Developments in the Built Environment 15 (2023) 100220

Contents lists available at ScienceDirect

Developments in the Built Environment

journal homepage: www.sciencedirect.com/journal/developments-in-the-built-environment

A metaheuristic-guided machine learning approach for concrete strength prediction with high mix design variability using ultrasonic pulse velocity data

S. Selcuk [a], P. Tang [b,*]

[a] Cankaya University Main Campus, Civil Engineering Department NA17, Ankara, Turkiye
[b] Carnegie Mellon University, Civil and Environmental Engineering Department, Porter Hall, 123F, Pittsburgh, PA, USA

ARTICLE INFO

Keywords:
ANN
Ultrasonic pulse velocity
Deep learning
Non destructive testing
Concrete strength assessment
Metaheuristic algorithms

ABSTRACT

Assessment of concrete strength in existing structures is a common engineering problem. Several attempts in the literature showed the potential of ML methods for predicting concrete strength using concrete properties and NDT values as inputs. However, almost all such ML efforts based on NDT data trained models to predict concrete strength for a specific concrete mix design. We trained a global ML-based model that can predict concrete strength for a wide range of concrete types. This study uses data with high variability for training a metaheuristic-guided ANN model that can cover most concrete mixes used in practice. We put together a dataset that has large variations of mix design components. Training an ANN model using this dataset introduced significant test errors as expected. We optimized hyperparameters, architecture of the ANN model and performed feature selection using genetic algorithm. The proposed model reduces test errors from 9.3 MPa to 4.8 MPa.

1. Introduction

Reliable concrete strength evaluation is critical for accurately assessing the conditions of existing structures. The compressive strength of concrete can vary considerably depending on concrete age and exposure to environmental conditions. Non-destructive testing (NDT) methods are available to assess in-situ concrete strength. However, using these NDT methods with little contextualized adjustments can be unreliable because collected data depends significantly on environmental conditions and is prone to human errors. Often NDT methods are used in collaboration with a destructive testing method for concrete strength evaluation because most standalone NDT results can be affected by ambient conditions and testing procedures (Breysse, 2012). It is, therefore, necessary to use a combination of destructive and NDT methods for improved accuracy when assessing in situ concrete strengths. However, the procedures for using destructive testing methods are usually costly, time-consuming, and rely on heavy equipment use. These destructive testing techniques also depend on the data interpretation by experienced staff and require repairs since they damage concrete.

Many non-destructive methods are available to evaluate in-situ strength, such as rebound hammer, Windsor probe, infrared thermography, and radiographic testing. Among these methods, the Ultrasonic Pulse Velocity (UPV) test is reliable, truly non-destructive, and easy to apply in multiple field conditions with relatively better repeatability. UPV is commonly used to evaluate concrete properties with a destructive testing method (Sbartaï et al., 2012). In the literature, there has been a search for evaluation procedures that could eliminate or reduce the use of destructive testing. Using more than one NDT method to evaluate material strength is also recommended so that each technique complements the others (Breysse, 2012) (Sbartaï et al., 2012).

Machine Learning (ML) methods, such as Artificial Neural Networks (ANN), have shown potential for improving concrete strength prediction based on NDT data, especially in the last decade (Sbartaï et al., 2012) (Yeh, 2009). ANN models have the advantage of detecting nonlinear relationships between inputs and outputs without having the need to assume any behavioral connections. Whereas their main disadvantage is finding optimum hyperparameters for the best accuracy without knowing the exact inner workings of the model. Some studies used ANN and NDT results to predict concrete strength. Few of these studies focus on using UPV as the NDT data source. The critical aspects of prediction models are their accuracy, systematic bias, and applicability to various concrete types and mix designs. Although high accuracies were reported in the literature, most studies lacked applicability for a wide range of the mix design parameters. Usually, models were trained on a single

* Corresponding author.
E-mail addresses: sselcuk@cankaya.edu.tr (S. Selcuk), ptang@andrew.cmu.edu (P. Tang).

https://doi.org/10.1016/j.dibe.2023.100220
Received 9 May 2023; Received in revised form 23 August 2023; Accepted 24 August 2023
Available online 25 August 2023

concrete type using low variance data resulting in a high bias when applied to a different mix design. The accuracy would suffer when such a model is tested on data with high variance in the mix design (Paixão et al., 2022).

The proposed model has the potential to be used as part of an alternative structural evaluation scheme for existing buildings. Such a global model can be used in applications where the tradeoff between speed of evaluation and accuracy is in favor of speed rather than top-notch accuracy. For regions that are known to be prone to disasters such as an earthquake, it is essential to evaluate the building stock for disaster preparedness; concrete strength evaluation is an integral part of this evaluation. Likewise, after a disaster, one of the most urgent and costly needs is to evaluate standing buildings for safety. An alternative procedure that integrates the current model can eliminate the use of destructive testing for concrete strength evaluation and reduce the cost and duration of the operations to a great extent. A striking example demonstrating the importance of utilizing machine learning to improve structural evaluation procedures would be the post-disaster evaluation of structures. After a disaster such as an earthquake, one of the most urgent and costly needs is to provide shelter for the survivors. Logistics and costs for short-term shelters can be hard to organize, while it can be a very high priority, especially in harsh climates. There are various reported instances where survivors took shelter in buildings that did not collapse, only to perish when they collapsed during the aftershocks. Suppose an evaluation procedure that utilizes either of the proposed models to evaluate concrete structures in the disaster region can be standardized. In that case, it can be a fast and economical solution that would reduce the need for short-term shelters and prevent loss of life during aftershocks.

We wanted to see if we could train an ML model with acceptable accuracy and could be applied directly to any field data during concrete strength assessment of buildings in a large region. Concrete used in buildings in any urban setting would have a wide range of mix designs and different ages. A strength prediction model needs further adjustments for handling such concrete mix diversity. It is common to use additional optimization methods to augment ANN algorithms to increase model accuracy on different mix design parameter ranges. Metaheuristics is among the most successful optimization methods integrated with ANN (Ojha et al., 2017), lacking such an application on concrete strength prediction. Using ANN and metaheuristic algorithms, the authors trained a global model that can predict concrete strength based on UPV data and concrete properties for a wide range of concrete types. The model can be utilized for applications where fast evaluation is necessary, and top-level accuracy is not critical, such as post-disaster response operations.

The work presented in this paper systematically investigates the augmentation of a machine learning algorithm (an ANN algorithm) with a metaheuristic component to assess concrete strength for various concrete types and mix designs. The authors found no studies in the literature on concrete strength prediction using ML for a wide range of concrete types and mix design parameter ranges. We compiled a diverse UPV concrete tests database through extensive searching and organization of UPV testing results. The database contains a wide range of mix design parameter ranges. This comprehensive UPV tests database provides a basis for training and testing a new machine learning model that can reliably predict the strengths of a wide range of the mix design parameter ranges.

The contributions of the work presented in this paper include 1) *a cleaned and organized concrete mix design dataset with corresponding UPV and compressive strength data for the research community to test various concrete strength prediction methods based on UPV data*; 2) *a new approach that uses metaheuristic algorithms to augment an ANN algorithm for producing a machine learning model that can reliably predict the strengths of a wide range of mix designs*. The following sections present a review of relevant studies (section 2), the methodology of compiling an extensive UPV tests database and training a new metaheuristic-guided ANN model (section 3), and the data compilation and the new algorithm's testing results (sections 4 and 5). The last two sections present the discussions and conclusions (sections 6 and 7).

2. Literature review on concrete strength prediction using ANN and UPV

This section reviews existing studies using various ANN approaches predicting concrete strength. As different studies could use varying error metrics to report the performance of the ML models, an objective review of these existing ML studies needs to summarize the metrics shared in previous studies. Coefficient of determination (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percent Error (MAPE) are the most commonly used metrics in the literature, as detailed in the "methodology" section below. There are many studies that predict concrete strength using ANN spreading through more than two decades. (Yeh, 1998; Duan and Poon, 2014; Naderpour et al., 2018). These studies mostly reported relatively high accuracies and low RMSE values for their testing datasets. A portion of the mentioned literature have small datasets that allow high accuracy predictions. Another large part of these literature focuses on particular concrete types such as foamed concrete, roller compacted concrete, recycled aggregate concrete, etc. Focusing on a specific concrete type allows low variations in mix design, thus, comparably high accuracies. A recent study addressed this situation where authors tested an ANN model that was trained and cross validated using a portion of Yeh's dataset (Yeh, 1998). The model's RMSE output increased from 3.4 MPa to 4.09 MPa when they used a test dataset gathered from the literature. They concluded regionalization and lack of heterogeneity of some datasets can lead to overly optimistic results when models are intended to be used universally for strength prediction or mix design (Paixão et al., 2022).

As mentioned in the previous section, we aim for a global ML-based model trained to predict concrete strength for a wide range of concrete types. High accuracy results reported in the current state of literature would be obsolete for high variability mix design data because the models were trained with low variability data. In order to obtain acceptable accuracy for such a large variability, we focused on three approaches that are known to improve accuracy for ML models in general, i.e., increasing training data size, including an input variable closely related to the target parameter, and hyper parameter optimization. The materials and methods section discusses the first of these three approaches. As for the second approach, we proposed using NDT data, which is known to improve strength prediction. As discussed previously, we decided to use UPV data as an input when training ANNs to improve the accuracy of concrete strength prediction. When the limited literature on ANN studies for concrete strength prediction using UPV data is investigated, we observed similar limitations in previous ANN studies. Atici (2011) predicted concrete strength using an ANN model with values of the mix design variables and UPV data as the inputs (Atici, 2011). However, their model had high R^2 values primarily due to the mix design parameters' relatively small value ranges.

Consequently, the trained ANN model could not reliably predict the strength of different concrete types or concretes with vastly different mix designs. In a similar study by Kewalramani and Gupta (2016), the authors used concrete specimens with different shapes and dimensions in their ANN model (Kewalramani and Gupta, 2006). The percent error measured was reported as 25%. They trained an individual ANN model for each specific shape and dimension. Specimens were tested at different ages and under various curing conditions. Still, the ages and curing conditions were not included as input features in the ANN models, which led to high test errors. As a result, the models had low accuracy and limited applicability for wide input parameter ranges.

