Construction and Building Materials 302 (2021) 124191

Contents lists available at ScienceDirect

Construction and Building Materials

journal homepage: www.elsevier.com/locate/conbuildmat

Machine learning in mix design of Miscanthus lightweight concrete

Patrick Pereira Dias, Laddu Bhagya Jayasinghe, Daniele Waldmann *

Technology and Medicine (FSTM), University of Luxembourg, Esch-sur-Alzette L-4364, Luxembourg

ARTICLE INFO

Keywords:
Miscanthus
Lightweight concrete
Machine learning
Gaussian process regression
Mix design

ABSTRACT

This research is carried out to investigate the Gaussian process regression (GPR) based on a machine learning model to predict the compressive strength of Miscanthus lightweight concrete (MLWC). A database of 414 experimental data, which includes nine input variables such as six mix constituents of concrete, form of specimen, curing time and pre-treatment condition and an output variable of compressive strength of MLWC, is constructed from the data collected by a series of experimental tests on MLWC. Two kernel functions, namely, the squared exponential and rational quadratic are used in the GPR model. It is found from experiments that the GPR model with rational quadratic kernel gives minimum errors for predicting compressive strength of MLWC. In addition, a user-friendly graphical user interface is created using MATLAB software to deploy the GPR model which can be used at an early stage of designing the Miscanthus concrete members instead of using costly experimental investigation.

1. Introduction

Concrete, a mixture of cement, water, fine aggregates (sand) and coarse aggregates, is the most common material used in the construction industry due to its various advantages over other materials such as durability, integrity and economy. However, the process of producing concrete causes several impacts on the environment, not only the large amount of CO_2 emissions, but natural resources such as sand and gravel also become depleted. The extraction of natural aggregates, often from environmentally sensitive areas such as river valleys, can lead to a destruction of an ecosystem. Using Miscanthus in concrete formulations is a possible solution to lower the use of natural aggregates in concrete [1]. In addition, if the Miscanthus can replace the use of aggregates in concrete, benefits can be gained through reducing energy consumption and carbon emission [2].

Many efforts have been made in the last decade to achieve improved Miscanthus fibre-binder bond characteristics in the concrete mix as well as to replace the conventional building materials with Miscanthus [3,4,5,6,7,8]. Pude et al. [9] stated that Miscanthus could be used as basic material for structures because it contains considerably strong fibre compounds made of silicon, cellulose and lignin. They investigated the suitability of four genotypes of Miscanthus stems as lightweight concrete aggregates. Acikel [10] showed that using the grinded Miscanthus with a diameter 4–8 mm and a length of 60–80 mm as lightweight aggregates in a concrete mixture, the strength of concrete in compression, tension and bending is increased by 4 to 28%, 9 to 25% and 4 to 9%, respectively. The pore structure of Miscanthus contributes to reduce the thermal conductivity of the Miscanthus concrete [11,12]. Chen et al. [13] analyzed the thermal and acoustic performance of Miscanthus lightweight concrete (MLWC). The test results showed that MLWC has better thermal and sound insulation characteristics compared to normal-weight concrete, which is due to the high porosity of MLWC. It was also concluded that the sound absorption coefficient increases from 0.28 to 0.63 with the increase of the volume of Miscanthus from 0 to 30%.

The use of Miscanthus as a lightweight aggregate in concrete is limited due to several reasons such as inconsistent properties of the porous composite and lack of research on the characteristics and performance of Miscanthus as lightweight aggregate in concrete. This research was designed to address the above issues by conducting laboratory experiments on MLWC to investigate the effects of different constituents and moisture content on the strength and performance characteristics of MLWC. It is also needed to know which mix proportions should be used to achieve the specific properties of concrete, especially the desired compressive strength. If the strength of concrete could be predicted beforehand, it could save a lot of time and cost. However, such a topic has never been studied on MLWC. Thus, as the first step of such a detailed investigation of MLWC, this paper focuses on developing an accurate and reliable prediction model to quantify accurately the volumes of the different constituents for an optimisation

* Corresponding author.
E-mail address: daniele.waldmann@uni.lu (D. Waldmann).

https://doi.org/10.1016/j.conbuildmat.2021.124191
Received 24 February 2021; Received in revised form 10 June 2021; Accepted 7 July 2021
Available online 16 July 2021

of the mixture regarding the compressive strength of MLWC.

Recently, machine learning (ML) techniques have been utilized increasingly for mixture design and optimization due to their excellent pattern recognition, auto-association and self-learning. The ML models can be used for predicting output data, based on a defined input dataset [14]. Various researchers used different ML techniques such as regression analysis [15,16,17], artificial neural networks (ANN) [18,19,20], fuzzy logic [21,22] for the prediction of compressive strength of concrete.

Linear regression (LR) is a basic and relatively simple supervised-ML algorithm that can be implemented very easily compared to other ML techniques. Due to its lower computational time when compared to other ML algorithms, it has been widely used to define the relationship between a dependent variable and one or more independent predictors [23]. Chopra et al. [15] used it for the prediction of compressive strength of concrete with and without fly ash. Huang et al. [24] used multiple linear regression (MLR) to estimate the bond strength of fibre reinforced polymer concrete.

In the last years, ANN structures are being used to solve many complex problems in civil engineering applications due to its capability of performing multiple tasks in parallel without affecting the system performance. Yeh [25] developed a model with eight input variables (i.e. cement, fly ash, blast furnace slag, water, superplasticizer, coarse aggregate, fine aggregate and age of testing) to optimize the high-performance concrete mix for given workability and strength using ANN. Alshihri et al. [26] predicted the compressive strength of structural lightweight concrete using ANN. They used laboratory test results to train and to validate the neural networks. Ni and Wang [18] showed that the single-layer ANN model could accurately predict the 28-day compressive strength of concrete. Bingol et al. [27] predicted the compressive strength of lightweight concrete with pumice aggregate subjected to high temperatures using ANN. Das et al. [28] developed an ANN model with seven inputs to predict the mix proportions of concrete. The hidden-layer neurons were adjusted by trial and error method to achieve optimal value. Topcu and Saridemir [21] gathered 52 different mixes with 180 specimens to develop a model to accurately predict the compressive strength of concrete containing fly ash using ANN and fuzzy logic.

Kewalramani and Gupta [29] developed MLR and ANN models to predict the long-term compressive strength of concrete. They concluded that ANN is more efficient in producing reliable strength compared to MLR. Similarly, Khademi et al. [30] developed three different models based on MLR, ANN and adaptive neuro-fuzzy inference system (ANFIS) to predict the 28-day compressive strength of concrete. They concluded that ANN and ANFIS models produced reliable results, while MLR model was not capable enough due to the nonlinear relationship between mix parameters. Omran et al. [31] compared the accuracy of different ML techniques for predicting the compressive strength of environmentally friendly concrete. They found that ensemble models such as additive regression and bootstrap aggregating was able to achieve good prediction performance. Their results have shown that the Gaussian process regression (GPR) model had the best prediction accuracy as an individual data-mining model. Pal and Deswal [32] also found that GPR model could provide a reliability response to an input data. Another advantage of the GPR model was the prior fitted function could be shaped by the choice of kernel function.

