**ORIGINAL PAPER**

# Incorporation of artificial neural network with principal component analysis and cross-validation technique to predict high-performance concrete compressive strength

**Mohammed Majeed Hameed[1] · Mohamed Khalid AlOmar[1] · Wajdi Jaber Baniya[2] · Mohammed Abdulhakim AlSaadi[3,4]**

## Abstract

Compressive strength is the most essential mechanical characterization for concrete due to its crucial role in stating the design standards. Therefore, early, and accurate evaluation of concrete compressive strength minimizes efforts, costs, and time. In this study, we investigate the ability of artificial neural network (ANN) incorporated with principal component analyses (PCA) and cross-validation (CV) techniques to forecast the high-performance concrete (HPC) compression strength. The obtained results from the ANN-CVPCA model showed a good agreement between predicted and actual values. The proposed model provides high accuracy prediction of HPC compressive strength. It also provided a higher correlation coefficient (0.96) and a lower value of mean absolute error (3.43mpa), root mean square error (4.64mpa) and normalized root mean square error (0.13). Moreover, a sensitivity analysis was carried out to identify the most influential parameters and the simulated results showed that the superplasticizer, blast furnace slag, and cement parameters respectively have great effects on the compressive strength of HPC. The performance of the ANN-CVPCA model compared with other models published in previous studies and achieved the desired superiority and more stable predictions due to the existence of PCA and CV which play a significant role in increasing the generalization ability as well as avoiding redundant data and reducing the uncertainty in modeling outcomes.

**Keywords** Neural network · Cross-validation · Principal component analyses · High-performance concrete · Compressive strength test

## Introduction

Concrete is the main component in the construction industry. However, ordinary concrete mixtures cannot fulfill the industrial demand for all structural projects, wherein, special combinations and performance is needed for some structural elements. Therefore, in some projects, the use of high-performance concrete (HPC) is enviable (Aïtcin, 2011). HPC excelled in many physical characteristics in comparison with the ordinary concrete throughout permeability, durability, bleeding, workability, and extreme compression strength.

Compressive strength is an extremely vital property of concrete. The adopted method of compressive strength test included soaking the sample in water for 28 days and a carefully controlled environment (Khademi, 2017). Numerous factors affect the compressive strength of concrete including compaction, age, additives, aggregate size, water-cement ratio, etc. The compressive strength assay and assessment is time-consuming especially for ongoing buildings and infrastructural projects (Khademi, 2016). The importance of conducting research targeting the development of useful toll to predict concrete essential properties comes from the

✉ Mohamed Khalid AlOmar
mohd.alomar@yahoo.com

[1] Department of Civil Engineering, Al-Maaref University College, Ramadi, Iraq

[2] Department of Civil Engineering, Cairo University, Cairo, Egypt

[3] National Chair of Materials Science and Metallurgy, University of Nizwa, Nizwa, Sultanate of Oman

[4] Nanotechnology and Catalysis Research Centre (NANOCAT), IPS Building, University of Malaya, 50603 Kuala Lumpur, Malaysia

continuous need to save energy and time which caused by conventional laboratory work. (Yu, 2018),

Recently, many fields of science are adopting artificial neural networks (ANN) (Cao, 2017; Fiyadh, 2017, 2018; Fiyadh, et al. 2017, 2018; Shahin et al. 2001; Shi, 2016), and concrete research is not an exception to this field. (Akbari & Jafari Deligani, 2020; Atici & U.J.E.S.w.a., 2011; Hola & Schabowicz, 2005; Lee, 2003; Li, 2019; Öztaş, 2006a; Reddy, 2018; Zhang, 2020). Several studies, conducted over the last two decades, have shown that the use of ANN approaches to estimate the physical and chemical properties of concrete has enhanced the quality control of this field (Ni & Wang, 2000). Machine learning approaches such as ANN have gain popularity in solving complex engineering problems and capture the nonlinear relations where the empirical and casual approach may not address these problems well (AlOmar et al. 2020; Kaveh et al. 2001, 2008; Rofooei et al. 2011). For the last two decades, the backpropagation algorithm is one of the most efficient algorithms used to train ANN and correct the weight and bias values (Iranmanesh & Kaveh, 1999; Kaveh & Dehkordi, 2002; Kaveh & Elmieh, 2001; Kaveh & Iranmanesh, 1998; Kaveh & Servati, 2001).

Many researchers used traditional ANN techniques for such purposes including, feed-forward neural network (Ni & Wang, 2000; Öztaş, 2006b; Słoński, 2010), the back-propagation (Naderpour et al. 2018), multiple linear regression, and fuzzy logic (Khademi, 2016, 2017). It is worth mentioning, that the prediction of compressive strength of concrete by conventional ANN is difficult and inaccurate due to the many parameters involved in building the model (Nikoo, et al. 2015). Unlike ordinary concrete, the compressive strength test of HPC and high-strength concrete (HSC) is more complicated due to many additive materials(Hameed, 2021). Thus, reliable forecasting of the HPC compressive strength is considered a major challenging mission facing researchers in the recent decade. The traditional method such as pure ANN models are no longer efficient in addressing concrete properties prediction. To overcome this issue, researchers start to use combined models and techniques (Nikoo, et al. 2015).

Eventually, in this study, the principal component analysis (PCA) method is adopted using experimental data containing concrete samples of different variables. The data set consists of more than one thousand HPC samples with a variety of ages; 1, 3, 7, 14, 28, 90, 270 and 365 days.

This research aims to constitute a robust predictive model that can forecast the compressive strength of concrete with fewer forecasted errors. This study suggests employing ANN coupled with principal components analyses (PCA) technique and cross-validation (CV) method to constitute a more reliable effective model abbreviated as (ANN-CVPCA). For further assessment, the proposed model performance was compared with several and novel models developed in previous studied to ensure the reliability of the proposed model. Besides, a sensitivity analysis procedure was applied to study the effect of each input parameter and to identify the one which has the greatest effect on the compressive strength test of HPC.

## Methodology

### The dataset

Modeling HPC is a difficult task due to the complexity of its components. In this study, dozens of HPC data collected from many sources by (Yeh (1998)), were used to inspect the reliability and accuracy of the model. The 1030 samples obtained contained several characteristics including blast furnace slag (kg/m$^3$), superplasticizer (kg/m$^3$), cement (kg/m$^3$), age (days), coarse aggregate (kg/m$^3$), water (kg/m$^3$), fine aggregate (kg/m$^3$), fly ash (kg/m$^3$), and concrete compressive strength (MPa). The details of all variables used in this study are presented in Table 1.$S_X, X_{min}, X_{max}, , C_{sx}, C_V, C_c$ denote the stander deviation, minimum, maximum, skewness, coefficient of variation, and coefficient of correlation with compressive strength of HPC.

