Computers and Concrete, *Vol. 21, No. 4 (2018) 463-470*
DOI: https://doi.org/10.12989/cac.2018.21.4.463

Comparison of machine learning techniques to predict compressive strength of concrete

Susom Dutta[*1], Pijush Samui[2a] and Dookie Kim[3b]

[1]School of Civil and Chemical Engineering (SCALE), VIT University, Vellore, Tamil Nadu 632014, India
[2]Department of Civil Engineering, NIT Patna, Patna 800005, Bihar, India
[3]Department of Civil Engineering, Kunsan National University, Kunsan, Jeonbuk, South Korea

(Received September 21, 2017, Revised December 26, 2017, Accepted January 16, 2018)

Abstract. In the present study, soft computing i.e., machine learning techniques and regression models algorithms have earned much importance for the prediction of the various parameters in different fields of science and engineering. This paper depicts that how regression models can be implemented for the prediction of compressive strength of concrete. Three models are taken into consideration for this; they are Gaussian Process for Regression (GPR), Multi Adaptive Regression Spline (MARS) and Minimax Probability Machine Regression (MPMR). Contents of cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate and age in days have been taken as inputs and compressive strength as output for GPR, MARS and MPMR models. A comparatively large set of data including 1030 normalized previously published results which were obtained from experiments were utilized. Here, a comparison is made between the results obtained from all the above mentioned models and the model which provides the best fit is established. The experimental results manifest that proposed models are robust for determination of compressive strength of concrete.

Keywords: concrete; compressive strength; Gaussian Process for Regression (GPR); Multi Adaptive Regression Spline (MARS); Minimax Probability Machine Regression (MPMR)

1. Introduction

Concrete is composed mainly of cement primarily Portland cement, water, aggregate, and chemical admixtures. Concrete is a versatile material that can be easily mixed to meet a variety of special needs and casted into virtually any shape. Hydration is a chemical process due to which concrete solidifies and hardens after mixing with water and placement. The water reacts with the cement, which bonds with the other components together, eventually creating a stone-like material. In case of concrete mix design and quality control, the uniaxial compressive strength of concrete is considered as the most crucial property which is determined by number of factors. Several factors affect the concrete mix design like to educe a concrete as High Performance Concrete, it should possess, in addition to good strength, several other favorable properties. To obtain good workability, it requires special additives in the concrete, along with a superplasticizer as the water/cement (*w/c*) ratio in the concrete is lower than normal concrete. Usually special cements are also involved. The nature of aggregate plays an important role to incur high strength. The gradation of the aggregates determines the workability. Also, the order in which the materials are mixed is also important for the workability of the concrete. From engineering point of view, strength is the most important property of structural concrete. The characteristics of the coarse aggregate, fine aggregate, mortar and the interface determines the strength of the concrete. Properties of concrete are influenced by the properties of each and every constituent added in it. For example, for the same quality mortar, different types of coarse aggregate with different shape, texture, mineralogy, and strength may result in different concrete strengths. The tests for compressive strength are generally carried out at about 7 or 28 days from the day when the concrete is casted. Generally, strength after 28-days is standard and therefore essential and if required strength for other ages can be carried out. Accidentally, if there is some experimental error in designing the mix, the test results will fall short of required strength, the entire process of concrete design has to be repeated which may be a costly and time consuming. The same applies to all types of concrete, i.e. normal concrete, self-compacting concrete, ready mixed concrete, etc. It is well acknowledged that prediction of the compressive strength of concrete is most important in modern concrete designing and in taking engineering decisions.

In the last decade due to the importance of the research topic Numerous Studies were concentrated on many linear and nonlinear regression equations. Modelling by using artificial intelligence (AI) has been a very active research area. According to previous studies, Although, AI

∗Corresponding author, Undergraduate Student
E-mail: susomdutta7@gmail.com
[a]Associate Professor
E-mail: pijush.phd@gmail.com
[b]Professor
E-mail: kim2kie@gmail.com

http://www.techno-press.org/?journal=cac&subpage=8
ISSN: 1598-8198 (Print), 1598-818X (Online)

techniques have proved their superior capability over traditional modelling methods and so Artificial Neural Network (ANN) was one the successful choice that used for prediction problems, it has the some following limitations (Navarro and Bennun 2014).

ANN does not provide information about the relative significance of the various parameters. A common criticism of neural networks is that they require a large diversity of training for operation. The knowledge acquired during the training of the model is stored in an implicit manner and hence it is hard to come up with reasonable interpretation of the overall structure of the network. In addition, ANN has some intrinsic disadvantages such as slow convergence speed, less generalizing performance, arriving at local minimum and over-fitting problems.

