Article

Classification-Based Regression Models for Prediction of Mechanical Properties of Roller Compacted Concrete Pavement

Ali Ashrafian[1], Mohammad Javad Taheri Amiri[2], Mahsa Asadi-shiadeh[3], Isa Yaghoobi-chenari[4], Amir Mosavi[5,6*], Narjes Nabipour[7*]

[1] Department of Civil Engineering, Tabari University of Babol, Iran, P.O. Box 47139-75689, Babol, Iran; Ali_ashrafian@yahoo.com
[2] Department of Civil Engineering, higher education institute of Pardisan, Freidonkenar, Iran, Jvd.taheri@gmail.com
[3] Department of Civil Engineering, Tabari University of Babol, Iran, P.O. Box 47139-75689, Babol, Iran; en.mahsa.asadi@gmail.com
[4] Department of Civil Engineering, Shomal University, Amol, Iran; mojtaba.ygb@yahoo.com
[5] Faculty of Civil Engineering, Technische Universität Dresden, 01069 Dresden, Germany;
[6] Kalman Kando Faculty of Electrical Engineering, Obuda University, 1034 Budapest, Hungary
[7] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
* Correspondence: amir.mosavi@kvk.uni-obuda.hu, narjesnabipour@duytan.edu.vn

Abstract: In the field of pavement engineering, the determination of the mechanical characteristics is one of the essential process for reliable material design and highway sustainability. Early determination of mechanical characteristics of pavement is highly essential for road and highway construction and maintenance. Tensile strength (TS), compressive strength (CS) and flexural strength (FS) of roller compacted concrete pavement (RCCP) are very crucial characteristics as they are necessitated for many data from mixture proportions as input variables. In this research, the classification-based regression models named Random Forest (RF), M5rule model tree (M5rule), M5prime model tree (M5p) and Chi-square Automatic Interaction Detection (CHAID) are developed for simulation of the mechanical characteristics of RCCP. A comprehensive and reliable dataset comprising 621, 326 and 290 data records for CS, TS and FS experimental cases extracted from several open sources over the literature. The mechanical properties are developed based on influential inputs combination that processed using Principle Component Analysis (PCA). The applied PCA method as feature selection is specified that volumetric/weighted content forms of experimental variables (e.g., coarse aggregate, fine aggregate, supplementary cementitious materials, water and binder) and specimens' age are the most effective inputs to generate the better performances. Several statistical metrics are measured to evaluate proposed classification-based regression models. RF model revealed an optimistic classification capacity of the CS, TS and FS prediction of the RCCP in comparison with the CHAID, M5rule, and M5p models. The research is extended for the results verification using Monte-carlo model for the uncertainty and sensitivity of variables importance analysis. Overall, the proposed methodology indicated a reliable soft computing model that can be implemented for the material engineering construction and design.

Keywords: Roller compacted concrete pavement; Classification-regression models; Feature selection; Mechanical properties; Monte-Carlo uncertainty

1. Introduction

In this technologically advanced world, along with the advances in various scientific fields, the concrete industry has also grown, and such advances have resulted in the production of roller-compacted concrete pavement (RCCP). In recent years, the construction and maintenance of road pavements has become an important challenge [1, 2]. The high cost of producing bituminous pavement and the quantity of petroleum contaminants of environment necessitate the use of alternative technologies in the road problems [3]. The lower cement paste content and higher aggregate volume in RCCP have led to its low consistency, which results in the longer durability of RCCP than bituminous asphalt. Also, higher temperature rise resistance, lower water absorption, better compressive strength and less deformation under loads in the long run are other advantages of RCCP. In the cold regions, RCCP is also resistant to frost cycles when facing possible damages [4]. In addition, due to the impermeability of the constituent materials, it acts as an environmentally friendly pavement and presents no problem in the used regions. The use of pozzolanic materials to ensure sufficient compaction in the mixtures with standard fine-grained aggregates in the production of RCCP has also attracted the interest due to lower production costs than cement and improved strength properties [5, 6]. Therefore, this study explores the RCCP mixtures containing pozzolan. Pozzolans are well mixed with the gels produced in the concrete and increases the concrete hydration, thereby increases the density of produced concrete and enhances the chemical and mechanical properties of RCCP.