**Table 1** Descriptive analysis of compressive strength of concrete variables

| Component | $X_{max}$ | $X_{min}$ | $X_{mean}$ | $S_x$ | $C_V$ | $C_{SX}$ | $C_C$ |
|---|---|---|---|---|---|---|---|
| Cement | 540 | 102 | 281.17 | 104.46 | 0.37 | 0.51 | 0.498 |
| Blast furnace slag | 359.4 | 0 | 73.90 | 86.24 | 1.17 | 0.80 | 0.135 |
| Fly Ash | 200.1 | 0 | 54.19 | 63.97 | 1.18 | 0.54 | −0.106 |
| Water | 247 | 121.75 | 181.57 | 21.35 | 0.12 | 0.07 | − 0.290 |
| Superplasticizer | 32.2 | 0 | 6.20 | 5.97 | 0.96 | 0.91 | 0.366 |
| Coarse Aggregate | 1145 | 801 | 972.92 | 77.72 | 0.08 | − 0.04 | 0.165 |
| Fine Aggregate | 992.60 | 594.00 | 773.58 | 80.14 | 0.10 | − 0.25 | −0.167 |
| Concrete compressive strength | 82.60 | 2.33 | 35.82 | 16.70 | 0.47 | 0.42 | 1 |

## Back propagation artificial neural network

In machine learning, the backpropagation (BP) algorithm is broadly used as a learning algorithm to train a single layer of ANN. During the training phase, the BP algorithm iteratively adjusts and updates the bias and connection weight values (Afan, 2015; Yaseen, 2016). Figure 1 shows the classical structure of ANN. The ANN training method can be carried out as follows, with the ANN initiating input parameters and the BP algorithm starting to compute the sum square error by calculating the gap between the actual and predicted ANN values. The BP algorithm minimizes the sum square error in the second stage by setting weights and bias values by propagating the expected error from the output layer back to the input layer. (Fiyadh, 2019a, b; Hameed, 2017; Heidari et al. 2017; Kewalramani & Gupta, 2006). Since the BP algorithm is known to be a gradient descent-based algorithm, the variance in weight values is directed to the gradient in the negative path, refer to Eq. (1).

$$\Delta W_k = \alpha \Delta W_{k-1} - \eta \frac{\partial E}{\partial W} \tag{1}$$

where W is the weight and bias values, $\alpha$ is the momentum parameter, $\Delta W_k$ and $\Delta W_{k-1}$ refer to the changes in the weight W at k, and k-1 stands for iteration, $\eta$ stands for the learning rate factor, and $E$ is the forecasted error.

## Model development

Principal Component Analysis (PCA) can be defined as a dimension reduction procedure usually used for converting the large input set variables to smaller ones that contain the most beneficial information from the whole large data set. PCA is a mathematical technique that converts high dimensional and correlated variables to smaller dimensional and uncorrelated (orthogonal) variables known as principal components (PCs) which are assigned as a linear combination's procedure of the original parameters. The eigenvalues of the PCs measured the associated variance while the sum of eigenvalues corresponded with the total number of parameters in the data set. The first PCs considered the most important ones since it contains most of the variability in the original data set and the following PCs considered for another possible residual variability. PCA decreases the characteristic space from a higher number of parameters to a smaller number of variables since it is considered a non-dependent method which means. Besides, PCA does not assume a dependent variable as defined. PCA is considered as data compression or dimensionality reducing method where the main goal for using this technique is to ensure the presented data can be effectively interpretable. In this study, the input variables are transformed into uncorrelated variables and compressed before introducing them to the ANN model.

One of the main steps in developing an advanced and accurate model is data partition. This reality arises from the complex phenomena of compressive strength of HPC, and it is therefore mostly inaccurate to construct a single predictive

**Fig. 1** Classical structure of ANN
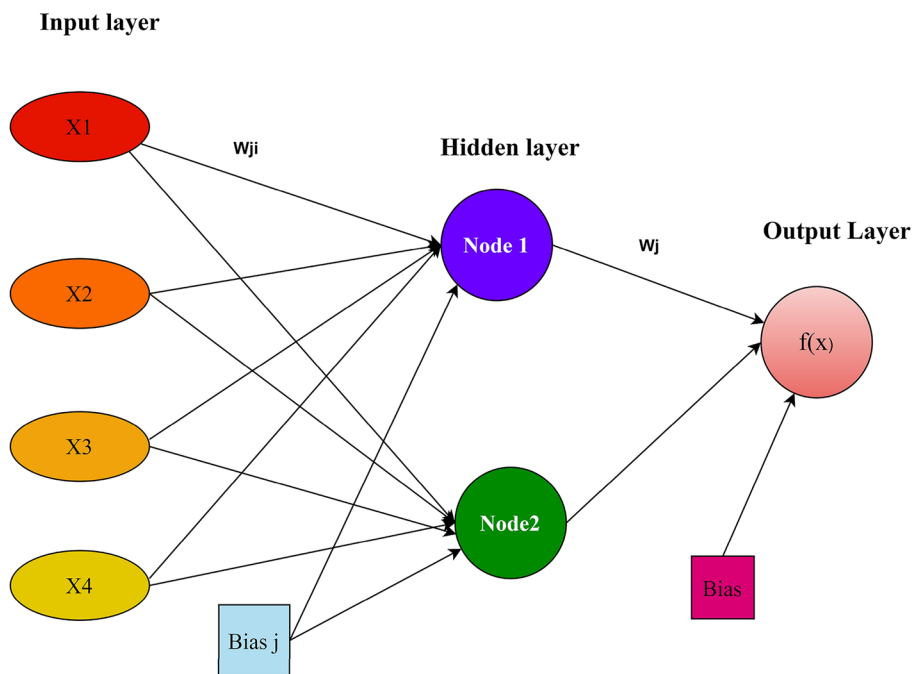
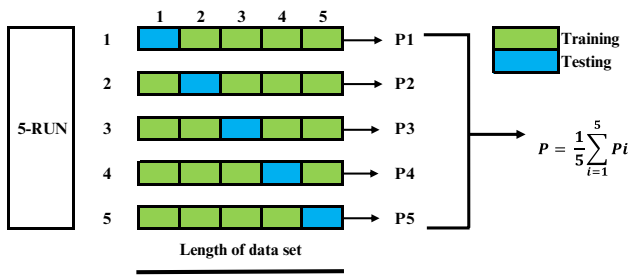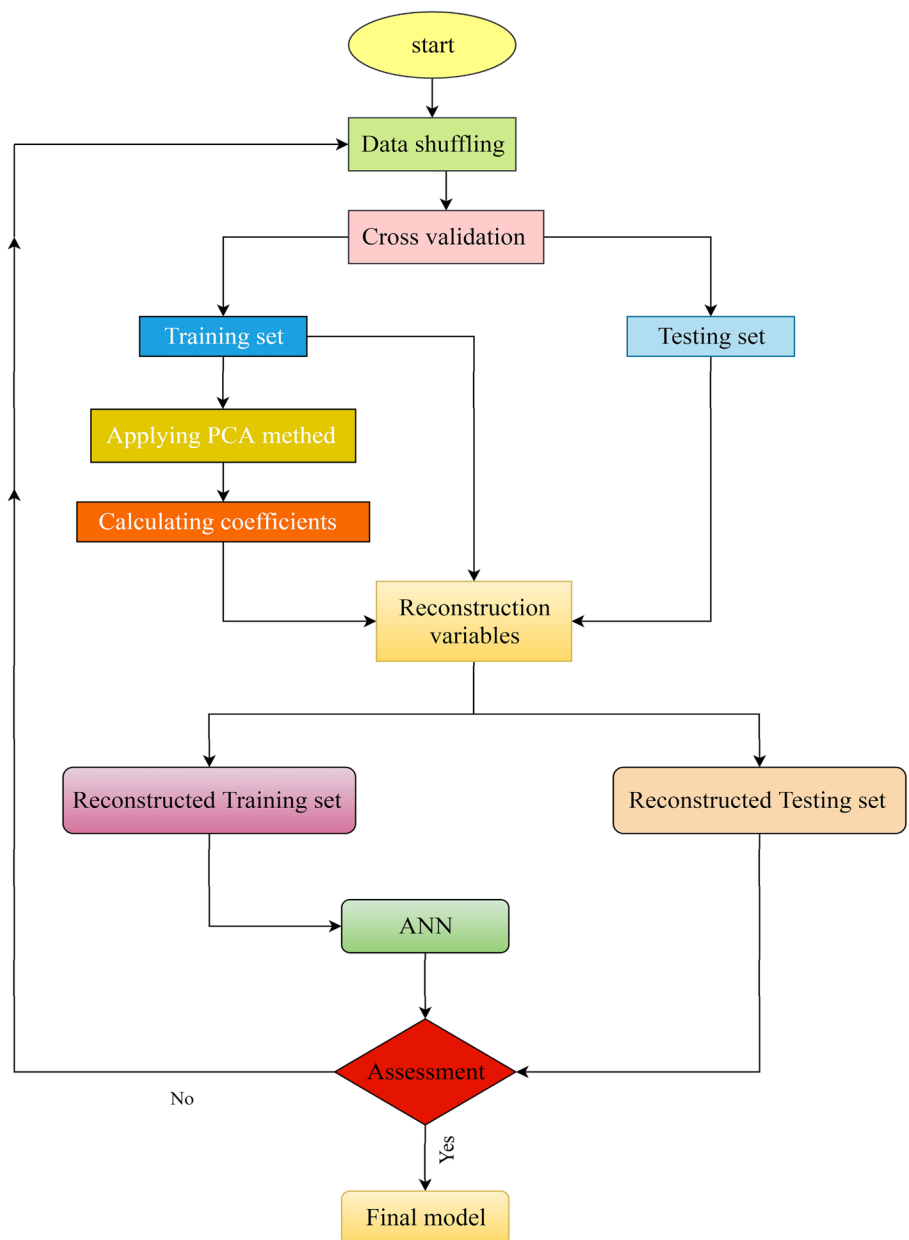model that is capable to capture and mimic the entire system characteristic using only one group. Accordingly, in