So to overcome the above limitations, we adopted Gaussian Process for Regression (GPR), Multivariate Adaptive Regression Spline (MARS), Minimax Probability Machine Regression (MPMR) for prediction of compressive strength of concrete. These methods have been used earlier individually for predicting the compressive strength of concrete (Razavi *et al.* 2012, Nedushan 2012, Ozturk and Turan 2012, Cheng and Cao 2014, Chou *et al.* 2015, Samui *et al.* 2015) but a comparison has never been made. A Gaussian process is an accumulation of random variables, any finite number of which has a joint Gaussian distribution. Gaussian Process is completely defined by a mean function and a positive definite covariance function. GPR is a non-parametric regression model (Rasmussen *et al.* 2005). For any input variables, GPR defines a Gaussian distribution over the output value. It is employed in many fields such as Noise Heart Rate Data (Stegle *et al.* 2008), nonstationary time series prediction (Belhouari and Bermak 2004), multivariate spectroscopic calibration (Bermak *et al.* 2004), single image super-resolution (He and Siu 2011). MARS is a flexible, more accurate, and faster simulation method for both regression and classification problems (Friedman 1991). It is capable of fitting complex, nonlinear relationships between output and input variables. Some examples of its usage are Biological conservations (Kandel *et al.* 2015), Ecological Modeling (Pickens *et al.* 2014), Transportation (Sun *et al.* 2013). MPMR is developed based on the concept of minimax probability machine classification (Strohmann and Grudic 2002). It does not assume any data distribution. MPMR is used in various fields like medical diagnosis (Huang *et al.* 2006), Porous membrane reactor (Trianto and Kokugan 2002).

2. Dataset employed

In all about 1200 concrete samples from the investigations were evaluated. During the evaluation, some of the concrete samples were deleted from the data due to larger size aggregates (larger than 20 mm), special curing conditions, etc. About 1030 concrete samples made with ordinary Portland cement and cured under normal conditions were evaluated. Different studies used specimens of different sizes and shapes. All of these specimen types were converted into 15-cm cylinders through accepted guidelines (IS 516 1959, GB 50205 2001). Each input feature is described using only a single term which actually represent a variety of forms. For example, a cement can be powdered to various degrees of finenesses and composed of several different chemical compositions. Apart from the component types, the properties of concrete are influenced by the mixing proportions and by the mixing preparation technique. Therefore, in this approach, the compressive strength of concrete is a function of the following eight input features: contents of cement (c), blast furnace slag (b), fly ash (f), water (w), superplasticizer (sp), coarse aggregate (ca), fine aggregate (fa) and age in days (d) which have been taken as inputs and compressive strength (f_{ck}) as output for the model. Table 1 shows the ranges of parameters of the database used. The database often contains unexpected inaccuracies, for instance, the class of fly ash is sometimes not reported. The greatest difficulty seems to be related to the application of superplasticizers. They are from different manufacturers, of different chemical compositions, and without details concerning the solid contents in the suspension.

Table 1 Ranges of input features of database

Input Features	Minimum (kg/m^3)	Maximum (kg/m^3)	Average (kg/m^3)
Contents of cement	71	600	232.2
Blast furnace slag	0	359	79.2
Fly ash	0	175	46.4
Water	120	228	186.4
Superplasticizer	0	20.8	3.5
Coarse aggregate	730	1322	943.5
Fine aggregate	486	968	819.9

The data (Yeh 1998, 1998, 1999, 2003, 2003, 2006) used in both the techniques are normalized against their maximum values. In carrying out the formulation, the data has been divided into two sub-sets:

(a) Training dataset: This is required to construct the model. In this study, 824 (80% of total data) out of the 1030 values are considered as training dataset.

(b) A testing dataset: This is required to estimate the model performance. In this study, the remaining 206 (20% of total data) values are considered as testing dataset.

3. GPR model

GPR is a non-parametric model (Trianto and Kokugan 2002). Let us consider the dataset

$$\{x_k, y_k\}_{k=1}^{N}, \quad x \in M^N, \quad y \in M, \tag{1}$$

where x is the input variable, y is the output variable, M^N is the N-dimensional vector space, M is the one dimensional vector space, k is the number of iterations from 1 to N and N is number of data samples. We use c, b, f, w, sp, ca, fa and d as input variables for predicting f_{ck}. The GPR output is f_{ck}. So, x=[c,b,f,w,sp,ca,fa,d] and y=[f_{ck}]. This dataset has been drawn from the noise process for i^{th} iteration shown in Eq. (2) where ε is a constant.

$$y_i = f(x_i) + \varepsilon, \;\; \varepsilon \sim N(0, \sigma^2) \tag{2}$$

The joint distribution of y is represented in Eq. (3).

$$P(y) = N(0, K(x,x) + \sigma^2 I) \quad (3)$$

where $K(x, x)$ is the kernel function and I is the identity matrix. For a given input x^*, GPR defines a Gaussian predictive distribution over the output Y^* with a mean (Eq. (4)) of

$$\mu = K(x^*, x)(K(x,x) + \sigma^2 I)^{-1} y \quad (4)$$

and variance (Eq. (5))

$$\Sigma = K(x^*, x^*) - \sigma^2 I - K(x^*, x)(K(x,x) + \sigma^2 I)^{-1} K(x, x^*) \quad (5)$$

We observe from Eq. (4) that the mean prediction is a linear combination of y. A suitable covariance function and its parameter are required to develop the GPR model. For a fixed value of Gaussian noise, GPR is trained by maximizing marginal likelihood.

4. MARS model

MARS is widely accepted by researchers and practitioners for the following reasons.

- MARS is capable of modeling complex non-linear relationship among variables without strong model assumptions.
- MARS can capture the relative importance of independent variables to the dependent variable when many potential independent variables are considered.
- MARS does not need long training process and hence can save lots of model building time, especially when the dataset is huge.

Finally, one strong advantage of MARS over other classification techniques is the resulting model can be easily interpreted. It not only points out which variables are important in classifying objects/observations, but also indicates a particular object/observation belongs to a specific class when the built rules are satisfied. The final fact has important managerial and interpretative implications and can help to make appropriate decisions.

