International Journal of Computers and Applications

ISSN: 1206-212X (Print) 1925-7074 (Online) Journal homepage: https://www.tandfonline.com/loi/tjca20

An efficient approach for cement strength prediction

Vilas H. Gaidhane, Nand Kumar, Ravi Kant Mittal & J. Rajevenceltha

To cite this article: Vilas H. Gaidhane, Nand Kumar, Ravi Kant Mittal & J. Rajevenceltha (2019): An efficient approach for cement strength prediction, International Journal of Computers and Applications, DOI: 10.1080/1206212X.2019.1673288

To link to this article: https://doi.org/10.1080/1206212X.2019.1673288

Published online: 08 Oct 2019.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tjca20

INTERNATIONAL JOURNAL OF COMPUTERS AND APPLICATIONS
https://doi.org/10.1080/1206212X.2019.1673288

Taylor & Francis
Taylor & Francis Group

Check for updates

An efficient approach for cement strength prediction

Vilas H. Gaidhane [a], Nand Kumar [b], Ravi Kant Mittal [c] and J. Rajevenceltha [a]

[a]Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, Dubai Campus, UAE; [b]Department of Computer Science, Birla Institute of Technology and Science Pilani, Dubai Campus, UAE; [c]Department of Mechanical Engineering, Birla Institute of Technology and Science, Pilani, India

ABSTRACT
In this paper, a simple and computationally efficient approach is proposed to predict the cement strength. It is based on the mathematical concept of covariance matrix and polynomial coefficients. The polynomial coefficients are used to represent the features of cement strength data set. The efficiency and feasibility of the proposed approach is demonstrated on the cement strength data set collected from the cement industry for 2 days, 7 days and 28 days samples. Based on the number of dynamic input variables of the cement strength, the different prediction models such as SOM, linear and nonlinear regression and artificial neural network are designed and the various experimentations are carried to evaluate the proposed approach. Experimental results have shown the effectiveness of the proposed approach in the form of RMSE, *R*-square coefficients and computational time. It is observed that the proposed polynomial coefficient-artificial neural network approach performs better and predict the cement strength efficiently as compared to other existing methods.

ARTICLE HISTORY
Received 19 February 2019
Accepted 23 September 2019

KEYWORDS
Cement strength; polynomial coefficients; SOM; regression model; artificial neural network; prediction

1. Introduction

Cement is a fundamental material used in construction works and civil engineering applications. High-quality control of the cement is of paramount importance and becoming a challenge for the manufacturing industry. The good quality cement depends on the various parameters such as capacity of equipment, human resources and the availability of the raw material. There are some factors that affect the cement manufacturing and construction engineering. These factors are classified into two categories, structured and unstructured. Structured factors are those which can be well studied and described by the fixed rules. These factors include mixed design, raw materials, and lean cement, etc. The unstructured factors such as control accuracy, manufacturing conditions, skilled manpower cannot be defined by the fixed rules. However, these types of factors greatly influence the production quality as well as quantity of the cement.

The compressive strength is one of the most important quality indicators in the cement industry. The prediction of the cement strength plays an important role and facilitates the flexibility to take the necessary action before the manufacturing. Therefore, an accurate and reliable strength prediction model is a necessary requirement of the manufacturing system. A good prediction model may help to save the manufacturing cost as well as time. In recent years, many researchers have carried out the numerous research works to predict the strength of the cement [1–4]. These approaches are generally presented using the simple linear and nonlinear regression models. However, these classical prediction approaches unable to achieve the accurate prediction results. It is due to the unequal raw material, inaccurate measurements, unskilled labors, and unstable operations. To overcome these issues, various researchers have carried out an extensive research work and more attention has been given to the artificial intelligence (AI) methods such as self-organizing feature map (SOM), artificial neural network, fuzzy logic, genetic algorithm, and hybrid models.

In recent years, Nikoo et al. [5] have used the SOM to calculate the cement compressive strength. In this, the structure of SOM is optimized using the genetic algorithm. A basic statistical principal analysis (PCA) technique is considered for feature extraction and combined with the SOM to evaluate the adhesion between the cement layers [6]. A back propagation artificial neural network (BP-ANN) has been used to predict the strength of concrete and proved that the ANN-based approach performs better than the linear and multimodal regression models [7,8]. Further, the various researchers have proposed the hybrid models based on the ANN and fuzzy logic models. Ozcan et al. [9] have used the ANN and fuzzy-based approach to predict the compressive strength of the silica additive cement. Sobhani et al. [10] have proposed an approach based on regression, adaptive fuzzy and neural network models to predict the compressive strength of slump-free concrete. Demir [11] developed a modified fuzzy-based model for the prediction of normal and high concrete strength.

Recently, the various hybrid models are employed with a combination of artificial neural network, fuzzy logic system, ANFIS and genetic algorithm [12–17]. These methods show that the ANFIS based system results into an efficient tool for prediction of the strength of the cement. However, the limited investigations have been focused on parameters setting which may be an important issue in prediction. Another type of hybrid system based on ANN and genetic algorithm has been proposed by various researchers and employed in different applications [18–20]. A genetic algorithm is used to optimize the weights and thresholds of ANN to minimize error between the actual and target output.

In all of the above systems, the hybrid techniques are based on the supervised learning approach. However, these empirical model-based approaches are unable to achieve the accurate prediction due to the nebulous input variables as well as the redundant data set. Moreover, the limited number of inputs has been used to avoid the

CONTACT Vilas H. Gaidhane vilasgd612@gmail.com Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai 345055, UAE

© 2019 Informa UK Limited, trading as Taylor & Francis Group

computational complexity. Therefore, a simple and computationally efficient method is needed for the cement strength prediction application.

After a detailed study on various techniques in control engineering and data compression, it is observed that the control techniques such as polynomial coefficient [21] can be useful in cement strength prediction applications. Therefore, in this paper an attempt has been made to use this technique for prediction. The proposed approach is based on the concept of two-dimensional normalized covariance matrix and polynomial coefficients. These coefficients represent the important features of data sets and further used to calculate reduced dimension input variables. This mathematical proposed approach increases the effectuality of the prediction system. The advantage of the proposed approach is that it can reduce the input data onto a two-dimensional structure with high correlation.

The rest of the paper is structured as follows. A brief overview of existing prediction methods and data collection is presented in Sections 2 and 3, respectively. The proposed approach is presented in Section 4. The result of experiments and analysis are explained in Section 5. Finally, conclusions are drawn in Section 6.

2. Overview of existing prediction methods

In recent years, SOM, linear and non-linear regression models and hybrid ANN methods are extensively used for cement strength prediction. The mathematical preliminaries of these methods are explained in following subsections.

2.1. Self-organizing map (SOM)

In self-organizing map, a competitive learning approach is used for the training which has been adopted based on the concept of the human brain. In the SOM network, the processors are represented by the nodes in one dimensional, two-dimensional or multidimensional or more-dimensional networks as shown in Figure 1 [6]. Each unit is regulated based on the input patterns. A competitive learning is exploited in such network so that at the end of a stage race, only one unit wins and its weight changes differently as compared to other units. Such type of learning is called as unsupervised learning.