To the best of our knowledge, the capability of GPR has rarely been investigated for the prediction of compressive strength of concrete. However, due to its methodological advantages, GPR models have been widely used in many fields [33,34]. GPR is a non-parametric regression model, and it is capable of fitting complex, nonlinear relationships between output and input variables [35]. GPR has its own intrinsic advantages over other ML techniques such as ANN and fuzzy logic model. In general, ANN has some shortcomings like slow convergence speed, less generalising performance, and over-fitting problems [36] whilst fuzzy logic model has disadvantages like requires lots of data, and needs high human expertise on determining fuzzy rules [37]. Compared to other ML methods, the major advantage of using GPR is that its ability to get the predictions using the mean and to capture the uncertainty of predicted values using variance. In addition, GPR can be applied effectively to solve regression problems with small datasets and high dimensions, by choosing a kernel function with the knowledge of data that best represents the data [38].

Accordingly, this paper utilizes state-of-the-art achievements in machine learning techniques for Miscanthus concrete mix design. The objective of this paper is to illustrate that the ML model that employs the GPR can be used to accurately predict the compressive strength of MLWC. The subsequent part of the paper is organized as follows. Section 2 discusses the materials and the experimental procedure used to produce required dataset. Section 3 then describes the ML technique and the evaluation method, followed by the discussion of the results. The conclusion of this study is presented in the final section.

2. Experimental procedure

The experimental tests have been conducted at the laboratory of Solid Structures of the University of Luxembourg. The main characteristics of materials and procedures used for the purpose of this study are presented in this section.

2.1. Materials

The Miscanthus concrete mixtures consisted of six components such as Miscanthus, cement, water, lime, a superplasticizer and a mineraliser. Miscanthus × giganteus (Mxg), the giant Miscanthus, which was produced by Luxembourgish farmers, was chosen because of its wide availability. In this study, dried Miscanthus with average density of 120 kg/m^3 was used. The length of the used Mxg fibres ranged from 15 to 60 mm with a cross-section of 1 to 6 mm. As Mxg has high water absorption capacity, pre-treatment condition of the Mxg will also have an impact on the compressive strength. Therefore, a silicate sealant was applied on the Mxg fibres before their application in some mixtures in order to investigate the effect a pre-treatment.

CEM I Portland cement of class 42.5R (CEM I 42.5R) as per EN 197-1:2011 [39] and natural hydraulic lime strength class 3.5 (NHL 3.5) as per EN 459-1 [40] were used as the binder materials. It is well known that the lime improves the workability and permeability of the lime-cement mortar, while increase of cement in the binder improves the mechanical strengths of the mortar [41,42]. Calcium Chloride ($CaCl_2$) was added as mineraliser to avoid the Miscanthus fibres to absorb the water needed for the hydration process. It leads to an increase of the compressive strength without increasing the amount of cement [43,44]. Since the Miscanthus fibres have a high water absorption, cement hydration process has to be increased before the Miscanthus fibres can absorb the water. On this purpose, MasterGlenium ACE 456 superplasticiser was introduced with the mortar due to its advantages of improving the rheology of the mixture and accelerating the cement hydration [1]. Normal tap water was used as mixing water. Table 1 summarises the properties of the materials used in the Miscanthus

Table 1
Properties of the materials used in the Miscanthus concrete mixture.

Material	Properties
Miscanthus × giganteus	Density: 120 kg/m^3 15-60 mm with a cross section of 1–6 mm fibres
Cement (CEM I 42.5R)	Density: 3100 kg/m^3
Natural hydraulic lime (NHL 3.5)	Density: 740 kg/m^3
$CaCl_2$	Density: 710 kg/m^3 Particle size: 2–5 mm
Superplasticizer ACE 456	Density: 1060 kg/m^3
Water	Density: 1000 kg/m^3 PH: 6.9

concrete mixture. More details on the materials and their preparation can be found in ref. [1].

2.2. Test details

Since this study focused to study the effect of mix proportions on the compressive strength of MLWC, 73 different kind of mixtures were used to prepare the test specimens. The amounts of Mxg and cement were varied from 67.8 to 300 kg/m^3 and 75 to 750 kg/m^3, respectively. Initially, a water/cement (W/C) ratio of 0.7 was used to prepare the mixtures. However, additional mixtures were prepared using different W/C ratios (between 0.5 and 1.8) to study their effect on the strength of MLWC. The amounts of NHL3.5, superplasticizer and $CaCl_2$ used in this study were between 0 and 483.15 kg/m^3, 0 and 8.7, and 6 and 19.65 in kg/m^3, respectively. The details of first 20 mix proportions used in this study are given in Table 2.

The mixing procedure to get the MLWC test specimens was as follows. First, the dry Mxg and the lime were mixed without water. Then, the water volume was subdivided into two equal parts, and the first half of the water volume was mixed with the superplasticizer (ACE 456) and the mineraliser ($CaCl_2$). Next, it was poured into the dry mixture of Mxg and the lime, followed by the cement. After that, the remaining water volume was added on the ongoing mixture and it was mixed for about 7.5 min until a homogenous mixture was obtained. Finally, mixture was casted in 40 mm × 40 mm × 160 mm prism moulds and 150 mm × 150 mm × 150 mm cubical moulds in two successive stages with a needle vibration according to EN 196-1:2005 [45]. From the mixtures, test prisms and test cubes were prepared for the compression tests.

The specimens were stored in their moulds for 1 day at room temperature (around 20 – 22 ℃). After 1 day, they were demoulded and wrapped in cellophane to avoid the loss of water. After the curing time reached 13 days, the cellophane was removed and the specimens cured for the remaining time at room temperature. The specimens were tested in compression according to EN 12390–7:2019 [46]. More details on the test details can be found in the ref. [1].

3. Methodology

3.1. Gaussian process regression (gpr)

ML models have been widely used in many fields to simulate material behaviour. In this study, GPR is used to predict the compressive strength of MLWC. GPR is an efficient, probabilistic and nonparametric

Table 2
Various mix proportions used in the study.

Weight proportions of the mixture						W/C ratio
Mxg	Cement	NHL3.5	ACE 456	$CaCl_2$	Water	
1	3.96	2.17	0.04	0.09	2.78	0.7
1	3.96	2.17	0.04	0.09	2.08	0.53
1	2.97	2.17	0.04	0.06	2.78	0.94
1	3.96	0	0.04	0.09	2.78	0.7
1	8.33	4.56	0.08	0.19	5.84	0.7
1	2.48	1.36	0.02	0.06	1.74	0.7
1	1.75	0.96	0.02	0.04	1.23	0.7
1	2.02	2.17	0.04	0.09	1.42	0.7
1	11.06	2.17	0.04	0.09	7.76	0.7
1	3.96	2.17	0.04	0.09	1.98	0.5
1	3.96	2.17	0.04	0.09	2.37	0.6
1	3.96	2.17	0.04	0.09	3.16	0.8
1	3.96	2.17	0.04	0.09	3.56	0.9
1	3.44	1.38	0.04	0.09	2.78	0.81
1	3.77	1.89	0.04	0.09	2.78	0.74
1	4.17	2.5	0.04	0.09	2.78	0.66
1	4.68	3.27	0.04	0.09	2.78	0.59
1	3.96	2.17	0.04	0.09	2.78	0.7
1	3.97	2.19	0.04	0.09	2.02	0.51
1	3.97	2.19	0	0.09	5.26	1.32

supervised ML approach for modelling nonlinear and complex functional mappings [47]. GPR implements Gaussian process (GP) for regression purposes.

GP is used to describe a distribution over functions. Theoretically, set of function values in a finite dimensional space can be sampled from a probability distribution of functions determined by a mean function and a covariance function. In GP, the covariance function is determined by a chosen kernel function that describes how much influence one point has on another, and hence, it is very efficient to handle nonlinear data [47].