Tang et al. (2007) used the mix design parameters and UPV values as inputs for an ANN model that predicts concrete strength (Tang et al., 2007). Additionally, they performed a sensitivity analysis for dimension reduction. They determined model topology and hyperparameters and

performed a sensitivity analysis using validation data. Similar to the previous studies, the listed data set had limited variability, which restricted the models' prediction ability to a specific type of concrete and a small parameter value range. Sadowski et al. (2019) used quartz and feldspar replacement rates and UPV measurement to predict compressive strength of concrete screeds with high volume of waste quartz mineral dust. They reported an R^2 of 0.91 with very limited mix design variability and small dataset size (Sadowski et al., 2019). Tenza-Abril et al. (2018) used inputs for lightweight concrete (Tenza-Abril et al., 2018). The coefficient of determination was 0.82. Trtnik et. et al.09) reported a high coefficient of determination; however, it is essential to note that parameter value ranges tested in that study were limited (Trtnik et al., 2009). The entire data set consisted of data from a single batch of concrete. Bilgehan and Turgut (2010) conducted a study using data from core specimens and trained an ANN model with the only input being UPV values (Bilgehan, 2011). They reported a high coefficient of determination. However, the dataset they created also had low variability. Such low variability has been the case for most literature on this topic. Biswas et al. (2019) used silica fume efficiency factor and silica fume content and UPV measurements to predict concrete strength with emotional neural networks. They reported an R^2 of 0.949 and RMSE of 1.52 MPa. They provided limited information on network architecture and the training dataset. (Biswas et al., 2019). The most recent literature on this topic is a paper by Albuthbahak et al. (Albuthbahak and Hiswa, 2019), where ANN yielded the highest coefficient of determination among various prediction techniques while still not handling high variability.

As seen from the literature review, the strength prediction capability of ANN models developed are mostly limited to a specific concrete type or mix design. This results in high accuracy in test datasets. The last of the three approaches known to improve accuracy is hyperparameter optimization. Structural parameters of an ANN, such as learning rate, architecture, etc., that are not trained during the model training are called hyperparameters. Setting or optimizing hyperparameters can change the training outcome to a great extent.

Model topology and hyperparameters were selected using validation data or a simple grid search for all these references. It has been more common to do a simple grid search for determining hyperparameters or using values from previously published literature. However, as trends in neural networks shifted to hybrid algorithms to increase accuracy, some researchers also applied this approach to concrete strength prediction. As mentioned earlier, we used Metaheuristics for hyperparameter optimization in this study. There are few recent studies that used metaheuristics for the optimization of ANN for concrete strength prediction. Zhang and Aslani (2021) used Genetic Algorithm optimization on a neural network to predict compressive strength of lightweight concrete using mix design properties and UPV values. They reported an 11% improvement for training RMSE. Test RMSE was 4.51 MPa for the GA-NN model (Zhang and Aslani, 2021). Ly et al. (2021) (Ly et al., 2021) reported a 0.12 MPa test RMSE that was improved to 0.093 MPa with metaheuristic optimization. Although the paper was limited to foam concrete, the improvement was significant with the PSO algorithm optimization. Another noticeable study by Ranjbar et al. (2020) (Ranjbar et al., 2020) reported 65.41 MPa test RMSE with ANN and 49.11 MPa RMSE with metaheuristic optimization of ANN. However, this paper also limited its prediction capabilities to roller compacted concrete. Hybrid ANN studies are also either trained and tested on small datasets and/or focused on a specific type of concrete, which limits their practical implications.

In order to obtain acceptable accuracies from an ML model for various concrete types, we focused on three strategies: a large dataset with high variability, using UPV data as an input known to be highly correlated to the strength of concrete, and hyperparameter optimization. We discussed the compilation of the dataset in Chapter 4. We examined different optimization approaches to the model topology and hyperparameters, as detailed in section 5.

3. Research methodology

The overall research approach has two components: 1) compilation of an extensive UPV tests database that capture the UPV data collected and the contextual information of those tests (e.g., dimensions of the tested concrete components, Age); 2) development and tests of a series of metaheuristics-guided ANN models using the extensive UPV tests database. We studied the second component in three stages: a) hyperparameter optimization, b) dimensionality reduction, and c) simultaneous optimization d) hyperparameter optimization with field applicable feaures. Fig. 1 shows the general scheme for the workflow of the study. The following paragraphs introduce these two components at the conceptual level. The sections following this section present how the research team carried out these research activities with technical details and the obtained results (sections 4 and 5).

In population-based metaheuristic algorithms, a predefined number of population members are created, and each population member is assigned various input parameters called "coordinates." They are also assigned a fitness value calculated based on their coordinates using a fitness function. In our case, this fitness value is the validation error of the ANN model trained with coordinates assigned to the population member. At each iteration, population members will explore the solution space and update their coordinates to get closer to the members with the smallest fitness values. The main parameters of a metaheuristic algorithm are population size and the number of generations. Population size refers to the number of population members in each generation. Setting the number of generations decides how many iterations a population will carry out.

The proposed metaheuristics-guided ANN models use contextual data such as dimensions, Age, and so on, together with raw ultrasonic pulse velocity (UPV) data as input features, each constructing a different machine learning (ML) model. The models aim to extract the complex relationship between these input features and the compressive strength of concrete as output. With the need for training and testing such ML models, we created a large dataset using literature data found through an extensive search for literature starting from 1993. The focus of literature search and data compilation are those published studies that detail the UPV data and contexts of the tests. Papers that introduced noise and variability from different sources of raw materials, raw material properties, mixing and placing techniques, mix designs, environmental exposures, etc., are included during database compilation. The total number of articles searched is 912. We demonstrated the variability of this dataset in chapter 4. We initially collected all papers that report UPV test data and compressive strength values through the database compilation. We only performed this initial compilation among peer-reviewed SCI and SCIE papers to ensure good data quality.

Data preprocessing overview. We prepared a total of 11,096 rows of data points from the literature to be used in the training, validation, and testing of the ANN model. This number was reduced to 5205 after the data preprocessing. The data processing includes 1) contextual core parameter identification through random forest-based feature importance analysis, 2) data size reduction by removing all data points that did not contain core parameter values, and 3) initial model training. Datapoints missing core parameter values were especially prevalent in the studies published before 1995. Most studies published before 1995 report limited information about the specimens.

To plan model training and testing, we explored error metrics used in past ML studies for concrete strength prediction using UPV data. As mentioned in the literature review chapter, Coefficient of determination (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percent Error (MAPE) are the most common error metrics in this domain. Mean Squared Error (MSE) is the averaged squared error, and as the name implies, the Root Mean Squared Error (RMSE) is calculated by taking the root of the mean squared error. Because both MSE and RMSE values depend on the target parameter's value range, prediction accuracy can be better perceived when the target parameter's

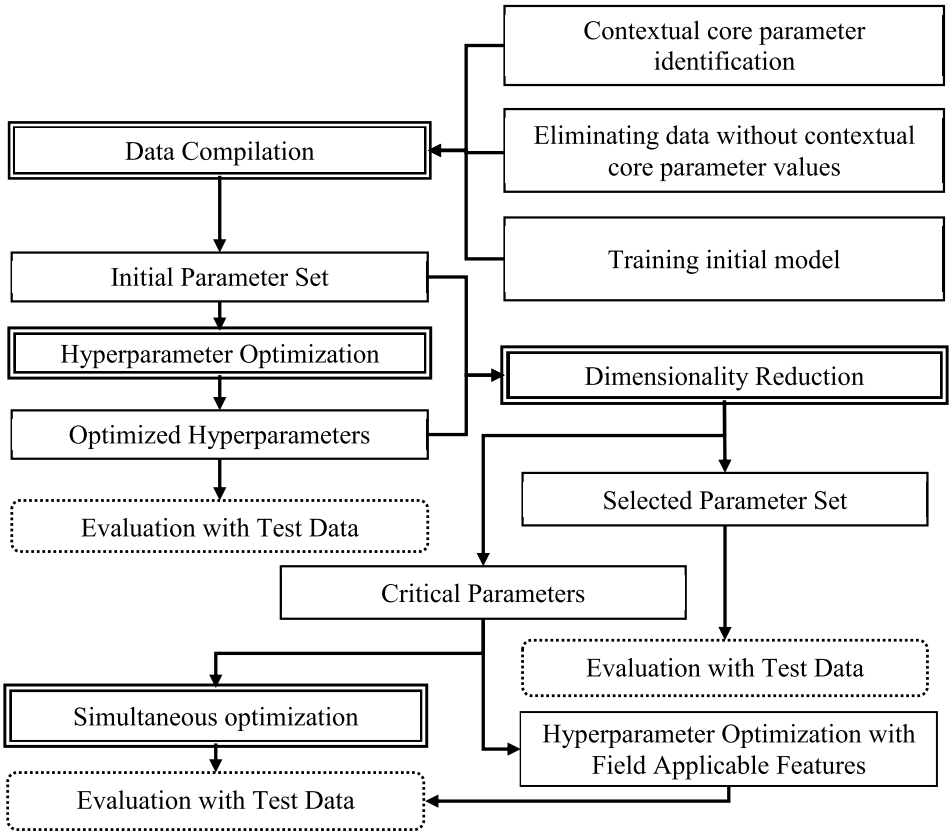

Fig. 1. Overview of the UPV tests data compilation and machine learning research based on the compiled data.

(compressive strength) value range is provided. Consequently, comparing different ML models with the Coefficient of Determination (R2) and Mean Absolute Percent Error (MAPE) is more straightforward. MAPE is calculated by calculating the percent deviation of error from the actual value of the target parameter. The coefficient of determination (R^2) is the square of the correlation between actual and predicted values, commonly used in machine learning studies. Accordingly, R^2 values close to 1 indicate better predictions.

Model training and testing overview: The development and testing of the metaheuristics-guided ML models have the following elements: 1) hyperparameter optimization, 2) dimensionality reduction, and 3) simultaneous optimization. Hyperparameter optimization aims to select hyperparameters that minimize the prediction error (RMSE). The purpose of dimensionality reduction is to reduce the number of input features of the model without decreasing accuracy to increase overall computational efficiency. Simultaneous optimization aims to select optimal hyperparameters while minimizing the number of input features to reduce the prediction error (RMSE) more efficiently.

In the *hyperparameter optimization part,* we performed hyperparameter tuning using metaheuristic algorithms with the initial ANN model, as seen in Fig. 2. We used different meta-heuristic algorithms designed to work together with ANN. We created an initial population for the metaheuristic algorithms using several hyperparameter values as the coordinates of the individuals in the population. The fitness value of each individual in the population was the RMSE of each ANN run for corresponding hyperparameter values. We performed a comprehensive search for the optimization of a set of hyper-parameters to minimize model errors.

In the d*imensionality reduction part,* we used meta-heuristic approaches again to decrease the number of features. Fig. 3 shows the process of dimensionality reduction. We utilized meta-heuristic algorithms with top-down and bottom-up procedures to achieve a more efficient model with fewer features. We included an entire set of features in the top-down approach in the initial model. As the optimization algorithm proceeds, a random number of features are removed from the features list. In the bottom-up approach, we started with a few features identified in the data preprocessing as contextual core features. As the optimization algorithm proceeds, a random number of features are added to the features list.

Consequently, we created an initial population by assigning each member a random number of features smaller than the total number of features. The population members have different features and also a different number of features at each meta-heuristic step for both approaches. The fitness value of the populations is the validation RMSE of the model with this smaller set of features.