**Fig. 2** K-fold cross validation approach

**Fig. 3** ANN-PCA joins with CV methodology

this study, the K-fold cross-validation approach was used for reducing the uncertainty in modeling outcomes. In this current study, the dataset was randomly divided into five equaled-size subsets (K) which were all used for training the model, except one, which is used for testing the model accuracy. The process repeated 5 times, and the final estimations computed by averaged the obtained results, refer to Fig. 2. The most accurate model is subsequently selected to simulate the compressive strength of HPC using the training and testing dataset.

The scheme in Fig. 3 illustrates the use of PCA coupled with ANN and CV technique to constitute the suggested model (ANN-CVPCA) in predicting HPC compressive strength. Thus, in this study, eight models were developed

to predict the compressive of high-strength concrete. The input parameters which already discussed were introduced to the PCA approach to get the principal component variables. Accordingly, the first model was developed using one principle component variable while the second model was also established using two principal components only. This process continued until establishing the eight models developed based on principal component variables obtained from the PCA algorithm. These essential variables were introduced to CV and ANN to build the predictive models.

It is important to mention that the input data were standardized before introducing it to the predictive model based on Eq. (2).

$$X_{new} = \frac{X - \overline{X}}{\sigma} \tag{2}$$

where, $X_{new}$, standardized value, $X$ is the original data value and $\overline{X}$ and $\sigma$ are mean and standard deviation, respectively.

On top of that, a single-layer feed-forward has been used equipped with a BP algorithm (Levenberg–Marquardt) to train the model and updates the weights and biases. In the hidden layer and output layer, the tangent sigmoid function and linear function have been used, respectively. The codes were written using MATLAB 2018b languages.

## Performance evaluation

The data set was utilized to construct a prediction model of the HPC compressive strength. ANN incorporated with the PCA and CV method to constitute (ANN-CVPCA) model for forecasting the HPC compressive strength. Later, the results compared with four different models were carried out in previous studies. The compared models were: the robust modeled called classification and regression tree coupled with support vector machine, linear regression and cross-validation (CART + SVM + LR) (Chou, 2014) and other classical and novel forecasting models such as extreme learning machine (ELM), regularized extreme learning machine (RELM) combined with cross-validation method, and other powerful modeling approached (Al-Shamiri & Yuan, 2020; Chou & Pham, 2013).

To assess the validity of the adopted model (ANN-CVPCA) against the referenced models, different statistical criteria have been applied namely; correlation coefficient (cc), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Normalized Root Mean Square Error (NRMSE)(AlOmar, 2020). Refer to Eqs, (3), (4), (5), and (6).

$$cc = \frac{\sum_{i=1}^n (m - m^-)(p - p^-)}{\sqrt{\sum_{i=1}^n (m - m^-)^2 \sum_{i=1}^n (m - p^-)^2}} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (m - p)^2} \tag{4}$$

$$MAE = \frac{\left| \sum_{I=1}^N (m - p) \right|}{n} \tag{5}$$

$$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (m - p)^2}}{m^-} \tag{6}$$

whereas the number of samples is represented by $n$, the measured and predicted data is represented by $m$ and $p$, while $\boldsymbol{m^-}$ and $\boldsymbol{p^-}$ refer to the mean value of actual and predicted data set respectively. For more detailed analyses, a graphical figure (boxplot) was used to provide more description of the prediction accuracy of each model and helps to get fast and summarized information which is very significant in the evaluation step(Tao, 2021).

Additionally, sensitivity data analyses have been carried out to evaluate the performance of the predictive model (ANN). Sensitivity analysis is an approach used to distinguish between the causes and consequences of the input and output phases. The basic principle of this approach is that the input parameters applied to the ANN system are slightly omitted and the resultant output adjustment is recorded as a percentage.

## Results and discussion

In this study, ANN combined with PCA and CV Technique constituted the ANN-CVPCA model that was employed to predict the compressive strength of HPC. PCA method is used to reduce and reconstruct the input variables and then introduce them to ANN. K-fold CV procedure used to separate data into two sets: a training set and testing set where training set make up 75% of the data and 30% of data were employed for testing the functioning of the model. It is important to mention that the trial and error method is used to determine the number of hidden nodes in the single layer of ANN since there is no specific approach can be applied to establish the best possible number of hidden neurons. The capability of artificial intelligence (AI) modeling techniques to predict the compressive strength of HPC is investigated for the set of data used in the current research. For verification purposes, the attained results from the adopted model (ANN-CVPCA) are compared

with the previous studies that used CART classification and regression tree + SVM(support vector machine) + LR(linear regression) (Chou, 2014), extreme learning machine(ELM), regularized extreme learning machine( RELM) and other novel predictive models (Al-Shamiri & Yuan, 2020; Chou & Pham, 2013; Hameed & AlOmar, 2020).