Mathematically, a Gaussian variable X can be defined in a normal distribution with its mean μ and covariance $\sum$ as,

$$X\ N\left(\mu, \sum\right) \tag{1}$$

GP is more rigorously defined as: f is a Gaussian process if for any vector of inputs $x = [x_1, \cdots\cdots, x_n]^T$, the vector of output $f(x) = [f(x_1), \cdots\cdots, f(x_n)]^T$ is Gaussian distributed.

$$f(x)\ N(m(x), k(x, x\acute{1})) \tag{2}$$

where $\times \in \mathbb{R}^D$, $m(x)$ and $k(x, x\acute{1})$ are the mean function and covariance function, respectively.

The mean function can be any function maps index points × onto real values. In GP, the mean function is often set to be zero to make notation easier as well as to compare different models together. In the modelling context, the kernel function, which is the covariance function, is the more important. It maps the input space onto a real value. The most commonly used kernel function in ML is the squared exponential function defined as,

$$k(x, x\acute{1}) = \sigma^2 \exp\left(-\frac{(x - x\acute{1})^2}{2l^2}\right) \tag{3}$$

where σ^2 and l are two hyper parameters. The length scale, l, defines the influence of the function value in length, while the variance, σ^2, specifies the height of the kernel.

It is common to many applications of regression that there is noise in the observations. Therefore, in GPR with noise, for a given training data set $D = \{(x_i, y_i | i = 1, \cdots\cdots, n)\}$, where $\times \in \mathbb{R}^{D\ x\ n}$ is the input data matrix and $y \in \mathbb{R}^D$ is the vector of output, the output y differs from the function values *f(x)* by the additive noise as [48],

$$y = f(x) + \varepsilon, \varepsilon\ N\left(0, \sigma_n^2\right) \tag{4}$$

where ε is the Gaussian noise with variance σ_n^2. Then, the new covariance function can be written as,

$$k(x, x\acute{1}) = \sigma^2 \exp\left(-\frac{(x - x\acute{1})^2}{2l^2}\right) + \sigma_n^2 \delta(x, x\acute{1}) \tag{5}$$

where $\delta(x, x\acute{1})$ represents the Kronecker delta function: i.e. $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

For a given training data × ($x \in \mathbb{R}^{D\ x\ n}$), corresponding observations y ($y \in \mathbb{R}^D$), and test data points x_* ($x \in \mathbb{R}^{D\ x\ n_*}$) which is needed to predict y_*, a joint distribution can be expressed as,

$$\begin{bmatrix} y \\ y_* \end{bmatrix} N\left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right) \tag{6}$$

where μ and μ_* are the means of training sets and test sets, respectively. K, K_* and K_{**} are the training set covariances, training-test set covariances and test set covariances, respectively.

$$\mu = m(x) \text{and} \mu_* = m(x_*) \tag{7}$$

$$K = K(x, x) = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix} \tag{8}$$

$$K_* = K_*(x_*, x) = \begin{bmatrix} k(x_*, x_1) & k(x_*, x_2) & \cdots & \cdots & k(x_*, x_n) \end{bmatrix} \tag{9}$$

$$K_{**} = k(x_*, x_*) \tag{10}$$

Since the values for the training set y are known, the conditional distribution of y_* given y can be expressed as,

$$p(y_*|y) = N(\mu_{2|1}, K_{2|1}) \tag{11}$$

where

$$\mu_{2|1} = \mu_* + K_* K^{-1}(y - \mu) \tag{12}$$

$$K_{2|1} = K_{**} + K_* K^{-1} K_*^T \tag{13}$$

If we assume the mean prior $\mu = 0$, then Eq. (12) can be rewritten as,

$$\mu_{2|1} = K_* K^{-1} y \tag{14}$$

By using the mean $\mu_{2|1}$ of the resulting distribution as a prediction, it is then possible to predict y_* corresponding to the test setsx_*.

3.2. Dataset

In this study, the ML model was trained and evaluated using the experimental dataset of compressive strength of MLWC, described in section 2.2. The compressive strengths of test specimens were evaluated at 7, 14 and 28 days after casting. The experimental dataset consisted of 414 specimens having different concrete mixture compositions. Some samples of the dataset are presented in Table 3.

It can be seen that each variable in the dataset has completely different range. For example, as mentioned in Section 2.2, the cement variable has a range from 75 to 750 kg/m^3 while another variable, like ACE 456 superplasticiser has a range from 0 to 8.7 kg/m^3. This will cause some issues in most of ML algorithms like GPR. The reason is most of ML models are based what is called the Euclidean distance. If one variable has much wider range of values, the Euclidean distance will be dominated by that variable. Therefore, the data needs to be scaled in order to understand how the variables move around their mean and thus standardize the effect of the movement of the variables. There are several ways of scaling the data. The very common one is the standardization which removes the dependence on arbitrary scales in the predictors and generally improves performance. In the developed ML model, each column of the predictor data is standardized so that they have mean 0 and standard deviation 1.

3.3. Cross-validation and performance evaluation

The most common model evaluation techniques are the train-test-split and k-fold cross-validation. In the train-test-split approach, the original dataset is randomly divided into training and testing sets. The training set is used to train the model and then testing set is used to test the model. In most of cases, two imbalanced datasets are used as training set and testing sets (80/20 or 70/30 proportions). Since these sets are randomly selected, if a small dataset is available, there is a huge possibility to miss out the interesting information about the data in the training set to learn an effective mapping of inputs to outputs. Therefore, keeping part of the data for testing would decrease the accuracy of the predictive model and hence, this approach is not appropriate when the original dataset is small.

As a solution, k-fold cross-validation technique is suitable when there is not a sufficiently large dataset available. In k-fold cross-validation, the original dataset is randomly divided into k number of sections (folds). In each of k folds of model building and validation, it chooses a different data subset for testing and training the model with the remaining k-1 data subsets. The appropriate error is simply calculated as a mean of all the k folds as a cross-validation error. The advantage of this approach is that all data in the dataset are eventually used for both training and testing.

In the present study, the 10-fold cross-validation algorithm ($k = 10$) was used to improve the model prediction in the ML model by using a MATLAB program [49]. This was achieved by using the inbuilt MATLAB function "Crossval" with scalar 10 [49]. In the 10-fold cross-validation, the dataset is randomly divided into ten non-overlapping groups (folds), as shown in Fig. 1. Each fold will have an equal number of data samples. In each of the ten validation rounds (iteration), nine of ten folds are used to train the ML model and the remaining fold is used to test the ML model. In the first iteration, the first fold is used as testing set while the rest are used to train the model. In the second iteration, the second fold is used to test the model and the rest serve as the training set. This process is repeated until each fold of the 10 folds has been used as the testing set.

The predictive accuracy of each validation round was evaluated in terms of errors in regression. Four of such common indicators such as the Mean Absolute Error (MAE), the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE) and this R-Squared (R^2) were used for this purpose.

MAE is the average of the absolute value of the error, which is very commonly used to measure the accuracy of predicted values. It is

Table 3
Some samples in the dataset of compressive strength of MLWC.