The important mechanical characteristics of concrete are highly influenced by the concrete mix design [7]. The parameters such as cement content, water-to-cement ratio, cement substitutes, and so on affect the mechanical properties of concrete, which makes it difficult to predict the mechanical properties of concrete due to the presence of numerous parameters. In the mix design methods, the cost of production is tried to be reduced. It is time consuming and costly to use the regulation methods for the calculation of the mix design and it is necessary to comply with the conditions and assumptions of the regulations for all constituent materials of concrete [8-10]. Therefore, different researchers have presented valuable models using the different mathematical techniques to estimate the concrete behavior, which have mainly been based on linear and nonlinear regressions. Nowadays, methods based on Machine Learning (ML) have been successfully used in this field and these models have generally stemmed from laboratory experiments and analyses.

Up to date, various ML techniques that have been focused to simulate the mechanical characteristics of concretes including Multivariate Adaptive Regression Splines (MARS) [11], Genetic Expression Programming (GEP) [12], Artificial Neural Network (ANN) [13], Adaptive Neuro-Fuzzy Inference Systems (ANFIS) [14] and Support Vector Machines (SVM) [15]. For instance, Ashrafian et al. developed an evolutionary method based on MARS integrated water cycle algorithm to propose non-linear relationship between mixture components and compressive strength of foamed cellular lightweight concrete [16]. Hardened strength estimation of recycled aggregate concrete using traditional ANN system was considered in deng et al. study [17]. Sun et al. proposed extended SVM model to estimate permeability coefficient and unconfined compressive strength [18]. Shahmansouri et al. applied GEP method to simulate hardened characteristics and electrical resistivity of zeolite based eco-friendly concrete [19]. Feng et al. implemented intelligent ML method named an adaptive boosting approach for estimating the compressive strength of concrete [20]. Iqbal et al. focused on comprehensive data to present simple and robust model to formulate mechanical characteristics of green concrete using GEP approach [21]. Asteris et al. used data driven methods for hardened properties of self-compacting concrete prediction as surrogate models [22].

Although, aforementioned ML methods provide reliable and robust tools for concrete properties modeling, they are complex and computationally costly during the learning phase. As such, classification-based regression methods as extended ensemble ML tools has the attractive concept of few setting parameters for developing of models and robust resistance to overfitting [23]. They have

become increasingly implemented for regression challenges because they are relatively simple, straightforward, flexible and have relatively low-cost computational process [24]. In this case, Behnood et al. formulated mechanical properties of poplar concretes based on tree method [25, 26]. Han et al. proposed improved RF model to simulate CS of high-performance concrete.[27]. Mohamed used the RF technique to approximate hardened properties of sustainable concrete [28]. Ashrafian et al. evaluated tree-based heuristic regression model named M5p model tree to predict properties of fiber reinforced concrete [29]. Gholampour et al. [25] applied M5 model tree to estimate the mechanical properties of coarse recycled aggregate concrete and reported the influential predictor variables [30].

The main goals of this study are considered: (1) development and evaluation of nonlinear decision tree-based classification methods including model tree M5rule (M5rule), Chi-square Automatic Interaction Detector (CHAID), RF and M5p to simulate mechanical characteristics of RCCP (e.g., CS, TS and FS). (2) improvement of proposed regression-based models using optimized technique named principal component analysis (PCA) for selection of better contribution of the predictor variables, (3) comparison of proposed performances and integration of the advantages of the decision tree-based classification methods to build and evaluate proposed models. 4) Another focus of this attempt was to present new ensemble-based method named CHAID for mechanical characteristics estimation of RCCP for the first time in concrete technology prediction that could potentially lead to enhance estimation capability.

This research is organized into four different sections. In the present study, an introduction that describe research significant and literature review (Section 1) followed by Section 2 which propose a materials and methods for RCCP background, experimental dataset and describes the details of the investigated methods. We then present modeling process, training and testing phase and comparison of developed models in Section 3. Finally, section 4 summarizing the research findings.

2. Materials and Methods

2.1. Theoretical background and data description

Proper blends designing is challenging problem in producing a high-quality concrete mixture [31]. There are a number of proportions that should be attentional to design RCCP blends and they consisting of mechanical characteristics, economical benefit and project constructability [32]. Among of types of concrete, RCCP has become very conventional due to the fact that it is a simple production and it can be located fast while producing a great surface. RCCP blends incorporating a fewer cement weight (110-120 kg/m^3), utilized natural aggregate of concreting quality, and has been specified by the American Concrete Institute (ACI) standard 325-10R-95 as concrete incorporating fewer water, cement and supplementary cementitious material in comprising of ordinary concrete [33].