In the *simultaneous optimization* part of the study, we implemented a genetic algorithm to optimize the ANN model for hyperparameters and dimension reduction simultaneously, as seen in Fig. 4. We only implemented the best working meta-heuristic method identified in previous steps to work with ANN. This time we used hyperparameters plus the name and number of features to be added to the model as coordinates of the individuals in the population. The fitness value of the population is the validation RMSE of the model run with the corresponding hyperparameters and features. We finally performed *hyperparameter optimization with field applicable features* where we eliminated features in the final model trained in simultaneous optimization that might be harder to access in a field operation. We performed the hyper-parameter optimization using GA on the final model with only selected features.

4. Compilation of an extensive UPV tests database

As shown in Fig. 1 earlier, the dataset compilation has three major steps. The first step is contextual core parameter identification using random forest feature importance. Next, we eliminated data that did not contain information on contextual core features. Finally, we ran an initial model using the most common values of hyperparameters

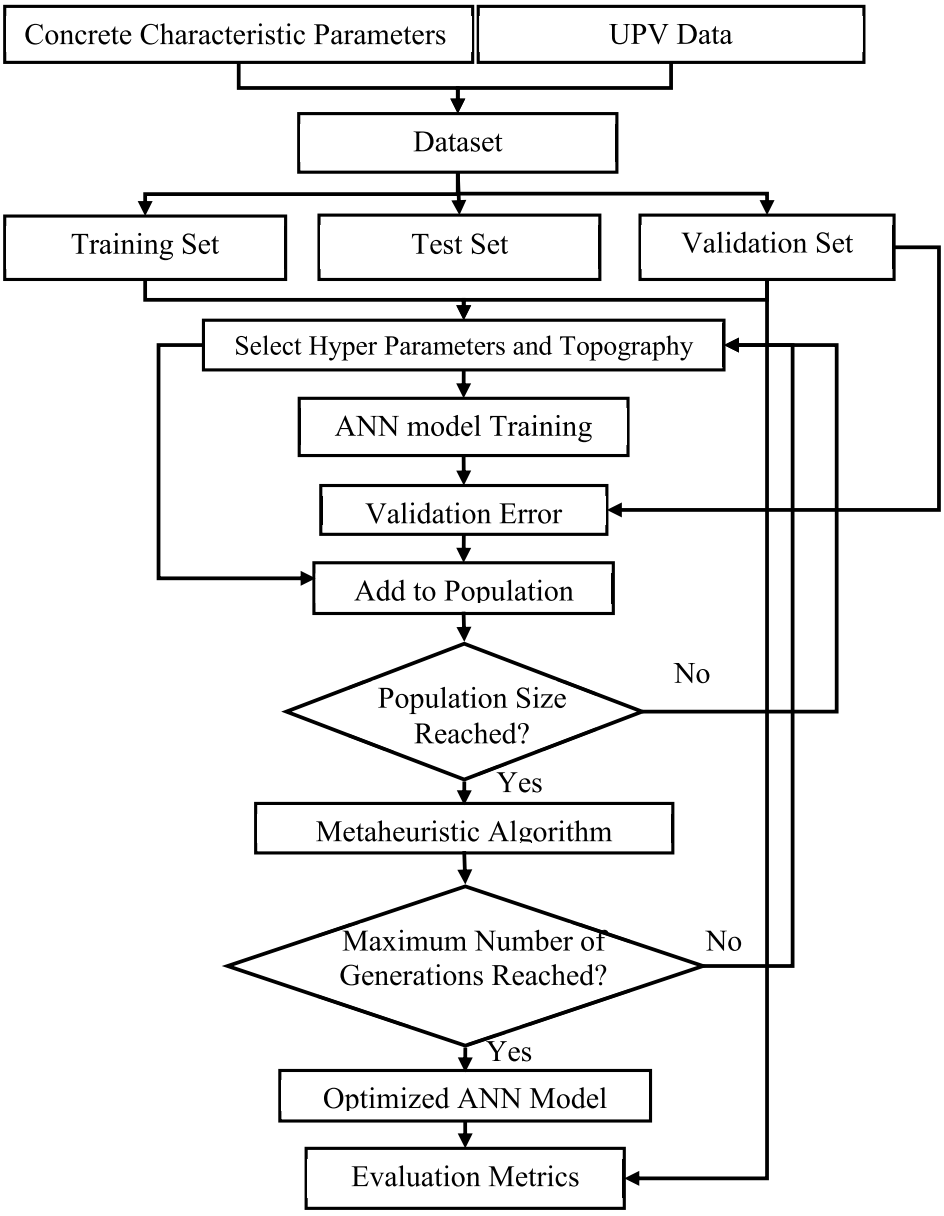

Fig. 2. Flow chart for hyperparameter optimization.

deduced from the literature. This section details these data preprocessing and cleaning steps.

We identified critical features using mean decrease impurity (MDI) as a part of dimension reduction. Random forest is a supervised learning method that combines bagging and decision tree methods. It is commonly used for feature selection because the random forest algorithm can produce MDI values for features. Mean decrease impurity is calculated based on how much each parameter decreases the weighted impurity of a tree in a random forest. Features can be ranked according to their importance using the average impurity decrease from each feature. This MDI is defined as "Feature Importance" in a random forest. We used the random forest method to list all the features according to their feature importance for strength prediction. We selected features with feature importance higher than 0.01 as core features in the initial data cleaning phase. Some data rows have missing values of these features that are known to be relatively more critical for strength prediction. We chose to eliminate those rows because high importance features are more relevant for predicting the target, and if highly relevant information is missing, that will compromise the model's accuracy. As a result, we eliminated all literature references that did not report any of the core features, which reduced the dataset size to 4062 rows and the number of references to 53. (Ly et al., 2021; Ranjbar et al., 2020; Güneyisi et al., 2009; Shishegaran et al., 2021; Fan et al., 2016; Farahani et al., 2017; Yusuf et al., 2016; Yaman, 2000; Yew et al., 2014; Yap et al., 2016; Yaqub and Bailey, 2016; Thirumurugan and Sivakumar, 2013; Willard, 2001; Uysal et al., 2012; Türkmen et al., 2003; Türkmen, 2003; Zhong and Yao, 2008; Sua-Iam and Makul, 2013; Al-Akhras, 1995; Aqel, 2016; Tanyidizi and Coskun, 2008; Goutham Sai and Singh, 2019; Şahmaran et al., 2006; Selcuk et al., 2012; Oeswein, 2000; Nacer, 2005; S Marikunte et al., 2010; Dolatabadi, 2013; Dervisoglu, 2002; Becker, 2000; Koh, 2014; Langevin, 1993; Latif, 1995; Hamood, 2014; Lai et al., 2001; Chao-Lung et al., 2011a; Lin et al., 2016; Jain et al., 2013; Irrigaray et al., 2016; Khan et al., 2007; Ikpong, 1993; Karagol et al., 2015; Panesar and Shindman, 2011; Pal, 2019; Chu, 2012; Ozbay et al., 2011; Owaid et al., 2017; Popovics, 1993; Sadeghi Nik and Lotfi Omran, 2013; Chao-Lung et al., 2011b; Asteris and Kolovos, 2017; Atahan et al., 2011; Asteris et al., 2017; Demirboğa et al., 2004). Dataset contains data from 1993 to 2021, based on more than 15 locations. Most of the data was

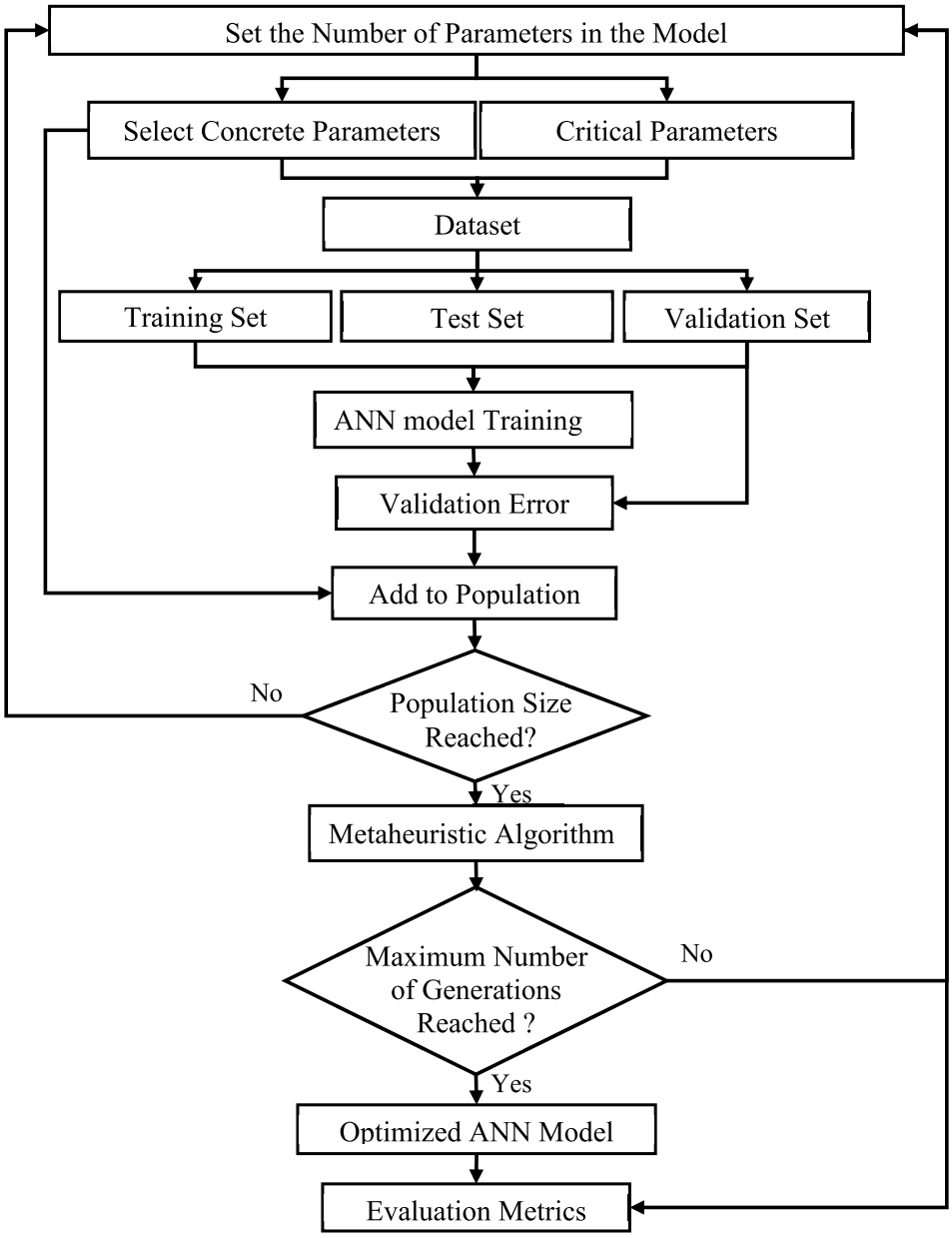

Fig. 3. Flow chart for dimensionality reduction.

obtained in the form of kg/m^3 of concrete in terms of units, a few of the data was converted from percent to kg/m^3 using information in the paper.

