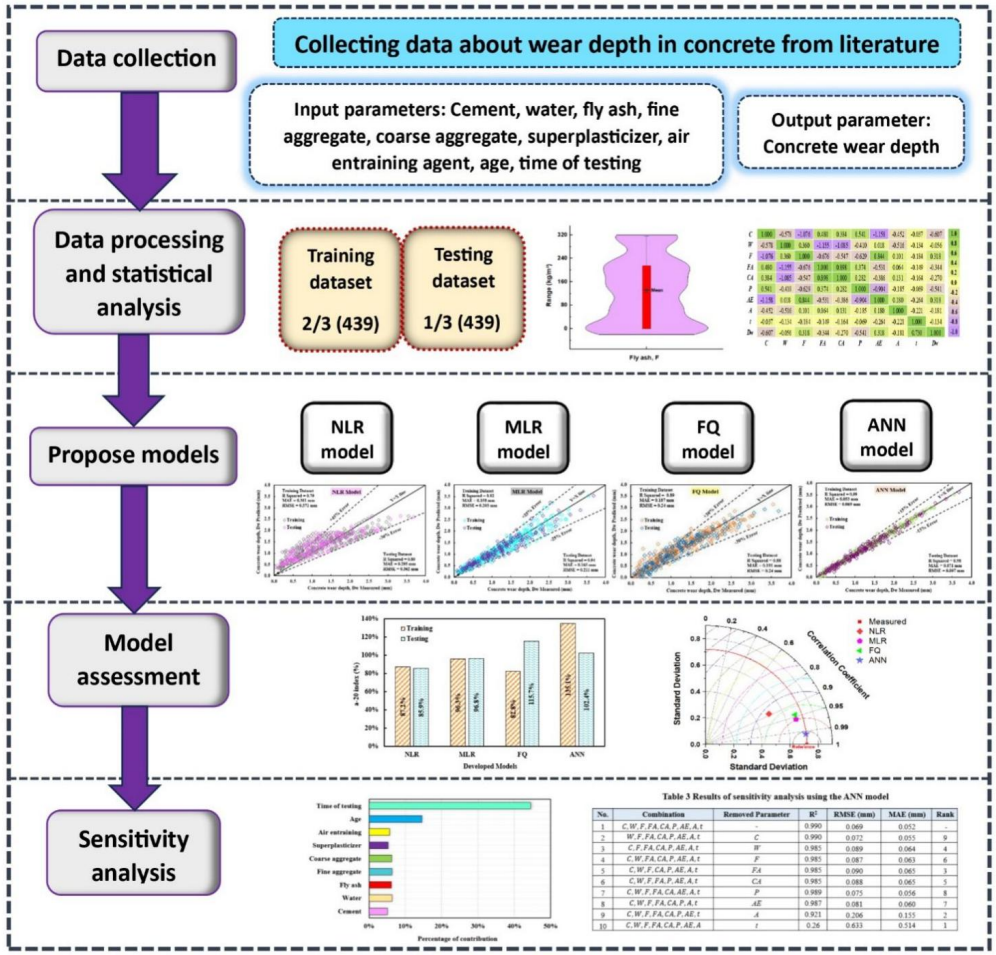

Figure 1. Flow chart of the study.

Before developing the model, we ran all extracted data through multiple quality-control steps. Variable names were standardized, measurement units were unified, duplicate records were removed, and inconsistent entries were checked against original publications. The excluded missing values when complete reconstruction of the mixture design wasn't feasible. We also converted all numerical variables to consistent engineering units before model development. The previously developed a machine learning framework to predict the wear depth of fly ash concrete using a literature-based database of 391 experimental observations. The dataset included mixture proportions (cement, water-to-binder ratio, Class C fly ash, fine and coarse aggregates, plasticizer, and air-entraining agent), along with curing time and abrasion testing duration as input variables, while wear depth (Dw) was the output parameter. The study evaluated four predictive models, namely Linear Regression (LR), Pure Quadratic (PQ), Interaction (IN), and M5P-tree, and reported that the M5P-tree model achieved the highest prediction accuracy. The present study builds upon a similar database and in-

put variables but employs different machine learning algorithms to further improve prediction accuracy, robustness, and the understanding of the factors influencing concrete wear depth[2].

2.2. Data Processing and Statistical Analysis

To understand the distribution and correlations, the collected data were organized and statistically analyzed. There were 439 datasets in the entire dataset. The dataset was separated into:

Training dataset: 439 × (2/3) = 2/3 of the data = 293 data.

Testing dataset: 439 × (1/3) = 1/3 of the data = 146 data.

(**Figure 1**, Testing = 1/3 (439), Training = 2/3 (439)).

In addition to the independent test set, we assessed model robustness using five-fold Group K-Fold Cross-Validation. The grouping variable was based on unique mixture identifiers, ensuring that observations from the same concrete mixture remained in the same fold. This approach prevents information leakage and yields statistically reliable estimates of predictive performance. The training dataset was used to develop and calibrate the predictive models, while the testing dataset, excluded from training, was used to assess the models' predictive performance on unseen data. The derived equations, based on the training dataset, were verified on the testing dataset. For the statistical analysis, the parameters obtained included Mean, Standard Deviation (SD), Variance (Var), Kurtosis (Kur), Skewness (Skew), Minimum (Min), and Maximum (Max). Preliminary data processing is essential to improve the overall accuracy and effectiveness of ML models. **Figure 2** shows a box plot summary of the independent and dependent parameters, confirming that the dataset contains diverse input conditions, with variables (such as F, FA, A, and *t*) showing high variability. The dependent variable (Dw) also varies widely. Consequently, when a study provided a variable as a range instead of a specific value, we used the average of the lower and upper bounds as the representative figure. This preprocessing step standardized the numerical format across the database and allowed us to include all observations in the machine learning models. Therefore, pre-processing was applied to improve prediction accuracy and model stability[28]. Statistical parameters such as Mean, Standard Deviation (SD), Variance (Var), Kurtosis (Kur), Skewness (Skew), minimum (Min), and Maximum (Max) are calculated to evaluate the distribution of the independent parameters relative to the dependent parameter and to examine the correlation between the independent variables and the dependent variable (concrete wear depth). **Table 1** tabulates the statistical analysis for the parameters. For kurtosis, negative values denote a short distribution tail, while positive values denote a longer distribution tail. Skewness describes the distribution shape, with the left tail representing negative values and the right tail representing positive values. **Figure 3** plots the distribution of the collected dataset for each parameter using violin plots, showing how each input parameter (independent variable) and the output (dependent variable), wear depth, are distributed in the dataset. The plots highlight the range, average values, and where most data points are concentrated for each variable. The violin plots show the concentration of data points. The internal box segments highlight the central tendency (mean) and dispersion of each parameter. Cement and fine aggregate show wide ranges, meaning the dataset includes many different proportions and mix designs. Water has a narrower distribution, indicating less variation compared to aggregates. Age shows a broad distribution, reflecting multiple curing periods. Concrete wear depth shows most values concentrated in the lower-to-mid ranges, with fewer high values. Overall, the results confirm that the dataset includes a wide variety of mixture proportions and testing configurations, improving predictive performance and enabling accurate prediction of concrete wear depth.

Table 1. Statistical analysis for parameters.

Statistical Analysis	Mean	Median	Mode	St. Dev.	Variance	Kurtosis	Skewness	Min.	Max.	No. of Data
Parameters										
Cement (kg/m^3)	277.7	320.0	174	110.13	12,127.68	−1.1730	−0.0950	107.0	470.0	439
Water (kg/m^3)	147.1	139.0	208	25.42	645.97	1.4625	1.6164	123.0	208.0	439
Fly ash (kg/m^3)	131.2	139.0	0	110.65	12,244.39	−1.2070	0.2612	0.0	316.0	439
Fine aggregate (kg/m^3)	662.3	659.0	620	35.47	1,257.84	−1.3561	−0.0430	607.0	715.0	439
Coarse aggregate (kg/m^3)	1,141.1	1,153.0	916	98.41	9,683.71	0.9688	−1.2486	916.0	1,264.0	439

Table 1. *Cont.*

Statistical Analysis	Mean	Median	Mode	St. Dev.	Variance	Kurtosis	Skewness	Min.	Max.	No. of Data
Superplasticizer (kg/m^3)	2.7	2.7	2.7	0.11	0.01	−1.1616	0.3030	2.6	2.9	439
Air entraining (l/m^3)	0.5	0.4	0	0.37	0.14	−0.0345	0.7926	0.0	1.4	439
Age (days)	144.0	91.0	28	144.49	20,878.67	−1.2050	0.8068	7.0	365.0	439
Time of testing (Minutes)	33.6	35.0	40	16.68	278.35	−1.1583	−0.0305	5.0	60.0	439
Depth of wear (mm)	1.2	1.2	0.9	0.72	0.52	−0.1810	0.5687	0.1	3.7	439

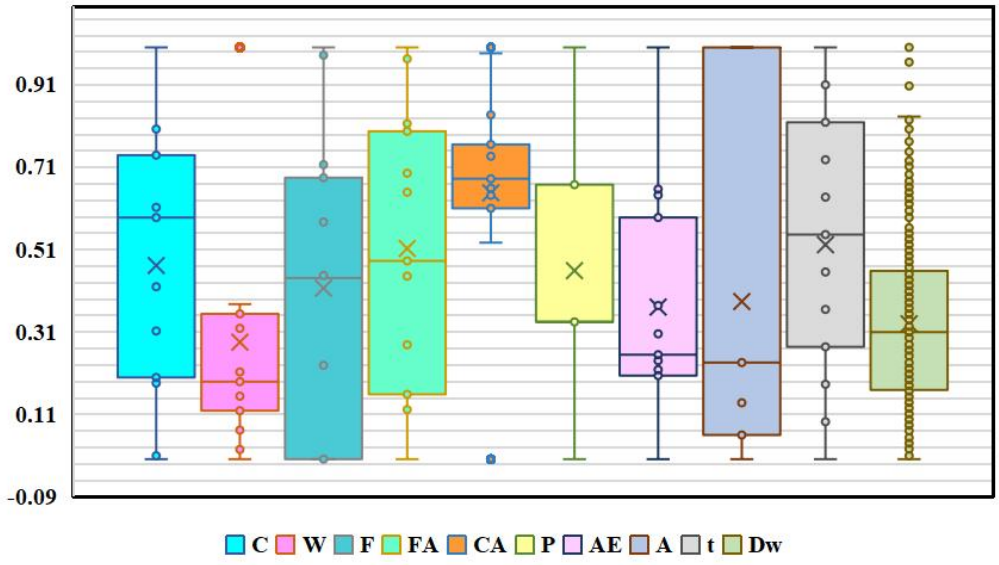

Figure 2. Box plot of independent and dependent parameters.