The MARS model splits the data into several splines on an equivalent interval basis (Friedman 1991). In every spline, MARS splits the data further into many subgroups (Yang *et al.* 2003). Several knots are created by MARS. These knots can be located between different input variables or different intervals in the same input variable, to separate the subgroups. The data of each subgroup are represented by a basis function (BF). The general form of a MARS predictor can be represented by Eq. (6) as

$$f(x) = \beta_0 + \sum_{j=1}^{P} \sum_{b=1}^{B} [\beta_{jb}(+) Max(0, x_j - H_{bj}) + \beta_{jb}(-) Max(0, H_{bj} - x_j)] \quad (6)$$

where x=input, $f(x)$=output, P=predictor variables and B=basis function. Max(0, x-H) and Max(0, H-x) are BF and do not have to each be present if their coefficients are 0. The H values are called knots. The spline function consists of two segments, i.e., truncated functions of the left-hand side of Eq. (7) and right-hand side Eq. (8) separated from each other by a so-called knot location (Veaux *et al.* 1993), as follows

$$b_q^-(x-t) = [-(x-t)]_+^q = \begin{cases} (t-x)^q, & if\ x > t \\ 0, & otherwise \end{cases} \quad (7)$$

$$b_q^+(x-t) = [+(x-t)]_+^q = \begin{cases} (x-t)^q, & if\ x > t \\ 0, & otherwise \end{cases} \quad (8)$$

where: t is the knot location and $b_q^-(x-t)$ & $b_q^+(x-t)$ are the spline functions. The MARS algorithm consists of (i) a forward stepwise algorithm to select certain spline basis functions, (ii) a backward stepwise algorithm to delete BFs until the "best" set is found, and (iii) a smoothing method which gives the final MARS approximation a certain degree of continuity. BFs are deleted in the order of least contributions using the generalized cross-validation (GCV) criterion (Craven and Wahba 1979). The GCV criterion is defined using Eq. (9)

$$GCV = \frac{\frac{1}{N}\sum_{i=1}^{N}[y_i - f(x_i)]^2}{\left[1 - \frac{C(B)}{N}\right]^2} \quad (9)$$

where N is the number of data and $C(B)$ is a complexity penalty that increases with the number of BF in the model and which is represented as Eq. (10)

$$C(B) = (B+1) + \lambda B \quad (10)$$

where λ is a penalty for each BF included into the model. It can be also regarded as a smoothing parameter. (Friedman 1991) provided more details about the selection of the λ.

5. MPMR model

MPMR uses kernel function for prediction of output(y). In MPMR, the relation between input(x) and y is given in Eq. (11).

$$y = \sum_{i=1}^{N} \beta_i K(x_i, x) + b \quad (11)$$

where N is the number of datasets, $K(x_i, x)$ is kernel function, β_i and b are outputs of MPMR. This article uses c, b, f, w, sp, ca, fa, d as inputs. The output of MPMR is f_{ck}. So, $x=[c,b,f,w,sp,ca,fa,d]$ and $y=[f_{ck}]$.

One data set is obtained by shifting all of the regression data $+\varepsilon$ along the output variable axis. The other dataset is obtained by shifting all of the regression data $-\varepsilon$ along the output variable axis. Regression model is the classification boundary between these two classes. More details of MPMR are given by (Strohmann and Grudic 2002).

The datasets are scaled between 0 and 1. Radial basis function $\left(\exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2\sigma^2}\right\}\right.$, where σ is the width of radial basis function) has been adopted as kernel function for the MPMR model.

In this study, a sensitivity analysis has been done to extract the cause and effect of relationship between the inputs and outputs of the GPR, MARS and MPMR models. The basic idea is that each input of the model is offset slightly and the corresponding change in the output is reported. The procedure has been taken from the work of Liong *et al.* (2000). According to Liong *et al.* (2000), the sensitivity(S) of each input parameter has been calculated

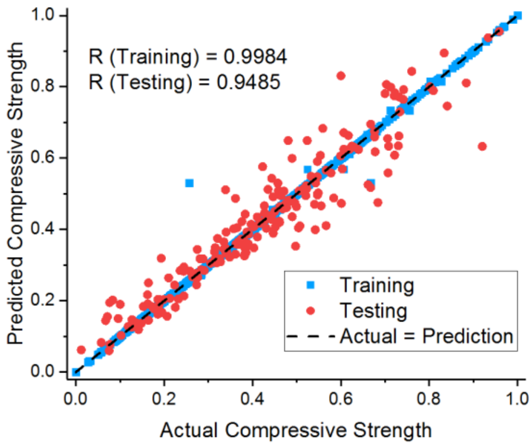

Fig. 1 Performance of training and testing dataset (GPR)

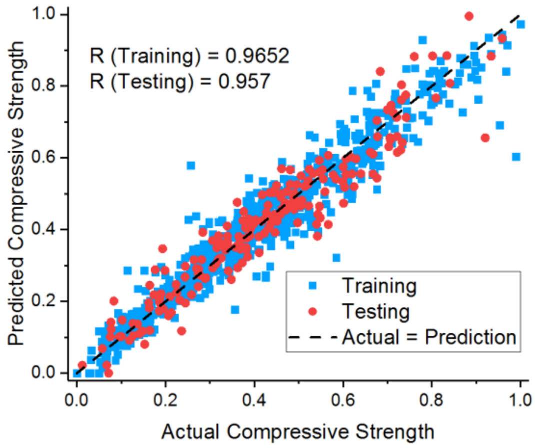

Fig. 2 Performance of training and testing dataset (MARS)

by the following formula:

$$S(\%) = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\%\ change\ in\ output}{\%\ change\ in\ input}\right)_j \times 100 \qquad (12)$$

where N is the number of data points. In this study, N=824. The analysis has been carried out on the trained model by varying each of the input parameters, one at a time, at a constant rate of 30%.