Mix proportions (kg/m^3)						W/C ratio	Form	Curing time (days)	Pre-treatment of Mxg	Compressive strength (MPa)
Mxg	Cement	NHL3.5	ACE 456	$CaCl_2$	Water					
130.92	518.23	284.09	5.21	11.60	363.64	0.7	Prism	7	Yes	6.364
144.04	569.97	312.56	5.65	12.78	299.91	0.53	Prism	14	Yes	15.187
144.04	427.44	312.57	5.60	8.00	399.96	0.94	Prism	14	Yes	5.037
144.04	570.03	0.00	5.66	12.68	399.95	0.7	Prism	14	Yes	8.260
75.00	624.45	342.30	6.15	13.95	438.15	0.7	Prism	28	No	10.509
225.00	558.75	306.30	5.55	12.45	391.95	0.7	Prism	28	No	2.474
300.00	525.75	288.30	5.25	11.70	369.00	0.7	Cube	14	No	0.416
222.75	450.00	483.15	8.70	19.65	315.75	0.7	Cube	28	No	0.262
67.80	750.00	147.30	2.70	6.00	526.20	0.7	Cube	28	No	13.371
162.45	642.90	352.50	6.30	14.40	321.45	0.5	Prism	14	No	2.798
155.85	616.50	337.80	6.15	13.80	369.90	0.6	Cube	14	Yes	3.556
144.00	569.70	312.15	5.70	12.75	455.70	0.8	Prism	28	No	19.388
138.75	548.85	300.75	5.40	12.30	493.95	0.9	Prism	28	No	15.642
172.05	591.75	236.70	6.75	15.15	477.60	0.81	Prism	7	Yes	9.552
156.87	591.75	295.88	6.14	13.85	435.52	0.74	Prism	7	Yes	6.625
141.75	591.75	355.05	5.55	12.45	393.45	0.66	Prism	28	Yes	13.367
126.55	591.75	414.23	4.95	11.17	351.35	0.59	Prism	14	Yes	13.198
144.06	570.01	312.53	5.64	12.72	399.95	0.7	Prism	14	Yes	13.984
113.00	449.00	247.00	4.49	10.00	228.00	0.51	Prism	28	Yes	5.648
113.00	449.00	247.00	0.00	10.00	594.00	1.32	Prism	28	Yes	4.560

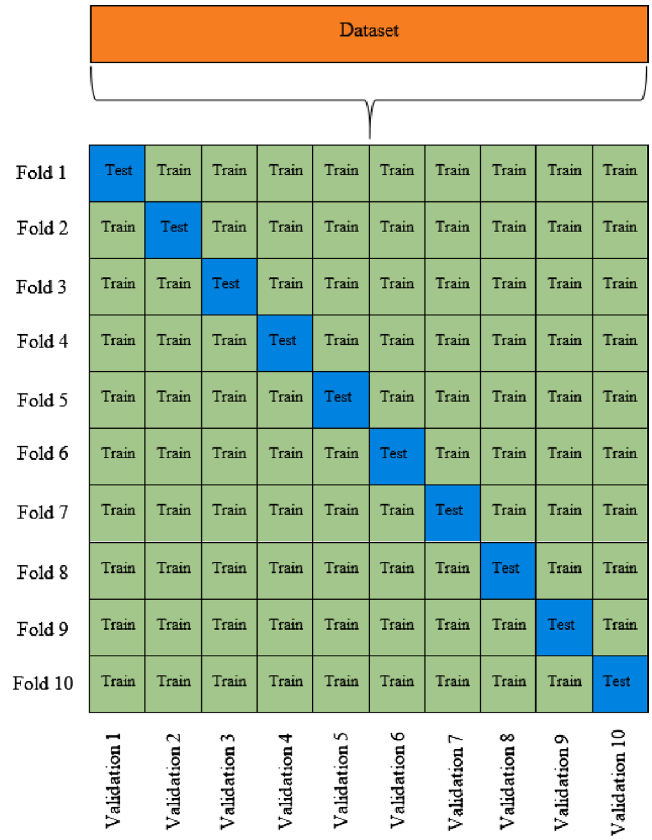


Fig. 1. 10-fold cross-validation algorithm.

defined as,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i^{'}\right| \tag{15}$$

where y_i' is the predicted value, y_i is the actual value and n is the number of observations.

MSE is the average of the squared value of the error. It is used as a default metric for the evaluation of the performance of most regression algorithms in MATLAB, and it is defined as,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2 \tag{16}$$

RMSE is the square root of the mean of the square of all of the error. It is defined as,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2} \tag{17}$$

R^2 represents the proportion of the variance in the dependent variable from the independent variables. It is defined as,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{18}$$

where $\overline{y}$ is the averaged actual value.

4. Results and discussion

The basic concept of this ML model is to establish a reliable nonlinear functional mapping between the compressive strength MLWC and its own mix proportions. The variables used in the model as predictors to relate with compressive strength of MLWC were: 1) six mix proportions (i.e. Miscanthus, cement, lime, superplasticizer, $CaCl_2$ and water), 2) form of specimen, 3) curing time, and 4) pre-treatment condition.

As described in Section 3.1, the predictive performance of GPR on a given data is highly dependent on the choice of the kernel function. The inbuilt MATLAB function "fitrgp" was used to fit a GP to the given data $\times$ and y [49]. There are a number of inbuilt kernel functions available in MATLAB. In this study, two kernel functions, namely, ARD squared exponential (ARD-SE) kernel and ARD rational quadratic (ARD-RQ) kernel were used by comparing the fit that can be achieved to determine which one best suits with the given data. Since more than one input (predictor) were used in the model, different length-scale parameters for each predictor are used in these two kernel functions. Here, ARD stands for automatic relevance determination and it implies that the estimation of the length scale parameters implicitly determines the relevance of each predictor.

ARD-SE kernel has the form [49]:

$$k(x,x') = \sigma^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{D}\frac{(x_i - x_i')^2}{l_i^2}\right) \tag{19}$$

where l_i is the length-scale parameter for predictor i and D is the number of predictor variables.

ARD-RQ kernel has the form [49]:

$$k(x,x') = \sigma^2\left(1 + \frac{1}{2\alpha}\sum_{i=1}^{D}\frac{(x_i - x_i')^2}{l_i^2}\right)^{-\alpha} \tag{20}$$

where α is a positive-valued scale-mixture parameter.

Fig. 2 shows the scatter plot of measured and predicted values of compressive strength of MLWC obtained from ML models for the dataset of 414 records. It reveals that both ML models have a good agreement between measured and predicted compressive strength of MLWC.

Fig. 3 shows the prediction outcome of the GPR between two kernel functions and actual values of the compressive strength of MLWC on the input dataset. Fig. 4 compares the predicted results from two developed models. These results emphasize that the prediction performances of the GPR rational quadratic and the GPR squared exponential models were almost similar. However, the validity of the proposed model for prediction of compressive strength of MLWC for any set of data was checked in terms of MAE, MSE, RMSE and R^2. The comparison of accuracy of the model is reported in Table 4. It can be seen that the GPR rational quadratic model has achieved a considerably good prediction result in all of the indicators compared to the other model.

The effect of each constituents of Miscanthus concrete mixture on the compressive strength of MLWC are illustrated in Fig. 5. These results emphasize that for both models the relationship between input variables of constituents of MLWC and the output of compressive strength of MLWC can be successfully mapped. In general, it can be seen that the compressive strength increases as the amount of each constituent increases. For example, compressive strength increases as the amount of cement increases. However, this is not true always but can be up to an extent, as if the W/C ratio keep unchanged, an increase of the cement induces also a proportional increase of the water volume. As a result, air voids in the hardened matrix are increased, which disrupt the bond between the Mxg and binder, and hence the compressive strength is decreased [1].

Since the GPR model can be successfully used to predict the compressive strength of the MLWC for different mix constituents of MLWC, a tool was developed using MATLAB [49] to deploy the model. This tool was developed to study the sensitivity of each variable on the compressive strength. It was developed with a simple user interface so that everyone can use it without knowledge in programming. The interface was divided into two sections as an input section to enter the mix proportions as required data and output section to display results for the entered data. The user interface of the developed tool for predicting the compressive strength of MLWC is shown in Fig. 6.