The comprehensive and integrate dataset is utilized for the building up of reliable and adoptable simulation models based on ML techniques. Hence, database was compiled from the open-source researches available in literature [34-55]. Among this database, the models of mechanical characteristics of RCCP were developed using 621, 326 and 290 data records for CS, TS and FS of RCCP, respectively, at ages 1, 3, 7, 28, 90 and 180 days. The gathered datasets containing information about mixture components of RCCP in different inputs combinations. For the ML techniques, the originally collected experimental data is randomized and categorized into two phases. The training (calibration) phase is implemented for learning procedure and constructed the model for CS, TS and FS properties. The testing (validation) phase is performed to evaluate the capability criteria of the presented RCCP characteristics models. For the development of the proposed methods, 75% of the data (466, 245 and 218 data records) for CS, TS and FS respectively, were prepared for the training phase while the remainder (155, 81 and

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

72data records) were used for testing phase of the classification-based regression methods. A schematic work flow for the simulation procedure of mechanical characteristics using ML-based models is presented in Fig. 1

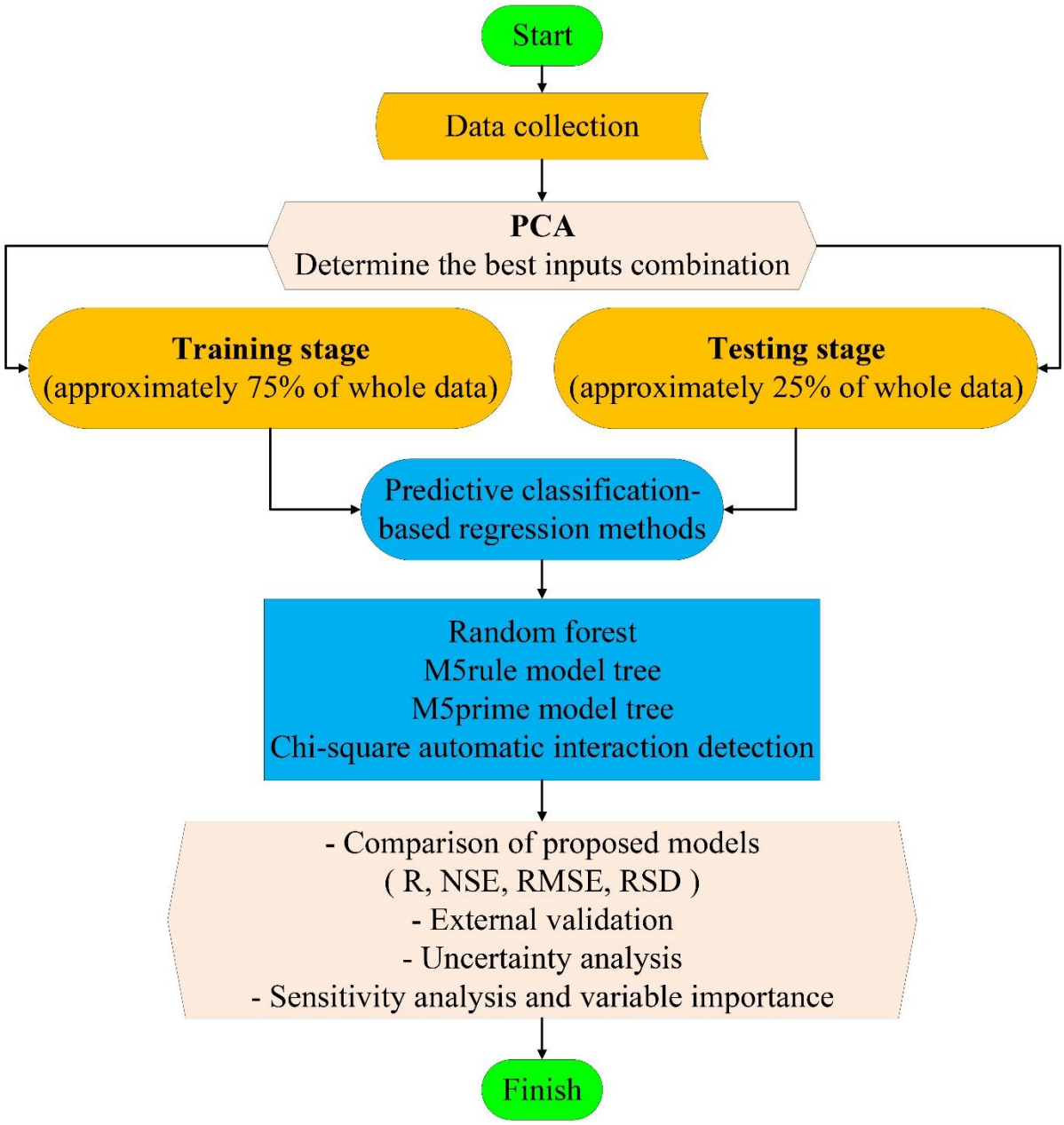

Fig. 1: Work flow for the procedure of this study.

2.2. Random forests, RF

Breiman [56] proposed the RF which is a nonparametric and classification-based regression methods [57, 58]. Instead of parametric models, many easy-to-interpret decision trees are incorporated in RF model. By integration of the decision tree models results a more comprehensive estimation technique could be attained. The proper objective of the research is estimation of the mechanical properties of RCCP through just discussing the regression approach. The steps of training performance in RF are as following [56-58].

(a) Based on dataset draw an instance which is chosen randomly with substitution.

(b) Using the bootstrap instance evolve a tree with these modifications: per each node select the best randomized subset is selected of m try descriptors (i.e. the number of predictors tried per each node). M try here has the role of a tuning parameter in the RF algorithm. The tree is generated to its maximum size without pruning it.

(c) stage (b) is iterated till the user-manual numbers of trees (ntree) are grown that are on the basis of the bootstrap instance of observations. The final prediction values are determined by combining all individual trees outcomes [56]. After growing the K trees {Tk(x)}, the regression explanatory variables in RF is stated by the following formula:

$$f(x) = [\sum_{K=1}^{K} T_k(x)]/K. \quad (1)$$

A new training set per each constructed RF regression tree is derived by replacing from the original calibrating phase. Thus, after constructing a regression tree at each time, through application of randomized training sample, the out-of-bag instance is utilized for validating its precision [56].

$$GI(t_{X(xi)} = 1 - \sum_{j=1}^{m} f(t_{X(xi),j})^2 \quad (2)$$

The validation features improve the robustness of random forests which is due to the use of independent test data. Also, random forests algorithm is a feasible method for classification and regression purposes and has many engineering applications such as forecasting the concrete properties [58].

2.3. M5 Rule model tree, M5rule

The complex or hidden information in a dataset can be explored using the IF-THEN rules-based M5rule model tree which is a commonly used model in machine learning for classification and regression tasks [59]. The M5rule model can create a single classification tree through repeated data splitting into groups while ensuring the uniformity in the output and applying some decision rules that are applicable to specific explanatory parameters [60]. The uniformity of the output can be estimated as the residual sum of the squares. The first stage involves the selection of the input variable for node splitting that ensures the maximum uniformity of the resulting child nodes from the original parent nodes. Then, the nest step is the section of the other input variables as the child nodes [61]. Having constructed the optimal regression tree, the next thing is to prune the tree to prevent over-fitting and for this purpose, a cross-validation process is applied for the selection of the model with the least prediction error.

2.4. M5 prime Model tree, M5p

M5p model, which is based on linear regressions and decision trees, was first developed in 1992. A binary decision tree consists of the primary terminal node with extra leaf nodes, which provide a connection between input (independent) and output (dependent) parameters [62]. It is essential to bear in mind that decision trees are generally applied for categorical type data while it is appropriate for quantitative type data [63]. M5p model can be summarized in two main steps as follows: step a) splitting input data to create a decision tree; it is reached when defining the standard deviation of each subset to find an appropriate primary node (parent node). Because of this step(splitting), the sd of child note would be smaller than the parent node.