6. Results and discussions

Error and Correlation Calculations

The validity of each and every model can be verified using these following formulas:

The mean absolute error (*MAE*), Eq. (13) is a quantity used to measure how close predictions are to the actual value.

$$MAE = \frac{\sum_{i=1}^{n}|w_{ai}-w_{pi}|}{n} \qquad (13)$$

Root-mean-square error (*RMSE*), Eq. (14) is used to measure the differences between predicted value by the models and the actual values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(w_{ai}-w_{pi})^2}{n}} \qquad (14)$$

Coefficient of correlation (*R*), Eq. (15) has been used as main criterion to examine the performance of the developed models. The value of *R* has been determined by using the following equation

$$R = \frac{\sum_{i=1}^{n}(w_{ai}-w_a)(w_{pi}-w_p)}{\sqrt{\sum_{i=1}^{n}(w_{ai}-w_a)}\sqrt{\sum_{i=1}^{n}(w_{pi}-w_p)}} \qquad (15)$$

ρ is known as the Performance Index, Eq. (16) is used to check the accuracy of the predicted values.

$$\rho = \frac{RMSE}{w_a}\frac{1}{R+1} \qquad (16)$$

where w_{ai} and w_{pi} are the actual and predicted W values, respectively, w_a and w_p are mean of actual and predicted W values corresponding to n patterns. For a predictive model of high accuracy, the value of R should be close to one.

For developing the GPR, the design values of Gaussian noise and σ have been determined by the trial and error approach. The GPR gives the best performance at Gaussian noise=0.1 and σ=0.001. Graphs are plotted between Actual Normalized Strength and Predicted Normalized Strength. Fig. 1 shows the performance of training and testing dataset respectively. After the compilation of the model, following results are obtained.

Training and testing performance are illustrated in the Table 2. As shown in Fig. 1, the value of R is very close to one for training but not so close for testing datasets. So, the developed GPR is less capable for prediction of compressive strength of concrete. The value of error and correlation functions for GPR is shown in Table 2.

For MARS model, during training, the forward stepwise procedure was carried out to select 70 basis functions (BF) to build the MARS model. This was followed by the backward stepwise procedure to remove redundant basis functions. The final model includes 61 basis functions, which are listed in Table A1 together with their corresponding equations and a_m.

The final equation for the prediction of strength (f_{ck}) based on MARS model is given Eq. (17)

$$f_{ck} = 0.7798 + \sum_{m=1}^{M} a_m B_m(x) \qquad (17)$$

where,

a_0=0.7798, i.e., coefficient of the constant basis function, or the constant term;

$\{a_m\}$=vector of coefficients of the non-constant basis functions, m=1, 2, …, M;

B_m are the basis functions that are selected for inclusion in the model.

The ANOVA decomposition is specified in row wise for each ANOVA function. The columns represent summary quantities for corresponding ones. The first column lists the function number. The second gives the standard deviation (STD) of the function. This gives indication of its (relative) importance to the overall model and can be interpreted in a manner similar to a standard regression coefficient in a linear model. The third column provides another indication of the importance of the corresponding ANOVA function, by listing the GCV score for a model with the entire basis functions corresponding to that particular ANOVA function removed. This can be used to judge whether this ANOVA

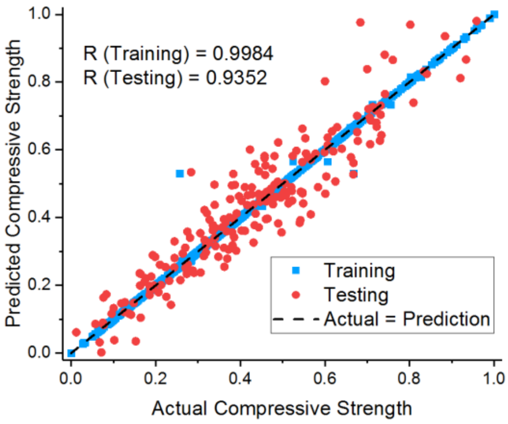

Fig. 3 Performance of training and testing dataset (MPMR)

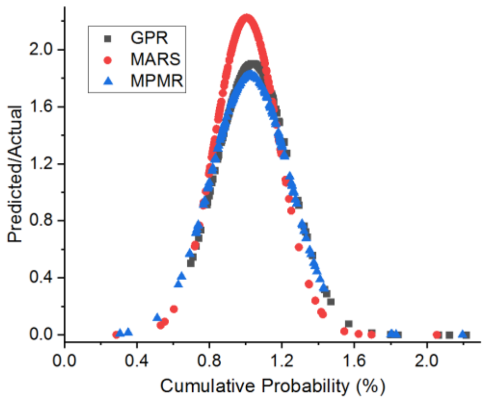

Fig. 5 Log normal distribution of Predicted/Actual Strengths for different models for testing data