It is important to note that concrete mix designs commonly include more than one chemical and/or mineral admixture. We listed three different chemical admixtures and two different mineral admixture features for admixture content and type. When a mix design has more than one chemical and/or mineral admixture type, we listed the admixture with the most significant amount as admixture 1. The other admixtures are listed as admixture 2 and 3, respectively, following the order of decreasing contents. After data cleaning, the remaining dataset contained both numerical and categorical data. Numerical and Categorical features and their ranges are listed in Table 1 and Table 2, respectively. We used one hot encoding to implement categorical features in ANN model training.

We trained an initial ANN model using hyperparameters and network topology common for concrete strength prediction in the literature (Atici, 2011; Kewalramani and Gupta, 2006; Tang et al., 2007; Sadowski et al., 2019; Tenza-Abril et al., 2018; Trtnik et al., 2009; Bilgehan, 2011; Biswas et al., 2019; Albuthbahak and Hiswa, 2019). The model's initial performance was not satisfactory as expected from a dataset containing various data resources. The coefficient of determinations observed from different run trials was around 0.70-0.76. We decided to use root mean squared error (RMSE) as the evaluation metric along with R^2 for the study because it is one of the most common metrics reported in the literature. We used a simple grid search to tune various hyperparameters using the validation data. The selected features after initial hyperparameter tuning are listed in Table 3. When these hyperparameters are used to run the model, the model runs around 25 microseconds for training with the current setup. Additional metrics are given in Table 4. As shown in Fig. 5, points that correlate the predicted and actual strength values lie roughly along the centerline, corresponding to the exact prediction of the actual strength.

In the initial model, we analyzed the data points with high errors for each parameter, as shown in Table 5 through 7. We identified two thresholds for the margin of prediction error, i.e., 5% and 20%. In the

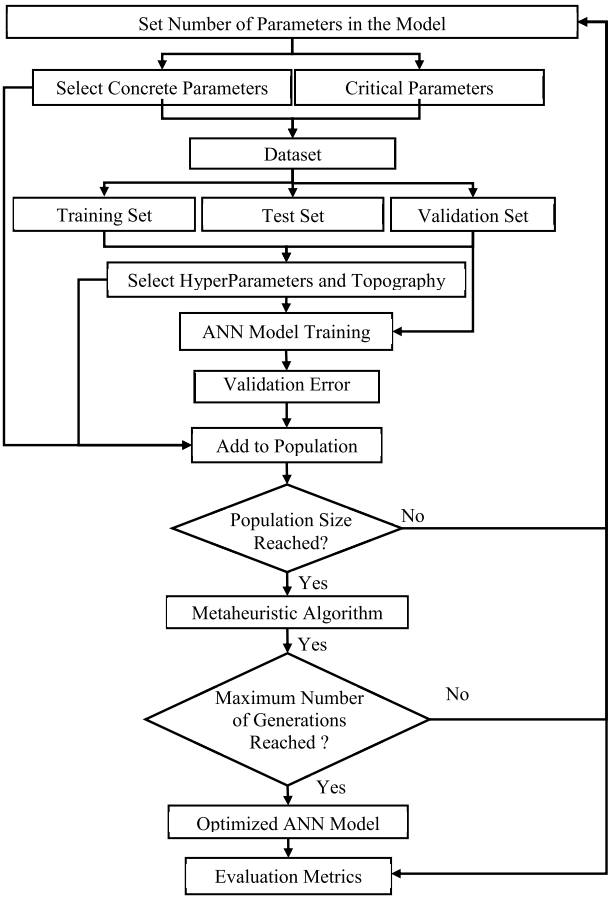

Fig. 4. Flow chart for simultaneous optimization.

Table 1
Data distribution for numerical features.

Parameter	Mean	Standard deviation	Min	Max
Age (Days)	58.96	105.03	0.83	570
UPV Reading (Km/s)	4.38	0.83	0.75	6.96
Compressive Strength (MPa)	47.56	23.92	0.88	134.03
Specimen Dimension 1 (mm)	118.99	71.48	50	914.4
Specimen Dimension 2 (mm)	164.54	58.66	50	304.8
Maximum Aggregate Size (mm)	17.58	9.72	0.01	38
Coarse/Fine Aggregate	1.29	0.66	0	4.75
Cement/Total Aggregate	0.32	0.36	0.06	5.47
Water/Binder	0.42	0.19	0.03	2
Chemical Admixture Count	0.94	0.72	0	3
Mineral Admixture Count	0.62	0.68	0	2
Chemical Admixture 1 Content (kg/m^3)	1.53	2.42	0	13.11
Mineral Admixture 1 Content (kg/m^3)	48.67	107.69	0	922
Chemical Admixture 2 Content (kg/m^3)	0.36	1.45	0	9
Chemical Admixture 3 Content (kg/m^3)	0.03	0.38	0	4.5
Mineral Admixture 2 Content (kg/m^3)	8.63	25.76	0	180
Mineral Admixture 1Calcium Content (%)	8.52	26.05	0	96
Mineral Admixture 1 Silica Content (%)	1.47	8.68	0	88
Air Entrainment %	0.26	1.24	0	11.5
Fiber Aspect	1.02	8.98	0	80
Fiber Content (% Vol)	0.01	0.02	0	0.3
Curing Temp (C)	36.52	63.31	-20	500
Trajectory Distance (mm)	163.67	82.18	50	914
Frequency (kHz)	62.95	44.06	50	340

Table 2
Categorical features.

Parameter	Number of Categories
Coarse Aggregate Type	4
Specimen Shape	3
Concrete Type	5
Cement Type	5
Chemical Admixture 1 Type	6
Mineral Admixture 1 Type	5
Chemical Admixture 2 Type	6
Chemical Admixture 3 Type	6
Mineral Admixture 2 Type	5
Fiber Type	3
Curing Method (Exposure)	4
UPV Configuration	3
Transducer Contact Material	3

Table 3
Hyperparameters for the initial model.

Optimization algorithm	Adam
Starting Learning Rate	0.06
Learning Rate Type	Adaptive
Hidden Layers	2
Nodes at Each Hidden Layer	20*20
Activation Function for Each Hidden Layer	tan h
Weight Initialization Method	Xavier
Epoch	2500
Early Stopping	Yes

Table 4
Performance of the initial model.

Run Time-Total	25 ms
Training Error (RMSE-MPa)	8.9
Validation Error (RMSE-MPa)	9.2
Training Coefficient of Determination (R2)	0.76
Validation Coefficient of Determination (R2)	0.72

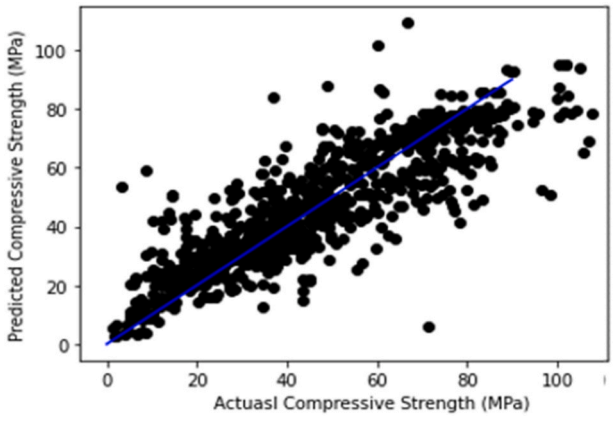

Fig. 5. Actual strength and predicted strength for the validation data in the initial model.

analysis, we identified and isolated data points where prediction errors were larger than either 5% or 20% in the two groups. Data points where the predicted strength is larger (or smaller) than 20% of the actual strength are the "extreme outliers" where the initial model failed with an unacceptable margin. Data points where the predicted strength is larger (or smaller) than 5% of the actual strength are "practical outliers" where the model's prediction performance is below standards used in practical applications. We divided features into three groups, i.e., categorical, discrete, and continuous. We identified the average values of each parameter for the entire dataset for the "extreme outliers' group" (20%

Table 5
Summary of outlier data points for the continuous features in the initial model compared to the entire dataset.

Parameter	Average value (Dataset)	Average value (Outlier groups)	
		20%	5%
UPV Reading (Km/s)	4.38	4.53	4.54
Compressive Strength (MPa)	47.56	41.47	35.45
Coarse/Fine Aggregate	1.29	1.15	1.24
Cement/Total Aggregate	0.32	0.27	0.29
Water/Binder	0.42	0.35	0.36
Chemical Admixture 1 Content (kg/m^3)	1.53	1.2	1.42
Mineral Admixture 1 Content (kg/m^3)	48.68	***111.61***	***97.17***
Chemical Admixture 2 Content (kg/m^3)	0.36	0.13	0.31
Mineral Admixture 2 Content (kg/m^3)	8.63	12.5	19.26
Mineral Admixture 1 Calcium Content (%)	8.52	17.94	21.26
Mineral Admixture 1 Silica Content (%)	1.47	0	2.1
Air Entrainment %	0.26	0.55	0.23
Fiber Content (% Vol)	0.01	0	0

Table 6
Summary of outlier data points for the discrete features in the initial model compared to the entire dataset.

Parameter	The most common value	% of the most common value	The most common value		% of the most common value	
	(Dataset)		(Outlier groups)			
			20%	5%	20%	5%
Age (Days)	28	40	28	28	39	38
Specimen Dimension 1 (mm)	100	61	100	100	75	81
Specimen Dimension 2 (mm)	200	39	200	200	50	75
Maximum Aggregate Size (mm)	13	18	13	13	28	38
Chemical Admixture Count	1	51	1	1	64	62
Mineral Admixture Count	0	***47 (41)***	***1***	***1***	***50***	***50***
Fiber Aspect	0	99	0	0	100	100
Curing Temp (C)	23	71	23	23	79	81

or more prediction error) and the 'practical outliers' group" (5% or more prediction error) for continuous features. For the continuous features, the average values of most features for both outlier datasets are close to that of the entire dataset, as shown in Table 5. The most significant difference between the dataset average and the average of outliers is for the mineral admixture content features. This fact implies that the model is prone to overestimate when the mineral admixture content of the concrete is high.

The initial model will predict strength more poorly when at least one type of mineral admixture is present in the mix design. We identified the most common values of each parameter for the dataset and the two outlier groups for discrete and categorical features. We also listed the occurrence of the most common values in percent for both the dataset and the outlier groups. As can be seen from Table 6, the most common values in both outlier sets are the same as the most common value of the entire dataset for all discrete features except the Mineral Admixture Count parameter. 50% of the outlier data points have a Mineral Admixture count of 1, compared to 41% in the total dataset. The most common value (47%) for mineral admixture count is zero in the entire dataset.