Step b) testing each node in the decision tree to diminish the error. The standard deviation is calculated as:

$$SDR = sd\ (T) - \sum \frac{|T_j|}{|T|} sd\ (T_j) \quad (3)$$

Where sd represents the standard deviation, T is a set of examples that reach the primary node, Tj represents the subset of patterns that possess the jth outcome of the potential set.

Thus, as stated above, based on different processes of splitting input data, the most probable error reducing node would be chosen. For the over-fitting problem in decision trees, pruning techniques were used for omitting sub-trees. This pruning technique is based on methods of linear regression functions. One of the strengths of this model over the M5rule model is its efficiency in learning and treating problems with high complexity. One of the features of this model is that its regression functions do not have many variables. The M5p model has widespread applications in engineering, medical and agricultural disciplines

2.5. Chi-square automatic interaction detector, CHAID

This CHAID model was first introduced by Kass for use in qualitative and classified quantitative variables [65]. As a modeling approach, this algorithm is suitable for establishing the relationship between a dependent parameter and several independent parameters. The CHAID model is mainly characterized by the following: (1) finding the influential parameters in the final result by applying chi-square test of independence; (2) useful in the combination of the effective variable groups [66]. This implies that CHAID employs Chi-squared independence test to examine the significance of independent parameters within a classification in comparison to the dependent parameters [67]. The Chi-square statistic is expressed as follows:

$$X^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4)$$

where O_{ij} is the observed value while E_{ij} is the predicted value. There are 3 stages in the CHAID model; these are merging, splitting, & stopping. The merging phase involves the application of the Chi-square test to test the significance of each independent parameter. Each pair of dependent & independent parameters, as well as the probable tables are subjected to this test. For the splitting stage, it initiates with the comparison of the calculated p-values of each independent parameter with the independent parameters that have the least P-value, followed by their selection as the node separator. in situations where no variable has a significant P-value, there will be no splitting stage and the final node will be accounted as the node that precedes no branching [68]. The last stage (the stopping stage) begins with a repeat of the combination and analysis stages of all subsets. The process is terminated after all the subsets have been analyzed [66].

The formation of different parts in the CHAID model is represented by a classification tree diagram, where each dependent parameter is represented by a root while the independent parameters are associated with significant P-values and are directly related with the root [69]. The weakness of this algorithm is that it cannot generate the best feasible divisions from the current parameters. More information on CHAID has been provided by [65-68].

2.6. Principal component analysis, PCA

Issues such as high dimensional input space, variables correlation, insufficient training samples can create problems in the learning process and the conditions might become worse when we want spatially interpolate values for various locations within a city but with few observation points [70]. So, it becomes inevitable to implement dimension reduction methods to reduce the number of many correlated variable into the uncorrelated ones. Through application of PCA, while maintaining the highest variation and dispersion in the data, one could transform the input variables into a set of new uncorrelated variables

called the principal components [71, 72]. Eqs. (5) and (6) are used to provide linear transformation from the input space to the principal component space. Here, orthogonal linear transformation matrix is defined by P, Z represents the original data matrix according to which each row denotes a variable, and Y represents the matrix of transformed. In this matrix each row denotes the uncorrelated principle components.

$$PZ = Y \tag{5}$$

$$\begin{bmatrix} P_{1.T_1} & \dots & P_{1.T_m} \\ & \vdots & \\ P_{m.T_1} & \dots & P_{m.T_m} \end{bmatrix} \begin{bmatrix} ZT_1(x_1) & \dots & ZT_1(x_n) \\ & \vdots & \\ ZT_m(x_1) & \dots & ZT_m(x_n) \end{bmatrix} = \begin{bmatrix} y_1(x_1) & \dots & y_1(x_n) \\ & \vdots & \\ y_m(x_1) & \dots & y_m(x_n) \end{bmatrix} \tag{6}$$

The transformation matrix (P) is obtained from the eigenvalues (λ1, λ2, … , λ1) of the covariance matrix of the original variables, by applying PCA. The rows of this matrix represent the corresponding eigenvector. The eigenvectors specify the directions of the new space, and the eigenvalues specify their magnitude [72]. In order to find which eigenvector (s) could be removed without much affecting the information needed for building a subspace with lower dimensions, we should inspect their corresponding eigenvalues. Those eigenvectors which have smaller corresponding eigenvalues are those that have lower information on the data distribution and can be removed.