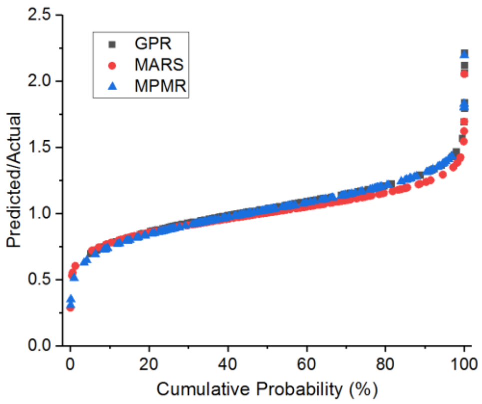

Fig. 4 Cumulative probability plots of Predicted/Actual Strengths for different models for testing data

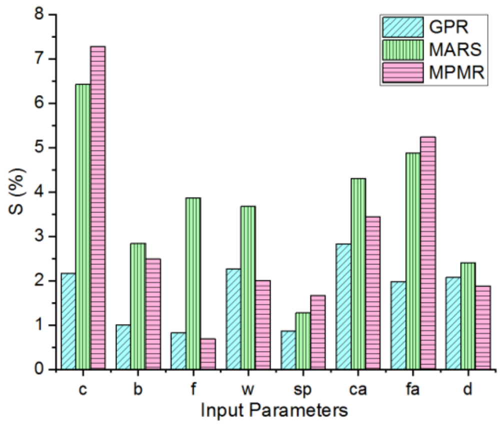

Fig. 6 Sensitivity analysis of the input parameters for GPR, MARS and MPMR models

function is making an important contribution to the model, or whether it just slightly helps to improve the global GCV score. The fourth column gives the number of basis functions comprising the ANOVA and the last column of Table A2 gives the particular predictor variables associated with the ANOVA function. Table A2 shows the ANOVA decomposition for Training dataset.

Fig. 2 depicts the performance of training and testing dataset. It is observed from figure that the value of *R* is very close to one for training as well as for testing datasets. Therefore, the developed MARS proves is highly capable for prediction of compressive strength of concrete. The value of error and correlation functions for MARS is shown in Table 2.

For developing MPMR, the design value of ε and width (σ) of radial basis function have been determined by trial and error approach. The design values of ε and σ are 0.1 and 0.01 respectively. The performance of training and testing datasets has been depicted in Fig. 3 respectively. The performance of MPMR has been assessed in terms of coefficient of correlation (*R*). For a good model, the value of *R* should be close to one. As shown in Fig. 3, the value of *R* is very close to one for training but not so close for testing datasets. So, the developed MPMR has low ability for prediction of compressive strength of concrete.

The approximates of error and correlation functions i.e., mean absolute error (*MAE*), root-mean-square error (*RMSE*), coefficient of correlation (*R*) and performance index (ρ) for all the models employed are consolidated in Table 2.

A comparative study has been carried out between the developed GPR, MARS and MPMR models. Figs. 1, 2 and 3 shows the graph of *R* value of the training and testing datasets for GPR, MARS and MPMR models respectively. Fig. 4 shows the cumulative probability plots of Predicted/Actual Strengths for different methods for the testing dataset. The lognormal distributions of the Predicted/Actual Strengths for different models of the testing data are shown in Fig. 5. Based on the plots it can be seen that MARS model is better than GPR and MPMR models. It can be inferred from Figs. 2, 4 and 5 that the performance of MARS outperforms the performance of GPR and MPMR model. It is also clear from Table 3 that the performance of MARS is best. Fig. 6 presents the results of sensitivity analysis. It can be seen that, contents of cement (*c*) has the most significant effect on the predicted compressive strength (f_{ck}) for MARS and MPMR models followed by fine aggregate (*fa*), coarse aggregate (*ca*), fly ash (*f*), water (*w*), blast furnace slag (*b*), age in days (*d*) and superplasticizer (*sp*) (specifically for MARS model). Superplasticizer seems to contribute the least for all the 3 models.

The performance of training and testing dataset is almost same for the GPR and MPMR models but MARS shows the best performance among the three models. Table

Table 2 Approximates of error and correlation functions

Models Employed	Mean absolute error (*MAE*)		Root-mean-square error (*RMSE*)		Coefficient of correlation (*R*)		Performance Index (ρ)	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
GPR	0.00101	0.04292	0.01199	0.06375	0.9984	0.9485	0.01449	0.07634
MARS	0.03978	0.04399	0.05477	0.05846	0.9652	0.957	0.06726	0.06970
MPMR	0.00099	0.05659	0.01199	0.07358	0.9984	0.9352	0.01448	0.08871