The SOM networks are classified into different classes such as Kohenen, MaxNet, Maxican Hat, and Hamming. Kohenen layer may be one, two or multiple dimensional neuron as shown in Figure 1. A probability density function $p(x)$ is described by creating a vector set w_i. Each neuron computes the distance of the input vector x from its weight w_i given as [22]

$$\delta_i = d(x, w_i) \tag{1}$$

where d is the distance measure function. A spherical arc distance $d(p,q) = 1 - \cos\,\theta$ or a Euclidian distance $d(p,q) = |p - q|$. This measure defined the closeness to the input vector x. The neuron with the closest weight to the input vector wins the computation for which Z_i is equal to one and other Z_i equal to zero. Based on these mathematical preliminaries, Kohenan defined the rule for automatic weight update given as

$$w_{i,\text{new}} = w_{i,\text{old}} + \alpha(x - w_{i,\text{old}})Z_i \tag{2}$$

where $0 < \alpha \leq 1$. Thus, the updated new weight can be represented as

$$w_{i,\text{new}} = \begin{cases} (1-\alpha)w_{i,\text{old}} + \alpha x & \text{for winner} \\ w_{i,\text{old}} & \text{for other neurons} \end{cases} \tag{3}$$

2.2. Linear and non-linear regression model

Regression is a statistical technique used to define the relationship between a dependent and multiple independent variables. The analysis of experimental, ordinal, or categorical data can be done using regression approach. In this paper, multiple regression analysis (linear and non-linear) has been carried out for the prediction of the cement strength. The cement strength can be predicted using first-order linear equation and second-order non-linear equation models. The first-order linear equation model is represented as [23,24]

$$Y = b_0 + \sum_{i=1}^{N} b_i X_i \tag{4}$$

Similarly, the second-order non-linear equation model can be written as [24]

$$Y = b_0 + \sum_{k=1}^{N} b_k X_k + \sum_{i=1}^{N} \sum_{j=1}^{N} b_i X_i X_j \tag{5}$$

where Y is the estimated cement strength, X_i are the input variables, b_0 is a constant, b_i are the coefficients, and N is the total number of input parameters.

2.3. Back propagation artificial neural network

A BP-ANN works on the basis of back propagation of an error between the target output and the input. The output to the neuron is

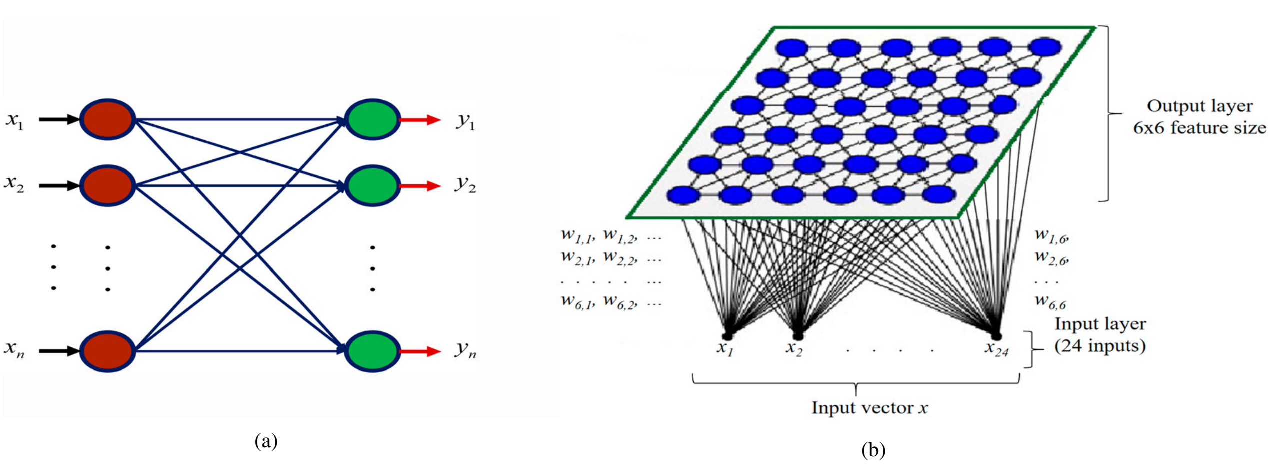

Figure 1. Structural model of Kohenan Map, (a) one-dimensional network (b) two-dimensional network.

obtained by multiplying the input of connected neuron by synaptic weights. The output of the neuron can be obtained as

$$O_{j,\text{net}} = f\left(\sum_{i=1}^{k} w_{ij}x_i + t_h\right) \tag{6}$$

where w_{ij} is the connected weight and t_h is the threshold, and k is the number of neurons in each layer. Such types of neural network work in two steps. In the first step, the data fed to the input neuron and process through the hidden layer until the network response is computed at the output layer. In the second stage, network response compared with the target output and error is generated. The error signal distributed back among the input neurons and connection weights between the neurons are updated until the network results into predefined target output. The back propagation algorithm can be used in this network, which adjusts the weights in the steepest descent direction and error converges rapidly. The error function E, which is defined by the sum of square of average between the desired output and the output of each neuron in the output layer can be calculated as,

$$E = \frac{1}{p}\sum_{p}\sum_{k} w_i g_i (d_{pk} - o_{pk})^2 \tag{7}$$

where p is the training vector index, k is the elements in output vectors, d_{pk} is the kth element of the pth desired pattern vector, and o_{pk} is the kth element of the output vector when pattern p is presented as input to the network. In recent years, many researchers have used an artificial neural network (ANN) to construct a prediction model [25,26,27,28]. However, the over fitting or the outliers in the input data is a main issue in such models. These over fitting data may result the neural network in slow learning and training. Hence, the prediction results may not be efficient and accurate.

The motivation of the proposed approach came from the fact that the cement strength prediction results mostly influenced by the input variables data set. The outlier and the redundant data of the input variables may result in inaccurate cement strength predicted values. Therefore, the traditional regression function models such as SOM, linear and nonlinear regression models and ANN-based methods may not work properly. Moreover, these models are tested on the offline dataset only. Therefore, a simple and computationally efficient cloud-based approach is proposed to handle these issues. A real-time reduced data set using the cloud model may improve the efficiency and production of the cement industry.

3. Data collection and design of prediction models

The cement preparation process depends on the various parameters which really affects the cement strength. The various input parameters may be responsible for cement strength. The input parameters responsible for cement concrete strength have been decided after the detailed discussion with experienced engineers and the personals of a cement industry. These decided input parameters are blaine, residue (at 90 or 45 Âµ), moisture, temperature, water spray, LOI, SO_3, C_3S, C_2S, C_3A, LSF, NC, Clinker liter weight, Clinker Feed temperature, GA dosage.

These input parameters used for the preparation of cement concrete and strength has been recorded for 2 days strength, 7 days strength, 28 days. The main aim of this research is to predict the cement strength for a sample of 2 days, 7 days and 28 days, respectively. These data have been collected from the same cement plant for one-year period January 2017 to December 2017. The recorded data (304 records excluding weekly and public holidays) have been

Table 1. Sample of recorded input variable data (16 input variables).

No. Days	Blaine cm2/gm (X1)	Residue 90 Âµ % (X2)	Residue 45 Âµ % (X3)	NC % (X4)	LOI % (X5)	SO3% (X6)	C3S % (X7)	C2S % (X8)	C3A % (X9)	LSF % (X10)	Moisture % (X11)	Clinker Liter weight gm/L (X12)	Clinker Feed Temp (X13)	Water Spray l/h (X14)	Cement Temp (X15)	GA Dosage g/mt (X16)
1	3236	1.4	13.4	28.8	2.52	2.72	58.46	12.46	7.347	95.772	0.19	1366	153	1000	100	496
2	3209	1.3	12.8	28.8	2.36	2.61	59.51	11.55	7.241	96.198	0.15	1337	167	1580	102	496
3	3236	1.1	11.9	28.8	2.32	2.55	59.68	11.42	7.241	96.264	0.13	1329	149	2250	101	496
4	3288	1	11.3	28.6	2.68	2.64	58.47	12.63	7.321	95.688	0.14	1319	170	1530	100	496
5	3249	1.1	11.7	28.8	2.37	2.62	58.63	12.22	7.178	95.843	0.16	1323	167	1710	99	496
6	3275	1	10.7	28.8	2.36	2.57	59.37	11.95	7.188	95.996	0.11	1303	154	2040	100	496
7	3301	1	10.3	28.8	2.45	2.62	58.42	12.78	7.241	95.595	0.14	1320	168	1960	102	496
8	3249	1.2	12.4	28.6	2.48	2.67	59.06	11.92	7.321	96.036	0.11	1308	170	1920	108	496
9	3314	1	12	28.8	2.76	2.42	59.38	11.77	7.099	96.059	0.12	1290	154	1960	108	496
304	3281	1.3	12.9	28.6	2.67	2.46	60.75	10.53	7.769	96.877	0.12	1370	132	2500	100	450

pre-processed for any missing and wrong data. Based on the collected data, the existing cement concrete strength prediction models such as SOM, linear and non-linear regression and ANN model are designed, and various experimentations are carried out to validate the effectuality of the proposed approach.