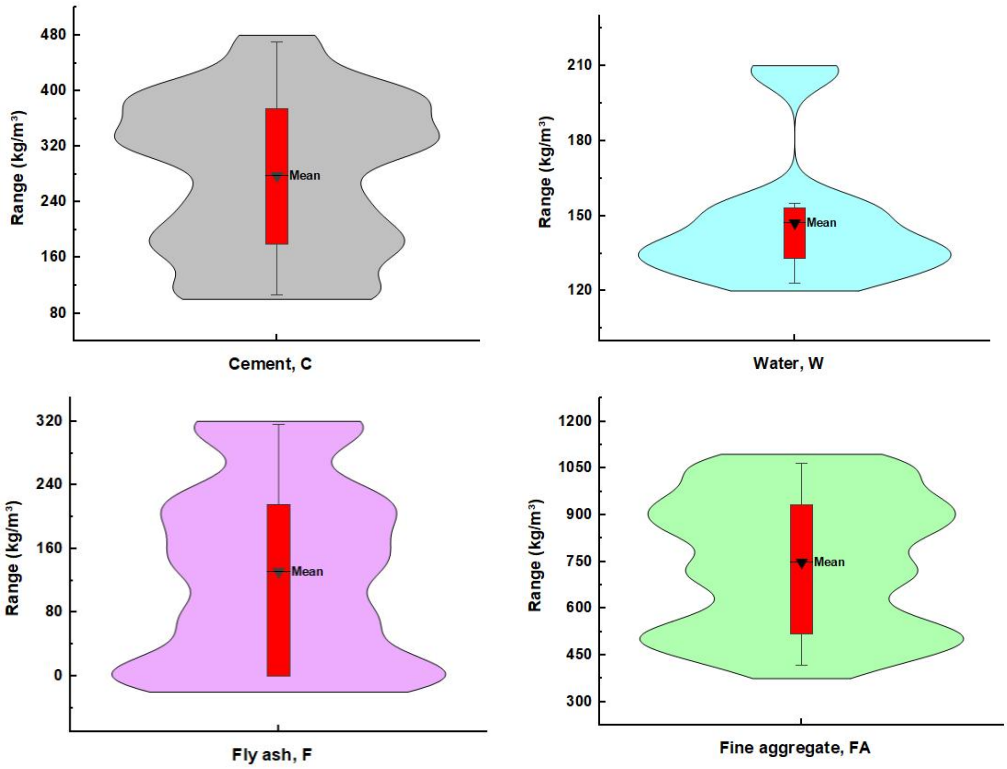

Figure 3. *Cont.*

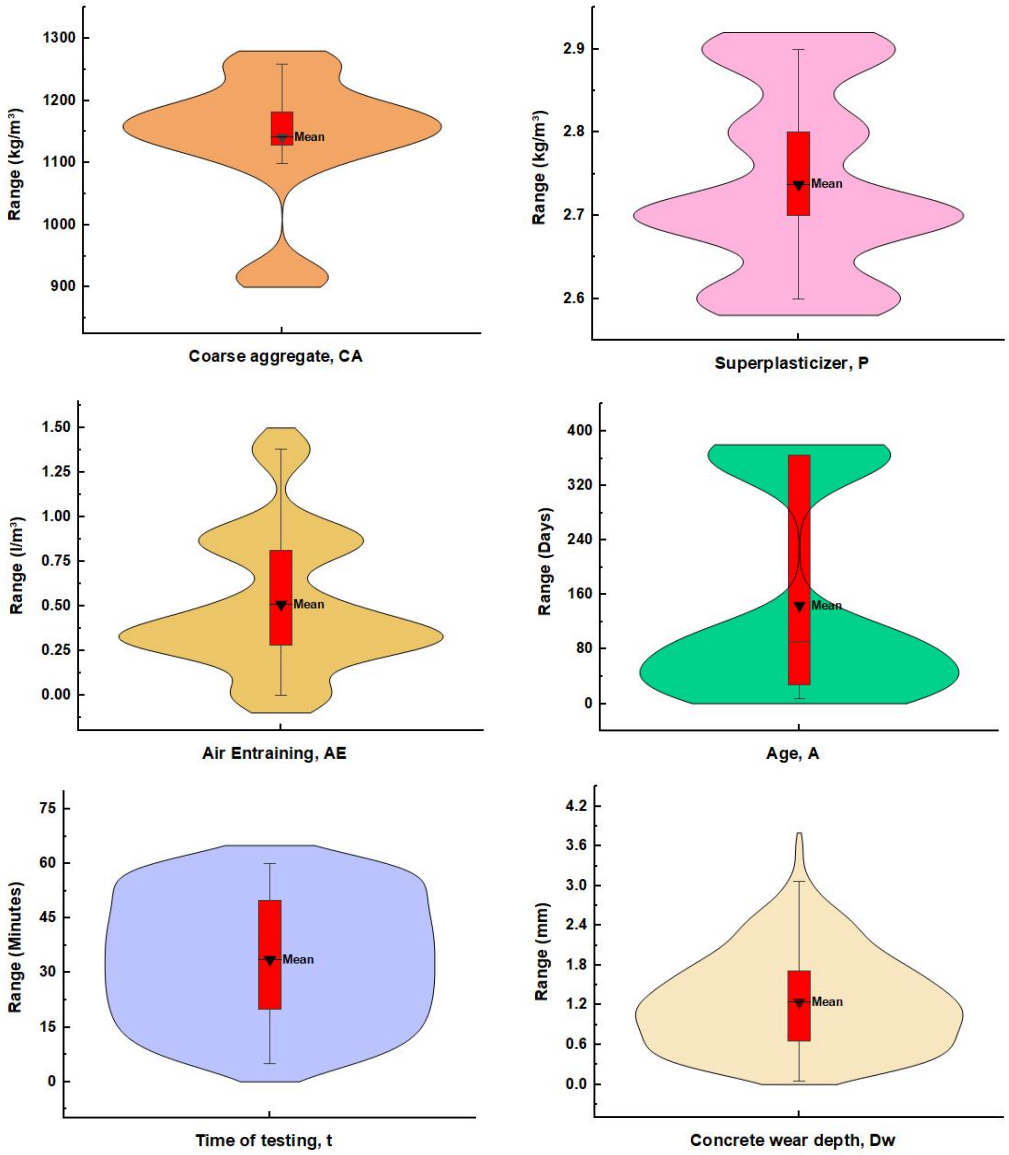

Figure 3. Dataset distribution through violin plots.

2.3. Correlation Analysis and Model Development

The correlation analysis shows a clear relationship between the amounts of concrete mixture constituents and concrete wear depth. The dataset was randomly split into training and testing subsets. The models were developed using the training dataset, while the testing dataset was used to evaluate their performance on unseen data.

Figure 4 presents the correlation analysis between the independent parameters and the dependent parameter (concrete wear depth, Dw) using two statistical approaches: (a) Pearson correlation, which measures linear relationships between variables, and (b) Spearman correlation, which evaluates monotonic relationships (increasing/decreasing trends) and is more reliable when there are deviations or non-linear data. The numerical values in the correlation tables, shown as colored cells, indicate the pattern and degree of the relationship between each pair of parameters. Negative coefficients indicate an inverse relationship, whereas positive coefficients indicate a direct relationship. Color intensity shows the degree of correlation. The color scale ranges from −1 to +1, where values near ±1 indicate stronger correlations and values near 0 indicate a weak relationship. The correlation is not uniformly strong across all variables. In the Pearson interpretation, the correlation is mostly weak to moderate for

many of the parameters. Only time of testing (t) shows a strong relationship with Dw. Spearman shows stronger correlations than Pearson for some variables (especially cement and superplasticizer), meaning their effect on Dw may be monotonic but not strictly linear. According to the correlation analysis, the time of testing has a substantial positive correlation (t = 749 and t = 0.730 in **Figure 4a,b**, respectively), indicating its dominant impact on wear formation, whereas most independent variables show weak- to moderate correlations with concrete wear depth. The models used included Full quadratic (FQ), nonlinear regression (NLR), multilinear regression (MLR), and artificial neural network (ANN).

	C	W	F	FA	CA	P	AE	A	t	Dw
C	1.000	0.105	-0.971	0.539	-0.012	0.560	-0.881	-0.153	0.104	-0.441
W	0.105	1.000	0.118	-0.710	-0.946	0.074	-0.253	-0.287	0.088	-0.095
F	-0.971	0.118	1.000	-0.685	-0.188	-0.566	0.832	0.087	-0.084	0.419
FA	0.539	-0.710	-0.685	1.000	0.773	0.404	-0.436	0.127	-0.007	-0.226
CA	-0.012	-0.946	-0.188	0.773	1.000	-0.042	0.200	0.274	-0.080	0.073
P	0.560	0.074	-0.566	0.404	-0.042	1.000	-0.700	-0.099	0.058	-0.339
AE	-0.881	-0.253	0.832	-0.436	0.200	-0.700	1.000	0.188	-0.103	0.463
A	-0.153	-0.287	0.087	0.127	0.274	-0.099	0.188	1.000	-0.012	0.045
t	0.104	0.088	-0.084	-0.007	-0.080	0.058	-0.103	-0.012	1.000	0.749
Dw	-0.441	-0.095	0.419	-0.226	0.073	-0.339	0.463	0.045	0.749	1.000

(**a**)

	C	W	F	FA	CA	P	AE	A	t	Dw
C	1.000	-0.578	-1.076	0.480	0.384	0.541	-1.158	-0.452	-0.037	-0.607
W	-0.578	1.000	0.360	-1.155	-1.085	-0.410	0.018	-0.516	-0.134	-0.056
F	-1.076	0.360	1.000	-0.676	-0.547	-0.629	0.844	0.101	-0.184	0.318
FA	0.480	-1.155	-0.676	1.000	0.898	0.374	-0.531	0.064	-0.149	-0.344
CA	0.384	-1.085	-0.547	0.898	1.000	0.282	-0.386	0.131	-0.164	-0.270
P	0.541	-0.410	-0.629	0.374	0.282	1.000	-0.904	-0.185	-0.069	-0.541
AE	-1.158	0.018	0.844	-0.531	-0.386	-0.904	1.000	0.180	-0.264	0.318
A	-0.452	-0.516	0.101	0.064	0.131	-0.185	0.180	1.000	-0.221	-0.181
t	-0.037	-0.134	-0.184	-0.149	-0.164	-0.069	-0.264	-0.221	1.000	-0.134
Dw	-0.607	-0.056	0.318	-0.344	-0.270	-0.541	0.318	-0.181	0.730	1.000

(**b**)

Figure 4. Correlation analysis of the experimental database: (**a**) Pearson correlation matrix and (**b**) recalculated Spearman rank correlation matrix. All Spearman correlation coefficients were verified to fall within the theoretical range of −1 to +1.