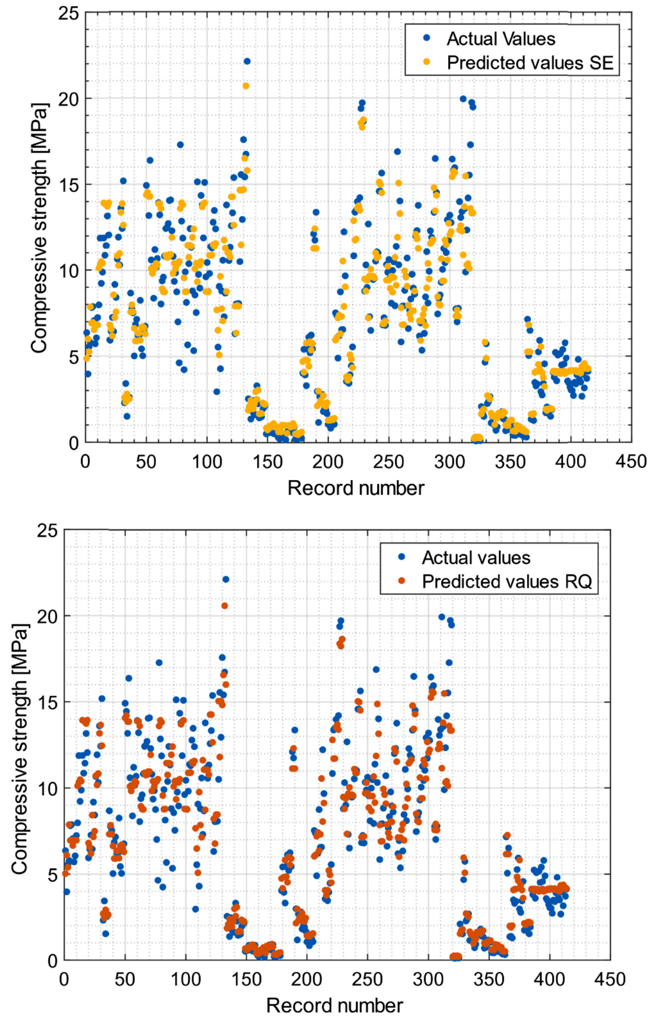


Fig. 2. Actual and predicted values of compressive strength of MLWC by using GPR models.

The main inputs of the tool are: the amount of Miscanthus, water, cement, lime, ACE456 (superplasticizer) and $CaCl_2$ in kg/m^3 as well as the form of the specimen, the curing time and the pre-treatment condition of Miscanthus. The users have the choice between two options for the form of the specimen, mainly prism or cube. The curing time can be picked to be 7, 14 or 28 days. The user can select pre-treated or not treated as pre-treatment condition. Moreover, the user may decide to use the RQ or SE kernel function for the GPR.

The developed tool can be used to check the different mixture combinations which will indicate the compressive strength for the given parameter combination. The influence of sensitivity of each input variable on the final prediction of the compressive strength can also be analysed using the graphs in the developed tool. It is achieved by randomly changing the values of the selected variable while keeping the remaining ingredients unchanged. The difference of the predictions when changing the value of the selected ingredient makes a sense to define the importance of each ingredient. If the difference in the predictions is high, that ingredient has a significant influence on the compressive strength. It was found that the cement content has the most significant influence on final compressive strength of MLWC. On the other hand, the amount of $CaCl_2$ is the least critical factor affecting the

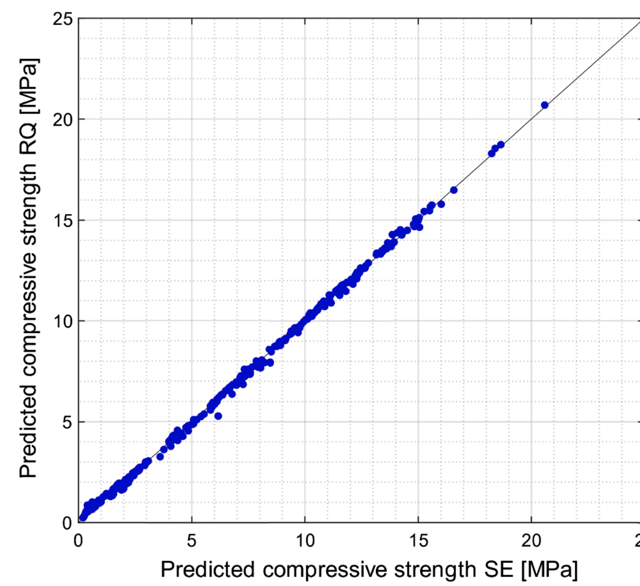


Fig. 4. Prediction results for the dataset by using GPR.

Table 4
Comparison of the accuracy of the model.

Indicator	GPR squared exponential	GPR rational quadratic
MAE	0.905	0.900
MSE	2.061	2.054
RMSE	1.436	1.433
R^2	0.917	0.918

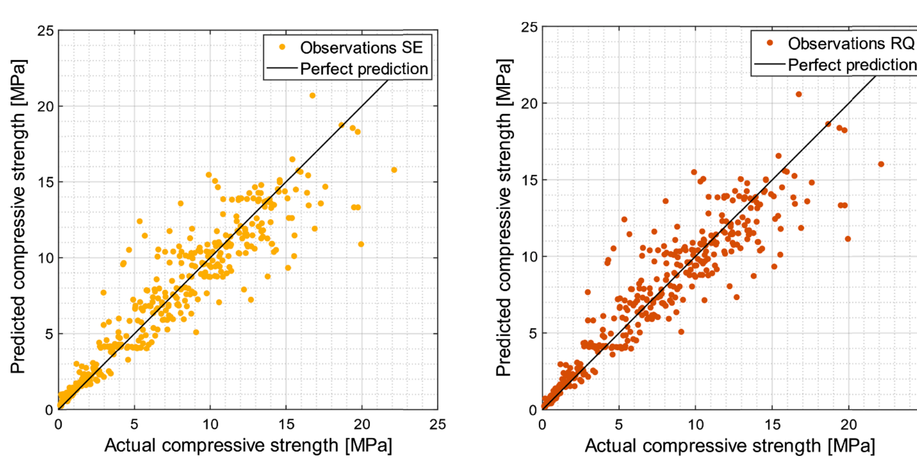


Fig. 3. Prediction results for the dataset by using GPR models.