In this research, the experimental data collected from a cement industry for 12 months are shown in Table 1. It is used for analysis and prediction of the cement strength. The recorded data (304 records, excluding weekly and public holidays) have been pre-processed for any missing or wrong data. Table 2 shows the three different data for 2 days, 7 days and 28 days cements strength and the minimum and maximum value of strength in each case. Based on the collected data, the cement strength prediction models such as SOM, linear and non-linear regression and ANN model are designed and various experimentations are carried out to validate the effectuality of the proposed approach.

The data have been used for training, validation and testing. The 70% of the total data (212 records) has been considered for training, 15% data (46 records) for validation and 15% data (46 records) for testing the network. The cloud-based MATLAB programming tool has been used to develop the model. The network training, validation, testing and prediction are also done using the same programming tool.

The SOM prediction model is designed with 16 input variables (x_1,x_2, . . . ,x_{16}) at the input layer as shown in Figure 1(b). A SOM architecture with 16 input and 1 output is used to predict the cement strength. This model is implemented in MATLAB software and used for the prediction. From literature, it is observed that the SOM results can be improved further using the regression models. Therefore, linear and non-linear regression-based models are designed for the cement strength prediction application. These regression models are designed using the mathematical Equations (4) and (5), respectively. The first order linear model is given as

$$\begin{aligned} Y = {} & -0.6651 - 0.93163X_1 + 0.93604X_2 \\ & - 1.7099X_3 - 1.0078X_4 + 1.777X_5 \\ & - 0.27765X_6 + 1.181X_7 + 0.91335X_8 - 0.54676X_9 \\ & + 0.68651X_{10} - 1.5485X_{11} \\ & + 2.6379X_{12} - 3.3361X_{13} + 2.4798X_{14} \\ & - 1.0646X_{15} + 8.4849X_{16} \end{aligned} \tag{8}$$

Similarly, the second order non-linear model is given as

$$\begin{aligned} Y = {} & 17394512.14 + (-92.85 \times X_1) + (-24166.73 \times X_2) + \cdots \\ & + (1344.67 \times X_{15}) \\ & + (-295.11 \times X_{16}) + (-359.26 \times X_1 \times X_1) \\ & + (-66.20 \times X_1 \times X_2) + \cdots + (-103.20 \times X_1 \times X_{15}) \\ & + (-112.02 \times X_1 \times X_{16}) + (359.26 \times X_2 \times X_1) \\ & + (24.80 \times X_2 \times X_2) + \cdots + (-31772.76 \times X_2 \times X_{15}) \\ & + (-19380.62 \times X_2 \times X_{16}) + \\ & \vdots \end{aligned}$$

Table 2. Cement strength and its min-max value for 304 days.

	Range of cement strength										
Types	1	2	3	4	5	6	7	8	304	Min	Max
2 days	23.56	23.63	23.82	23.41	23.62	24.26	23.94	23.51	24.45	20.08	28.87
7 days	38.75	38.38	38.94	38.45	38.85	39.12	38.36	39.2	40.95	35.38	45.06
28 days	51.33	51.53	52.18	52.01	51.82	52.97	51.54	52.6	53.88	50.69	60.34

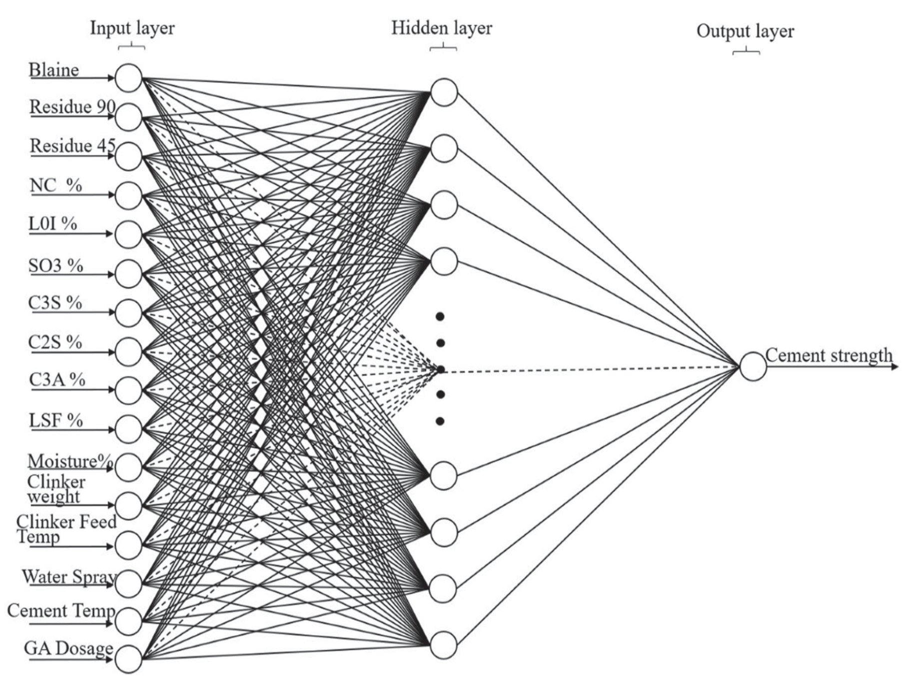

Figure 2. The architecture of ANN used for cement strength prediction.

$$(112.02 \times X_{16} \times X_1) + (19382.47 \times X_{16} \times X_2) + \cdots + (0.16 \times X_{16} \times X_{16}) \quad (9)$$

In the literature, it is stated that the ANN outperformed the SOM as well as linear and non-linear models. Therefore, an artificial neural network-based prediction model is considered for the experimentations. The neural network architecture depends on the input variable of the data set used in the experiment and the number of outputs required for the observations. In this research work, 16 major input variables are considered in the preparation of the cement concrete in the industry. Moreover, the main task is to predict the concrete strength only. In feed-forward neural network architecture the number of input and output neurons are decided based on the variables. Therefore, a simple ANN architecture is proposed which has 16-input neurons and 1-output neuron which is shown in Figure 2. Moreover, the number of hidden layers may be selected randomly and decided based on the final performance of the system.

4. Proposed approach for cement strength prediction

In [22], a method is proposed to determine the features of the system based on the common eigenvalues using Hessenberg matrix reduction method. However, it is observed that the calculation of Hessenberg matrices with specified conditions is more complicated. To overcome these issues, Gaidhane et al. [29] proposed an approach based on the concept of the companion matrix for feature extraction and employed successfully in face recognition and PCB surface defect detection applications. Moreover, a polynomial coefficient and spectral radius-based technique is used for image sharpness measure [30]. From literature, it is observed that the concept of covariance matrix along with polynomial coefficient can be used for feature extraction of the data [21,29,30]. After collection of cement strength data set, it is observed that a numerous redundant data is collected which can be removed easily from the collected data. Therefore, based on the concept of the covariance matrix and polynomial coefficients, a simple and computationally efficient approach is proposed for cement strength prediction. In this, covariance matrix plays an important role in data reduction. It helps to form a reduced dimension 2D square matrix. Moreover, the normalization of such 2D square matrix results in less variation in the data set values.