2.3.1. NLR Model

The nonlinear dependence of concrete wear depth (Dw) on mixture components and testing circumstances has been observed using NLR. Because of the combined effects of binder concentration, admixtures, curing age, and exposure length, NLR can capture curvature and non-proportional effects anticipated in abrasion/wear behavior, unlike linear models. The training dataset was used to calibrate the NLR model in the present study, while the testing dataset was used to validate it and assess its performance on new observations. A common form to generate a nonlinear regression model[33] for concrete wear depth prediction is as outlined in Equation (1):

$$Dw = \alpha_1(CA)^{\alpha 2} + \alpha_3(S)^{\alpha 4} + \alpha_5(C)^{\alpha 6} + \alpha_7(FA)^{\alpha 8} + \alpha_9(CR)^{\alpha 10} + \alpha_{11}\left(\frac{w}{b}\right)^{\alpha 12} + \alpha_{13}(t)^{\alpha 14} \quad (1)$$

2.3.2. MLR Model

MLR was developed as a standard statistical model to estimate concrete wear depth as a linear combination of the independent variables. By measuring each factor's independent contribution while keeping the other variables constant, the MLR model provides a clear, interpretable relationship. Although the MLR model offers advantages such as a simple mathematical formulation and ease of implementation, it improves prediction performance compared to simple linear regression by incorporating multiple input variables. However, the MLR model is highly dependent on the quantity of data; fewer data points result in lower accuracy[32,33]. The training dataset was used to develop the model, and the testing dataset was used to evaluate its performance on unseen data. Equation (2) shows the exponential and constant prod-

uct of the variables influencing the concrete's compressive strength.

$$Y = aX_1^{b1}X_2^{b2} \tag{2a}$$

$$Dw = \alpha_1(C)^{\alpha_2}(W)^{\alpha_3}(F)^{\alpha_4}(FA)^{\alpha_5}(CA)^{\alpha_6} (P)^{\alpha_7}(AE)^{\alpha_8}(A)^{\alpha_9}(t)^{\alpha_{10}} \tag{2b}$$

2.3.3. FQ Model

By combining linear, quadratic, and interaction terms, the FQ model was developed to capture nonlinear interactions between the independent variables and concrete wear depth. Because the FQ model accounts for interactions among mixture parameters and their quadratic effects, it provides a more flexible description of complex material behavior than linear regression [34]. This method preserves the interpretability of the model framework while improving prediction accuracy, meaning it can predict more accurately than a simple linear model while remaining understandable. The relationship between concrete wear depth, the first- and second-degree terms of each independent variable, and the interactions among independent variables [35,36] is represented by Equation (3), which displays the full quadratic formula.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \beta_{ii} X_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} X_i X_j \tag{3a}$$

where β_0 is the intercept, β_i are linear coefficients, β_{ii} represent quadratic coefficients, and β_{ij} denote pairwise interaction coefficients.

$$\begin{aligned} Dw = {} & \beta_1 + \beta_2(C) + \beta_3(W) + \beta_4(F) + \beta_5(FA) + \\ & \beta_6(CA) + \beta_7(P) + \beta_8(AE) + \beta_9(A) + \\ & \beta_{10}(t) + \beta_{11}(C)^2 + \beta_{12}(W)^2 + \beta_{13}(F)^2 + \\ & \beta_{14}(FA)^2 + \beta_{15}(CA)^2 + \beta_{16}(P)^2 + \\ & \beta_{17}(AE)^2 + \beta_{18}(A)^2 + \beta_{19}(t)^2 \end{aligned} \tag{3b}$$

2.3.4. ANN Model

Concrete wear depth (Dw) was predicted using an Artificial Neural Network (ANN) model because of its exceptional ability to capture complex, highly nonlinear interactions between mixture parameters and material performance. ANNs process data in ways similar to the human brain. In engineering applications, this model serves as a machine learning system for a variety of numerical forecasts and problems. The network consists of three main parts: an input layer, one hidden layer, and an output layer (**Figure 5**). The input layer includes nine neurons, each representing one variable: cement (C), water (W), fly ash (F), fine aggregate (FA), coarse aggregate (CA), superplasticizer (P), air-entraining agent (AE), age (A), and time of testing (*t*). The output layer has one neuron that produces the predicted value of concrete wear depth (Dw). The network was trained on the training dataset, and its performance was tested on a separate testing dataset not used during training to evaluate how well the model predicts new, unobserved data. Equation (4) presents a typical process for ANN calculation in a single hidden layer.

$$Output = f\left(\sum_{j=1}^{n} w_j x_j + bias\right) \tag{4}$$

2.4. Model Evaluation Metrics

Several statistical metrics, including the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), scatter index, Taylor diagrams, and radar charts, were used to evaluate model performance. Together, these metrics may offer a thorough assessment of prediction accuracy, consistency, and reliability across training and testing datasets [37,38].

2.5. Sensitivity Analysis

A sensitivity analysis was carried out using the ANN model to evaluate the relative importance of each input parameter to concrete wear depth prediction. This analysis validates the interpretability of the developed models and provides insight into parameter significance.

3. Results and Discussion

3.1. Performance of the Developed Models

3.1.1. NLR Model

NLR was one of the models used to predict concrete wear depth. Equation (5) shows the NLR model output, and **Figure 6** compares measured concrete wear depth with NLR-predicted values. Although there is noticeable scattering around the ideal 1:1 line, especially at greater wear depths, the data indicate that NLR can capture typical nonlinear relationships between inputs, including mixture parameters, testing conditions, and Dw. However, NLR remains

limited by predetermined functional forms that might not accurately capture the nonlinear interactions among mixture proportions, curing age, and testing length, even though it can capture curvature in the response. A single nonlinear expression may not adequately explain the complex relationships between binder content, aggregate strength, air-void structure, and time-dependent damage buildup that characterize concrete wear.

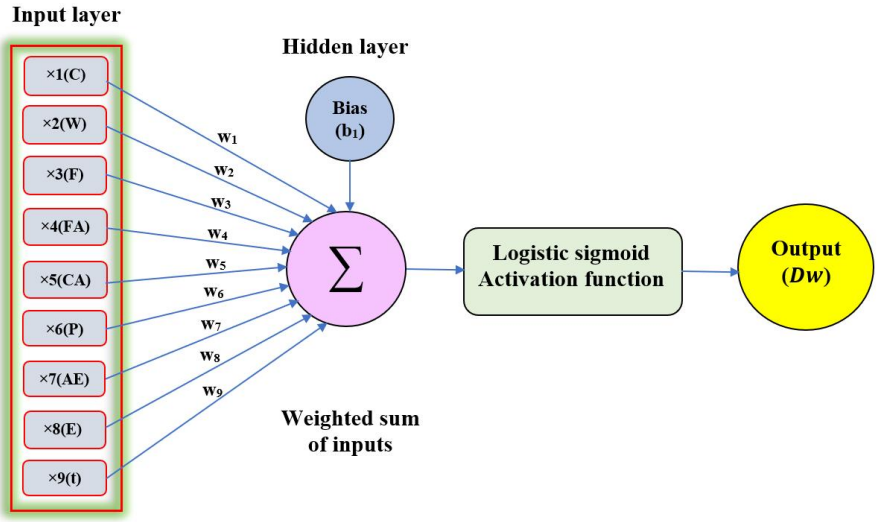

Figure 5. Typical ANN output procedure in a single node.

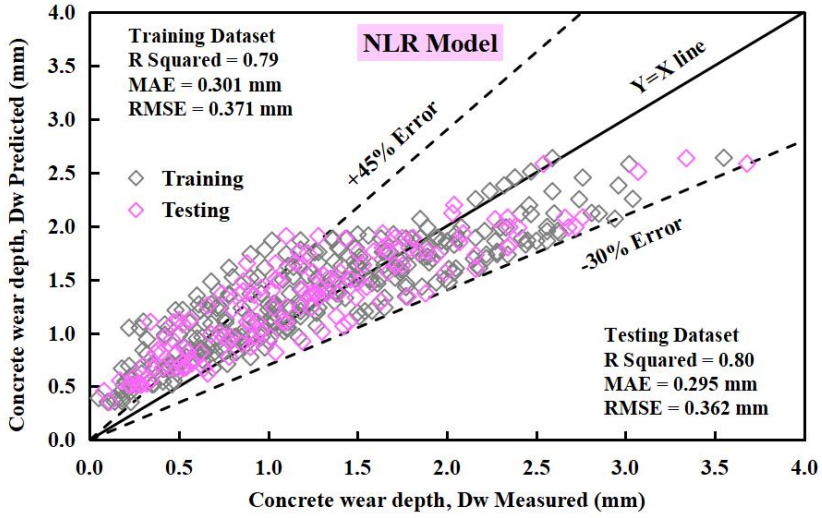

Figure 6. Measured versus predicted compressive strength for the NLR model. The solid line represents the line of perfect agreement (1:1), while the dashed lines indicate the standardized ±10% prediction tolerance used consistently across all predictive models.

In addition, the moderate degree of agreement between the observed and projected values shows how sensitive the NLR model is to changes in the input parameters. The fixed nonlinear structure of the NLR model may not adapt to all possible input combinations because the dataset includes a wide range of mixture proportions and testing periods. This limitation is evident when comparing the wider distribution of data points to the ANN model outcomes presented later in the study.

$$\begin{aligned} Dw = {} & 0.03(C)^{-0.31} + 0.03(W)^{0.03} + \\ & 0.03(FA)^{-0.03} + 0.03(CA)^{-0.02} + \\ & 0.03(P)^{-0.02} + 0.03(AE)^{0.37} + \\ & 0.03(A)^{-0.55} + 0.05(t)^{0.87} \end{aligned} \tag{5}$$

3.1.2. MLR Model

The MLR is another machine learning model used to predict concrete wear depth with various parameters. **Figure**

7 illustrates the relationship between measured and predicted concrete wear depth (Dw) obtained using the MLR model. Because of its simple mathematical formula, it is regarded as one of the least effective models. The MLR model formula includes terms raised to constant powers, as shown in Equation (6). Compared to the NLR plot, the MLR plot shows a more compact alignment around the 1:1 line. However, some scatter still persists at higher Dw values, with many points lying inside the ±25% error zone, suggesting that nonlinear wear development and interaction effects, such as the combined influence of binder content and testing time, cannot be well described by a completely linear model. The testing performance is close to training (R^2 = 0.82 train vs. 0.81 test), indicating limited overfitting and suggesting the model learned a general pattern that works well on new data. The prediction errors are small and consistent. This means the model is reliable and does not overfit the data. Additionally errors is clearly smaller than NLR (RMSE = 0.305 mm train, 0.311 mm test; MAE = 0.158 mm train, 0.165 mm test).

$$\begin{aligned} Dw = &-0.27(C)^{0.8}\times(W)^{-0.0001}\times(F)^{0.69} \\ &\times(FA)^{-0.64}(CA)\times^{-1.27}\times(P)^{0.08} \\ &\times(AE)^{-0.07}\times(A)^{0.84}\times(t)^{0.03} \end{aligned} \tag{6}$$

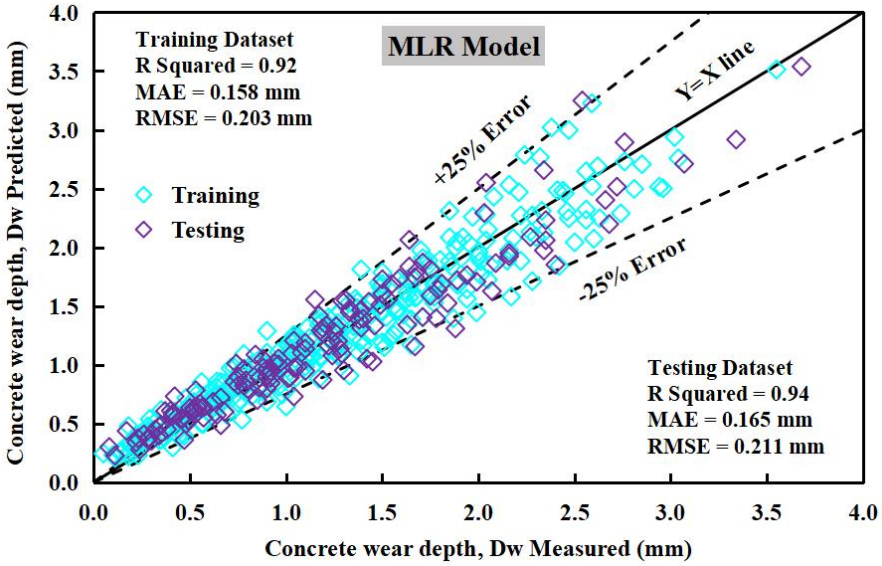

Figure 7. Measured versus predicted compressive strength for the MLR model. The solid line represents the line of perfect agreement (1:1), while the dashed lines indicate the standardized ±10% prediction tolerance used consistently across all predictive models.