K-fold CV method was used to split the data into two phases namely, training set for calibration and testing set were utilized to test the accuracy of the ANN-CVPCA model. Hameed & AlOmar investigate the prediction of compressive strength of HPC using ANN coupled with CV method which is used to select the best quality data set for the training and testing phase (Hameed & AlOmar, 2020). The mentioned study showed that the ANN-CV can forecast efficiently the compressive strength test compared with the Multilinear Regression model. Although this study presented high-accuracy forecasting results, it lacks some important things that may benefit researchers in the field of developing concrete mixtures. One of the most important issues that concrete mixture developers may encounter is to determine which factors with the most significant impact on the compressive strength of high-performance concrete. In this study, the sensitivity analysis method was applied to the ANN model in order to uncover more concrete properties and determine any of the eight variables, mentioned in the methodology part, that have an effective impact on the HPC compression strength. Figure 4 showed the result of sensitivity analyses for the best predictive ANN model. Based on Fig. 4, the ANN showed all variables are significant and highly associated with the target. Moreover, the most two efficient parameters are superplasticizer and cement, respectively. These two parameters made up about 28%. The observation according to this percentage confirms that the sensitivity analysis of each parameter is consistent with the

complexity HPC. The relationship between these variables is characterized as highly nonlinear since most input factors have an important influence on the compressive strength of HPC. Therefore, it is very essential to use sophisticated and advanced techniques for getting a comprehensive understanding of the compacted behavior of HPC mixture. Additionally, superplasticizers and cement parameters can play an important role in affecting HPC compressive strength. Therefore, taking into consideration these factors is vital in designing a more efficient HPC mixture.

Table 2 illustrates the predicted results of the proposed modeling approach. In the first seven models listed in Table 2, the PCA approach was utilized to minimize and compress the original input variables whereas, the PCA approach was used for the reconstruction of the input variables in the last model. It is noticeable, the accuracy of ANN-CVPCA models improved by increasing the number of PCs. For example, when only one PC variable is used as input to ANN, the model led to inadequate predicted values according to statistical criteria (RMSE = 15.33(Mpa), MAE = 12.39(Mpa), NRMSE = 0.43 and CC = 0.18). On the other hand, the performance of ANN-CVPCA is improved when 7 PCs are used and recorded less error (RMSE = 5.33(Mpa), MAE = 4.07(Mpa), NRMSE = 0.16 and CC = 0.95). According to Table 2, it can be observed that M8 with using 8 PCs provides the best accuracy with less error reported in the testing set. For more comparative information, Fig. 5 exhibits a graphical presentation using a boxplot diagram. The mentioned figure summarizes the forecasting accuracy for each predictive model and can be concluded that the M8 has the lowest calculated error among the other eight models. Moreover, Using the PCA technique for compressing the variables and reconstructing them enhance the accuracy of the predictive model and eventually
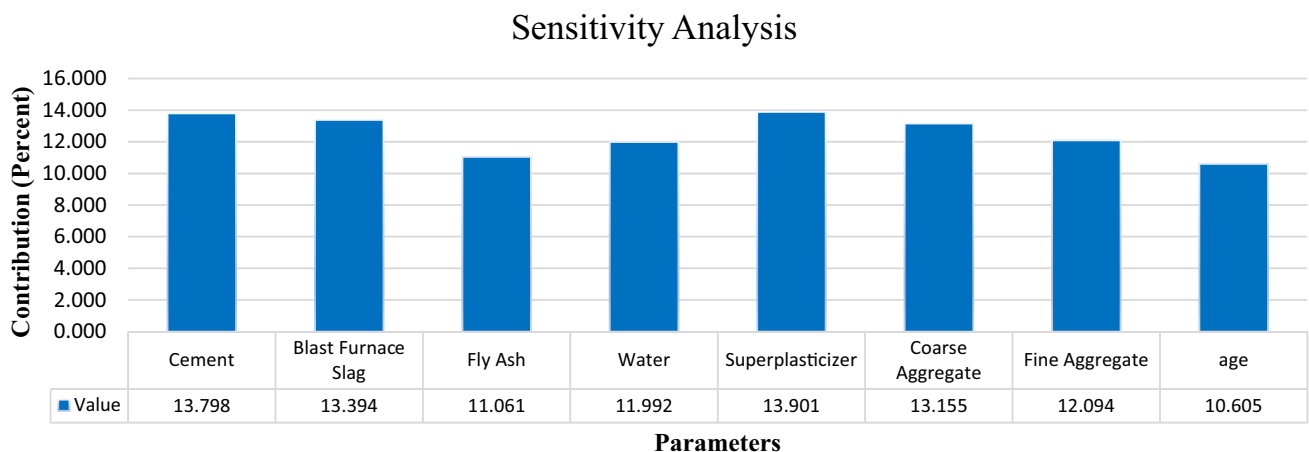


**Sensitivity Analysis**

| Parameters | Cement | Blast Furnace Slag | Fly Ash | Water | Superplasticizer | Coarse Aggregate | Fine Aggregate | age |
|---|---|---|---|---|---|---|---|---|
| ■ Value | 13.798 | 13.394 | 11.061 | 11.992 | 13.901 | 13.155 | 12.094 | 10.605 |

**Fig. 4** Sensitivity analysis percentage for each parameter contribute to HPC compressive strength: ANN approach with Cross-Validation (4 k-fold)

**Table 2** evaluating ANN-PCA models with cross validation: Testing set

| Model | Architecture | Principal components | RMSE (mpa) | MAE (mpa) | NRMSE | CC |
|---|---|---|---|---|---|---|
| Model1 | 1–3-1 | 1 | 15.33 | 12.39 | 0.43 | 0.18 |
| Model2 | 2–2-1 | 2 | 14.58 | 12.10 | 0.42 | 0.41 |
| Model3 | 5–3-1 | 3 | 11.86 | 9.79 | 0.34 | 0.68 |
| Model4 | 4–7-1 | 4 | 11.04 | 8.75 | 0.32 | 0.76 |
| Model5 | 5–10-1 | 5 | 9.76 | 7.70 | 0.27 | 0.80 |
| Model6 | 6–12-1 | 6 | 6.50 | 5.07 | 0.18 | 0.92 |
| Model7 | 7–15-1 | 7 | 5.33 | 4.07 | 0.16 | 0.95 |
| **Model8** | **8–8-1** | **8** | **4.64** | **3.43** | **0.13** | **0.96** |
| *Individual approaches*(Chou & Pham, 2013) | | | | | | |
| ANN | – | | 6.329 | 4.421 | – | 0.93 |
| CART | – | | 9.703 | 6.815 | – | 0.84 |
| CHAID | – | | 8.983 | 6.088 | – | 0.861 |
| LR | – | | 11.243 | 7.867 | – | 0799 |
| GENLIN | – | | 11.375 | 7.867 | – | 0.779 |
| SVM | - | | 6.911 | 4.764 | - | 0.923 |
| *Ensemble approaches*(Chou & Pham, 2013) | | | | | | |
| ANN + CHAID | – | | 7.028 | 4.668 | – | 0.922 |
| ANN + SVM | – | | 6.174 | 4.236 | – | 0.939 |
| ANN + SVM + CHAID | – | | 6.692 | 4.580 | – | 0.929 |
| *Individual approaches* (Chou, 2014) | | | | | | |
| MLP | – | | 15.66 | 19.7 | – | – |
| CART | | | 5.71 | 4.26 | – | – |
| SVM | | | 5.75 | 4.01 | – | – |
| LE | | | 10.13 | 7.88 | | |
| *Ensemble approaches* (Chou, 2014) | | | | | | |
| CART + SVM + LR | | – | 5.08 | 3.52 | – | – |
| MLP + CART + LR | | – | 5.08 | 7.03 | | |
| MLP + SVM + +LR | | – | 7.66 | 5.92 | – | – |
| MLP + CART + SVM | | – | 9.48 | 7.25 | – | – |
| *Ensemble approaches*(Al-Shamiri & Yuan, 2020) | | | | | | |
| CV + ELM | | – | 6.0377 | 4.4419 | – | 0.929 |
| CV + RELM | | – | 5.5075 | 3.9745 | – | 0.9403 |