Consider the data set X for different input variables and days. The data set is arranged in n columns (number of input parameters) and m rows (number of samples). In proposed method, each row is then converted into 2D matrix x_i order of $p \times q$. Suppose $\mathrm{R}(x,y)$ is a 2D matrix such that $\mathrm{R}(x,y) \in \Re^{n\times n}$, then their characteristics polynomial can be represented as $\det(\lambda I - A)$, where A is a non-negative $n \times n$ definite square matrix. Here, matrix A is also called covariance matrix, obtained from 2D matrix x_i. Then the term $\det(\lambda I - A)$ can be computed as

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda - a_{11} & a_{12} & a_{13} & \cdots & c_{1n} \\ a_{21} & \lambda - a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & \lambda - a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \lambda - a_{nn} \end{vmatrix} \quad (10)$$

For Equation (10), the characteristics polynomial $p(A)$ can be written as

$$\mathrm{P}(A) = \lambda^n - a_n\lambda^{n-1} - a_{n-1}\lambda^{n-2} - \cdots - a_1 \quad (11)$$

where $a_i \forall\, i = 1, 2, 3, \ldots, n$ are the coefficients of polynomial $p(A)$. These coefficients of polynomial can be obtained as

$$\begin{aligned} a_n &= \sum_{i=1}^{n} a_{ii} \\ a_{n-1} &= \sum_{i<j} \begin{vmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{vmatrix} \\ a_{n-2} &= \sum_{i<j<k} \begin{vmatrix} a_{ii} & a_{ij} & a_{ik} \\ a_{ji} & a_{jj} & a_{jk} \\ a_{ki} & a_{kj} & a_{kk} \end{vmatrix} \\ &\vdots \\ a_1 &= det\,(A) \end{aligned} \quad (12)$$

Thus, $a_1, a_2, a_3, a_4, \ldots, a_n$ are the polynomial coefficients represent the features of the matrix A_i. These mathematically calculated features are used as an input to the neurons at the input layer. The polynomial coefficients preserve the significant information of the collected data set without loss of originality. A data set of size $n \times n$ is now reduced to n only. Now the processing of any prediction model with n samples become simpler than the approach with whole data samples. Therefore, based on the above mathematical preliminaries, a new approach for cement strength has been proposed. It is summarized in Algorithm 1.

Algorithm 1: Proposed approach for cement strength prediction

Step 1: Given the collected data from cement industry dataset X
Step 2: Convert dataset X into 2D matrix x_i of order $p \times q$
Step 3: Calculate the covariance matrix A of the size of $n \times n$
Step 4: Normalized the A by its energy defining as

$$A_P(x,\ y) = \frac{A(x,y)}{\sqrt{\sum_{x,y=0}^{N-1} [A(x,y)]^2}} \text{ such that } \sum_{x,y=0}^{N-1} [A(x,y)]^2 = 1$$

Step 5: Calculate the extracted features using characteristics polynomial $p(A)$ for each input variable using Eq. 11.
Step 6: Set prediction model (SOM, linear, non-linear and ANN) and apply extracted feature to the input of each model.
Step 7: Train the models and calculate the E using Eq. 7.
Step 8: Compare the performance index R and $R^2$for each model.

Based on the above mathematical proposed algorithm, the collected concrete cement strength data set has been reduced to $n$samples. Further, the reduced data set then is used for prediction of concrete cement strength using the various well-known prediction models such as SOM, linear and non-linear regression models and artificial neural network. It is observed that the proposed polynomial coefficient approach removes the redundant data from the input data sets. Thus, the over-fitting and redundant data issue can be avoided which can result in faster learning, training and accurate prediction results.

5. Experimental results and discussion

A cloud server has been set up with the following details to host the high-end software and web application. The server hardware is a Dell EMC machine with two processors having 3.6 Ghz frequencies and 16 cores each, 256 GB RAM with 2 v100 GPUs. The server platform is a windows server family operating server (OS) (windows server 2012 R2), it is used as host OS. The type-1 hypervisor is used on top of the host OS to virtualize the server for multiple virtual server entities

Table 3. The relationship between the number of hidden nodes and RMSE for ANN architecture.

	Number of hidden nodes										
Data type	1	2	4	6	8	10	12	14	16	18	20
2 days	1.517	0.856	0.527	0.403	**0.369**	0.590	0.429	0.669	0.758	0.964	1.140
7 days	0.825	0.694	0.521	0.490	**0.355**	0.483	0.448	0.694	0.559	0.889	0.935
28 days	0.829	0.624	0.525	0.466	**0.352**	0.394	0.425	0.508	0.745	0.799	0.869

where softwares or web applications are hosted. The internet information services are enabled for this server and connected with a 10 GBPS speed of Ethernet. This server setup is used to host MATLAB R2017a software and web application for cement strength prediction.

5.1. Performance indexes

In this section, the performance of the proposed approach is tested by performing a series of experiments on real-time data obtained from the cement industry. The performance analysis has been carried out using the performance indexes, root mean square error (RMSE) and R-square coefficient (R^2). The RMSE is the square root of the mean square error. The RMSE defines the average distance of a data point from the fitted line measured. Ideally the value RMSE should be close to zero. Mathematically, it can be defined as

$$R = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(O_i - t_i)^2} \tag{13}$$

where t_i is the strength of cement, O_i is the predicted strength output and $M$is the total number of data points considered in the experimentation.

Another performance index R-square coefficient (R^2) is a measure of how efficiently the independent variables considered account for the dependent variables. In ordered to calculate the goodness of fit of the models, it is generally common to calculate the R-square coefficient which can represent the percentage of variance in the data set. It compares the predicted output with the observed values. It can be represented as

$$R^2 = \frac{\frac{1}{M}\sum_{i=1}^{M}(O_i - t_i)^2}{\frac{1}{M}\sum_{i=1}^{M}(O_i)^2} \tag{14}$$

The value of R-square coefficient should be closed to 1 for matching of the predicted and original data.

5.2. Computation of hidden nodes

After a detailed study and extensive analysis, it is observed that the number of hidden nodes in neural network architecture is important. The number of these nodes depends on the number of input neurons. It is reported in literature that the hidden nodes can vary from $\boldsymbol{n}$ to $n+1$,where $\boldsymbol{n}$ represents the number of input neurons. Therefore, the various experimentations are carried out to calculate the number of hidden nodes for a minimum value of RMSE. The relationship between the number of input neurons, hidden nodes, output neurons and the root mean square error (RMSE) is summarized in Table 3. It is observed from Table 3 that the training error (RMSE) decreased as the number of hidden nodes increased. However, it is not always true for more hidden nodes. In this experiment, when the number of hidden nodes increased above 8 the RMSE also increased again. Therefore, in this paper an ANN architecture with 16 inputs, 8 hidden nodes and 1 output (16-8-1 architecture) has been exploited for the various experimentations.

5.3. Performance analysis of existing methods

Based on the above performance indices, the existing approaches such as SOM, linear, nonlinear and ANN are tested on the real-time data set obtained from the cement industry. A SOM architecture as shown in Figure 1(b) is used with 16 input variables (x_1,x_2, ... ,x_{16}).The various experimentations are carried out on the cement strength data set (2, 7 and 28 days sample) and the performance of the SOM in the form of RMSE and R-square coefficient(R^2)is obtained. The value of RMSE is found to be 0.587, 0.589 and 0.594 for 2, 7 and 28 days samples, respectively, Moreover, the R-square coefficient is also calculated which is 0.494, 0.498 and 0.519 for 2, 7 and 28 days samples, respectively.

Further, the linear regression model is trained with 16 inputs as defined in Equation (8). Moreover, a nonlinear regression model is trained as represented by Equation (9) with 16 input variables. The performance of the linear and nonlinear regression models in the form of RMSE and R-square coefficient(R^2)is obtained and results are summarized in Table 4.