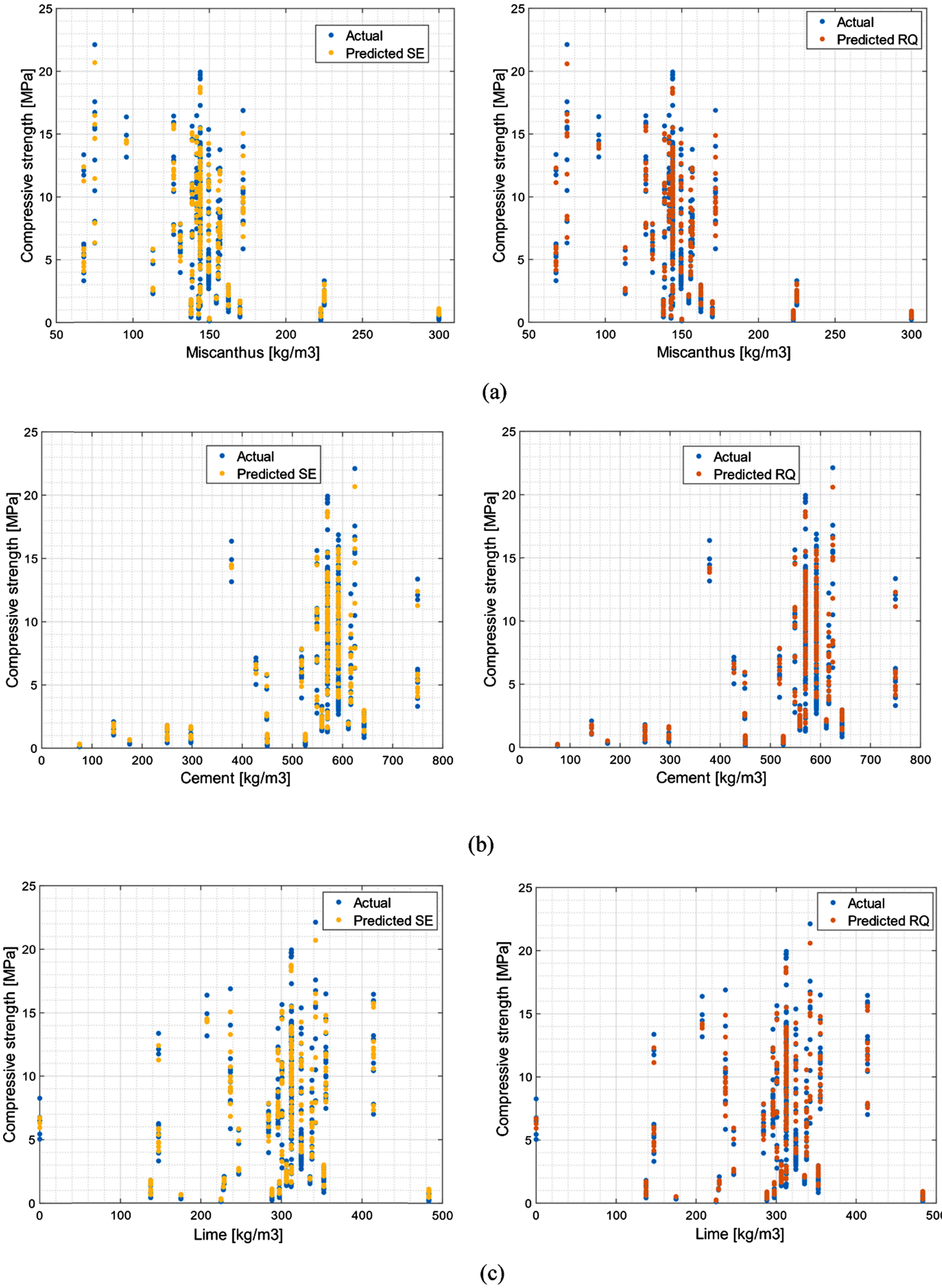
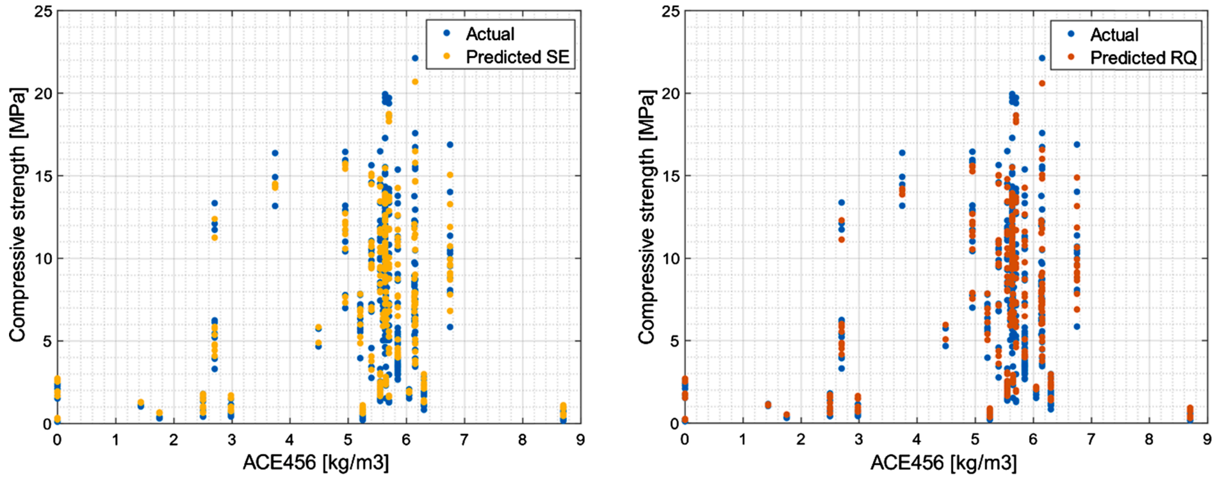
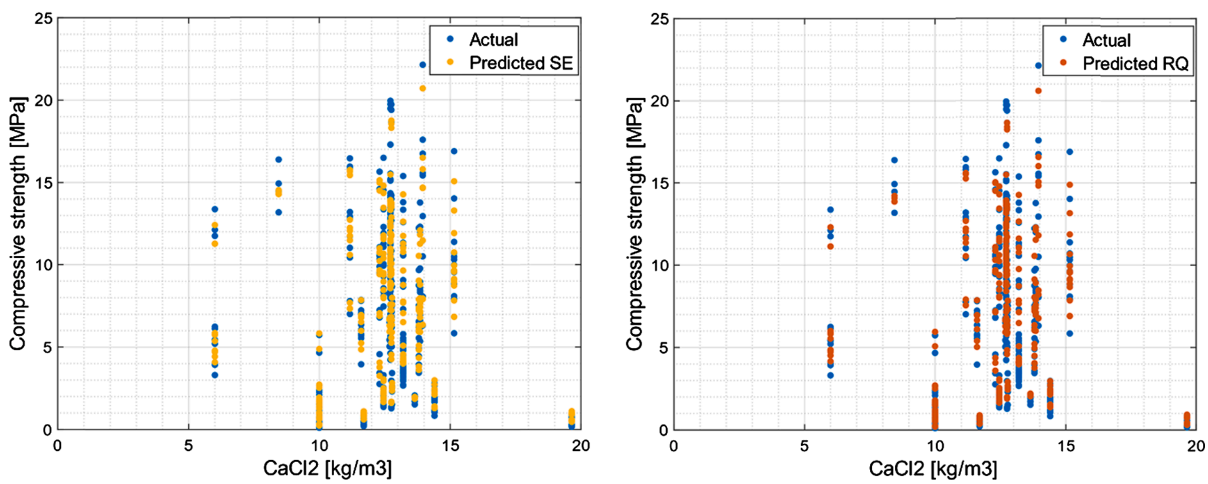


Fig. 5. The scatter plots of compressive strength of MLWC versus the input variables of (a) Miscanthus (b) cement (c) lime (d) ACE456 (e) $CaCl_2$ and (f) water.