Table 3 Actual results and predicted f_{ck} by MARS, MPMR & GPR

Cement (component 1) (kg in a m^3 mixture)	Blast Furnace Slag (component 2) (kg in a m^3 mixture)	Fly Ash (component 3) (kg in a m^3 mixture)	Water (component 4) (kg in a m^3 mixture)	Superplasticizer (component 5) (kg in a m^3 mixture)	Coarse Aggregate (component 6) (kg in a m^3 mixture)	Fine Aggregate (component 7) (kg in a m^3 mixture)	Age (day)	Actual Concrete compressive strength (MPa, megapascals)	Predicted Concrete compressive strength (MPa, megapascals) MARS	Predicted Concrete compressive strength (MPa, megapascals) MPMR	Predicted Concrete compressive strength (MPa, megapascals) GPR
424.0	22.0	132.0	178.0	8.5	822.0	750.0	7	39.03	38.04	49.03	44.85
157.0	236.0	0.0	192.0	0.0	935.4	781.2	28	33.66	33.03	26.58	31.19
374.0	189.2	0.0	170.1	10.1	926.1	756.7	7	46.20	47.25	47.56	49.7
349.0	0.0	0.0	192.0	0.0	1056.0	809.0	90	40.66	40.02	37.57	39.33
168.0	42.1	163.8	121.8	5.7	1058.7	780.1	28	24.24	24.69	21.05	21.94

3 shows the actual results and predicted f_{ck} from the developed MARS, MPMR and GPR for five random testing values from the dataset. It shows that the predicted f_{ck} from MARS match well with the actual f_{ck}. The developed models do not show overtraining. Therefore, the developed models have good generalization capability. Datasets are normalized between for developing the GPR, MARS and MPMR models. The developed models do not make assumption about the dataset. The developed MARS gives equation for prediction of strength. However, MARS do not use statistical parameters of the dataset for prediction. MARS adopts basis function for final prediction. MPMR uses kernel function for prediction of output. GPR defines a Gaussian predictive distribution over the output.

When the results of GPR, MARS and MPMR are compared with the results of previously used model for prediction i.e., ANN (Artificial Neural Networks) (Yeh 1998, 1999, 2003, 2003, 2006, Moretti *et al.* 2016, Erdala *et al.* 2013, Chou and Pham 2013, Kong and Chen 2015, Kumar *et al.* 2014, Viswanathan *et al.* 2015) it has been observed that these three models generate more accurate results than ANN. The coefficient of correlation for testing dataset obtained using ANN is 0.914 which is less compared to the three models used here. Therefore, among the four models, MARS gives the best result. MARS gives the equation to predict f_{ck} whereas ANN does not provide any equation. Moreover, MARS model is easier to build. And also, it automatically takes care of unimportant and redundant predictor values whereas ANN fails to do so. The developed MARS does not use any strong model assumptions. The main advantage of the developed MARS is that it can determine the contributions of basis functions for a particular problem. It can find out interaction between variables. It can easily handle the large variation of variables. The developed MARS is more flexible than the GPR and MPMR models. Outliers have no effect on the developed MARS.

7. Conclusions

This study has described the application of GPR, MARS and MPMR models for the prediction of compressive strength of concrete. Sensitivity analysis indicates that contents of cement has the most significant effect on the predicted compressive strength of concrete. The performance of MARS is better than GPR and MPMR model. User can use the developed model for prediction of compressive strength of concrete. The developed models can be used as a quick tool for prediction of compressive strength of concrete. The developed equation is very useful for estimation of compressive strength of concrete. Experiment is not required for determination of compressive strength of concrete. Hence, the developed models are cost effective. This paper shows that the developed MARS is a robust model for prediction of compressive strength of concrete.

References

Belhouari, S.B. and Bermak, A. (2004), “Gaussian process for nonstationary time series prediction”, *Comput. Stat. Data Anal.*, **47**(4), 705-714.

Chen, T., Morris, J. and Martin, E. (2007), “Gaussian process regression for multivariate spectroscopic calibration”, *Chem. Intell. Lab. Syst.*, **87**(1), 59-71.

Cheng, M.Y. and Cao, M.T. (2014), “Evolutionary multivariate adaptive regression splines for estimating shear strength in reinforced-concrete deep beams”, *Eng. Appl. Artif. Intell.*, **28**, 86-96.

Chou, J., Ngo, N. and Pham, A. (2015), “Shear strength prediction in reinforced concrete deep beams using nature-inspired metaheuristic support vector regression”, *J. Comput. Civil Eng.*, **30**(1), 04015002.

Chou, J.S. and Pham, A.D. (2013), “Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength”, *Constr. Build.*

Mater., **49**, 554-563.

Craven, P. and Wahba, G. (1979), "Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation", *Numer. Math.*, **31**, 317-403.

De Veaux, R.D., Psichogios, D.C. and Ungar, L.H. (1993), "A comparison of two nonparametric estimation schemes: MARS and neural networks", *Comput. Chem. Eng.*, **17**(8), 819-837.

Erdala, H.I., Karakurtb, O. and Namlic, E. (2013), "High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform", *Eng. Appl. Artif. Intell.*, **26**(4), 1246-1254.

Friedman, J.H. (1991), "Multivariate adaptive regression spline", *Ann. Stat.*, **19**, 1-141.

GB 50205 (2001), Code for Acceptance of Construction Quality of Steel Structures, GB National Standard.

He, H. and Siu, W.C. (2011), "Single image super-resolution using gaussian process regression", *IEEE Conference: Computer Vision and Pattern Recognition (CVPR),* 449-456.

Henry, N. and Leonardo, B. (2014), "Descriptive examples of the limitations of artificial neural networks applied to the analysis of independent stochastic data", e-print, arXiv:1404.5598.

Huang, K., Yang, H., King, I. and Lyu, M.R. (2006), "Maximizing sensitivity in medical diagnosis using biased minimax probability machine", *IEEE Eng. Med. Bio. Soc.*, **53**(5), 821-831.