Bold numbers indicate the best model accuracy

made the current model superior to other forecasting models reported in publication (Chou (2014); Chou & Pham, 2013; Al-Shamiri & Yuan, 2020; Hameed & AlOmar, 2020). The main reason behind the superiority of the pressed model (ANN-CVPCA) is the existence of CV and PCA approaches. Based on the given results, the ANN-CVPCA model was stable and capable to capture nonlinear relations between target and input parameters. Moreover, the two approaches enhance the generalization efficiency and reduce the uncertainty in modeling results. The PCA method improved the prediction capacity of the proposed model by introducing a non-redundant variable to the model hence, providing high accurate prediction estimations. Additionally, comparing the obtained results to other models, reported in other

publications, showed an improvement in the accuracy of the current ANN-CVPCA model based on many statistical matrices. The current ANN-CVPCA model outperformed different models such as ANN + CHAID (artificial neural network with chi-squared automatic interaction detector), ANN + SVM (ANN with support vector machine), GEN-LIN (generalized linear model), and other models have been reported in the literature as listed in Table 2. PCA method played an important role in reducing the sum square errors compared to several powerful models like ANN + SVM, and GENLIN. The evidence of the better performance obtained by ANN-CVPCA is shown based on several statistical measures such as and fewer RMSE (4.64 Mpa) is
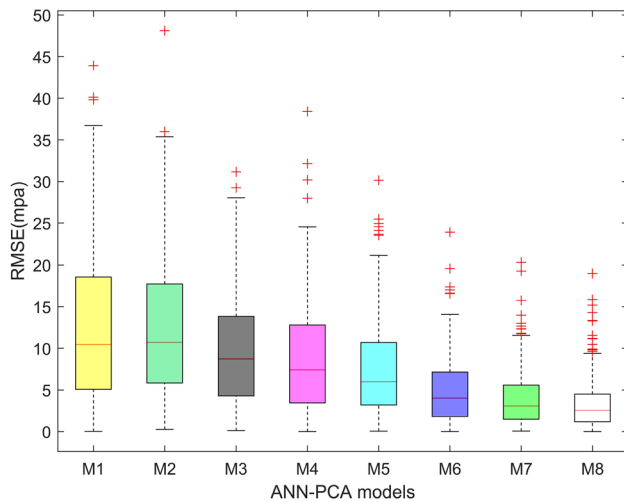
**Fig. 5** Boxplot diagram describing the comparative analyses for eight different models

obtained compared to ANN + SVM (6.174 Mpa) and GEN-LIN (11.375 Mpa).

Moreover, the improvement in prediction results compared also with previous studies, in terms of reducing sum square error, could also be noticed when the CV method is combined with ANN-CVPCA. Consequently, the proposed model (ANN-CVPCA) also outperforms complicated sophisticated models (ANN and CART (classification and regression tree) + SVM(support vector machine) + LR(linear regression)) (Chou, 2014). The ANN-CVPCA model recorded the uppermost value of the CC and the lowermost value of RMSE (see Table 2). In summary, the used ANN-CVPCA model yielded lower value RMSE than the advanced forecasting model (CART + SVM + LR). Wherein, there is a good and promising improvement in reduction RMSE which reached 9.48%.

Additionally, the ANN-CVPCA model outperformed and achieved an excellent improvement in generating more accurate results in comparison with recent and novel predictive models such as extreme learning machine (ELM), regularized extreme machine (RELM) approaches combined with CV strategy (Al-Shamiri & Yuan, 2020), and other production models which can be summarized in Table 2. It can be seen there was a satisfactory enhancement in minimizing RMSE index, thereby, improving the accuracy of ANN-CVPCA performance by 23.15% and 15.75% over CV + ELM (cross-validation with extreme learning machine) and CV + RELM (cross-validation with regularized extreme learning machine), respectively. The comparison with robust models gave the impression that the suggested model in this study gave satisfactory estimates. Besides, the proposed model was found to produce more accurate prediction than 19 models with had done in previous studies. In general

these models gave correlation of coefficient (CC) values ranged from 0.779 to 0.9403. However, the capacity of the proposed model for this study generate excellent prediction with higher value of CC (0.96). CC criteria is considered the most significant parameters than other statistical criteria when it comes to compare a certain model with others were developed in previous studies. The reason behind this fact that the most statistical parameters except CC are highly affected with the range of dataset and normalization approach. Therefore, these criteria such as RMSE, MSE, and so on may give misleading comparisons.

It is believed that removing the correlation between input variables, by using the PCA method, could be an important reason for improving the result. Once the correlation between raw input variables is removed, these variables are transformed as PCs components. Later, the redundant data are generally removed. The performance of the ANN technique seemed to be improved because the forecasting model considered all input variables since they became independent variables that have significant impacts on the targets. Additionally, the CV method could select the proper data set to train ANN, which led to minimizing the sum square error of the predicted values of the HPC compressive strength. Selecting good quality data for both training and testing phases by applying the CV approach is very important for improving the accuracy of ANN.

Figure 6a and b shows the performance of all the forecasted results for the testing phase obtained by the predictive models. It is noticeable that the estimation of the ANN-CVPCA model using eight PCs is less scattered and closer to the trend line compared to other models. For M8, the first PC accounts for about 28.337% of the total variability, and the first 7 PCs can summarize the majority of the input data set and explains about 99.610% of the total variability as shown in Table 3.

For a more detailed analysis of the ANN-CVPCA model based on 8 PCs, quantitative analysis has been carried out to identify the performance of ANN in predicting extreme values of HPC compressive strength. Figure 7 illustrates the peak ten predicted values of HPC compressive strength obtained by using ANN coupled with the CV method. The peak predicted readings are very close to actual ones. For instance, the three highest actual (predicted) values of HPC compressive strength are 77.3, 77.3, and 76.80 (73.04, 73.04, and 73.6) MPa. Generally, there is a good match between predicted and actual results. Also, the relative error percentage indicator (RE), which can be expressed by Eq. 7, is generally small (see Fig. 7).