3.1.3. FQ Model

The FQ model displays more accurate predictive performance compared to the MLR and NLR models, justified by the majority of data staying within the ±30% error zone and points clustering closer to the 1:1 line (fewer points crossing boundaries, especially at higher Dw), indicating that most predictions stay between 0.70 and 1.30 of the measured-to-predicted wear depth ratios. The model achieves lower error levels (RMSE = 0.24 mm train and 0.24 mm test; MAE = 0.137 mm train and 0.101 mm test) and a greater degree of fit (R^2 = 0.89 train, 0.88 test). **Figure 8** shows the correlation between measured and predicted concrete wear depth values. Equation (37) presents the FQ model for predicting concrete wear depth. The incorporation of squared variables and interaction terms enables the FQ model to represent nonlinear and combined effects among variables such as cement content, fly ash, and testing duration more accurately. As a result, both training and testing performance improves, with reduced RMSE and MAE compared to linear regression approaches. Although the FQ model outperforms MLR and NLR, the ±30% error range indicates that considerable scatter persists, especially at greater wear depths.

$$\begin{aligned} Dw = &\ 34.89 - 0.002(C) - 0.002(W) + 0.001(F) - \\ &\ 0.16(FA) + 0.04(CA) - 0.01(P) - \\ &\ 0.31(AE) - 0.01(A) + 0.04(t) + \\ &\ 0.008(P)^2 + 0.008(AE)^2 \end{aligned} \tag{7}$$

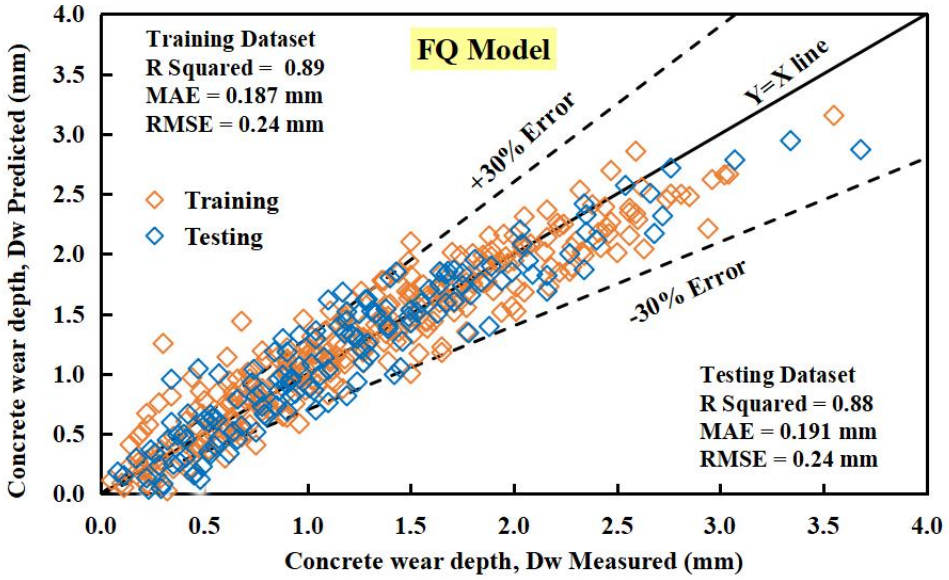

Figure 8. Measured versus predicted compressive strength for the FQ model. The solid line represents the line of perfect agreement (1:1), while the dashed lines indicate the standardized ±10% prediction tolerance used consistently across all predictive models.

3.1.4. ANN Model

The ANN model has the lowest dispersion relative to the error bounds and the closest conformity to the ideal 1:1 line. The graph shows ±15% deviation limits, indicating that the ratio of predicted to measured wear depth falls between 0.85 and 1.15 for 85% of the data. The final model for predicting Dw is the optimal ANN network. As shown in **Figure 9**, five optimal trials were selected based on RMSE and MAE values. Among them, the best trial was the single-layer network with 10 hidden neurons, which achieved the lowest RMSE (0.069) and MAE (0.053). Accordingly, the ANN network with one hidden layer and 10 neurons is selected (**Figure 10**). Equation (8) shows the mathematical formula for the ANN, including all learned parameters. Additionally, **Figure 11** presents the variation between measured and predicted Dw values based on the training dataset. With consistently excellent performance across all stages (R^2 = 0.99 for training and testing), the prediction accuracy is the highest of all models. There is a significant decrease in error magnitudes (RMSE = 0.069 mm train, 0.097 mm test; MAE = 0.055 mm train, 0.071 mm test). The ANN model's better performance is expected because it can learn complex relationships between the input variables and Dw directly from the data. In reality, concrete wear is influenced by several factors simultaneously, such as paste quality, aggregate hardness, air-void structure, the presence of supplementary cementitious materials (SCMs), and especially the testing duration. Since these factors interact in complex, often nonlinear ways, traditional equation-based models may struggle to represent them accurately, whereas ANN can adapt more easily to these patterns.

$$\begin{bmatrix} 0.068 & 3.99 & 1.53 & 1.36 & -0.72 & -0.89 & -0.09 & 5.12 & 1.075 & -2.27 \\ 0.92 & -0.14 & -1.73 & -3.62 & 2.32 & 3.54 & 2.8 & 1.68 & -0.42 & 4.39 \\ 0.17 & 0.85 & 0.14 & -0.09 & -0.59 & 0.25 & 0.24 & 1.1 & -0.52 & -1.67 \\ 0.065 & 0.4 & 0.13 & 0.17 & -0.34 & 0.26 & 0.5 & 0.04 & 0.02 & -1.87 \\ 0.74 & 0.09 & 0.58 & -0.05 & -1.8 & -0.22 & -0.98 & -2.15 & -1.33 & -1.44 \\ 0.3 & -0.88 & 0.68 & 0.95 & 2.94 & 0.27 & -0.81 & 1.54 & -0.74 & -2.9 \\ -1.38 & 4.14 & 2.56 & 0.33 & 0.37 & -1.73 & -1.96 & 3.95 & 1.56 & -2.01 \\ -0.14 & 0.47 & 0.24 & -0.3 & -0.58 & 0.42 & 0.92 & -0.65 & 0.46 & -2.05 \\ -0.05 & 0.62 & 0.14 & -0.15 & 0.05 & 1.8 & 2.57 & -1.02 & 0.76 & -2.85 \\ -0.4 & 1.45 & 1.56 & 1.62 & -0.13 & -1.19 & 0.13 & -0.07 & -1.66 & -2.52 \end{bmatrix} \times \begin{bmatrix} C \\ W \\ F \\ FA \\ CA \\ P \\ AE \\ A \\ t \\ 1 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \\ \beta_{10} \end{bmatrix}$$

$$CS = \frac{1.92}{1+e^{-\beta_1}} - \frac{2.18}{1+e^{-\beta_2}} - \frac{0.87}{1+e^{-\beta_3}} + \frac{0.17}{1+e^{-\beta_4}} - \frac{0.81}{1+e^{-\beta_5}} - \frac{0.69}{1+e^{-\beta_6}} - \frac{1.04}{1+e^{-\beta_7}} + \frac{0.83}{1+e^{-\beta_8}} + \frac{2.15}{1+e^{-\beta_9}} - \frac{1.38}{1+e^{-\beta_{10}}} + 2.04 \quad (8)$$

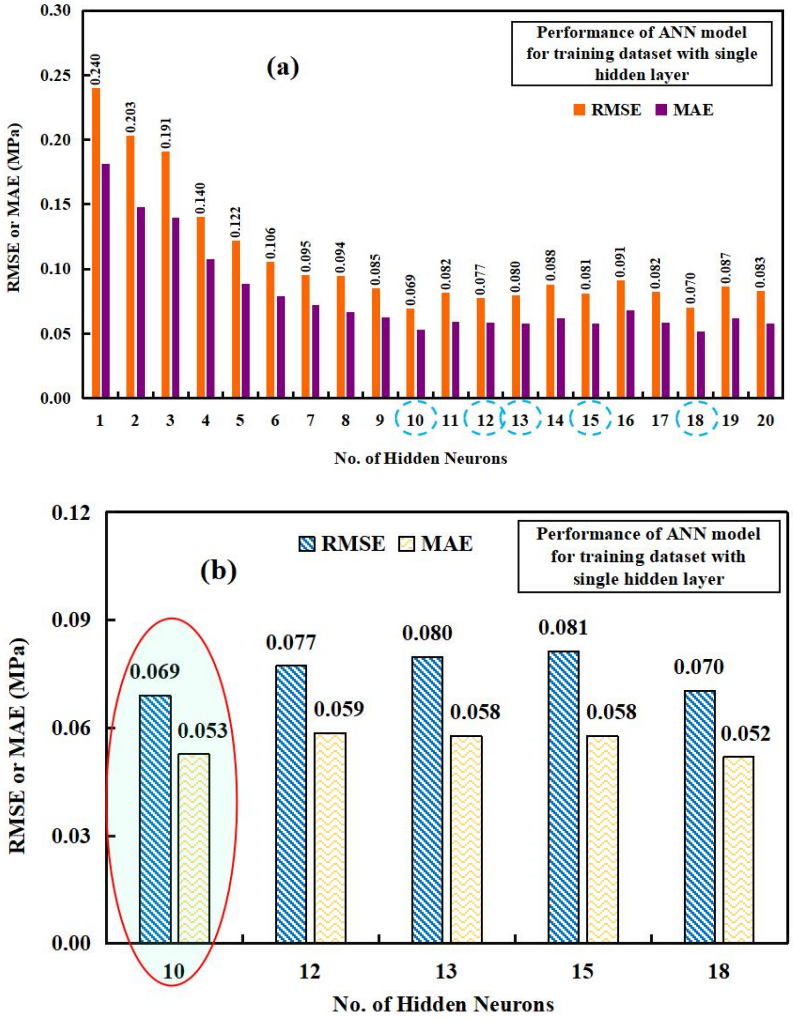

Figure 9. The RMSE and MAE values of predicted concrete wear depth to select the optimum ANN. (**a**) 20 hidden neurons; (**b**) 5 hidden neurons.