IS 516 (1959), Method of Tests for Strength of Concrete, Bureau of Indian Standards.

Kandel, K., Huettmann, F., Suwal, M.K., Ram Regmi, G., Nijman, V., Nekaris, K.A.I., Lama, S.T., Thapa, A., Sharma, H.P. and Subedi, T.R. (2015), "Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (Ailurus fulgens) in the Hindu-Kush Himalaya region", *Biolog. Conserv.*, **181**, 150-161.

Kong, L. and Chen, X. (2015), "Influence mechanism of lightweight aggregate on concrete impermeability: prediction by ANN", *Mag. Concrete Res.*, **67**(1), 17-26.

Kumar, M., Aiyer, B.G. and Samui, P. (2014), "Machine learning techniques applied to uniaxial compressive strength of oporto granite", *Int. J. Performa. Eng.*, **10**(2), 189-195.

Liong, S.Y., Lim, W.H. and Paudyal, G.N. (2000), "River stage forecasting in Bangladesh: neural network approach", *J. Comput. Civil Eng.*, **14**(1), 1-8.

Moretti, J.F., Minussi, C.R., Akasaki, J.L., Fioriti, C.F., Melges, J.L.P. and Tashima, M.M. (2016), "Prediction of modulus of elasticity and compressive strength of concrete specimens by means of artificial neural networks", *Acta Scientiarum. Technol.*, **38**(1), 65-70.

Nedushan, B.A. (2012), "An optimized instance based learning algorithm for estimation of compressive strength of concrete", *Eng. Appl. Artif. Intell.*, **25**(5), 1073-1081.

Ozturk, A.U. and Turan, M.E. (2012), "Prediction of effects of microstructural phases using generalized regression neural network", *Constr. Build. Mater.*, **29**, 279-283.

Pickens, B.A. and King, S.L. (2014), "Linking multi-temporal satellite imagery to coastal wetland dynamics and bird distribution", *Ecol. Model.*, **285**, 1-12.

Rasmussen, C.E. and Williams, C.K.I. (2006), *Gaussian Processes for Machine Learning*, MIT Press.

Razavi, S.V., Jumaat, M.Z., El Shafie, A. and Mohammadi, P. (2012), "Using generalized regression neural network (GRNN) for mechanical strength prediction of lightweight mortar, *Comput. Concrete*, **10**(4), 379-390.

Samui, P., Dalkiliç, Y.H., Rajadurai, H. and Jagan, J. (2015), "Minimax probability machine: A new tool for modeling", *Handbook of Research on Swarm Intelligence in Engineering*, 182-210.

Stegle, O., Fallert, S.V., MacKay, D.J.C. and Brage, S. (2008), "Gaussian process robust regression for noisy heart rate data", *IEEE Tran. Biomed. Eng.*, **55**(9), 2143-2151.

Strohmann, T.R. and Grudic, G.Z. (2002), *A Formulation for Minimax Probability Machine Regression*, Advances in Neural Information Processing Systems (NIPS) 14, MIT Press.

Sun, L., Pan, Y. and Gu, W. (2013), "Data mining using regularized adaptive B-splines regression with penalization for multi-regime traffic stream models", *J. Adv. Tran.*, **48**(7), 876-890.

Trianto, A. and Kokugan, T. (2002), "Method for improving the performance of porous membrane reactor", *J. Chem. Eng. JPN*, **34**(2), 199-206.

Viswanathan, R., Kurup, P. and Samui, P. (2015), "Examining efficacy of metamodels in predicting ground water table", *Int. J. Performa. Eng.*, **11**(3), 275-281.

Yang., C.C., Prasher, S.O., Lacroix, R. and Kim, S.H. (2003), "A multivariate adaptive regression splines model for simulation of pesticide transport in soils", *Biosyst. Eng.*, **86**(1), 9-15.

Yeh, I.C. (1998), "Modeling concrete strength with Augment-Neuron networks", *J. Mater. Civil Eng.*, **10**(4), 263-268.

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

Yeh, I.C. (1999), "Design of high performance concrete mixture using neural networks", *J. Comput. Civil Eng.*, **13**(1), 36-42.

Yeh, I.C. (2003), "A mix proportioning methodology for fly ash and slag concrete using artificial neural networks", *Chung Hua J. Sci. Eng.*, **1**(1), 77-84.

Yeh, I.C. (2003), "Prediction of strength of fly ash and slag concrete by the use of artificial neural networks", *J. Chin. Inst. Civil Hydra. Eng.*, **15**(4), 659-663.

Yeh, I.C. (2006), "Analysis of strength of concrete using design of experiments and neural networks", *J. Mater. Civil Eng.*, **18**(4), 597-604.