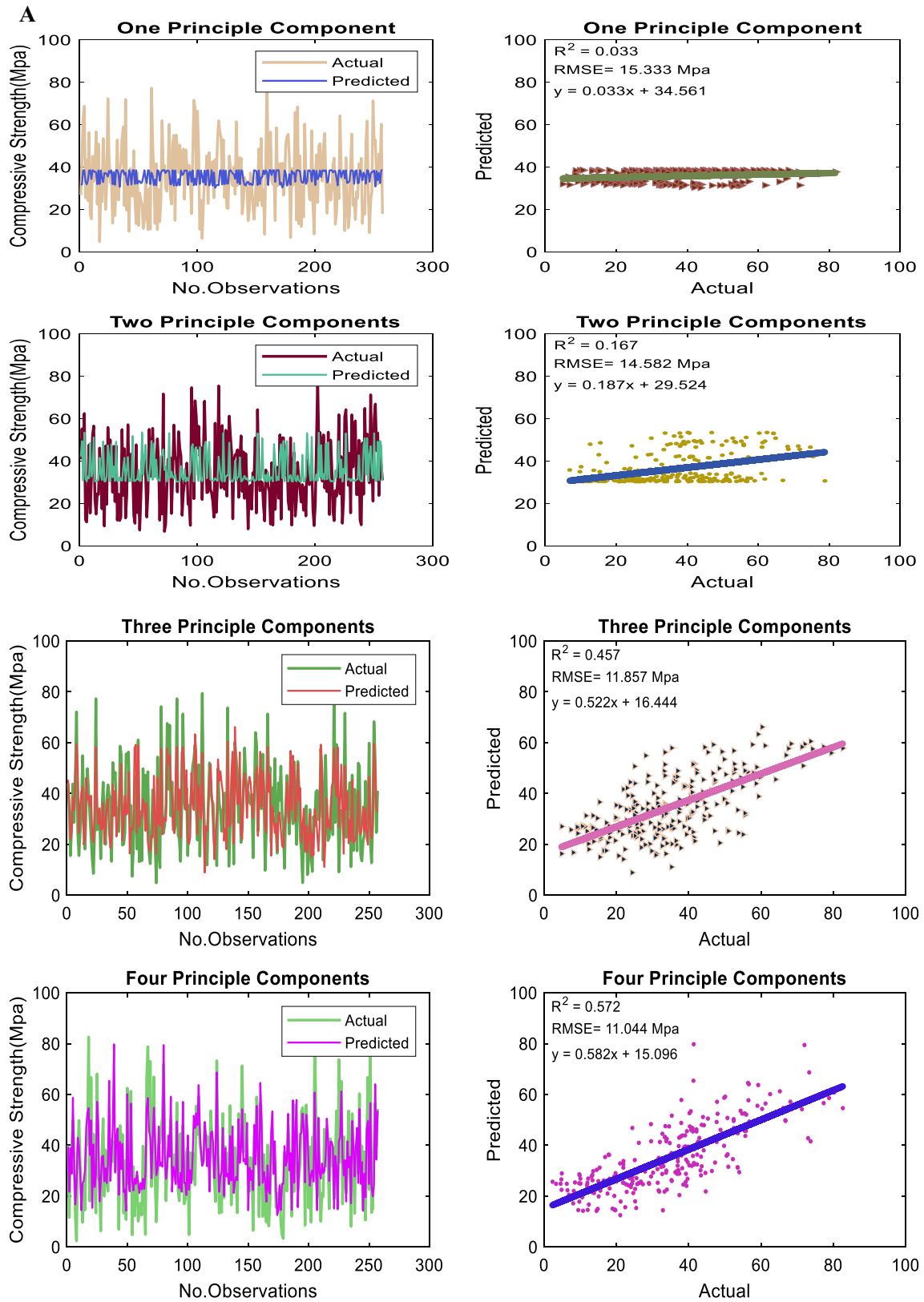$$RE\% = 100 \times \frac{\text{actual value - predective value}}{\text{actual value}} \quad (7)$$

**Fig. 6** **a** Comparative and scatter plots between observed and predictive compressive strength test of concrete using different input (PCs) variables: Testing set of 1, 2, 3 and 4 PCs. **b** Comparative and scatter plots between observed and predictive compressive strength test of concrete using different input (PCs) variables: Testing set of 5, 6, 7, and 8 PCs
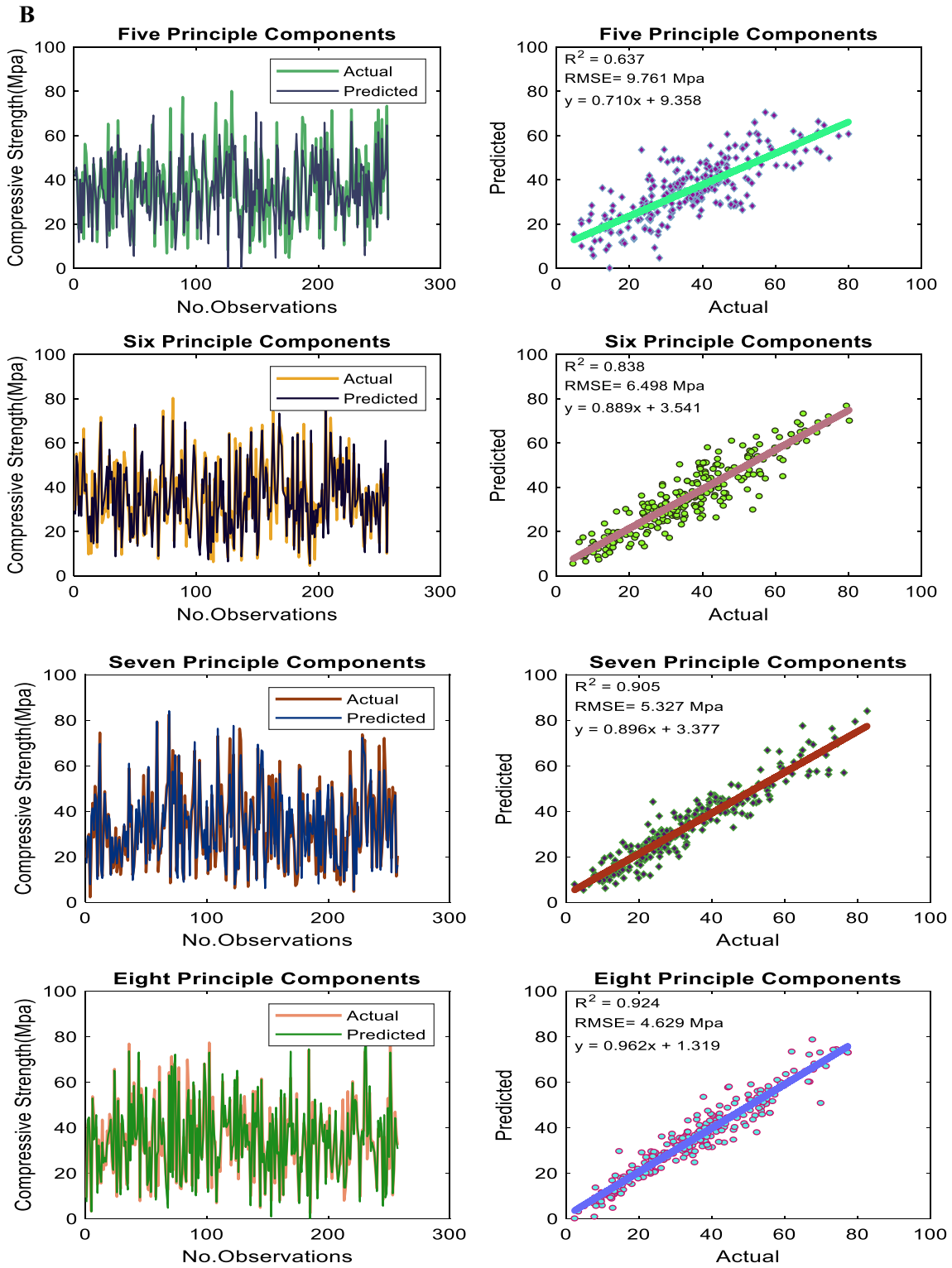
**Fig. 6** (continued)