Finally, the fourth existing method based on ANN is simulated with 16 inputs (x_1,x_2, ... ,x_{16}), 8-hidden layers and 1-output neuron as architecture shown in Figure 2. The various experimentations are carried out to decide the number of hidden layer and finally observed the 8 layers for best results. After extensive analysis and simulations, the performance of the ANN in the form of RMSE and R-square coefficient(R^2)is summarized in Table 4. It is observed from Table 4 that the RMSE using ANN is 0.378, 0.357 and 0.351 for 2 days, 7 days and 28 days data set, respectively. The RMSE value for ANN is lower than the SOM, linear and nonlinear which indicate the better results of ANN prediction model. Similarly, the R-square coefficient (R^2) for ANN approach is 0.616, 0.651 and 0.662 for 2, 7 and 28 days respectively, which is higher than SOM, linear and nonlinear techniques. Ideally, the value of RMSE should be closed to zero and (R^2) approached to unity. However, it is observed that the RMSE value is higher and closed to 1 where as the value of R-square coefficient is away from unity. Such type of results may give the false cement strength prediction. Moreover, in these methods the entire data has been used which is consisting of large redundant data. To overcome this issue, a simple and efficient polynomial coefficient-based approach is proposed. The proposed approach helps to remove the redundant data from the collected data set and thus, improve the performance of the system.

5.4. Performance analysis of proposed approach

In this experiment, the various experimentations are carried out using the polynomial coefficients-based proposed approach. The cement data set collected for 2 days, 7 days and 28 days is represented in 2D matrix 16 × 304 size. Then the polynomial coefficients are calculated using the Algorithm 1. These polynomial coefficients-based input data set are then applied to the designed prediction models. These models are defined as PC-SOM, PC-linear, PC-nonlinear and PC-ANN. The performance parameters RMSE and R-square coefficient (R^2)is recorded in the Table 5. It is observed from Table 5 that the proposed approach (PC-ANN) performs better as compared to PC-SOM, PC-linear and PC-nonlinear approaches. The RMSE using

Table 4. RMSE and (R^2) coefficient for existing prediction method.

	SOM		Linear regression		Non-linear regression		ANN	
Data type	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
2 Days	0.587	0.494	0.442	0.548	0.425	0.564	0.378	0.616
7 Days	0.589	0.498	0.424	0.571	0.415	0.597	0.357	0.651
28 Days	0.594	0.519	0.414	0.583	0.409	0.602	0.351	0.662

Table 5. RMSE and (R^2) coefficient for proposed prediction models.

	PC-SOM		PC-Linear regression		PC-Non-linear regression		PC-ANN	
Data type	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
2 Days	0.437	0.676	0.391	0.713	0.380	0.737	0.269	0.874
7 Days	0.402	0.685	0.382	0.719	0.371	0.748	0.255	0.879
28 Days	0.397	0.693	0.376	0.721	0.364	0.752	0.252	0.883

the proposed approach is less as compared to the other prediction models.

It is observed that the RMSE has been reduced to 0.269, 0.255 and 0.250 for 2, 7 and 28 days, respectively. Moreover, the R-square coefficient (R^2)for a proposed approach (PC-ANN) is higher and close to 1 (0.874, 0.879 and 0.883 for 2, 7 and 28 days, respectively). Thus, the proposed approach (PC-ANN) performs better than the existing PC-SOM, PC-linear and PC-nonlinear prediction algorithm.

Further, the cement strength predicted by the SOM, linear, non-linear regression, ANN models and proposed approach for 28 days data set is compared with the experimental cement strength and results are shown in Figure 3.

In the first case, the prediction is carried out using the SOM and PC-SOM and the predicted strength is compared with the experimental cement strength as shown in Figure 3(a). It is observed from Figure 3(a) that the experimental cement strength value is slightly

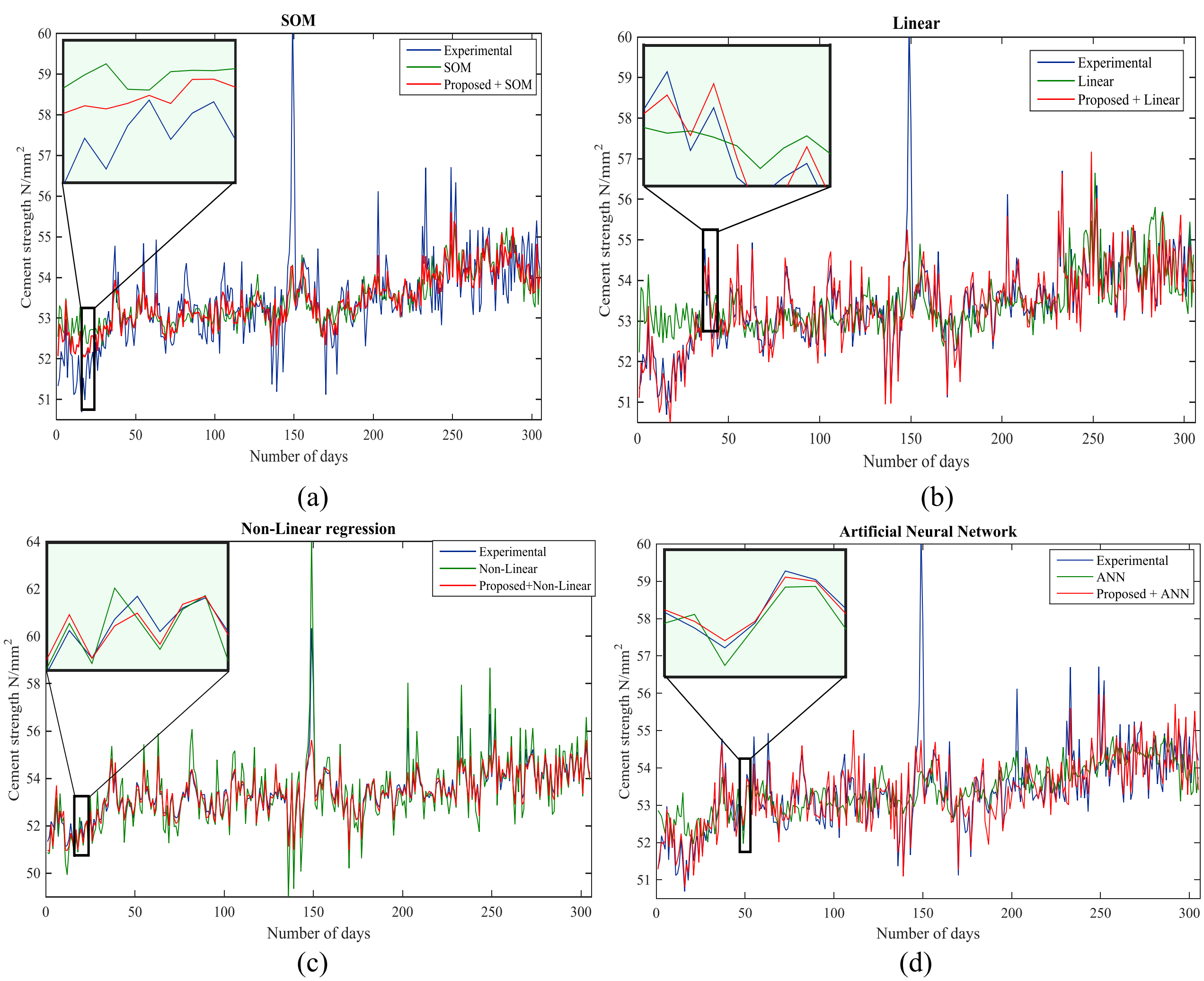

Figure 3. Comparison of the experimental and predicted values of cement strength using existing and proposed approach. (a) SOM, (b) linear regression model, (c) nonlinear regression model and (d) ANN model.