Table 7
Summary of outlier data points for the categorical features in the initial model compared to the entire dataset.

Parameter	Most common category	% of the most common category	Most common category		% of the most common category	
	(Dataset)		(Outlier groups)			
			20%	5%	20%	5%
Coarse Aggregate Type	2	***29 (26)***	***4***	***4***	***44***	***39***
Specimen Shape	1	52	1	1	67	61
Concrete Type	1	40	1	1	44	39
Cement Type	19	37	19	19	44	46
Chemical Admixture 1 Type	5	***43 (21)***	***2***	***2***	***44***	***36***
Mineral Admixture 1 Type	7	48	7	7	33	36
Chemical Admixture 2 Type	3	81	3	3	89	82
Chemical Admixture 3 Type	2	99	2	2	100	100
Mineral Admixture 2 Type	4	88	4	4	89	86
Fiber Type	0	99	0	0	100	100
Curing Method (Exposure)	12	***50 (15)***	***2***	***12***	***33***	***39***
UPV Configuration	0	97	0	0	100	100
Transducer Contact Material	1	94	1	1	89	96

Analyzing the distributions of categorical features in the entire and outlier datasets reveals that the chemical admixture type and curing method (exposure) both are important sources of error in the initial model. Among the categorical features, the most common categories differ from the original dataset for three features: coarse aggregate type, chemical admixture type, and curing method (exposure). As can be seen from Table 7 for the coarse aggregate type parameter, the percent of the most common category in the total dataset, i.e., category 2, is %29 for the entire dataset. The most common category in the outlier dataset for the same parameter is category 4, which makes up 26% (denoted in parenthesis) of the whole dataset. Since both categories have almost equal samples in the dataset, coarse aggregate type is not a dominant parameter for overestimated data points. However, we cannot repeat the same statement for the curing method (exposure) and chemical admixture type features. Both features have dominant categories in the outlier data points different from the dominant categories of the entire dataset. 43% of the chemical admixture type parameter is set to category 5 in the whole dataset, whereas only 21% is set to category 2. In both outlier datasets, category 2 is the majority for the same parameter. Likewise, the majority (50%) of the curing method (exposure) parameter is category 12 in the total dataset. Whereas prediction errors larger than %20 had the most common category as category 2 (33% of the outlier dataset).

5. Model training, concrete strength prediction results, and discussion

5.1. Hyperparameter optimization

We used three major meta-heuristic algorithms with the initial ANN

model to perform hyperparameter tuning. In this hybrid approach, we created an initial population using hyperparameter values as the coordinates of the individuals in the population. We assigned each individual's fitness value as RMSE resulting from ANN run with the hyperparameters assigned to that individual. We used a total of five different hyper-parameters in the optimization process. Hyper-parameters included in the solution space were the optimization algorithm, learning rate, number of hidden layers, nodes at each hidden layer, and activation function.

We applied the most commonly used meta-heuristic algorithms to the ANN model that trains with training data. We ranked each individual at each step of the meta-heuristic algorithm using the RMSE values obtained from the validation data. As the number of steps increases for a given epoch number of ANN, we observed lower RMSE values, indicating that the combination of the current hyper-parameters outperformed the previous ones. The ranges used for hyper-parameters are listed in Table 8.

We performed a series of initial trials with other meta-heuristic algorithms, both swarm-based and evolutionary. We observed that three of these algorithms have higher efficiency compared to the rest, namely, the genetic algorithm (GA), firefly algorithm (FA), and particle swarm optimization (PSO). Accordingly, we based the actual comparison of algorithms on these three algorithms for the current problem. Table 9 shows the minimum validation RMSEs observed for each meta-heuristic algorithm and corresponding hyper-parameter values for all three selected algorithms.

As seen from the tables, the best-performing optimization algorithms and activation functions were similar for all algorithms. However, the best-performing algorithm appeared to be the genetic algorithm, which performs the best in exploring the search space. As a result, we selected the ANN-GA hybrid for hyperparameter optimization.

5.2. Dimensionality reduction

Dimension reduction is commonly performed to eliminate the redundant features so the model's computation efficiency will not be compromised. A reduced number of features with fewer information redundancies is also critical for future practical applications of the model. We used the same three meta-heuristic algorithms for hyper-parameter optimization to perform dimension reduction. The initial number of features was 37, but it is necessary to decrease this number as much as possible for the model's computational efficiency. We utilized all three approaches with top-down and bottom-up procedures to achieve a more efficient model with minimum features. In the top-down approach, we dropped several features from individuals in the population. In the bottom-up approach, we added several features to individuals in the population. We rerun the model with the new features chosen for each individual. At the end of each run, we obtained RMSE values to be used as the fitness value of that individual in the population. We created a population of different models with different features at each meta-heuristic step.

As mentioned above, we performed a random forest analysis to detect essential core features for model performance. We based our dimensionality reduction methods on core features that have higher than 0.05. Consequently, all models include UPV Reading, Age, and Water/Binder Ratio features. In the bottom-up approach, the initial model starts with only core features; however, we add a random number of features to the model and calculate losses. Similarly, we did not allow the core features to be dropped in the top-down approach. When both top-down and bottom-up approaches were coded and run, we observed that the bottom-up approach yielded higher errors for the same number of steps.

After a trial run, we decided to use the top-down approach in the hybrid model because the top-down hybrid yielded smaller errors, and the efficiency of these two approaches was not significantly different. Tables 10 and 11 show the best performing models for each meta-heuristic hybrid; we ran the models using the hyperparameters determined in the hyperparameter tuning section. We listed the features selected by the optimization process for each meta-heuristic method. These tables show that the common features exist for all three hybrid models. Consequently, we decided these features (denoted in italics) must be in the final model. The convergence of the optimization algorithm is similar to the one in the hyperparameter optimization step. RMSE values were reduced to a minimum after the 12th step and stayed constant afterward. Likewise, R^2 values also reached a maximum at the same step.

5.3. Simultaneous optimization

We optimize the ANN model for the hyper-parameters and dimension reduction simultaneously as the final part of the study. The best meta-heuristic method for the hybrid model was GA for both steps. Accordingly, we only implemented the ANN-GA hybrid algorithm in this

Table 8
Range for values of hyper-parameters in the search space.

Optimization algorithms	Adam, LFBGS, Stochastic Gradient Descent,
Learning Rate	0.001-0.1
Hidden Layers	2-6
Nodes at Each Hidden Layer	5-100 (increments of 5) nodes at each row
Activation Function for Each Hidden Layer	Tan h, Relu, Sigmoid

Table 9
The best performing models for the optimization algorithms studied and the corresponding hyperparameter values.

Best Model Performance	Genetic Algorithm	Firefly Algorithm	Particle Swarm Optimization
Run Time-Total (sec, Avg for one step)	57	38	95
Training Error (RMSE-MPa)	5.1	5.9	4.9
Validation Error (RMSE-MPa)	5.3	6.1	5.2
Training Coefficient of Determination (R^2)	0.94	0.90	0.92
Validation Coefficient of Determination (R^2)	0.92	0.88	0.91
Corresponding Hyperparameters			
Optimization algorithm	Adam	Adam	Adam
Learning Rate	0.007	0.005	0.007
Hidden Layers	3	3	4
Nodes at Each Hidden Layer	85	90	80
Activation Function for Each Hidden Layer	Tan h	Tan h	Tan h

Table 10
The best performing model for metaheuristic algorithms.

Best Model Performance	Genetic Algorithm	Firefly Algorithm	Particle Swarm Optimization Algorithm
Run Time (Avg for one step-sec)	370	590	430
Training Error (RMSE-MPa)	3.2	4.6	4.2
Validation Error (RMSE-MPa)	3.84	5.1	4.5
Training Coefficient of Determination (R^2)	0.95	0.92	0.91
Validation Coefficient of Determination (R^2)	0.93	0.89	0.88

Table 11
Corresponding features that generate the best performance of each meta-heuristic algorithm.

Genetic Algorithm	*Specimen Age*	Chemical Admixture Count
	Water/Binder Ratio	Mineral Admixture Count
	Chem. Admixture Content	Mineral Admixture Content
	Max. Aggregate Size	Air Entrainment
	Specimen Dimension	Trajectory Distance
	CA/FA Ratio	Frequency
	Cement/Aggregate Ratio	Coarse Aggregate Type
	Water/Binder Ratio	Specimen Shape
		Curing Method (Exposure)
Firefly Algorithm	*Specimen Age*	Specimen Dimension
	Water/Binder Ratio	Fiber Content by Volume
	Chem. Admixture Content	CA/FA Ratio
	Max. Aggregate Size	Trajectory Distance
Particle Swarm Optimization Algorithm	*Specimen Age*	Mineral Adm. Si Content
	Water/Binder Ratio	Air Entrainment
	Chem. Admixture Content	Chemical Admixture Type
	Max. Aggregate Size	Mineral Admixture Type
		Mineral Admixture Content

last step. We utilized a top-down approach and prevented the algorithm from removing critical features determined in the dimension reduction section, i.e., UPV, Age, water binder ratio, maximum aggregate size, and chemical admixture content. According to the hyperparameter optimization part, the activation function and optimization algorithm were tanh and Adam for all best performing models, respectively.

Additionally, we set the number of layers to 3 because best performing models predominantly had 3-layer architecture. Also, 3-layer architecture yields the highest accuracy with GA. In this section, we only identified the number of nodes at each layer and the learning rate as hyperparameters. Each individual in the population has a number of parameters to be added, the name of features included, and two hyperparameter values as coordinates.

We also increased the range used for the number of nodes at each layer. The original range was 5-100 nodes per layer in the hyper-parameter optimization part of the study, which we increased to 100-1000 in the current step.

The final model yielded both a reduced parameter list and optimized hyper-parameters. As shown in Table 12, the validation errors in the current model are smaller than those in hyper-parameter optimization and dimensionality reduction. As expected, the total run time for one step is higher than the previous models.

Finally, after saving the best-performing model for the simultaneous optimization, we ran this model and output models from previous optimization steps using the test data. Table 13 shows the results from the test data. The simultaneous optimization model for training and validation data produced the highest coefficient of determination and lowest errors. The simultaneous optimization model had the smallest test errors (RMSE: 4.8 MPa). Run times averaged over multiple runs, and we observed that they are similar for all three models.