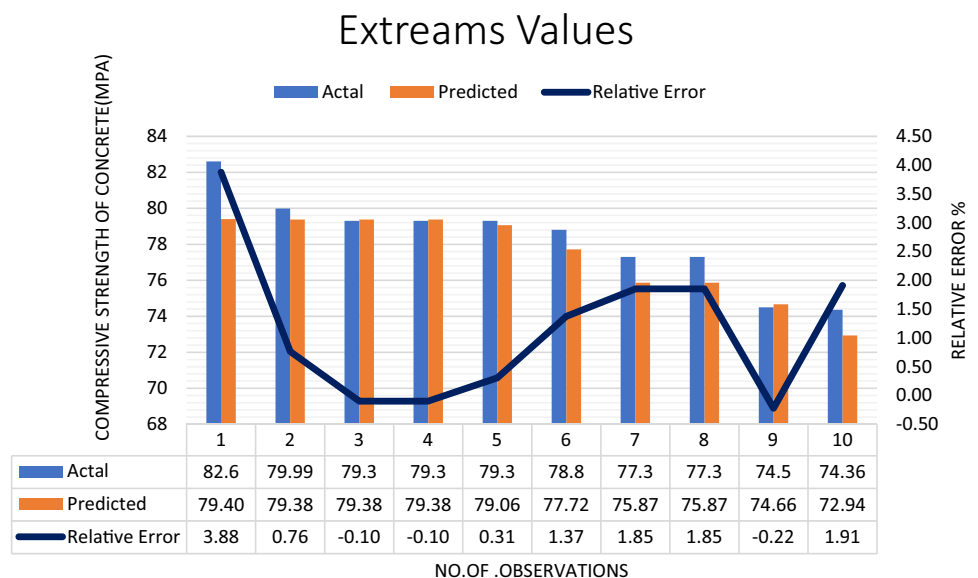**Table 3** Descriptive statistics of eight principal components

| Principal components | Variance proportion (%) | Accumulative variance proportion (%) | Eigen values |
|---|---|---|---|
| PC1 | 28.337 | 28.337 | 2.236 |
| PC2 | 17.966 | 46.303 | 1.418 |
| PC3 | 16.929 | 63.231 | 1.336 |
| PC4 | 12.776 | 76.008 | 1.008 |
| PC5 | 12.199 | 88.207 | 0.963 |
| PC6 | 9.111 | 97.318 | 0.719 |
| PC7 | 2.292 | 99.610 | 0.181 |
| PC8 | 0.390 | 100.000 | 0.031 |

## Conclusion

This study investigated the ability of the artificial intelligence technique to predict the compressive strength test of HPC based on several input parameters such as coarse aggregate, fine aggregate, water, cement, age as well as some chemical additives like blast furnace slag, fly ash, and superplasticizer. ANN was incorporated with the PCA method and CV constituted (ANN-CVPCA) model, which was utilized for predicting the compressive strength of HPC. The obtained results illustrated that there was a fair match between forecasted and real values of the HPC compressive strength test. The target values of the employed model were compared with four models from the literature that used the same data set. For evaluating and assessing the adopted ANN-CVPCA model, the accuracy of the

adopted model was validated against many different models including, (GENLIN, ANN + SVM, CART + SVM + LR), and other recent and novel models like (CV + ELM and CV + RELM) according to serval statistical measures. The result showed that the accuracy of the proposed model (ANN-CVPCA) outperformed all the other models. The adopted model showed superiority in comparison with the conventional and novel models in published literature due to the combination of ANN with PCA as well as the contribution of CV in data organization which eliminated the correlation between input data hence, providing significant and non-redundant information to the model. Moreover, the two approaches, CV and PCA play an efficient role in increasing the generalization ability of the adopted model and reducing the uncertainty in the modeling results compare to other AI models used in this current study. The given results of the suggested model are more stable and accurate. Additionally, the sensitivity analysis has been carried out to provide an important overview of the most important variables that may be significant to develop a concrete mixture thus, attaining a high compression strength. This analysis showed that the superplasticizer and cement are the most effective parameters on the compressive strength of HPC, with 28% effectiveness. This study recommends intensifying the PCA approach with ANN which is trained by bio-inspiration algorithm (Bui, 2018, 2020; Zhang, 2021) to enhance the performance capacity of the hybrid models thereby, obtaining more accurate estimates.

**Fig. 7** Peak values estimated using ANN-PCA model



| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Actal | 82.6 | 79.99 | 79.3 | 79.3 | 79.3 | 78.8 | 77.3 | 77.3 | 74.5 | 74.36 |
| Predicted | 79.40 | 79.38 | 79.38 | 79.38 | 79.06 | 77.72 | 75.87 | 75.87 | 74.66 | 72.94 |
| Relative Error | 3.88 | 0.76 | -0.10 | -0.10 | 0.31 | 1.37 | 1.85 | 1.85 | -0.22 | 1.91 |

NO.OF .OBSERVATIONS

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

Afan, H. A., et al. (2015). ANN Based sediment prediction model utilizing different input scenarios. *Water Resources Management, 29*(4), 1231–1245.

Aïtcin, P.-C., *High performance concrete*. 2011: CRC press.

Akbari, M. and V. Jafari Deligani, *Data driven models for compressive strength prediction of concrete at high temperatures.* Frontiers of Structural and Civil Engineering, 2020.

AlOmar, M. K., et al. (2020). Data-Driven Model for the Prediction of Total Dissolved Gas: Robust Artificial Intelligence Approach. *Advances in Civil Engineering, 2020,* 6618842.

AlOmar, M. K., Hameed, M. M., & AlSaadi, M. A. (2020). Multi hours ahead prediction of surface ozone gas concentration: Robust artificial intelligence approach. *Atmospheric Pollution Research, 11*(9), 1572–1587.

Al-Shamiri, A. K., & Yuan, T.-F. (2020). Non-tuned machine learning approach for predicting the compressive strength of high-performance concrete. *Materials, 13*(5), 1023.

Atici, U.J.E.S.w.a., *Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network*. 2011. **38**(8): p. 9609–9618.

Bui, D.-K., et al. (2018). A modified firefly algorithm-artificial neural network expert system for predicting compressive and tensile strength of high-performance concrete. *Construction and Building Materials, 180,* 320–333.

Bui, D.-K., et al. (2020). An artificial neural network (ANN) expert system enhanced with the electromagnetism-based firefly algorithm (EFA) for predicting the energy consumption in buildings. *Energy, 190,* 116370.

Cao, M., et al. (2017). Neural network ensemble-based parameter sensitivity analysis in civil engineering systems. *Neural Computing and Applications, 28*(7), 1583–1590.

Chou, J.-S., & Pham, A.-D. (2013). Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength. *Construction and Building Materials, 49,* 554–563.

Chou, J.-S., et al. (2014). Machine learning in concrete strength simulations: Multi-nation data analytics. *Construction and Building Materials, 73,* 771–780.