CC

Appendix

Table A1 list of basis functions which give the best performance

Basis Function	Equation	a_m
BF1	max(0, d -0.1511)	+0.153
BF2	max(0, 0.1511 -d)	-1.379
BF3	max(0, c -0.9657)	-0.070
BF4	max(0, 0.9657 -c)	-0.605
BF5	max(0, b -0.0528)	+0.938
BF6	max(0, 0.0528 -b)	+0.818
BF7	max(0, sp -0.3388)	-0.212
BF8	max(0, 0.3388 -sp)	+0.001
BF9	max(0, 0.7413 -w)	+0.365
BF10	max(0, 0.0357 -d)	-7.783
BF11	max(0, d -0.0357) * max(0, 0.1103 -fa)	-13.318
BF12	max(0, b -0.8707)	-0.412
BF13	max(0, 0.8707 -b)	-0.146
BF14	BF2 * max(0, 0.3614 -b)	+0.514
BF15	BF13 * max(0, 0.0054 -d)	-15.995
BF16	BF10 * max(0, 0.7853 -c)	+7.990
BF17	max(0, d -0.0357) * max(0, fa -0.1103) * max(0, ca -0.9418)	+1.763
BF18	max(0, d -0.0357) * max(0, fa -0.1103) * max(0, 0.9418 -ca)	-0.891
BF19	BF8 * max(0, ca -0.0491)	+0.053
BF20	BF8 * max(0, 0.0491 -ca)	-40.497
BF21	BF1 * max(0, 0.1906 -fa)	+7.242
BF22	BF19 * max(0, b -0.2114)	-4.737
BF23	BF19 * max(0, 0.2114 -b)	-5.154
BF24	BF8 * max(0, w -0.1856)	+0.295
BF25	BF8 * max(0, 0.1856 -w)	-11.501
BF26	BF9 * max(0, c -0.2545)	+8.166
BF27	BF9 * max(0, 0.2545 -c)	-1.323
BF28	BF22 * max(0, c -0.6347)	392.870
BF29	BF22 * max(0, 0.6347 -c)	+5.651
BF30	BF23 * max(0, w -0.7045)	+36.233
BF31	BF23 * max(0, 0.7045 -w)	+11.775
BF32	BF4 * max(0, b -0.1321)	-0.248
BF33	BF9 * max(0, d -0.0164)	+0.787
BF34	BF9 * max(0, 0.0164 -d)	-15.965
BF35	BF24 * max(0, ca -0.0755)	+0.569
BF36	BF24 * max(0, 0.0755 -ca)	+0.968
BF37	BF10 * max(0, b -0.2728)	-8.360
BF38	BF10 * max(0, 0.2728 -b)	+9.614
BF39	BF5 * max(0, b -0.1224)	-0.553
BF40	BF5 * max(0, 0.1224 -b)	-1.250
BF41	BF39 * max(0, w -0.5257)	+1.120
BF42	BF39 * max(0, 0.5257 -w)	+10.875
BF43	BF42 * max(0, c -0.1827)	-38.064
BF44	BF42 * max(0, 0.1827 -c)	-53.159
BF45	BF13 * max(0, ca -0.9418)	+0.499
BF46	BF13 * max(0, 0.9418 -ca)	-0.075
BF47	BF42 * max(0, fa -0.4804)	+14.161
BF48	BF42 * max(0, 0.4804 -fa)	-21.935
BF49	BF46 * max(0, c -0.7351)	-0.586

Table A1 Continued

BF50	BF46 * max(0, 0.7351 -c)	+0.128
BF51	BF4 * max(0, sp -0.1055)	-0.380
BF52	BF4 * max(0, 0.1055 -sp)	-0.485
BF53	BF26 * max(0, b -0.3948)	+37.225
BF54	BF26 * max(0, 0.3948 -b)	-20.455
BF55	BF26 * max(0, b -0.1224)	-33.035
BF56	BF53 * max(0, d -0.0741)	+17.613
BF57	BF53 * max(0, 0.0741 -d)	-36.959
BF58	BF26 * max(0, d -0.1510)	-1.340
BF59	BF26 * max(0, 0.1510 -d)	+3.712
BF60	BF4 * max(0, sp -0.3068)	+0.289

Table A2 ANOVA decomposition for Training dataset

Function Number	Standard Deviation	GCV	Basis Function	Parameters	Variable(s)
1	0.142	0.160	2	2.0	1
2	0.191	0.087	2	2.0	2
3	0.045	0.015	2	2.0	3
4	0.057	0.019	1	1.0	4
5	0.016	0.004	2	2.0	5
6	0.165	0.058	3	3.0	8
7	0.033	0.005	1	1.0	1, 2
8	0.650	0.698	2	2.0	1, 4
9	0.017	0.005	3	3.0	1, 5
10	0.039	0.006	1	1.0	1, 8
11	0.042	0.007	2	2.0	2, 3
12	0.038	0.006	3	3.0	2, 8
13	0.015	0.004	2	2.0	3, 6
14	0.005	0.004	1	1.0	3, 8
15	0.029	0.005	2	2.0	4, 5
16	0.038	0.005	2	2.0	4, 8
17	0.010	0.004	2	2.0	5, 6
18	0.035	0.005	2	2.0	7, 8
19	0.637	0.684	3	3.0	1, 3, 4
20	0.017	0.004	2	2.0	1, 3, 6
21	0.033	0.005	2	2.0	1, 4, 8
22	0.053	0.007	2	2.0	2, 3, 4
23	0.114	0.029	2	2.0	2, 5, 6
24	0.017	0.024	2	2.0	4, 5, 6
25	0.026	0.004	2	2.0	6, 7, 8
26	0.021	0.004	2	2.0	1, 2, 3, 4
27	0.049	0.007	2	2.0	1, 2, 5, 6
28	0.012	0.004	2	2.0	1, 3, 4, 8
29	0.015	0.004	2	2.0	2, 3, 4, 7
30	0.046	0.007	2	2.0	2, 4, 5, 6