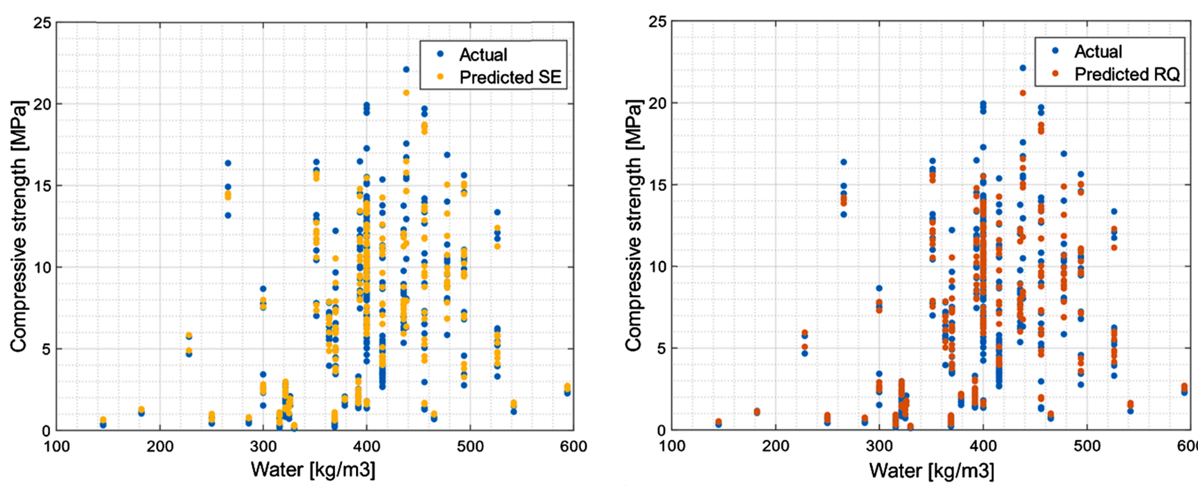
compressive strength of MLWC. Also, a given mixture can be optimised using the graphs in the developed tool. Thus, this tool can be used as a design tool in the construction industry as well as a support tool in research institutions.



(d)



(e)



(f)

Fig. 5. (*continued*).

5. Conclusion

This study proposed a ML model for predicting compressive strength of MLWC. For that, 414 sets of compressive test results were collected. Test variables were divided into nine inputs as mix constituents of concrete (cement, Miscanthus, lime, water, $CaCl_2$ and superplasticizer), form of specimen, curing time and pretreatment condition and one output variable (compressive strength of MLWC). In this paper, two GPR models with ARD-SE and ARD-RQ were used for the purpose of predicting the compressive strength of different MLWC mix designs. Four

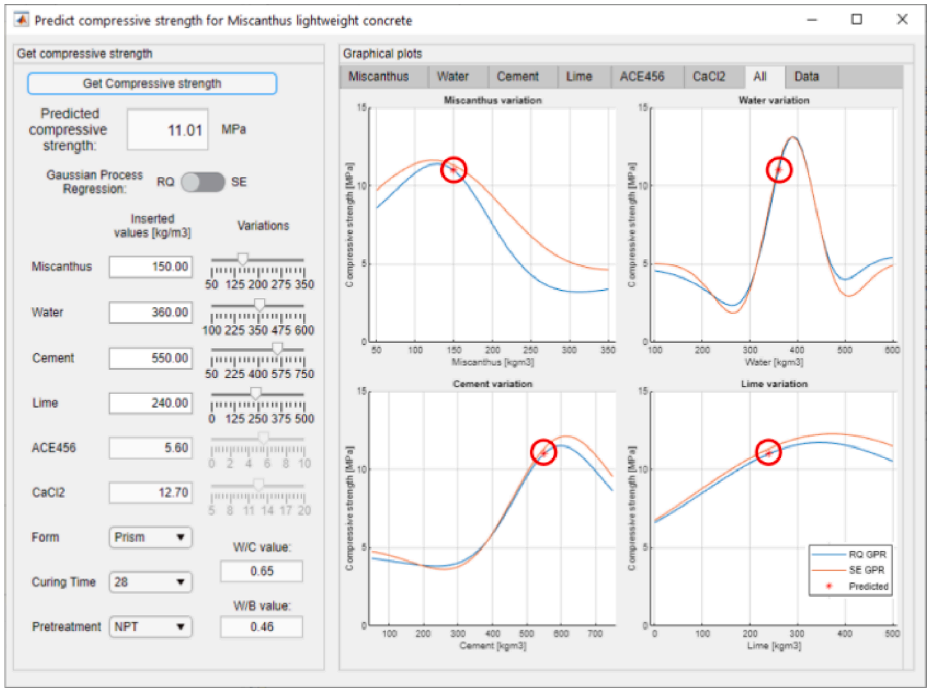


Fig. 6. The user interface of the tool.

indicators such as MAE, MSE, RMSE and R^2 were used to express the model accuracy. The overall performance of GPR model with ARD-RQ kernel function was more accurate to be applied to predict the compressive strength of MLWC.

A tool was developed to deploy the GPR model, which can be used to check the sensitivity of each constituents in the MLWC mixture on the compressive strength. It was revealed that cement, W/C ratio and Miscanthus are the most significant parameters on the final predictions of compressive strength of MLWC.

The results of this study provide helpful information on engineers to select the optimal mix proportions of the Miscanthus concrete, and the developed tool can be used at an early stage of designing the Miscanthus concrete members instead of using costly experimental investigation.

CRediT authorship contribution statement

Patrick Pereira Dias: Conceptualization, Methodology, Investigation, Resources, Formal analysis, Validation, Writing - original draft. **Laddu Bhagya Jayasinghe:** Investigation, Resources, Formal analysis, Validation, Writing - original draft. **Daniele Waldmann:** Conceptualization, Investigation, Resources, Supervision, Project administration, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P.P. Dias, D. Waldmann, Optimisation of the mechanical properties of Miscanthus lightweight concrete, Construction and Building Materials 258 (2020), 119643.
[2] L. Courard, V. Parmentier, Carbonated miscanthus mineralized aggregates for reducing environmental impact of lightweight concrete blocks, Sustainable Buildings 2 (2017) 3, https://doi.org/10.1051/sbuild/2017004.
[3] F. Tröger, G. Wegener, C. Seemann, Miscanthus and flax as raw material for reinforced particleboards, Industrial Crops and Products 8 (2) (1998) 113–121.
[4] A. Bourmaud, S. Pimbert, Investigations on mechanical properties of poly (propylene) and poly(lactic acid) reinforced by miscanthus fibers, Composites Part A: Applied Science and Manufacturing 39 (9) (2008) 1444–1454.
[5] J. Girones, L. Vo, S. Arnoult, M. Brancourt-Hulmel, P. Navard, Miscanthus stem fragment – Reinforced polypropylene composites: Development of an optimized preparation procedure at small scale and its validation for differentiating genotypes, Polymer Testing 55 (2016) 166–172.
[6] R. Muthuraj, M. Misra, A.K. Mohonty, Biocomposite consisting of miscanthus fiber and biodegradable binary blend matrix: compatibilization and performance evaluation, RSC advances 7 (2017) 27538–27548.
[7] J.A. Velsquez, F. Ferrando, J. Salvad, Binderless fiberboard from steam exploded Miscanthus Sinensis: optimization of pressing and pretreatment conditions, Wood Science and Technology 37 (3-4) (2003) 279–286.
[8] A. Eschenhagen, M. Raj, N. Rodrigo, A. Zamora, L. Labonne, P. Evon, H. Welemane, Investigation of Miscanthus and Sunflower Stalk Fiber-Reinforced Composites for Insulation Applications, Advances in Civil Engineering 2019 (2019) 1–7.
[9] R. Pude, C.-H. Treseler, G. Noga, Morphological, chemical and technical parameters of Miscanthus genotypes, Journal of Applied Botany and Food Quality 78 (1) (2004) 58–63.
[10] H. Acikel, The use of miscanthus (Giganteus) as a plant fiber in concrete production, Scientific Research and Essays 6 (13) (2011) 2660–2667.
[11] R. Pude, P. Banaszuk, R. Trettin, G.J. Noga, Suitability of Phragmites for lightweight concrete, Journal of Applied Botany and Food Quality 79 (2) (2005) 141–146.
[12] S. Elfordy, F. Lucas, F. Tancret, Y. Scudeller, L. Goudet, Mechanical and thermal properties of lime and hemp concrete ("hempcrete") manufactured by a projection process, Construction and Building Materials 22 (10) (2008) 2116–2123.
[13] Y. Chen, Q.L. Yu, H.J.H. Brouwers, Acoustic performance and microstructural analysis of bio-basedlightweight concrete containing miscanthus, Construction and Building Materials 157 (2017) 839–851.
[14] M. Kurpinska, L. Kulak, Predicting Performance of Lightweight Concrete with Granulated Expanded Glass and Ash Aggregate by Means of Using Artificial Neural Networks, Materials 12 (12) (2019) 2002.
[15] P. Chopra, R.K. Sharma, M. Kumar, Regression models for the prediction of compressive strength of concrete with & without fly ash, International Journal of Latest Trends in Engineering and Technology 3 (4) (2014) 400–406.
[16] M. Nikoo, P. Zarfam, H. Sayahpour, Determination of compressive strength of concrete using Self Organization Feature Map (SOFM), Engineering with Computers 31 (1) (2015) 113–121.
[17] R. Aggarwal, M. Kumar, R.K. Sharma, M.K. Sharma, Predicting Compressive Strengthof Concrete, International Journal of Applied Science and Engineering 13 (2) (2015) 171–185.
[18] H.-G. Ni, J.-Z. Wang, Prediction of compressive strength of concrete by neural networks, Cement and Concrete Research 30 (8) (2000) 1245–1250.
[19] F. Altun, Ö. Kişi, K. Aydin, Predicting the compressive strength of steel fiber added lightweight concrete using neural network, Computational Materials Science 42 (2) (2008) 259–265.
[20] I.-C. Yeh, Modeling of strength of high-performance concrete using artificial neural networks, Cement and Concrete Research 28 (12) (1998) 1797–1808.