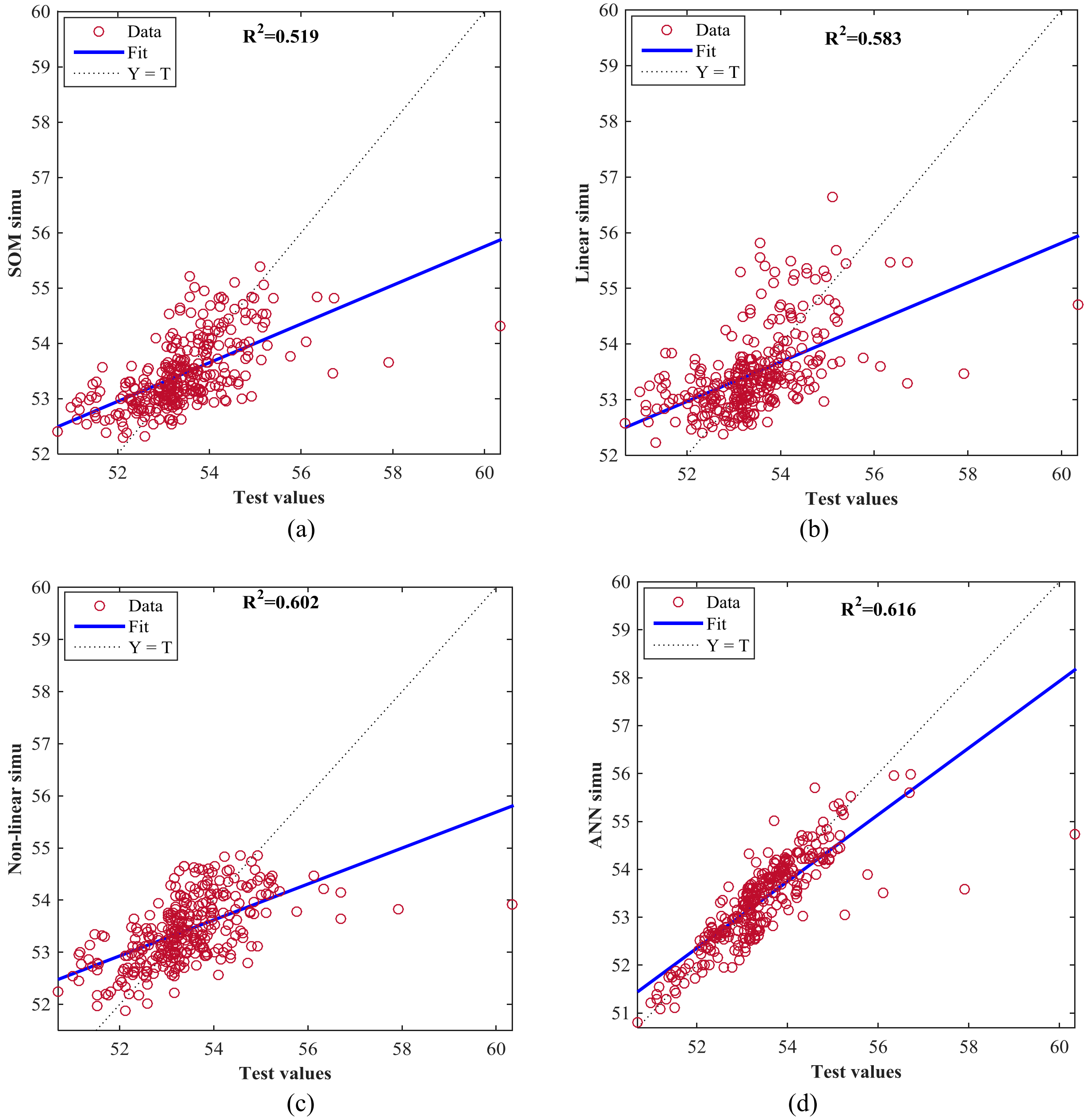

Figure 4. Comparison of *R*-square coefficient. (a) SOM, (b) linear regression model, (c) nonlinear regression model and (d) ANN model.

fitted with PC-SOM model. In the second case, the predicted strength obtained using the linear and proposed linear regression model is drawn in Figure 3(b).

Similarly, the experimentations are carried for proposed nonlinear and ANN models and the results are presented in Figure 3(c) and Figure 3(d), respectively. It is observed from Figure 3(d) that the cement strength predicted by the proposed (PC-ANN) model is very much fitting on the experimental data as compared to the linear and nonlinear regression models. From above experimentations, it is observed that the proposed ANN prediction model outperform the SOM, linear and nonlinear regression prediction models.

5.5. Training performance

In this experiment, effectuality of the proposed approach is tested using the training performance of the various models. The *R*-square coefficient (R^2) defines the goodness of fit of the models. It is generally used to calculate the *R*-square coefficient that represents the percentage of variance in the data set. In the first experiment, the test is carried out using the predicted and experimental data for SOM, linear, nonlinear and ANN prediction model. In this, the original cement strength data set is used to test the training performance of all four prediction models. The data fit training results are shown in Figure 4(a–d). It is well known that the *R*-square coefficient should be closed to 1 for best fit of the data set. Figure 4(a) shows the training performance for SOM model. It is observed from the Figure 4(a) that the data is mostly scattered and away from the data fit line and value of $R^2 = 0.519$. Further, Figure 4(b,c) represents the training performance for the linear and nonlinear regression models. It is observed from Figure 4(b,c) that the linear and nonlinear regression model fit the data better as the value of $R^2 = 0.583$ and $R^2 = 0.602$, respectively, more closed to 1 as compared to SOM model. The training performance of the Moreover, ANN model has the best fit among all