We also analyzed the outliers of the final model from simultaneous optimization and compared these with the outliers of the initial model. Similar to Tables 5–7, we prepared Tables 14–16 by isolating data points where compressive strength was predicted to be 5% and 20% higher (or lower) than the actual values. The most common values of both outlier groups are the same as those of the entire database for all discrete features. The initial model without metaheuristic optimization had a mineral admixture count parameter with a different most common value for the outliers. This fact indicates that the imbalance in the initial model specified at the end of chapter 4 is resolved in this final model. The final model successfully maps the relationship between the mineral admixture count and compressive strength. Likewise, the most common values are the same for both outlier datasets with the total dataset for all categorical features. The initial model's sensitivity to chemical admixture type and curing method (exposure) is also eliminated in the final model. When the continuous features are analyzed, most average parameter values in both outlier groups are closer to that of the entire dataset. Specifically for the mineral admixture content parameter, the average value for the extreme outlier group reduced from 111 kg/m3 to 42 kg/m3. Considering the average of this parameter for the entire dataset is 48 kg/m3, the sensitivity of the initial model to mineral admixture content was eliminated in the final model.

The above analysis of the outlier data proves that the model loss was reduced successfully through metaheuristic algorithms that optimize the hyperparameters of the ANN. However, another critical point for the estimation is the overestimation/underestimation of the compressive strength. Underestimation of the strength does not pose a risk to the safety of any structure; it might only be a consideration of project cost. Especially considering post-disaster evaluation or disaster preparedness, the cost disadvantage can be ignored, given the underestimation range is no higher than 10-15%. However, overestimation can have serious safety consequences, so the model must be checked for biases that might result in many overestimated strength values.

Table 12
The best performing model for simultaneous optimization.

Run Time-Total (sec)	720 (Avg for one step)
Training Error (RMSE-MPa)	2.3
Validation Error (RMSE-MPa)	3.6
Training Coefficient of Determination (R^2)	0.98
Validation Coefficient of Determination (R^2)	0.96
Learning Rate	0.03
Nodes at Each Hidden Layer	400,200,60

Corresponding Features:	
Specimen Age	Mineral Admixture Type
Water/Binder Ratio	Trajectory Distance
Mineral Admixture Silica Content	Frequency
Air Entrainment	Max. Aggregate Size
Mineral Admixture Content	Specimen Shape
Chemical Admixture Content	Curing Method (Exposure)

Table 13
Test runs for all three final models.

The best performing model from hyper-parameter optimization - test data performance	
Run Time-Total (Avg)	35 s (one step)
Test Error (RMSE-MPa)	5.2
Test Coefficient of Determination (R^2)	0.92
Mean Absolute Percentage Error (MAPE)	11.2%
The best performing model from dimension reduction - test data performance	
Run Time-Total (Avg)	40 s (one step)
Test Error (RMSE-MPa)	4.9
Test Coefficient of Determination (R^2)	0.93
Mean Absolute Percentage Error (MAPE)	10.9%
The best performing model from simultaneous optimization - test data performance	
Run Time-Total (Avg)	51 s (one step)
Test Error (RMSE-MPa)	4.8
Test Coefficient of Determination (R^2)	0.93
Mean Absolute Percentage Error (MAPE)	10.7%

Table 14
Summary of outlier data points for the discrete features in the final model compared to the entire dataset.

Parameter	Most common value	% of most common value	Most common value		% of most common value	
	(Dataset)		(Outlier groups)			
			20%	5%	20%	5%
Age (Days)	28	40	28	28	40	38
Specimen Dimension 1 (mm)	100	61	100	100	62	60
Specimen Dimension 2 (mm)	200	39	200	200	45	40
Maximum Aggregate Size (mm)	13	18	13	13	21	18
Chemical Admixture Count	1	51	1	1	52	54
Mineral Admixture Count	0	47	0	0	55	46
Fiber Aspect	0	99	0	0	100	99
Curing Temp (C)	23	71	23	23	67	73

Table 15
Summary of outlier data points for the continuous features in the final model compared to the entire dataset.

Parameter	Average value (Dataset)	Average value (Outlier groups)	
		20%	5%
UPV Reading (Km/s)	4.38	4.54	4.48
Compressive Strength (MPa)	47.56	22.02	32.16
Coarse/Fine Aggregate	1.29	1.26	1.34
Cement/Total Aggregate	0.32	0.32	0.37
Water/Binder	0.42	0.45	0.41
Chemical Admixture 1 Content (kg/m^3)	1.53	1.74	1.84
Mineral Admixture 1 Content (kg/m^3)	48.68	41.97	50.89
Chemical Admixture 2 Content (kg/m^3)	0.36	0.11	0.21
Mineral Admixture 2 Content (kg/m^3)	8.63	6.35	7.75
Mineral Admixture 1 Calcium Content (%)	8.52	9.07	16.95
Mineral Admixture 1 Silica Content (%)	1.47	1.29	1.35
Air Entrainment %	0.26	0.02	0.24
Fiber Content (% Vol)	0.01	0	0

Strength reduction factors in **ACI 318-19** (Building Code Requirements for Structural Concrete) vary from 0.9 to 0.6 for different concrete members, accounting for uncertainties in materials, possible design, and construction errors. Accordingly, an overestimation of around 5% can be acceptable, considering there should still be room for construction and design errors. We created overestimated and underestimated groups with strength values that are larger than 105% and 115% of the actual strength. When overestimated and underestimated groups were compared, it was observed that the data was balanced. Both overestimated and underestimated strength predictions were around 11% of the whole test set when the prediction difference was 5% or more. This means 93% of the strength predictions were safe to use in any structural analysis. As can be seen from Fig. 6, underestimated strength values are mainly accumulated between 20 and 50 MPa, whereas overestimated values lie primarily under the 30 MPa region.

Consequently, compressive strength is more likely to be overestimated for low-strength concrete. Especially large overestimations are more likely if the concrete strength is lower than 20 MPa. The reason in this overestimation can be attributed to the under representation of low strength concrete in the dataset, only 6% of the dataset contains concrete with strengths lower than 20 MPa. Considering that concrete strength lower than 20 MPa is not used as structural concrete, the risk of overestimation is less than 7% for most existing structures. Consequently, the model can be used when reliable UPV data is available, and it has the potential to estimate the strength of structural concrete with both relatively low error and low risk of overestimation.

Table 16
Summary of outlier data points for the categorical features in the final model compared to the entire dataset.

Parameter	Most common value	% of most common value	Most common value		% of most common value	
	(Dataset)		(Outlier groups)			
			20%	5%	20%	5%
Coarse Aggregate Type	2	29	2	2	24	25
Specimen Shape	1	52	1	1	57	50
Concrete Type	1	40	1	1	33	43
Cement Type	19	37	19	19	33	37
Chemical Admixture 1 Type	5	43	5	5	48	45
Mineral Admixture 1 Type	7	48	7	7	36	44
Chemical Admixture 2 Type	3	81	3	3	83	85
Mineral Admixture 2 Type	4	88	4	4	90	90
Fiber Type	0	99	0	0	100	98
Curing Method (Exposure)	12	50	12	12	57	59
UPV Configuration	0	97	0	0	100	96
Transducer Contact Material	1	94	1	1	93	91

6. Discussion and practical implications

The model obtained from simultaneous optimization can potentially be used as part of a material assessment method for existing structures. However, when the input features in this final model are considered, we observed that some of these features in the final model might be harder to access in a field operation, especially for older structures. Consequently, as an additional consideration, the final features listed in Table 12 were evaluated regarding accessibility for existing structures in particular application scenarios. Features that are most likely to be accessible for older structures were concrete age, air entrainment percent, trajectory distance for UPV measurement, frequency of the UPV measurement, maximum aggregate size, specimen shape, curing method (exposure), along with the obvious UPV reading. Assuming air entrainment and maximum aggregate size of a concrete specimen can be assessed using either a secondary non-destructive method or a semi-destructive method, the rest of the features listed above will be accessible in the majority of structural evaluation. As a result, we trained, optimized, and evaluated another model using the aforementioned "field applicable" features as a final consideration. We performed the hyper-parameter optimization using GA once again on this model, and the results were assessed. We observed the training and validation RMSE to be 4.7 and 5.3 MPa for this model, respectively. Table 17 compares the test RMSE of field applicable models with and without optimization. When we compared test RMSE values, the model from simultaneous optimization had a 48% improvement over the initial ANN model without optimization. This model optimized with field-applicable

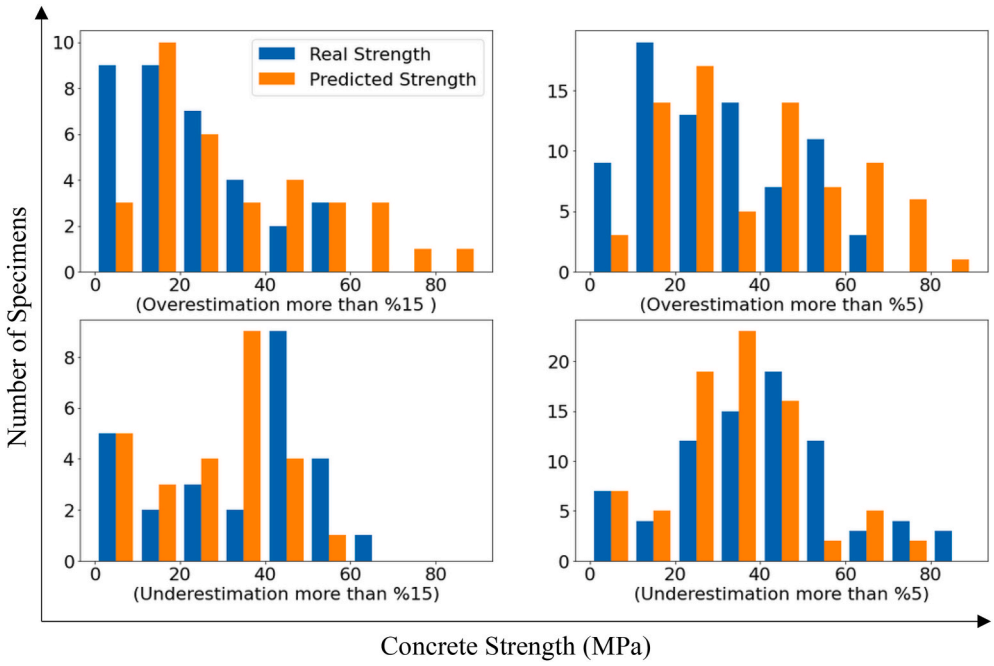

Fig. 6. Histograms of overestimated and underestimated strength predictions comparing the number of specimens for each concrete strength range - both predicted and actual.

Table 17
Comparing the initial model and the optimized field applicable model performances over the test data.

Initial ANN model without optimization	
Test Error (RMSE-MPa)	9.3
Test Coefficient of Determination (R^2)	0.71
Hybrid ANN with MetaHeuristic Optimization-Field Applicable Features	
Test Error (RMSE-MPa)	5.6
Test Coefficient of Determination (R^2)	0.89

features still has a 39% improvement over the ANN model without optimization.