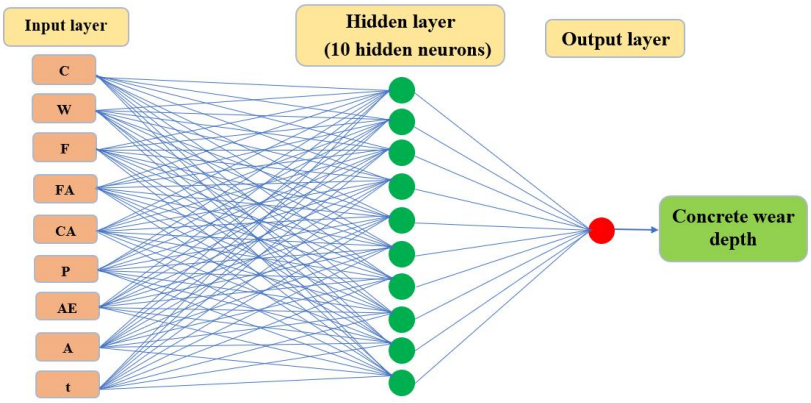

Figure 10. Optimized artificial neural network (ANN) architecture comprising one hidden layer with 10 neurons for predicting concrete wear depth.

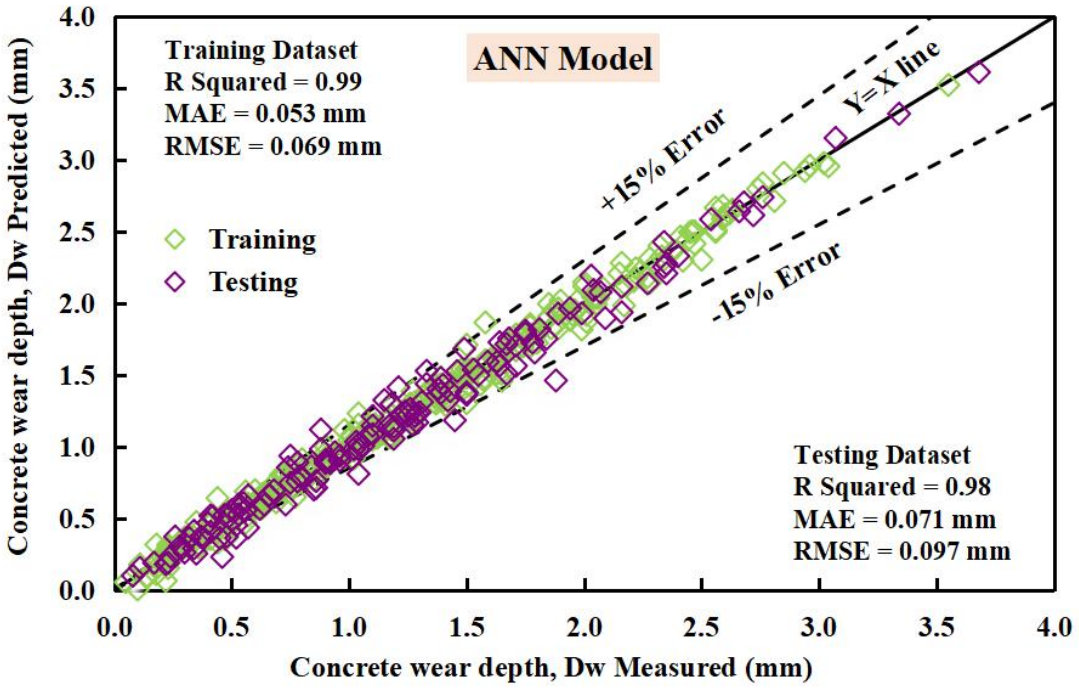

Figure 11. Measured versus predicted concrete wear depth for the ANN model. The solid line represents the line of perfect agreement (1:1), while the dashed lines indicate the standardized ±10% prediction tolerance applied consistently across all predictive models.

The ANN was designed as a feedforward multilayer perceptron featuring a single input layer, one hidden layer, and an output layer. The input layer included nine neurons for the predictor variables, while the hidden layer had ten neurons, selected via hyperparameter optimization. The output layer consisted of one neuron that predicted wear depth. The repeated grouped cross-validation results show that the ANN model consistently outperformed other machine learning methods, with low variability across different validation folds. The small standard deviations suggest stable predictions and strong generalization. Additionally, statistical tests confirmed that ANN were significant.

3.2. Comparison and Assessment of the Developed Models

This study attempted to forecast concrete wear depth across a range of parameters using four machine learning models: NLR, MLR, FQ, and ANN. Several statistical measures, such as the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), scatter index (SI), objective function (OBJ), and the a-20 index, were used to assess the developed prediction models. Together, these criteria provide a reliable assessment. Graphical assessment methods such as residual error analysis, radar chart comparison, and the Taylor diagram further supported the comparison of predictions using both the training and testing datasets, based on R^2, RMSE, and MAE statistics. **Table 2** summarizes the statistical results for the developed models.

Figure 12 illustrates the relationship between the observed and predicted concrete wear depth for the testing dataset. This illustration offers a clear visual comparison of how well the developed models predict unseen data. Predictions from reliable models should ideally align closely with the 1:1 reference line, indicating strong agreement between measured and predicted values. The ANN model clusters data points closest to the ideal line, suggesting higher prediction accuracy. The NLR model, on the other hand, shows greater dispersion of data points, particularly at higher wear depths, indicating a poorer ability to represent complex relationships between mixture parameters and abrasion behavior. The performance of the MLR and FQ models is mediocre at best, with the FQ model outperforming the MLR model in terms of alignment since it incorporates quadratic and interaction factors.

Table 2. An overview of the developed models' performance evaluation criteria.

Model	Figure (No.)	Equation (No.)	Training R^2	Training RMSE (mm)	Training MAE (mm)	Testing R^2	Testing RMSE (mm)	Testing MAE (mm)	Ranking
NLR	**Figure 6**	Equation (1)	0.816	0.98	0.78	0.47	1.26	0.90	4
MLR	**Figure 7**	Equation (2)	0.82	0.305	0.158	0.81	0.311	0.165	3
FQ	**Figure 8**	Equation (3)	0.89	0.24	0.137	0.88	0.24	0.101	2
ANN	**Figure 11**	Equation (4)	0.99	0.069	0.055	0.99	0.097	0.071	1

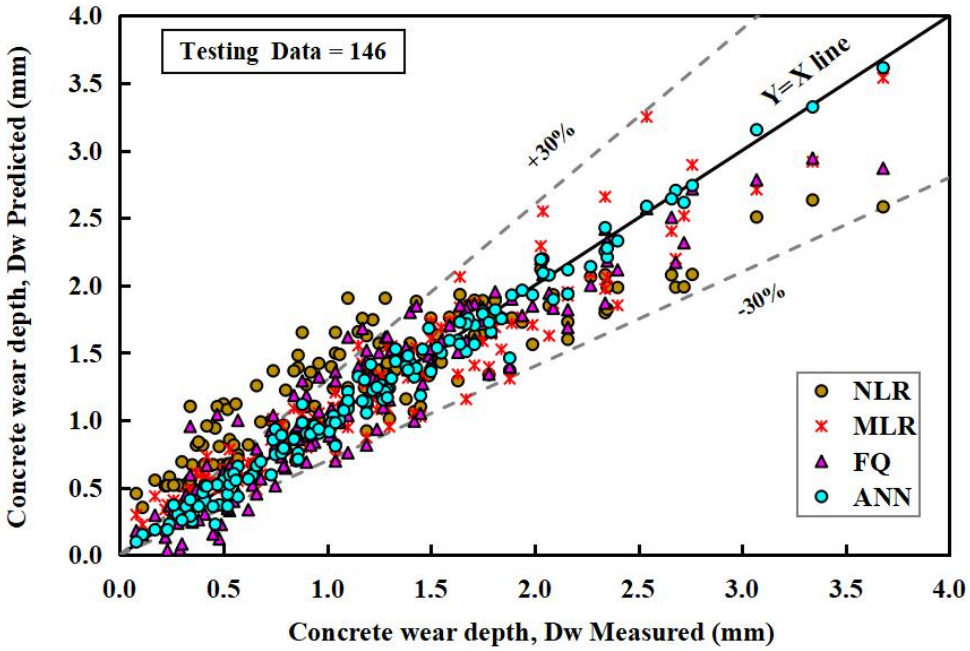

Figure 12. Relationship between measured and predicted Dw based on the testing dataset.

The objective function values shown in **Figure 13** are used to further evaluate the models' performance. An overall performance indicator is computed by combining error measures in the objective function. Lower objective values indicate better model performance and reduced prediction errors. The figure confirms the ANN model's superior predictive performance, as it has the lowest objective value among the models. The MLR and NLR models have higher objective values, indicating greater discrepancies between predicted and measured wear depth, whereas the FQ model is the second-best-performing model. **Figure 14** shows the scatter index (SI) values for the developed models. The SI reflects the degree of dispersion between predicted and measured values and serves as a normalized error indicator. Lower SI values indicate reduced prediction variation and improved agreement. Among all the models, the ANN model has the smallest SI values, indicating that its predictions exhibit the least dispersion relative to the measured data. Once more, the FQ model outperforms MLR and NLR, demonstrating the advantages of including nonlinear and interaction effects in the model formulation. The A-20 index, shown in **Figure 15**, is another essential evaluation metric used in this study. The A-20 index represents the percentage of predicted values that fall within ±20% of the measured values. Because it reflects the prediction model's reliability, this index is very helpful in engineering applications. For both the training and testing datasets, the results show that the ANN model achieves the highest A-20 values, indicating that a significant percentage of its predictions fall within the acceptable error range. While the MLR and NLR models provide lower A-20 values, suggesting worse prediction reliability, the FQ model likewise performs well. **Table 3** summarizes additional statistical indices used in this research: A-20 index, Objective Function (OBJ), and Scatter Index (SI). These indices offer a more comprehensive assessment of model performance and complement standard metrics (R^2, RMSE, and MAE).

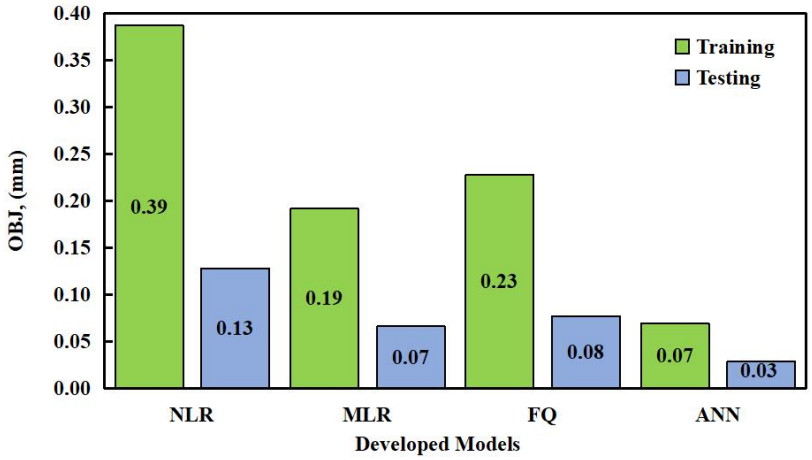

Figure 13. Comparison of the Objective Function (OBJ) values for all predictive models. The histogram values are directly derived from the calculated statistical performance metrics presented in **Table 3**.