[21] İ.B. Topçu, M. Sarıdemir, Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic, Computational Materials Science 41 (3) (2008) 305–311.
[22] C. Ozel, Prediction of compressive strength of concrete from volume ratio and Bingham parameters using adaptive neuro-fuzzy inference system (ANFIS) and data mining, International Journal of Physical Sciences 6 (31) (2011) 7078–7094.
[23] M. F.M. Zain, S. M. Abd, Multiple Regression Model for Compressive Strength Prediction of High Performance Concrete, Journal of Applied Sciences 9 (1) (2008) 155–160.
[24] J.L. Huang, P.Y. Huang, Z.W. Li, X. Zheng, Prediction of Bond Strength of FRP-Concrete Based on Multiple Linear Regression Method, Advanced Materials Research (2010) 3623–3628.
[25] I.-C. Yeh, Design of high-performance concrete mixture using neural networks and nolinear programming, Journal of Computing in Civil Engineering 13 (1) (1999) 36–42.
[26] M.M. Alshihri, A.M. Azmy, M.S. El-Bisy, Neural networks for predicting compressive strength of structural lightweight concrete, Construction and Building Materials 23 (2009) 2214–2219.
[27] A.F. Bingöl, A. Tortum, R. Gül, Neural networks analysis of compressive strength of lightweight concrete after high temperatures, Materials & Design 52 (2013) 258–264.
[28] S. Das, P. Pal, R.M. Singh, Prediction ofConcrete Mix Proportion using ANN Technique, International Research Journal of Engineering and Technology 2 (5) (2015) 820–825.
[29] M.A. Kewalramani, R. Gupta, Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks, Automation in Construction 15 (3) (2006) 374–379.
[30] F. Khademi, M. Akbari, S.M. Jamal, M. Nikoo, 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) (2017) 90–99.
[31] B.A. Omran, Q. Chen, R. Jin, Comparison of Data Mining Techniques for Predicting Compressive Strength of Environmentally Friendly Concrete, Journal of Computing in CivilEngineering 30 (6) (2016) 04016029.
[32] M. Pal, S. Deswal, Modelling pile capacity using Gaussian process regression, Computers and Geotechnics 37 (7-8) (2010) 942–947.
[33] R.R. Richardson, M.A. Osborne, D.A. Howey, Gaussian process regression for forecasting battery state of health, Journal of Power Sources 357 (2017) 209–219.
[34] A.A. Biaz, H. Ahmadi, F. Shariatmadari, M.A.K. Torshizi, A Gaussian process regression model to predict energy contents of corn for poultry, Poultry Science 99 (11) (2020) 5838–5843.
[35] S.-S. Jin, Composition kernel learning using tree-based genetic programming for Gaussian process regression, Strucutral and Multidisciplinary Optimization 62 (2020) 1313–1351.
[36] S. Dutta, P. Samui, D. Kim, Comparion of machine learning techniques to predict compressive strength of concrete, Computers and Concrete 21 (4) (2018) 463–470.
[37] F. Özcan, C.D. Atiş, O. Karahan, E. Uncuoğlu, H. Tanyildizi, Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete, Advances in Engineering Software 40 (9) (2009) 856–863.
[38] L. Tabor, J.A. Goulet, J.-P. Charron, C. Desmettre, Probabilistic modeling of heteroscedastic laboratory experiments using gaussian process regression, Journal of Engineering Mechanics 144 (6) (2018) 04018038.
[39] B.S. En, 197–1,, Cement. Composition, specifications and conformity criteria for common cements, BSI (2011).
[40] B.S. En, 459–1, "Building lime. Definitions, specifications and conformity criteria", BSI (2015).
[41] J.S. Pozo-Antonio, Evolution of mechanical properties and drying shrinkage in lime-based and lime cement-based mortars with pure limestone aggregate, Construction and Building Materials 77 (2015) 472–478.
[42] M.J. Mosquera, B. Silva, B. Prieto, E. Ruiz-Herrera, Addition of cement to lime-based mortars: Effect on pore structure and vapor transport, Cement and Concrete Research 36 (9) (2006) 1635–1642.
[43] D. Waldmann, V. Thapa, F. Dahm, C. Faltz, in: Perennial Biomass Crops for a Resource-Constrained World, Springer International Publishing, Cham, 2016, pp. 273–295, https://doi.org/10.1007/978-3-319-44530-4_23.
[44] Xueyu Pang, Peter Boul, Walmy Cuello Jimenez, Isothermal calorimetry study of the effect of chloride accelerators on the hydration kinetics of oil well cement, Construction and Building Materials 77 (2015) 260–269.
[45] B.S. En, 196–1, "Methods of testing cement. Determination of strength", BSI (2005).
[46] B.S. En, 12390–7, "Testing hardened concrete. Density of hardened concrete", BSI (2019).
[47] Nhat-Duc Hoang, Anh-Duc Pham, Quoc-Lam Nguyen, Quang-Nhat Pham, Estimating Compressive Strength of High Performance Concrete with Gaussian Process Regression Model, Advances in Civil Engineering 2016 (2016) 1–8.
[48] Mengwu Guo, Jan S. Hesthaven, Reduced order modeling for nonlinear structural analysis using Gaussian process regression, Computer Methods in Applied Mechanics and Engineering 341 (2018) 807–826.
[49] Mathworks,, Statistics and Machine Learning Toolbox, The MathWorks (2021).