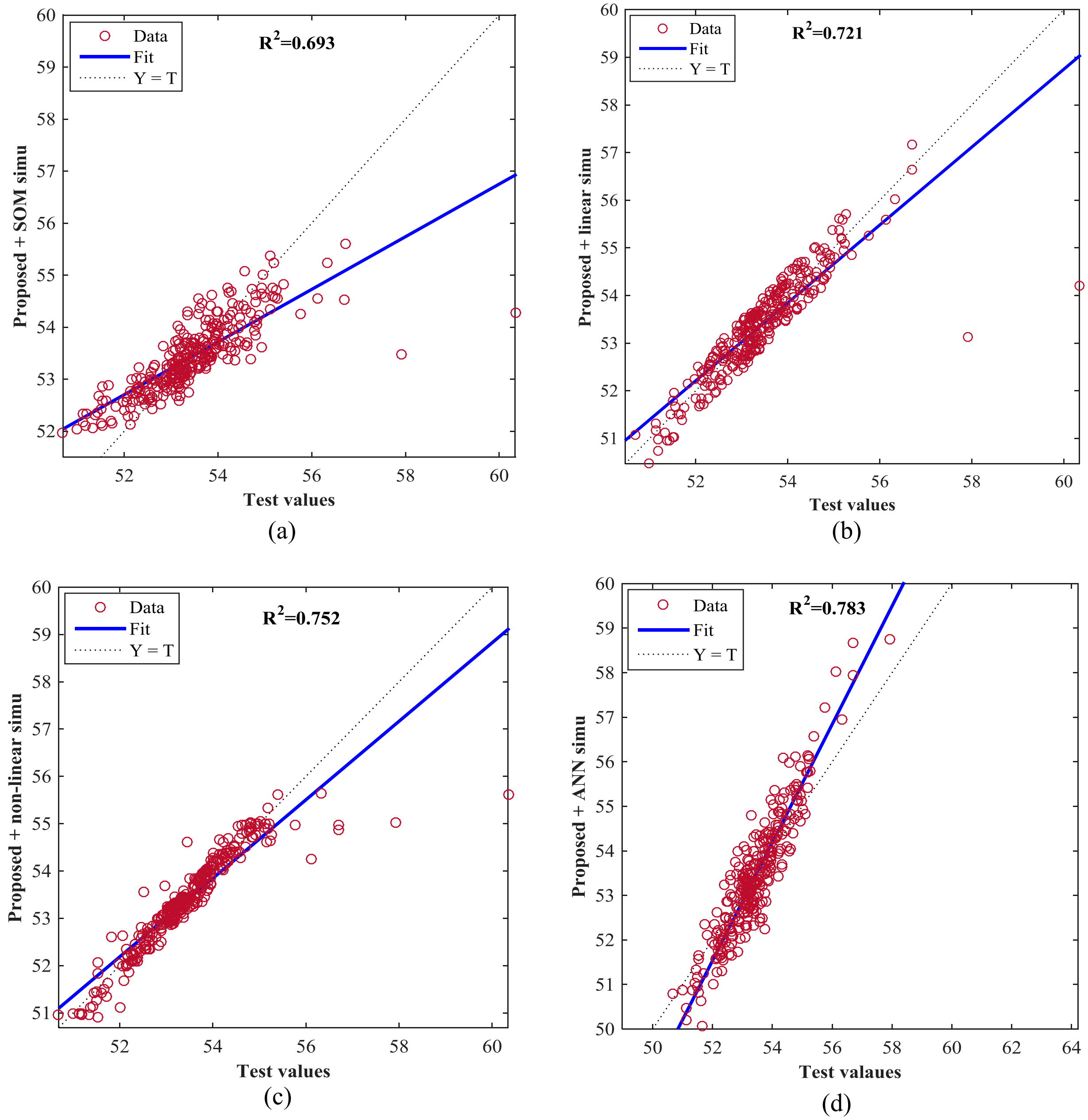

Figure 5. Comparison of *R*-square coefficient. (a) PC-SOM, (b) PC-linear regression model, (c) PC-nonlinear regression model and (d) PC-ANN model.

the models($R^2 = 0.616$). It is observed from Figure 4 that the results can be improved further using the data reduction approach.

Further, the training test is performed using the polynomial coefficient-based approach and the *R*-square coefficient results are shown in Figure 5(a–d). It is observed from Figure 5(d) that the predicted cement strength data is well fitted for the proposed-ANN approach as compared to the propose-SOM, proposed-linear and proposed-nonlinear existing approaches. The value of data fitting shown by *R*-square coefficient is improved and become 0.721 which is higher than the other models.

In order to evaluate the performance of the polynomial coefficient-based proposed approach in predicting the cement strength, PC-ANN results are compared with the results of PC-SOM, PC-linear and PC-nonlinear models. The predicted data using these models is shown in Figure 6. It is observed from Figure 6 that the predicted cement strength data using the PC-SOM is scattered more and diverted from the actual experimental data set. The cement strength predicted using the PC-linear and PC-nonlinear approach is closer to experimental strength values. Moreover, the prediction using the PC-ANN has shown the more effectiveness of the proposed algorithm and the predicted data is less scattered as compared to the other methods. It has been found that software-based application for the prediction of the cement strength of different composition of ingredients and time durations can reduce the time and cost of production. The screenshot of the developed software-based cement strength prediction application is shown in Figure 7.

5.6. Computational time

Computational complexity is one of the important evaluation parameters for any algorithm. Therefore, in this, the computational time of existing methods is calculated and compared with the computational time of proposed approach. Table 6 shows the execution

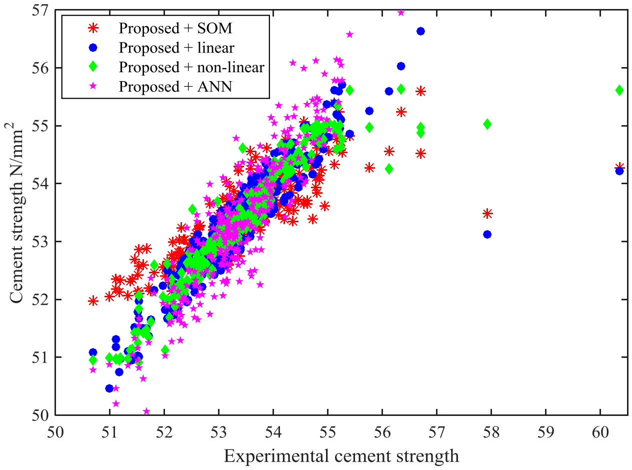

Figure 6. Comparison of predicted cement strength using different proposed models.

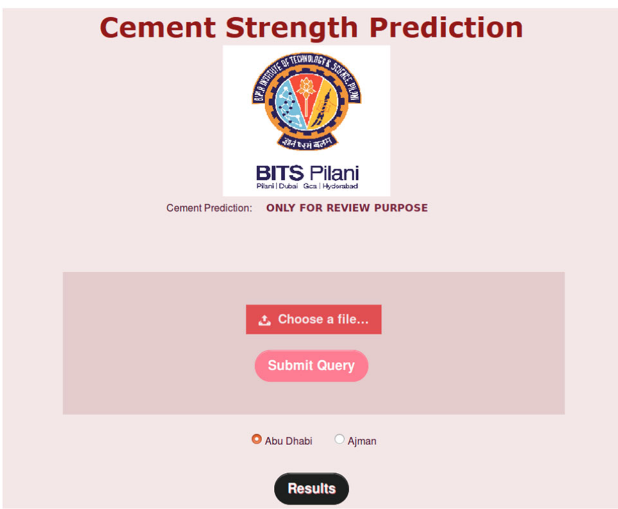

Figure 7. Screenshot of developed application for cement strength prediction.

time of the experimental data for existing models as well as proposed models. It is observed from Table 6 that the time in seconds taken by the ANN-based model is little more than SOM and multi-linear regression models. It is due to the iterative nature of the error optimization algorithm. However, the Proposed PC-ANN approach requires less time (2.09871) as compared to ANN model (2.19405). It is due to the fact that the input data set has been reduced to 75% using the proposed polynomial coefficient-based data reduction approach. However, the more computational time for PC-ANN (2.09871) compared to PC-SOM (1.65836) and PC-linear (1.35865), PC-non-linear (0.87035) regression models can be compensated by the effectiveness in cement concrete strength prediction.

6. Conclusion

In this research paper, a simple and computationally efficient approach is proposed to estimate the cement strength. Since, proposed approach is based on the covariance matrix concept, it can be easily used for the feature extraction. In this approach, the polynomial coefficients obtained from the covariance matrix represent the features of cement strength data set. The effectuality and feasibility of the proposed approach is demonstrated on three different cement strength data sets collected from the cement industry for 2 days, 7 days and 28 days. Based on the 16 input parameters of cement strength, the different prediction models such as SOM, linear, nonlinear and ANN models are designed and various experimentations are carried out to evaluate the presented approach. It gives the better performance in the terms of RMSE, R-square coefficients and computational time in comparison to the existing methods. It is concluded that the proposed polynomial coefficient-based ANN model predicts the cement strength more accurately compared to the SOM, multiple linear regression and artificial neural network model.

Notes on contributors

Vilas H. Gaidhane received his B.E. degree in 2000 from the Department of Electronics, Nagpur University, Nagpur and the M. Tech. degree in 2010 from the Department of VLSI Design, UPTU Lucknow, India. He also received Ph. D degree in Instrumentation and Control Engineering, University of Delhi in 2013. He is currently working as an Assistant Professor in Electrical and Electronics Engineering Department, Birla Institute of Technology and Science (BITS), Pilani, Dubai Campus, Dubai, UAE. His research interest includes image processing, pattern recognition, and computer vision, control system, and Microelectronics.