When literature is scanned for compressive strength prediction based on mix design properties of concrete using neural networks, RMSE values obtained from those studies range anywhere between 0.018 and 9.71 MPa (Nunez et al., 2021). It is undeniable that these RMSE values are dependent on various parameters as we discussed earlier. Most of these studies either focus on specific concrete type (limited mix design variability) and/or dataset sizes are very small. However, the median of the reported RMSE values in this specific review paper was around 4-5 MPa. On another note, there are a limited number of studies on compressive strength prediction of concrete using neural networks optimized with metaheuristic algorithms. Two of these studies stand out with large dataset sizes. Bui et al. (2018) focuses on high performance concrete with 1133 rows of data. They reported an RMSE of 4.85 MPA with a Modified firefly optimization algorithm and neural network hybrid (Bui et al., 2018). Galofshani et al. (2020) reported an RMSE of 4.69 MPa again focusing on high performance concrete with 2817 rows of data (Golafshani et al., 2020).

In both of these studies inputs are only mix design parameters, however they did not use any additional NDT data such as UPV measurements. Consequently, we can roughly state that 4 -5 MPa of RMSE for concrete strength prediction is optimal when neural networks are utilized. This conclusion is somewhat more accurate when a specific type of concrete strength is targeted. We overcome the increased errors due to various concrete types and high mix design variability by utilizing UPV measurements and optimization algorithms. Thus our hybrid model can predict strength with the same range of error reported in the literature when predicting low mix design variation and/or a specific concrete type.

The final output of the study is two ANN/meta-heuristic models that were trained and optimized for concrete strength prediction. The first model has higher accuracy for strength prediction and can be used for existing structures where mix design information is available. This model can also be used for construction planning, such as calculating the cost/benefit ratio before mix design, selecting a curing method, etc. The second model has relatively lower accuracy; however, it can be applied to existing structures where there is very little information available on the properties of concrete.

7. Conclusion and future research recommendations

Current evaluation methods can be time-consuming and costly. Relatively high accuracy in the models shows that advanced techniques relying on machine learning can easily replace current practices. As the state of the literature on concrete strength prediction based on UPV is considered, this study is the first of its kind where relatively high accuracy is obtained from a dataset of significant variability. Metaheuristic optimization was effective for increasing accuracy. We observed a significant improvement between 39% and 48% in test RMSE values when we used metaheuristics to optimize artificial neural networks when predicting the compressive strength of concrete.

As mentioned in the discussion, we considered two models as the final output. The first one is the direct output of simultaneous optimization, which yields the lowest prediction errors (RMSE 4.8 MPa, R^2 0.93, MAPE 10.7%). The second model only uses features that might be easier to access in a field operation. RMSE of this model increased to 5.6 MPa (R^2 0.89, MAPE 11.7%). We selected these features as; specimen age, air entrainment percent, trajectory distance for UPV measurement, frequency of the UPV measurement, maximum aggregate size, specimen shape, curing method (exposure), and UPV reading.

Considering the variety of cement types, aggregate types, mixing and curing practices, etc., and the level of noise created in the data by these variables, it might be safe to assume the models provide the lowest possible prediction error. Both models perform within acceptable limits

for a wide range of concrete types, proving their potential convenience in practical applications.

The model with field-applicable features has the potential to be used as part of an alternative structural evaluation scheme for existing buildings. An alternative evaluation procedure that integrates the current model can eliminate destructive testing for concrete strength evaluation and be useful during disaster response operations.

Additionally, a two-stage procedure for structural evaluation can be formulated using the models developed in this study. The first stage would cover determining the mix parameters as inputs of the model. A second non-destructive or a simple semi-destructive test can be used if the features are not readily available. The second stage would predict concrete strength using the outputs of the first stage using the ANN model. Additionally, if a second NDT method is not possible in the first stage, the second ANN model with 5.6 MPa RMSE would be used for strength prediction. Such a modified procedure for strength assessment can increase accuracy and decrease procedural fieldwork simultaneously.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Seda Selcuk reports financial support was provided by The Scientific and Technological Research Council of Turkey (TUBITAK).

Data availability

Data will be made available on request.

Acknowledgment

This work was supported through The Scientific and Technological Research Council of Turkey 2219-International Postdoctoral Research Fellowship Program for Turkish Citizens [grant number 2018/2, 1059B191802214].

References

Al-Akhras, N.M., 1995. Characterization and deterioration detection of Portland cement concrete using ultrasonic waves. Virginia Tech 204. M.Sci. Thesis.

Albuthbahak, O.M., Hiswa, A.M.R., 2019. Prediction of concrete compressive strength using supervised machine learning models through ultrasonic pulse velocity and mix parameters. Romanian Journal of Materials 49 (2), 232–243.

Aqel, M.A., 2016. Steam Cured Self-Consolidating Concrete and the Effects of Limestone Filler. University of Toronto, M.Sci. Thesis, p. 241.

Asteris, P.G., Kolovos, K.G., 2017. Data on the physical and mechanical properties of soilcrete materials modified with metakaolin. Data Brief 13, 487–497. https://doi.org/10.1016/j.dib.2017.06.014.

Asteris, P.G., Roussis, P.C., Douvika, M.G., 2017. Feed-forward neural network prediction of the mechanical properties of sandcrete materials. Sensors 17 (6). https://doi.org/10.3390/s17061344.

Atahan, H.N., Oktar, O.N., Tademir, M.A., 2011. Factors determining the correlations between high strength concrete properties. Construct. Build. Mater. 25 (5), 2214–2222. https://doi.org/10.1016/j.conbuildmat.2010.11.005.

Atici, U., 2011. Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network. Expert Syst. Appl. 38 (8), 9609–9618. https://doi.org/10.1016/j.eswa.2011.01.156.

Becker, A.N., 2000. Investigation of Pulse-Velocity and Maturity of Portland Cement Concrete, 229. University of Louisville, M.Sci. Thesis.

Bilgehan, M., 2011. A comparative study for the concrete compressive strength estimation using neural network and neuro-fuzzy modelling approaches. Nondestr. Test. Eval. 26 (1), 35–55. https://doi.org/10.1080/10589751003770100.

Biswas, R., Samui, P., Rai, B., 2019. Determination of compressive strength using relevance vector machine and emotional neural network. Asian Journal of Civil Engineering 20, 1109–1118. https://doi.org/10.1007/s42107-019-00171-9.

Breysse, D., 2012. Non-destructive evaluation of concrete strength: an historical review and a new perspective by combining NDT methods. Construct. Build. Mater. 33, 139–163. https://doi.org/10.1016/j.conbuildmat.2011.12.103.

Bui, D.K., Nguyen, T., Chou, J.S., Nguyen-Xuan, H., Ngo, T.D., 2018. A modified firefly algorithm-artificial neural network expert system for predicting compressive and tensile strength of high-performance concrete. Construct. Build. Mater. 180, 320–333. https://doi.org/10.1016/j.conbuildmat.2018.05.201.

Chao-Lung, H., le Anh-Tuan, B., Chun-Tsun, C., 2011a. Effect of rice husk ash on the strength and durability characteristics of concrete. Construct. Build. Mater. 25 (9), 3768–3772. https://doi.org/10.1016/j.conbuildmat.2011.04.009.

Chao-Lung, H., le Anh-Tuan, B., Chun-Tsun, C., 2011b. Effect of rice husk ash on the strength and durability characteristics of concrete. Construct. Build. Mater. 25 (9), 3768–3772. https://doi.org/10.1016/j.conbuildmat.2011.04.009.

Chu, P.Y., 2012. Carbon steel slag as cementitious material for self-consolidating concrete. Material Recycling - Trends and Perspectives 323–334 ([Online]).

Demirboğa, R., Türkmen, I., Karakoç, M.B., 2004. Relationship between ultrasonic velocity and compressive strength for high-volume mineral-admixtured concrete. Cement Concr. Res. 34 (12), 2329–2336. https://doi.org/10.1016/j.cemconres.2004.04.017.

Dervisoglu, O., 2002. Use of Attenuation to Monitor Load-Induced Damage in Plain & Fiber Reinforced Mortars, 121. University of Windsor, M.Sci. Thesis.

Dolatabadi, M.H., 2013. Properties and Performance of Photocatalytic Concrete, 121. University of Toronto, M.Sci. Thesis.

Duan, Z.H., Poon, C.S., 2014. Properties of recycled aggregate concrete made with recycled aggregates with different amounts of old adhered mortars. Mater. Des. 58, 19–29.

Fan, C.C., Huang, R., Hwang, H., Chao, S.J., 2016. Properties of concrete incorporating fine recycled aggregates from crushed concrete wastes. Construct. Build. Mater. 112, 708–715. https://doi.org/10.1016/j.conbuildmat.2016.02.154.

Farahani, J.N., Shafigh, P., Alsubari, B., Shahnazar, S., bin Mahmud, H., 2017. Engineering properties of lightweight aggregate concrete containing binary and ternary blended cement. J. Clean. Prod. 149, 976–988. https://doi.org/10.1016/j.jclepro.2017.02.077.

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. Construct. Build. Mater. 232, 117266 https://doi.org/10.1016/j.conbuildmat.2019.117266.

Goutham Sai, J., Singh, V.P., 2019. Prediction of compressive strength using support vector regression. Mendeliana 25 (1), 51–56. https://doi.org/10.13164/mendel.2019.1.051.

Güneyisi, E., Gesoğlu, M., Özbay, E., 2009. Effects of marble powder and slag on the properties of self compacting mortars. Materials and Structures/Materiaux et Constructions 42 (6), 813–826. https://doi.org/10.1617/s11527-008-9426-2.

Hamood, A., 2014. Sustainable Utilisation of Raw Sewage Sludge (RSS) as A Water Replacement in Cement-Based Materials Containing Unprocessed Fly Ash. MSc thesis. University of Wolverhampton, p. 284.

Ikpong, A.A., 1993. The relationship between the strength and non-destructive parameters of rice husk ash concrete. Cement Concr. Res. 23 (2), 387–398. https://doi.org/10.1016/0008-8846(93)90104-H.

Irrigaray, M.A.P., Pinto, R.C. de A., Padaratz, I.J., 2016. A new approach to estimate compressive strength of concrete by the UPV method. Revista IBRACON de Estruturas e Materiais 9 (3), 395–402. https://doi.org/10.1590/s1983-41952016000300004.

Jain, A., Kathuria, A., Kumar, A., Verma, Y., Murari, K., 2013. Combined use of non-destructive tests for assessment of strength of concrete in structure. Procedia Eng. 54, 241–251. https://doi.org/10.1016/j.proeng.2013.03.022.

Karagol, F., Demirboga, R., Khushefati, W.H., 2015. Behavior of fresh and hardened concretes with antifreeze admixtures in deep-freeze low temperatures and exterior winter conditions. Construct. Build. Mater. 76, 388–395. https://doi.org/10.1016/j.conbuildmat.2014.12.011.

Kewalramani, M.A., Gupta, R., 2006. Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks. Autom. ConStruct. 15 (3), 374–379. https://doi.org/10.1016/j.autcon.2005.07.003.