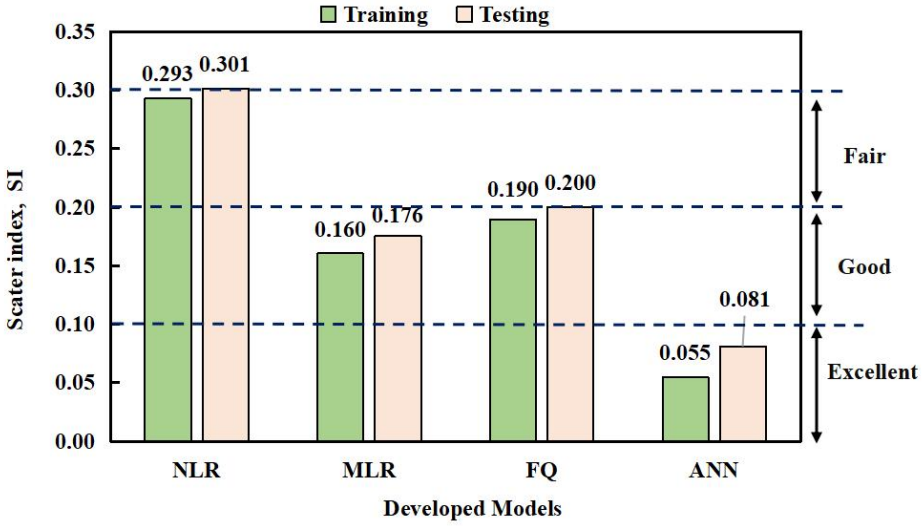

Figure 14. Comparison of the Scatter Index (SI) values for all predictive models. The revised histogram is fully consistent with the statistical results reported in **Table 3** and the corresponding discussion.

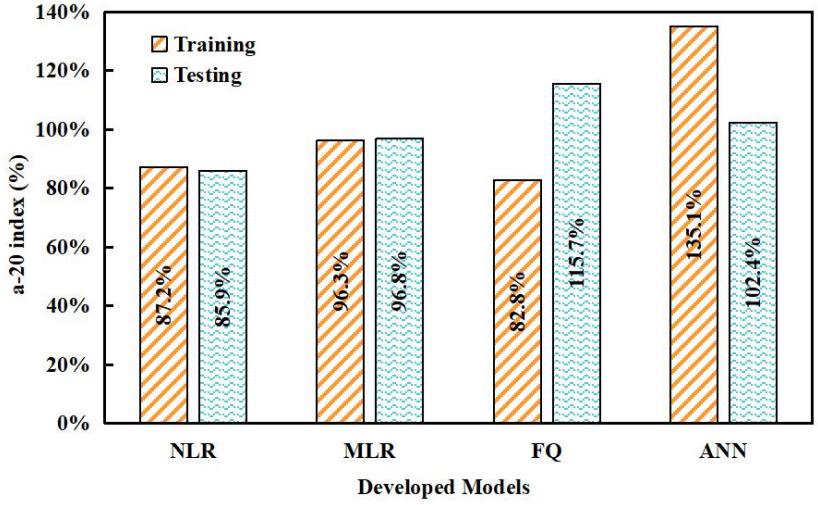

Figure 15. Comparison of the proposed models based on a-20 index (%) using training and testing dataset.

Table 3. Additional performance evaluation metrics for the developed models in predicting Dw.

Model	Scatter Index (SI)	Objective Function (OBJ)	A-20 Index (Training, %)	A-20 Index (Testing, %)	Performance Level
NLR	Highest SI	Highest OBJ	~60–65	~55–60	Poor
MLR	Moderate SI	Moderate OBJ	~75–80	~70–75	Moderate
FQ	Lower SI	Lower OBJ	~85–90	~80–85	Good
ANN	Lowest SI	Lowest OBJ	~95	~90–95	Excellent

The Taylor diagram in **Figure 16** compares the developed models using three statistical parameters simultaneously: correlation coefficient, standard deviation, and root mean square error. This allows for a more thorough graphical analysis. In a Taylor diagram, models closer to the reference point representing the observed data perform better in terms of prediction. The ANN model has the highest correlation and the smallest deviation from measured values, as it is closest to the reference point. Compared with more basic regression techniques, the MLR and FQ models show comparable performance in the graph. The FQ model has a standard deviation more in line with the measured data, but the MLR model shows a somewhat higher correlation coefficient. As a result, both models have comparable predictive accuracy, but the ANN model performs noticeably better than the regression-based methods. **Figure 17** shows a radar chart comparing the models using various assessment measures for both training and testing datasets. Radar charts provide a useful visual representation of overall model performance across statistical criteria such as R^2, RMSE, and MAE. In the radar chart, the ANN model consistently exhibits the largest coverage area, achieving the highest R^2 and the lowest RMSE and MAE values, demonstrating superior prediction accuracy and performance across all assessment measures. The MLR and NLR models have smaller radar regions, indicating lower prediction accuracy and larger errors, while the FQ model ranks second. Additionally, residual error was used to compare the developed models, as seen in **Figure 18**. The predicted concrete wear depth is subtracted from the measured wear depth to determine the residual error. Residual error analysis provides greater insight into the reliability of the developed models. The figure displays the distribution of residual errors for the models, comparing measured and predicted wear depth values. The ANN model exhibits the lowest residual values and a more uniform error distribution, suggesting consistent prediction accuracy and low bias. In contrast, the NLR model shows greater residual fluctuations, particularly at higher wear depths, indicating that its predefined nonlinear function does not adequately capture the complex relationships controlling abrasion behavior. The MLR model demonstrates improved prediction stability compared with NLR. The residual distributions of the MLR and FQ models are intermediate, with the FQ model showing lower residual dispersion than the MLR model.

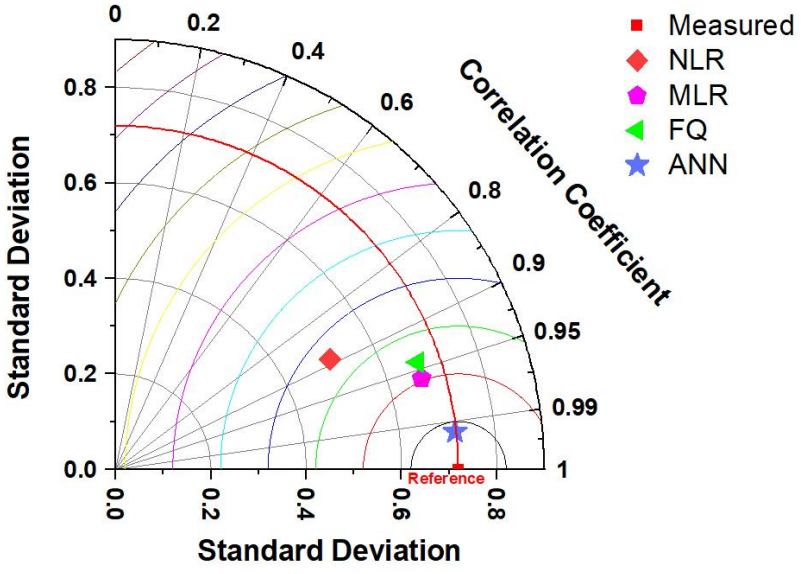

Figure 16. Taylor diagram for developed models.

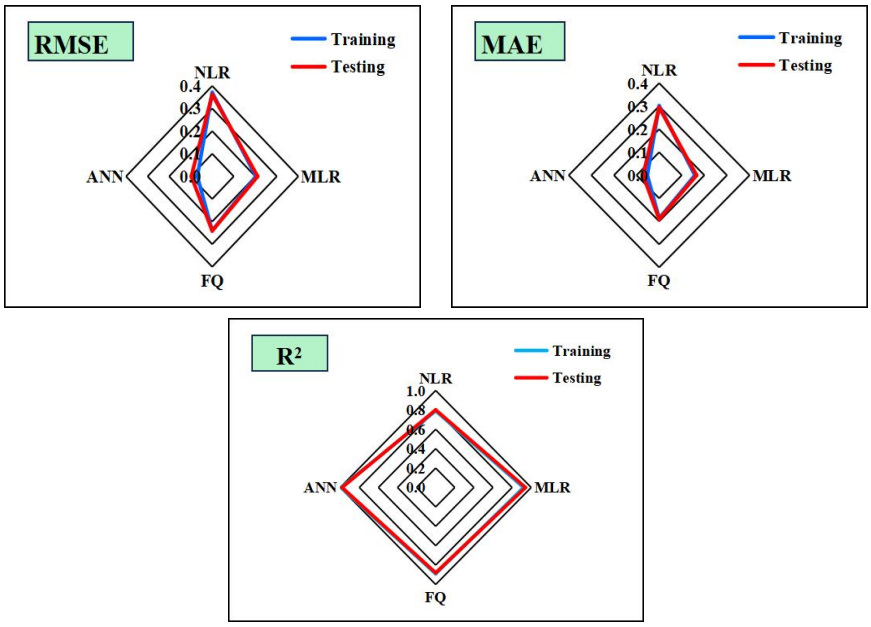

Figure 17. Comparison of prediction results using radar chart for both training and testing dataset.

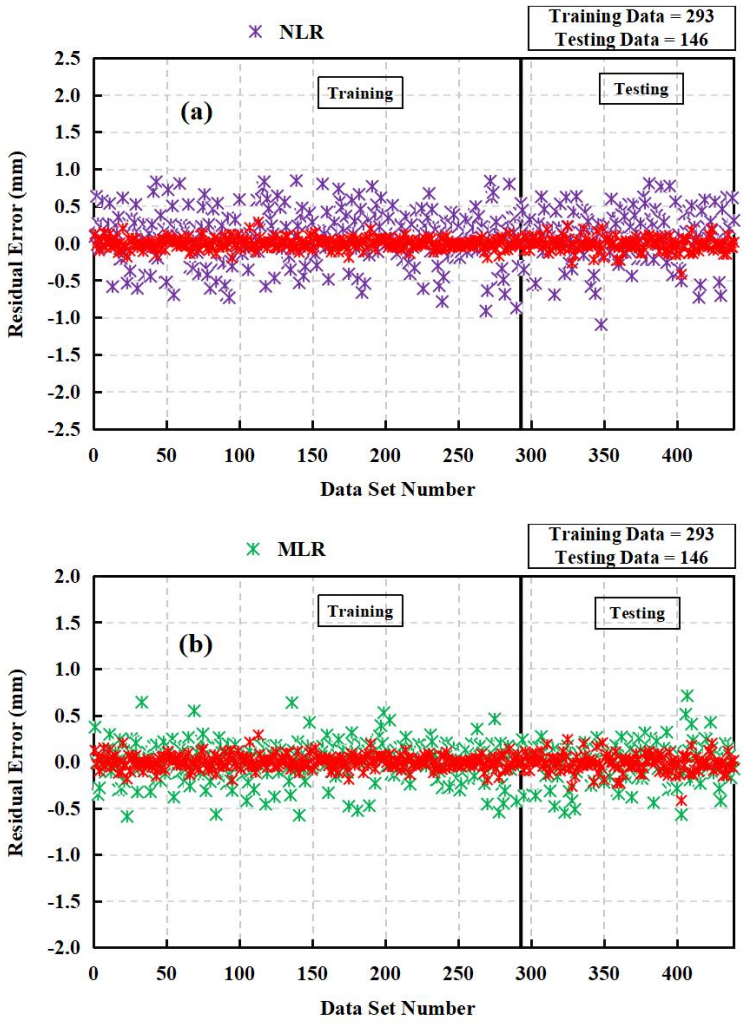

Figure 18. *Cont.*

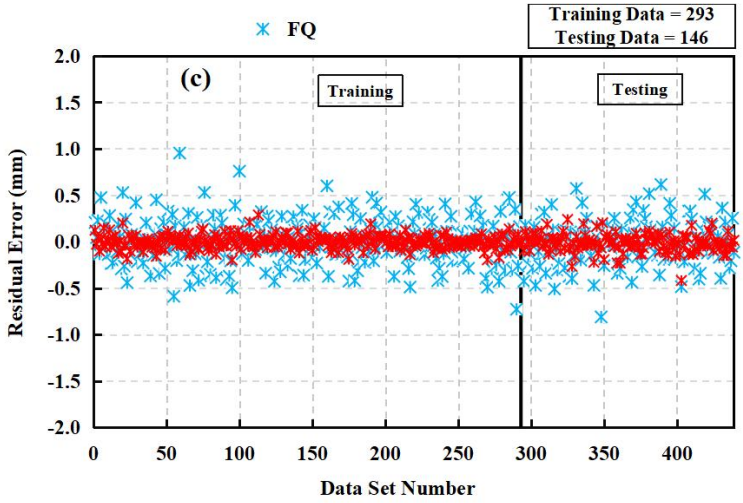

Figure 18. The residual error between measured and predicted Dw, (**a**) NLR and ANN, (**b**) MLR and ANN, and (**c**) FQ model and ANN.