Nand Kumar received his M.Sc. and M.E. degree from Birla Institute of Technology and Science, Pilani, Pilani Campus, India in 2005 and 2009, respectively. He is currently perusing PhD and working as Lecturer in the computer science department of Birla Institute of Technology and Science, Pilani, Dubai Campus, DIAC, Dubai, UAE. His research interests include Cloud Setup, Cloud security, Database efficiency, Adaptive website designing.
Email: nandkumar@dubai.bits-pilani.ac.in

Ravi Kant Mittal received his B.E. and M.E. degree in Mechanical Engineering from Birla Institute of Technology and Science, Pilani, Pilani Campus, India in 1973 and 1975, respectively. He received his Ph.D. degree in software engineering as well from the same institute in year 1990. He is currently Senior Professor and Director Academics with Sree Guru Gobind Singh Tricentenary University, Gurugram, India. He has held several administrative positions at BITS Pilani in last 40 years. His research area includes software testing, robust robot design, robot path planning, MEMS, nanotechnology and e-waste.
Email: rkm.bits@gmail.com

Table 6. Comparison of the computational time (seconds) of the proposed models and existing cement strength prediction models.

	Existing methods				Proposed Approach			
Data type	SOM	Linear	Nonlinear	ANN	PC-SOM	PC-linear	PC-Nonlinear	PC-ANN
2 Days	1.40936	1.30966	0.84754	2.01089	1.35811	1.25856	0.82982	2.00996
7 Days	1.53356	1.33320	0.85991	2.02558	1.41121	1.31149	0.83360	2.01702
28 Days	1.74965	1.64960	0.88884	2.19405	1.65836	1.35865	0.87035	2.09871

J. Rajevenceltha received the B.E. degree from the Department of Biomedical Engineering, PSG College of Technology in 2014 and the Master of Engineering degree from the Department of Electronics and Communication engineering, Amrita School of Engineering in 2016. Currently, she is research scholar in the Department of Electrical and Electronics Engineering, Birla Institute of Technology and Sciences (BITS), Pilani, Dubai campus, UAE. Her research interests include image and signal processing, machine learning, pattern recognition and computer vision. Email: vilasgla98@gmail.com

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Vilas H. Gaidhane http://orcid.org/0000-0001-7174-2739
Nand Kumar http://orcid.org/0000-0001-7747-0586
J. Rajevenceltha http://orcid.org/0000-0002-2590-9873

References

[1] Kheder GF, Al Gabban AM, Abid SM. Mathematical model for the prediction of cement compressive strength at the ages of 7 and 28 days within 24 hours. Mater Struct. 2003;36(10):693–701.
[2] Baykasoğlu A, Dereli T, Tanış S. Prediction of cement strength using soft computing techniques. Cem Concr Res. 2004;34(11):2083–2090.
[3] Zelić J, Rušić D, Krstulović R. A mathematical model for prediction of compressive strength in cement–silica fume blends. Cemand Concr Res. 2004;34(12):2319–2328.
[4] Tsivilis S, Parissakis G. A mathematical model for the prediction of cement strength. Cemand Concr Res. 1995;25(1):9–14.
[5] Nikoo M, Zarfam P, Sayahpour H. Determination of compressive strength of concrete using self-organization feature map (SOFM). Eng Comput. 2015;31(1):113–121.
[6] Sadowski L, Nikoo M, Nikoo M. Principal component analysis combined with a self organization feature map to determine the pull-off adhesion between concrete layers. Constr Build Mater. 2015;78(1):386–396.
[7] Yeh I-C. Modeling of strength of high-performance concrete using artificial neural networks. Cem Concr Res. 1998;28(12):1797–1808.
[8] Wei Y, Xue Y, Yin J, et al. Prediction of municipal solid waste generation in China by multiple linear regression method. Int J Comput Appl. 2013;35(3):136–140.
[9] Ozcan F, Atis CD, Karahan O, et al. Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete. Adv Eng Softw. 2009;40(9):856–863.
[10] Sobhani J, Najimi M, Pourkhorshidi AR, et al. Prediction of the compressive strength of no-slump concrete: a comparative study of regression, neural network and ANFIS models. Constr Build Mater. 2010;24(5): 709–718.
[11] Demir FA. New way of prediction elastic modulus of normal and high strength concrete-fuzzy logic. Cem Concr Res. 2005;35(8):1531–1538.
[12] Emiroğlu M, Beycioğlu A, Yildiz S. ANFIS and statistical based approach to prediction the peak pressure load of concrete pipes including glass fiber. Exp Syst Appl. 2012;39(3):2877–2883.
[13] Tesfamariam S, Najjaran H. Adaptive network–fuzzy inferencing to estimate concrete strength using mix design. J Mater Civil Eng. 2007;19(7): 550–560.
[14] Bilgehan M. A comparative study for the concrete compressive strength estimation using neural network and neuro-fuzzy modelling approaches. Nondestruct Test Eval. 2011;26(1):35–55.
[15] Yuan Z, Wang LN, Ji X. Prediction of concrete compressive strength: research on hybrid models genetic based algorithms and ANFIS. Adv Eng Softw. 2014;67:156–163.
[16] Rebouh R, Boukhatem B, Ghrici M, et al. A practical hybrid NNGA system for predicting the compressive strength of concrete containing natural pozzolan using an evolutionary structure. Constr Build Mater. 2017;149:778–789.
[17] Umrao RK, Sharma LK, Singh R, et al. Determination of strength and modulus of elasticity of heterogenous sedimentary rocks: An ANFIS predictive technique. Measurement. 2018;126:194–201.
[18] Fu Z, Mo J, Chen L, et al. Using genetic algorithm-back propagation neural network prediction and finite-element model simulation to optimize the process of multiple-step incremental air-bending forming of sheet metal. Mater Des. 2010;31(1):267–277.
[19] Lin H, Zhao J, Sun L, et al. Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis. Innov Food Sci Emerg. 2011;12(2):182–186.
[20] Tan EM, Li QQ, Jiang JG. Weighted cellular automata design for test pattern generator based on genetic algorithm. Appl Mech Mater. 2013;273: 840–844.
[21] Gaidhane VH, Hote YV, Singh V. An efficient approach for face recognition based on common eigenvalues. Pattern Recognit. 2014;47(5):1869–1879.
[22] Kohonen T. Self-organization and associative memory. Berlin: Springer; 2012.
[23] Chandwani V, Agrawal V, Nagar R. Applications of artificial neural networks in modeling compressive strength of concrete: a state of the art review. Int J Curr Eng Technol. 2014;4(4):2949–2956.
[24] Thirumalai C, Chandhini SA, Vaishnavi M. Analysing the concrete compressive strength using Pearson and Spearman. Proceedings of the International Conference Electronics Communication and Aerospace Technology, Vol. 2; 2017. p. 215–218.
[25] AbdElaty M. Comprehensive strength prediction of Portland cement concrete with age using a new modal. HBRC J. 2014;10(2):145–155.
[26] Chopra P, Sharma RK, Kumar M. Prediction of compressive strength of concrete using artificial neural network and genetic programming. Adv Mater Sci Eng. 2016;7648467:1–10.
[27] Kosmatka SH, Kerkhoff B, Panarese WC. High-Performance Concrete,Design and control of concrete mixtures. Skokie: Portland Cement Association; 2002.
[28] Shetty MS. Concrete Techonology Theory and Practice. New Delhi: S. Chand & Company Ltd; 2006. Chapter 7.
[29] Gaidhane VH, Hote YV, Singh V. An efficient similarity measure approach for PCB surface defect detection. Pattern Anal Appl. 2018;21(1):277–289.
[30] Gaidhane VH, Hote YV, Singh V. Image focus measure based on polynomial coefficients and spectral radius. Signal Image Video Process. 2015;9(1):203–211.