Khan, S.R.M., Noorzaei, J., Kadir, M.R.A., Waleed, A.M.T., Jaafar, M.S., 2007. UPV method for strength detection of high performance concrete. Struct. Surv. 25 (1), 61–73. https://doi.org/10.1108/02630800710740985.

Koh, C.J., 2014. Characterisation of Shape of Recycled Crushed Coloured Glass and the Effect on the Properties of Structural Concrete when Used as a Fine Aggregate Replacement. May, pp. 154–218.

Lai, C.-P., Lin, Y., Yen, T., 2001. Behavior and estimation of ultrasonic pulse velocity in concrete. Structural Engineering, Mechanics and Computation 2, 1365–1372. https://doi.org/10.1016/b978-008043948-8/50152-1.

Langevin, M., 1993. Rheological and Mechanical Behavior of Microfine Cement-Based Grouts, pp. 1–273. November.

Latif, R.A.W.M., 1995. Early Age Ultrasonic Pulse Monitoring of Cement Paste. mortar and concrete, p. 287.

Lin, Y.C., Lin, Y., Chan, C.C., 2016. Use of ultrasonic pulse velocity to estimate strength of concrete at various ages. Mag. Concr. Res. 68 (14), 739–749. https://doi.org/10.1680/jmacr.15.00025.

Ly, H.-B., Nguyen, M.H., Pham, B.T., 2021. Metaheuristic optimization of Levenberg–Marquardt-based artificial neural network using particle swarm optimization for prediction of foamed concrete compressive strength. Neural Comput. Appl. 33 (24), 17331–17351.

Nacer, S., 2005. "Optimization of silica fume content and water to enhance performance of concrete ". Southern Illinois University at Carbondale 85. M.Sci. Thesis.

Naderpour, H., Rafiean, A.H., Fakharian, P., 2018. Compressive strength prediction of environmentally friendly concrete using artificial neural networks. J. Build. Eng. 16, 213–219.

Nunez, I., Marani, A., Flah, M., Nehdi, M.L., 2021. Estimating compressive strength of modern concrete mixtures using computational intelligence: a systematic review. Construct. Build. Mater. 310, 125279 https://doi.org/10.1016/j.conbuildmat.2021.125279.

Oeswein, J.V., 2000. Durability Assessment of Portland Cement Concrete Using Accelerated Freeze-Thaw Testing, 138. University of Louisville, M.Eng. Thesis.
Ojha, V.K., Abraham, A., Snášel, V., 2017. Metaheuristic design of feedforward neural networks: a review of two decades of research. Eng. Appl. Artif. Intell. 60 (January), 97–116. https://doi.org/10.1016/j.engappai.2017.01.013.
Owaid, H.M., Hamid, R.B., Taha, M.R., 2017. Variation of ultrasonic pulse velocity of multiple-blended binders concretes incorporating thermally activated alum sludge ash. KSCE J. Civ. Eng. 21 (4), 1235–1246. https://doi.org/10.1007/s12205-016-1382-8.
Ozbay, E., Gesoglu, M., Guneyisi, E., 2011. Transport properties based multi-objective mix proportioning optimization of high performance concretes. Materials and Structures/Materiaux et Constructions 44 (1), 139–154. https://doi.org/10.1617/s11527-010-9615-7.
Paixão, R.C.F., Penido, R.E.K., Cury, A.C., Mendes, J.C., 2022. "Comparison of machine learning techniques to predict the compressive strength of concrete and considerations on model generalization". Rev. IBRACON Estrut. Mater. 15 (5), e15503 https://doi.org/10.1590/S1983-41952022000500003.
Pal, P., 2019. Dynamic Poisson's ratio and modulus of elasticity of pozzolana Portland cement concrete. International Journal of Engineering and Technology Innovation 9 (2), 131–144.
Panesar, D.K., Shindman, B., 2011. Elastic properties of self consolidating concrete. Construct. Build. Mater. 25 (8), 3334–3344. https://doi.org/10.1016/j.conbuildmat.2011.03.024.
Popovics, S., 1993. Portland cement-fly ash-silica fume systems in concrete. Adv. Cement Base Mater. 1 (2), 83–91. https://doi.org/10.1016/1065-7355(93)90013-E.
Ranjbar, A., Barahmand, N., Ghanbari, A., 2020. Hybrid artificial intelligence model development for roller-compacted concrete compressive strength estimation International Journal of Engineering. Transactions: Basics 33 (10), 1852–1863.
S Marikunte, S., Phelps, R.J., Moutairou, I.O., 2010. Optimization of water and metakaolin content to achieve high performance in concrete. Proceedings of Transportation Research Board 89th Annual Meeting 84. May.
Sadeghi Nik, A., Lotfi Omran, O., 2013. Estimation of compressive strength of self-compacted concrete with fibers consisting nano-SiO2 using ultrasonic pulse velocity. Construct. Build. Mater. 44, 654–662. https://doi.org/10.1016/j.conbuildmat.2013.03.082.
Sadowski, Ł., Piechówka-Mielnik, M., Widziszowski, T., Gardynik, A., Mackiewicz, S., 2019. Hybrid ultrasonic-neural prediction of the compressive strength of environmentally friendly concrete screeds with high volume of waste quartz mineral dust. J. Clean. Prod. 212, 727–740. https://doi.org/10.1016/j.jclepro.2018.12.059.
Şahmaran, M., Christianto, H.A., Yaman, I.Ö., 2006. The effect of chemical admixtures and mineral additives on the properties of self-compacting mortars. Cement Concr. Compos. 28 (5), 432–440. https://doi.org/10.1016/j.cemconcomp.2005.12.003.
Sbartaï, Z.M., Breysse, D., Larget, M., Balayssac, J.P., 2012. Combining NDT techniques for improved evaluation of concrete properties. Cement Concr. Compos. 34 (6), 725–733. https://doi.org/10.1016/j.cemconcomp.2012.03.005.
Selcuk, L., Gökce, H.S., Kayabali, K., Simsek, O., 2012. A non-destructive testing technique: nail penetration test. ACI Struct. J. 109 (2), 245–252. https://doi.org/10.14359/51683635.
Shishegaran, A., Varaee, H., Rabczuk, T., Shishegaran, G., 2021. High correlated variables creator machine: prediction of the compressive strength of concrete. Comput. Struct. 247.
Sua-Iam, G., Makul, N., 2013. Use of recycled alumina as fine aggregate replacement in self-compacting concrete. Construct. Build. Mater. 47, 701–710. https://doi.org/10.1016/j.conbuildmat.2013.05.065.
Tang, C.W., Lin, Y., Kuo, S.F., 2007. Investigation on correlation between pulse velocity and compressive strength of concrete using ANNs. Comput. Concr. 4 (6), 477–497. https://doi.org/10.12989/cac.2007.4.6.477.Z.
Tanyidizi, H., Coskun, A., 2008. Determination of the principal parameter of ultrasonic pulse velocity and compressive strength of lightweight concrete by using variance method. Russ. J. Nondestr. Test. 44 (9), 639–646. https://doi.org/10.1134/S1061830908090088.
Tenza-Abril, A.J., Villacampa, Y., Solak, A.M., Baeza-Brotons, F., 2018. Prediction and sensitivity analysis of compressive strength in segregated lightweight concrete based on artificial neural network using ultrasonic pulse velocity. Construct. Build. Mater. 189, 1173–1183. https://doi.org/10.1016/j.conbuildmat.2018.09.096.
Thirumurugan, S., Sivakumar, A., 2013. Synergistic interaction of polypropylene fibres in latex modified high strength concrete. Arch. Civ. Eng. 59 (3), 321–335. https://doi.org/10.2478/ace-2013-0018.
Trtnik, G., Kavčič, F., Turk, G., 2009. Prediction of concrete strength using ultrasonic pulse velocity and artificial neural networks. Ultrasonics 49 (1), 53–60. https://doi.org/10.1016/j.ultras.2008.05.001.
Türkmen, I., 2003. Influence of different curing conditions on the physical and mechanical properties of concretes with admixtures of silica fume and blast furnace slag. Mater. Lett. 57 (9), 4560–4569. https://doi.org/10.1016/S0167-577X(03)00362-8, 29.
Türkmen, I., Gavgali, M., Gül, R., 2003. Influence of mineral admixtures on the mechanical properties and corrosion of steel embedded in high strength concrete. Mater. Lett. 57 (13–14), 2037–2043. https://doi.org/10.1016/S0167-577X(02)01136-9.
Uysal, M., Yilmaz, K., Ipek, M., 2012. Properties and behavior of self-compacting concrete produced with GBFS and FA additives subjected to high temperatures. Construct. Build. Mater. 28 (1), 321–326. https://doi.org/10.1016/j.conbuildmat.2011.08.076.
Willard, K.C., 2001. Correlation of Pulse Velocity and Compressive Strength for Early Age PCC, p. 101.
Yaman, I.O., 2000. Finite element simulation of wave propagation in concrete for the evaluation of ultrasonic testing procedures. ProQuest Dissertations and Theses 147 ([Online]).
Yap, S.P., Alengaram, U.J., Jumaat, M.Z., 2016. The effect of aspect ratio and volume fraction on mechanical properties of steel fibre-reinforced oil palm shell concrete. J. Civ. Eng. Manag. 22 (2), 168–177. https://doi.org/10.3846/13923730.2014.897970.
Yaqub, M., Bailey, C.G., 2016. Non-destructive evaluation of residual compressive strength of post-heated reinforced concrete columns. Construct. Build. Mater. 120, 482–493. https://doi.org/10.1016/j.conbuildmat.2016.05.022.
Yeh, I.-C., 1998. Modeling of strength of high-performance concrete using artificial neural networks. Cement Concr. Res. 28 (12), 1797–1808.
Yeh, I.C., 2009. Optimization of concrete mix proportioning using a flattened simplex - centroid mixture design and neural networks. Eng. Comput. 25 (2), 179–190. https://doi.org/10.1007/s00366-008-0113-2.
Yew, M.K., bin Mahmud, H., Ang, B.C., Yew, M.C., 2014. Effects of oil palm shell coarse aggregate species on high strength lightweight concrete. Sci. World J. 2014. https://doi.org/10.1155/2014/387647.
Yusuf, I.T., Jimoh, Y.A., Salami, W.A., 2016. An appropriate relationship between flexural strength and compressive strength of palm kernel shell concrete. Alex. Eng. J. 55 (2), 1553–1562. https://doi.org/10.1016/j.aej.2016.04.008.
Zhang, Y., Aslani, F., 2021. Compressive strength prediction models of lightweight aggregate concretes using ultrasonic pulse velocity. Construct. Build. Mater. 292, 123419 https://doi.org/10.1016/j.conbuildmat.2021.123419.
Zhong, W., Yao, W., 2008. Influence of damage degree on self-healing of concrete. Construct. Build. Mater. 22 (6), 1137–1142. https://doi.org/10.1016/j.conbuildmat.2007.02.006.