Complex interactions among parameters such as paste quality, aggregate hardness, air-void system, supplementary cementitious materials, curing age, and abrasion duration influence concrete wear. The effect of one parameter frequently depends on the level of another, and these processes interact in a highly nonlinear way. These complex multivariable interactions may not be well represented by conventional regression models like MLR and FQ, which rely on preset mathematical relationships. ANN models, by contrast, do not use a predetermined equation form; instead, they learn these nonlinear correlations directly from the data. This flexibility enables ANN to capture subtle correlations between mixture parameters and test conditions, improving prediction accuracy and lowering error levels.

Concrete wear depth depends not only on its mix but also on the specific abrasion testing procedures used. Factors like the type of abrasion device, load applied, rotational speed, abrasive medium, specimen shape, surface prep, aggregate mineralogy, fly ash properties, and curing process can all affect the recorded wear depth. Even though the current database was assembled from studies using similar standardized tests, the published literature often lacked complete details on these testing parameters. As a result, these factors could not be explicitly included in the machine learning models. Future research should develop standardized, open-access databases that include detailed experimental metadata to enhance the robustness and transferability of materials informatics models.

A strict preprocessing protocol was implemented to ensure the database was consistent and suitable for machine learning. Normalization parameters were estimated solely from the training data and then applied to the validation and testing datasets, preventing information leakage and enhancing the reliability of the predictive results.

While fly ash is well-known as a supplementary cementitious material that can lower Portland cement use and reduce greenhouse gas emissions, this study specifically aims to predict wear depth based on experimentally reported mixture compositions. It does not quantify environmental indicators like embodied carbon, cement savings, CO_2 emissions, or life-cycle impacts. As a result, the findings serve as a predictive framework for material performance rather than a direct evaluation of environmental sustainability.

Although superplasticizer dosage was included as an input variable because it is part of the concrete mixture, its impact on the prediction models was relatively minor. This is mainly because its variation range (2.6–2.9 kg/m^3) is limited in the database. Machine learning models tend to rely more on variables with wider variability for predictions; thus, the small influence of superplasticizer here reflects its narrow dosage range, not its significance in concrete technology.

4. Sensitivity Analysis

To determine how each independent variable affected the modeled dependent variable, a sensitivity analysis was conducted. Sensitivity analysis is an essential step in machine learning research because it provides a physical inter-

pretation of the model and helps identify the most significant variables influencing the prediction results. One parameter was extracted from the training data at a time to calculate the sensitivity index for each parameter in the developed model. Because the ANN was determined to be the most effective model for predicting concrete wear depth compared with the other models tested, it was selected for the sensitivity analysis. For each trial, the R^2, RMSE, and MAE values were reported. **Table 4** presents the sensitivity results. The results indicate that testing time (t) has the most dominant influence on predicting concrete wear depth (**Figure 19**). Model accuracy is significantly reduced when this parameter is removed, confirming that the duration of wear exposure strongly influences wear depth. This result aligns with the physical characteristics of abrasion processes, where longer testing times result in increased material loss and progressive surface deterioration. The second most influential factor is curing age (A). The ongoing hydration and microstructural densification of concrete over time, which raise surface hardness and strengthen abrasion resistance, can account for this outcome. As a result, compared with younger concrete, older concrete specimens typically show lower wear rates [2]. The other factors have a moderate to low impact on the predicted wear depth, including cement content (C), water (W), fly ash (F), fine aggregate (FA), coarse aggregate (CA), superplasticizer (P), and air-entraining agent (AE). These factors mainly affect the internal structure and mechanical characteristics of concrete, including pore structure, aggregate–paste bonding, and paste quality, all of which have an indirect effect on abrasion resistance. However, compared with testing duration and curing age, their effect appears less significant. Overall, the sensitivity analysis indicates that time-dependent factors are the most important in predicting the depth of concrete wear, with mixture component parameters making secondary contributions.

Table 4. Results of sensitivity analysis using the ANN model.

No.	Combination	Removed Parameter	R^2	RMSE (mm)	MAE (mm)	Rank
1	$C, W, F, FA, CA, P, AE, A, t$	-	0.990	0.069	0.052	-
2	$W, F, FA, CA, P, AE, A, t$	C	0.990	0.072	0.055	9
3	$C, F, FA, CA, P, AE, A, t$	W	0.985	0.089	0.064	4
4	$C, W, FA, CA, P, AE, A, t$	F	0.985	0.087	0.063	6
5	C, W, F, CA, P, AE, A, t	FA	0.985	0.090	0.065	3
6	C, W, F, FA, P, AE, A, t	CA	0.985	0.088	0.065	5
7	$C, W, F, FA, CA, AE, A, t$	P	0.989	0.075	0.056	8
8	C, W, F, FA, CA, P, A, t	AE	0.987	0.081	0.060	7
9	$C, W, F, FA, CA, P, AE, t$	A	0.921	0.206	0.155	2
10	$C, W, F, FA, CA, P, AE, A$	t	0.26	0.633	0.514	1

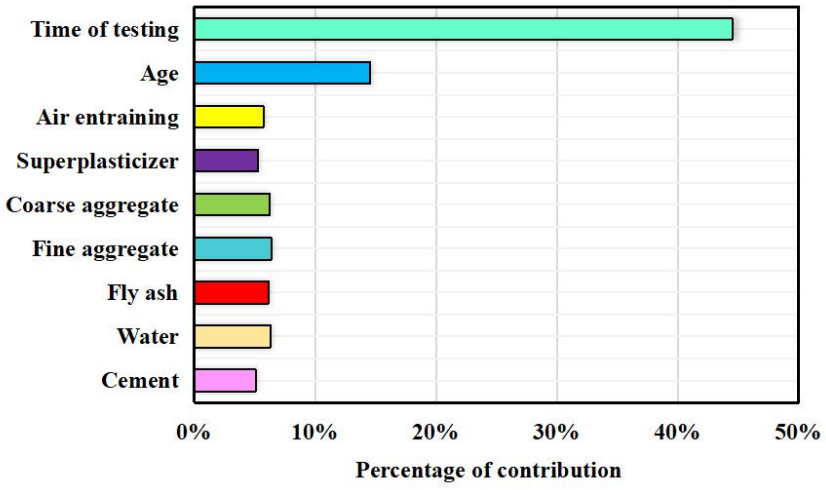

Figure 19. The percentage contribution of input parameters in predicting concrete wear depth using the ANN model.

This feature-removal analysis aims to assess how much predictor variables contribute to the ANN model's predictive ability, not to determine causal physical relationships. Since some mixture parameters are statistically correlated,

the importance values should be viewed as model-dependent indicators rather than direct reflections of material behavior.

5. Conclusions

1. To develop a reliable and efficient model to predict the wear depth of fly-ash-incorporated cement concrete, 439 data samples with different mixture proportions were collected from previous literature. The following conclusions were drawn based on the datasets and results obtained from the models used in this study.
2. Among the developed models, the ANN model demonstrated the highest prediction accuracy, achieving $R^2 \approx 0.99$ and the lowest RMSE and MAE values for both the training and testing datasets.
3. The FQ model ranked second, showing improved performance over MLR and NLR due to its ability to capture nonlinear and interaction effects between variables.
4. The MLR model's moderate accuracy suggests that the complex behavior of concrete wear cannot be well represented by solely linear relationships.
5. The NLR model performed the worst, demonstrating how difficult it is to capture complicated multivariable relationships using specified nonlinear functions.
6. The advantage of ANN over regression-based models was repeatedly demonstrated by comparative assessment using R^2, RMSE, MAE, SI, OBJ, and A-20 index, as well as graphical means (Taylor diagram, radar chart, and residual analysis).
7. Concrete wear is strictly governed by time-dependent processes, as shown by sensitivity analysis, which identified testing duration (t) and curing age (A) as the most important parameters.
8. The proposed machine learning framework aims to assist engineers in initial material screening and mixture assessment. However, it is not meant to substitute standardized laboratory testing, which is still essential for confirming the structural integrity and durability of concrete mixtures.
9. The predictive framework shows strong potential to reduce experimental effort and support the preliminary assessment of wear resistance in fly ash concrete. Nevertheless, the models should be treated as complementary tools, and their predictions should be experimentally verified before use in final engineering design or material qualification.

6. Limitations

I. A limitation of this study is that some experimental details, such as abrasion apparatus specs, rotational speed, abrasive medium properties, aggregate mineralogy, fly ash type, and detailed chemical makeup, were not consistently reported in all source publications. Including these details in future machine learning datasets is likely to improve prediction accuracy and model robustness.

II. Although SHAP and permutation importance significantly enhance model interpretability, they measure statistical contributions instead of direct causal relationships. To determine the causal mechanisms behind abrasion resistance, controlled experiments combined with explainable machine learning methods are necessary.

III. The prediction models developed are only applicable to the concrete mixtures and experimental conditions included in the database. For new materials or conditions beyond this range, predictions should be made with caution and confirmed through proper laboratory tests.

IV. The relatively narrow range of superplasticizer dosage reported in the available literature may have limited its statistical contribution to the prediction models. Future studies incorporating datasets with a wider range of superplasticizer contents could provide a more comprehensive assessment of its influence on wear depth.

Author Contributions

Investigation, methodology, experimental work, data analysis, and writing of the original draft, P.I.A. and H.S.M.A.; reviewing, editing, supervision, and manuscript revision, A.S.M. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The dataset (391) used in this study was obtained entirely from the published dataset reported by Mahmood et al.[2]. No new experimental data were generated. We reused the existing dataset to develop, train, validate, and compare different predictive models, including the Artificial Neural Network (ANN), nonlinear regression (NLR), multilinear regression (MLR), and full quadratic (FQ) models. Researchers interested in the original dataset are encouraged to refer to Mahmood et al.[2] (https://doi.org/10.1007/s41939-024-00564-7). The rest was taken from Siddique et al.[27] (https://doi.org/10.1016/j.conbuildmat.2011.09.011).