Fiyadh, S.S., et al., *The modelling of lead removal from water by deep eutectic solvents functionalized CNTs: artificial neural network (ANN) approach*. 2017. **76**(9): p. 2413–2426.

Fiyadh, S.S., et al., *BTPC-Based DES-Functionalized CNTs for A s 3+ Removal from Water: NARX Neural Network Approach*. 2018. **144**(8): p. 04018070.

Fiyadh, S. S., et al. (2017). The modelling of lead removal from water by deep eutectic solvents functionalized CNTs: artificial neural network (ANN) approach. *Water Science and Technology, 76*(9), 2413–2426.

Fiyadh, S. S., et al. (2018). Arsenic removal from water using N, N-diethylethanolammonium chloride based DES-functionalized CNTs:(NARX) neural network approach. *Journal of Water Supply: Research and Technology-Aqua, 67*(6), 531–542.

Fiyadh, S. S., et al. (2019a). *Lead removal from water using DES functionalized CNTs: ANN modeling approach., 150,* 105–113.

Fiyadh, S. S., et al. (2019b). *Mercury removal from water using deep eutectic solvents-functionalized multi walled carbon nanotubes: Nonlinear autoregressive network with an exogenous input neural network approach., 38*(6), e13261.

Fiyadha, S.S., et al., *The modelling of arsenic removal from water by deep eutectic solvents functionalized CNTs: Artificial neural network (ANN) approach.*

Hameed, M.M., et al., *Prediction of high-strength concrete: high-order response surface methodology modeling approach.* Engineering with Computers, 2021.

Hameed, M. M., & AlOmar, M. K. (2020). *Prediction of Compressive Strength of High-Performance Concrete: Hybrid Artificial Intelligence Technique*. Springer International Publishing.

Hameed, M., et al. (2017). Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region Malaysia. *Neural Computing and Applications, 28*(1), 893–905.

Heidari, A., M. Hashempour, and D. Tavakoli, *Using of backpropagation neural network in estimation of compressive strength of waste concrete %J Journal of Soft Computing in Civil Engineering*. 2017. **1**(1): p. 54–64.

Hola, J. and K.J.A.M.J. Schabowicz, *Methodology of neural identification of strength of concrete.* 2005. **102**(6): p. 459.

Iranmanesh, A., & Kaveh, A. (1999). Structural optimization by gradient-based neural networks. *International journal for numerical methods in engineering, 46*(2), 297–311.

Kaveh, A. and M.R. Dehkordi, *RBF and BP neural networks used for the design of domes*, in *Space Structures 5*. 2002, Thomas Telford Publishing. p. 1: 469–478.

Kaveh, A., & Elmieh, R. (2001). *and H*. Servati.

Kaveh, A., & Iranmanesh, A. (1998). Comparative study of backpropagation and improved counterpropagation neural nets in structural analysis and optimization. *International Journal of space structures, 13*(4), 177–185.

Kaveh, A., & Servati, H. (2001). Design of double layer grids using backpropagation neural networks. *Computers & Structures, 79*(17), 1561–1568.

Kaveh, A., H. Servati, and D.D. FAZEL, *Prediction of moment-rotation characteristic for saddle-like connections using FEM and BP neural networks*. 2001.

Kaveh, A., Gholipour, Y., & Rahami, H. (2008). Optimal design of transmission towers using genetic algorithm and neural networks. *International Journal of Space Structures, 23*(1), 1–19.

Kewalramani, M. A., & Gupta, R. (2006). Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks. *Automation in Construction, 15*(3), 374–379.

Khademi, F., K.J.I.U.o.S. Behfarnia, and Technology, *Evaluation of concrete compressive strength using artificial neural network and multiple linear regression models.* 2016. **6**(3): p. 423–432.

Khademi, F., et al. (2016). Predicting the strength of recycled aggregate concrete using Artificial Neural Network, Adaptive Neuro-Fuzzy Inference System and Multiple Linear Regression. *International Journal of Sustainable Built Environment, 5*(2), 355–369.

Khademi, F., et al. (2017). Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive

strength of concrete. *Frontiers of Structural and Civil Engineering, 11*(1), 90–99.

Lee, S.-C.J.E.S., *Prediction of concrete strength using artificial neural networks.* 2003. **25**(7): p. 849–857.

Li, X., et al. (2019). Evaluation of data-driven models for predicting the service life of concrete sewer pipes subjected to corrosion. *Journal of Environmental Management, 234,* 431–439.

Naderpour, H., Rafiean, A. H., & Fakharian, P. (2018). Compressive strength prediction of environmentally friendly concrete using artificial neural networks. *Journal of Building Engineering, 16,* 213–219.

Ni, H.-G., & Wang, J.-Z. (2000). Prediction of compressive strength of concrete by neural networks. *Cement and Concrete Research, 30*(8), 1245–1250.

Nikoo, M., et al., *Prediction of Concrete Compressive Strength by Evolutionary Artificial Neural Networks %J Advances in Materials Science and Engineering.* 2015. **2015**: p. 8.

Öztaş, A., et al. (2006a). *Predicting the compressive strength and slump of high strength concrete using neural network., 20*(9), 769–775.

Öztaş, A., et al. (2006b). Predicting the compressive strength and slump of high strength concrete using neural network. *Construction and Building Materials, 20*(9), 769–775.

Reddy, T. C. S. (2018). Predicting the strength properties of slurry infiltrated fibrous concrete using artificial neural network. *Frontiers of Structural and Civil Engineering, 12*(4), 490–503.

Rofooei, F., Kaveh, A., & Farahani, F. M. (2011). Estimating the vulnerability of the concrete moment resisting frame structures using artificial neural networks. *Iran University of Science & Technology, 1*(3), 433–448.

Shahin, M. A., Jaksa, M. B., & Maier, H. R. (2001). Artificial neural network applications in geotechnical engineering. *Australian geomechanics, 36*(1), 49–62.

Shi, X. (2016). Experimental and modeling studies on installation of arc sprayed Zn anodes for protection of reinforced concrete structures. *Frontiers of Structural and Civil Engineering, 10*(1), 1–11.

Słoński, M. (2010). A comparison of model selection methods for compressive strength prediction of high-performance concrete using neural networks. *Computers & Structures, 88*(21), 1248–1253.

Tao, H., et al. (2021). Global solar radiation prediction over North Dakota using air temperature: Development of novel hybrid intelligence model. *Energy Reports, 7,* 136–157.

Yaseen, Z. M., et al. (2016). RBFNN versus FFNN for daily river flow forecasting at Johor River Malaysia. *Neural Computing and Applications, 27*(6), 1533–1542.

Yeh, I.-C. (1998). Modeling of the strength of high-performance concrete using artificial neural networks. *Cement and Concrete research, 28*(12), 1797–1808.

Yu, Y., et al. (2018). A novel optimised self-learning method for compressive strength prediction of high performance concrete. *Construction and Building Materials, 184,* 229–247.

Zhang, G., et al., *Reinforced concrete deep beam shear strength capacity modelling using an integrative bio-inspired algorithm with an artificial intelligence model.* Engineering with Computers, 2020.

Zhang, H., et al., *A generalized artificial intelligence model for estimating the friction angle of clays in evaluating slope stability using a deep neural network and Harris Hawks optimization algorithm.* Engineering with Computers, 2021: p. 1–14.