Acknowledgments

The authors thank their respective universities for providing a supportive research environment that facilitated the preparation and publication of this work.

Conflicts of Interest

The authors declare no conflict of interest.

AI Use Statement

During the preparation of this manuscript, the authors used ChatGPT solely for language refinement and grammatical editing to improve clarity and readability. The authors did not use artificial intelligence tools for study design, data collection, data analysis, interpretation of results, generation of scientific content, or formulation of conclusions. The authors carefully reviewed, revised, and approved all AI-assisted edits. The authors take full responsibility for the work. Additionally, the manuscript has undergone extensive language revision to enhance originality, reduce repetitive phrasing, and better align with the journal's AI similarity requirements.

References

[1] Subedi, A., Kim, H., Lee, S.J., et al., 2025. Assessing Abrasion Resistance in Concrete Pavements: A Review. Applied Sciences. 15(4), 2101. DOI: https://doi.org/10.3390/app15042101

[2] Mahmood, W.I., Abdulrahman, P.I., Kakasor, D., et al., 2024. Multiscale modeling for accurate forecasting of concrete wear depth: A comprehensive study on mixture proportions and environmental factors. Multiscale and Multidisciplinary Modeling, Experiments and Design. 7(6), 5971–5989. DOI: https://doi.org/10.1007/s41939-024-00564-7

[3] Khan, A., Khan, M., Ali, M., et al., 2024. Predictive modeling for depth of wear of concrete modified with fly ash: A comparative analysis of genetic programming-based algorithms. Case Studies in Construction Materials. 20, e02744. DOI: https://doi.org/10.1016/J.CSCM.2023.E02744

[4] Amini, K., Vosoughi, P., Ceylan, H., et al., 2019. Effect of mixture proportions on concrete performance. Construction and Building Materials. 212, 77–84. DOI: https://doi.org/10.1016/J.CONBUILDMAT.2019.03.294

[5] Ayanlere, S.A., Ajamu, S.O., Odeyemi, S.O., et al., 2023. Effects of water-cement ratio on bond strength of concrete. Materials Today: Proceedings. 86, 134–139. DOI: https://doi.org/10.1016/J.MATPR.2023.04.686

[6] Shah, H.A., Yuan, Q., Zuo, S., 2021. Air entrainment in fresh concrete and its effects on hardened concrete-a review. Construction and Building Materials. 274, 121835. DOI: https://doi.org/10.1016/J.CONBUILDMAT.2020.121835

[7] Bediako, M., 2024. Superplasticized vs. Conventional Concrete: A Comparative Review. Materials Sciences and Applications. 15(9), 271–284. DOI: https://doi.org/10.4236/msa.2024.159019

[8] Cheng, D., Reiner, D.M., Yang, F., et al., 2023. Projecting future carbon emissions from cement production in developing countries. Nature Communications. 14(1), 8213. DOI: https://doi.org/10.1038/s41467-023-43660-x

[9] Hanifa, M., Agarwal, R., Sharma, U., et al., 2023. A review on CO_2 capture and sequestration in the construction industry: Emerging approaches and commercialised technologies. Journal of CO_2 Utilization. 67, 102292. DOI: https://doi.org/10.1016/J.JCOU.2022.102292

[10] Schneider, M., 2019. The cement industry on the way to a low-carbon future. Cement and Concrete Research. 124, 105792. DOI: https://doi.org/10.1016/J.CEMCONRES.2019.105792

[11] John, E., Lothenbach, B., 2023. Cement hydration mechanisms through time – a review. Journal of Materials Science. 58(24), 9805–9833. DOI: https://doi.org/10.1007/s10853-023-08651-9

[12] ASTM, 2019. ASTM C618M − 19. Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete. ASTM: West Conshohocken, PA, USA. DOI: https://doi.org/10.1520/C0618-19

[13] Kanthe, V.N., Kulkarni, T.A., Padalkar, P.A., et al., 2024. Strength and Durability Performance of Fly Ash Concrete with and without Superplasticizer. Journal of Physics: Conference Series. 2779(1), 012001. DOI: https://doi.org/10.1088/1742-6596/2779/1/012001

[14] Amran, M., Fediuk, R., Murali, G., et al., 2021. Fly Ash-Based Eco-Efficient Concretes: A Comprehensive Review of the Short-Term Properties. Materials. 14(15), 4264. DOI: https://doi.org/10.3390/ma14154264

[15] Binod, K., Tike, G.K., Nanda, P.K., 2007. Evaluation of Properties of High-Volume Fly-Ash Concrete for Pavements. Journal of Materials in Civil Engineering. 19(10), 906–911. DOI: https://doi.org/10.1061/(ASCE)0899-1561(2007)19:10(906)

[16] Sumer, M., 2012. Compressive strength and sulfate resistance properties of concretes containing Class F and Class C fly ashes. Construction and Building Materials. 34, 531–536. DOI: https://doi.org/10.1016/J.CONBUILDMAT.2012.02.023

[17] Nayak, D.K., Abhilash, P.P., Singh, R., et al., 2022. Fly ash for sustainable construction: A review of fly ash concrete and its beneficial use case studies. Cleaner Materials. 6, 100143. DOI: https://doi.org/10.1016/j.clema.2022.100143

[18] Jiang, J., Lin, J., Jin, L., et al., 2026. Advancements in Machine Learning for the Development of Cementitious Composites Toward an Intelligent and Green Lifecycle: A State-of-the-Art Review. Engineering. in Press. DOI: https://doi.org/10.1016/J.ENG.2026.02.012

[19] Siddique, R., Aggarwal, P., Aggarwal, Y., 2011. Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks. Advances in Engineering Software. 42(10), 780–786. DOI: https://doi.org/10.1016/J.ADVENGSOFT.2011.05.016

[20] Aiyer, B.G., Kim, D., Karingattikkal, N., et al., 2014. Prediction of compressive strength of self-compacting concrete using least square support vector machine and relevance vector machine. KSCE Journal of Civil Engineering. 18(6), 1753–1758. DOI: https://doi.org/10.1007/S12205-014-0524-0

[21] Sathiparan, N., Jeyananthan, P., 2023. Soft computing techniques to predict the compressive strength of groundnut shell ash-blended concrete. Journal of Engineering and Applied Science. 70(1), 134. DOI: https://doi.org/10.1186/s44147-023-00302-9

[22] Ziyad Sami, B.H., Ziyad Sami, B.F., Kumar, P., et al., 2023. Feasibility analysis for predicting the compressive and tensile strength of concrete using machine learning algorithms. Case Studies in Construction Materials. 18, e01893. DOI: https://doi.org/10.1016/J.CSCM.2023.E01893

[23] Moghaddas, S.A., Bao, Y., 2025. Explainable machine learning framework for predicting concrete abrasion depth. Case Studies in Construction Materials. 22, e04686. DOI: https://doi.org/10.1016/J.CSCM.2025.E04686

[24] Malazdrewicz, S., Sadowski, Ł., 2021. An intelligent model for the prediction of the depth of the wear of cementitious composite modified with high-calcium fly ash. Composite Structures. 259, 113234. DOI: https://doi.org/10.1016/J.COMPSTRUCT.2020.113234

[25] Merabti, S., Meziane, K., Rouabah, S., 2025. Prediction of Shrinkage in Waste Cork Cementitious Composites Using Multilayer Perceptron Neural Networks (MLP). Romanian Journal of Materials/Revista Română de Materiale. 55(3), 171.

[26] Bezari, S., Adda, A., Merabti, S., et al., 2025. Artificial Neural Network Model for Microclimate Performance of Solar Greenhouse with Thermal Storage. Thermal Science. 29(5A), 3603–3614.

[27] Siddique, R., Kapoor, K., Kadri, E.H., et al., 2012. Effect of polyester fibres on the compressive strength and abrasion resistance of HVFA concrete. Construction and Building Materials. 29, 270–278. DOI: https://doi.org/10.1016/J.CONBUILDMAT.2011.09.011

[28] Sargam, Y., Wang, K., Cho, I.H., 2021. Machine learning based prediction model for thermal conductivity of concrete. Journal of Building Engineering. 34, 101956. DOI: https://doi.org/10.1016/J.JOBE.2020.101956

[29] Ghorbanzadeh, S., Armaghani, D.J., Salimi, M., et al., 2026. Predictive models for dynamic properties of soils using machine learning approaches: A comprehensive review. Engineering Applications of Artificial Intelligence. 163, 113014. DOI: https://doi.org/10.1016/j.engappai.2025.113014

[30] Merabti, S., Bezari, S., Boudina, A., 2025. Prediction of the compressive strength in high-strength concrete cores through the destructive method. Sigma: Journal of Engineering & Natural Sciences/Mühendislik ve Fen Bilimleri Dergisi. 43(2), 541.

[31] Shah, M.I., Amin, M.N., Khan, K., et al., 2021. Performance Evaluation of Soft Computing for Modeling the Strength Properties of Waste Substitute Green Concrete. Sustainability. 13(5), 2867. DOI: https://doi.org/10.3390/su13052867

[32] Kakasor Ismael Jaf, D., Ismael Abdulrahman, P., Salih Mohammed, A., et al., 2023. Machine learning techniques and multi-scale models to evaluate the impact of silicon dioxide (SiO2) and calcium oxide (CaO) in fly ash on the compressive strength of green concrete. Construction and Building Materials. 400, 132604. DOI:

https://doi.org/10.1016/J.CONBUILDMAT.2023.132604

[33] Fan, H., Xu, J., Armaghani, D.J., 2025. Addressing environmental challenges in blasting: A dual-method data generation framework for peak particle velocity prediction. Applied Soft Computing. 186(Part D), 114263.

[34] Akeed, M.H., Qaidi, S., Ahmed, H.U., et al., 2022. Ultra-high-performance fiber-reinforced concrete. Part III: Fresh and hardened properties. Case Studies in Construction Materials. 17, e01265. DOI: https://doi.org/10.1016/J.CSCM.2022.E01265

[35] Rai, B., Kumar, S., Satish, K., 2014. Effect of Fly Ash on Mortar Mixes with Quarry Dust as Fine Aggregate. Advances in Materials Science and Engineering. 2014(1), 626425. DOI: https://doi.org/10.1155/2014/626425

[36] Marangu, J.M., 2020. Prediction of Compressive Strength of Calcined Clay Based Cement Mortars Using Support Vector Machine and Artificial Neural Network Techniques. Journal of Sustainable Construction Materials and Technologies. 5(1), 392–398. DOI: https://doi.org/10.29187/jscmt.2020.43

[37] Sharaky, I.A., Ghoneim, S.S.M., Abdel Aziz, B.H., et al., 2021. Experimental and theoretical study on the compressive strength of the high strength concrete incorporating steel fiber and metakaolin. Structures. 31, 57–67. DOI: https://doi.org/10.1016/J.ISTRUC.2021.01.061

[38] Golafshani, E.M., Behnood, A., Arashpour, M., 2020. Predicting the compressive strength of normal and High-Performance Concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer. Construction and Building Materials. 232, 117266. DOI: https://doi.org/10.1016/J.CONBUILDMAT.2019.117266