2.7. Statistical criteria

In the present research the following performance metrics (Eqs. 7-10) were applied which are: The correlation coefficient (R), Nash-Sutcliffe Efficiency (NSE), Root Mean Square Error (RMSE), Ratio of RMSE to standard deviation (RSD), [21,57–59]:

$$R = \frac{\sum_{i=1}^{N}(t_{exp} - \overline{t_{exp}}).(t_{pre} - \overline{t_{pre}})}{\sqrt{\sum_{i=1}^{N}(t_{exp} - \overline{t_{exp}})^2 \sum_{i=1}^{N}(t_{pre} - \overline{t_{pre}})^2}} \tag{7}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(t_{pre} - t_{exp})^2}{\sum_{i=1}^{N}(t_{exp} - \overline{t_{exp}})^2} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_{pre} - t_{exp})^2} \tag{9}$$

$$RSD = \frac{RSME}{\sum_{i=1}^{N}(t_{exp} - \overline{t_{exp}})} \tag{10}$$

In the above equations t_{exp} and t_{pre} denote the experimental and predicted target variable values, respectively. $\overline{t_{\text{exp}}}$ and $\overline{t_{\text{pre}}}$ are the mean of experimental and predicted target variable values, respectively. Also, N denotes the total number of data. The R index which is in the range of (0,1) (with R=1 as the ideal value) shows the selected predictors suitability in predicting the target variable. NSE with the range of (-∞, 1) and ideal value equal to unity is used for assessing the capability of proposed methods. Therefore, a value equal to unity shows perfect fitting between the actual and measured target values ad a negative value means bad performance of the model with respect to the arithmetic mean of used models. RMSE and RSD with the range of (0, +∞) and ideal value of zero are used to assess the accuracy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0233.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

3. Application results and discussion

3.1. Selection of the input variables using PCA technique

In this paper, to propagate the most effective combination of inputs for the simulation matrix of the mechanical characteristics, Principal Components Analysis (PCA) based on dimensionality reduction was performed. The predictor variables affecting mechanical characteristics of RCCP in different ages are described as bellow:

$$f_c = f(CA, FA, C, SCM, B, W, \frac{W}{C}, \frac{W}{B}, \frac{SCM}{B}, \frac{CA}{FA}) \quad (11)$$

where *CA* (Kg/m^3) , *FA* (Kg/m^3), *C* (Kg/m^3), *SCM* (Kg/m^3), *B* (Kg/m^3), *W* (Kg/m^3), *W/C, W/B*, SCM/B and CA/FA are the coarse aggregate content, fine aggregate content, cement content, supplementary cementitious material content, binder content, water content, ratio of water to cement, ratio of water to binder, ratio of supplementary cementitious material to binder and ratio of coarse to fine aggregate, respectively. Table 1 reported the results of analysis consisting the contribution of 10 inputs to 10 PCs, the explained variance (EV) of each PCm, and the cumulative sum (CS) of EV. As a result. PC1 represents 51.3% and the first 4 PCs deals 99.1% of total variance. The higher EV, the more dimension of the best combination of inputs is manifested. The PCA as dimension-reduction tool pre-processed dataset and determined most influential contribution of model's development.

Table 1: Principal component analysis results to select optimal inputs combination.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC8	PC9	PC10
CA	**0.262**	-0.942	0.187	0.086	-0.028	-0.001	0.000	0.000	0.000
FA	**-0.959**	-0.244	0.054	0.113	-0.065	-0.007	0.000	0.000	0.000
C	0.011	0.168	0.777	0.151	0.108	0.007	0.003	0.000	0.001
W/B	0.000	0.000	0.000	-0.001	0.000	0.007	-0.113	-0.993	0.019
SCM	**0.046**	-0.043	-0.546	0.600	0.069	0.004	-0.002	0.000	-0.002
W	**0.072**	0.082	0.081	0.187	-0.971	-0.056	-0.003	0.000	0.001
B	**0.057**	0.125	0.231	0.750	0.178	0.011	0.001	-0.001	-0.001
W/C	0.000	0.000	-0.003	0.001	-0.004	0.001	0.986	-0.115	-0.117
SCM/B	0.000	0.000	-0.002	0.001	0.000	0.000	0.118	0.006	0.993
CA/FA	-0.003	0.000	0.000	0.000	-0.058	0.998	0.000	0.007	0.000
EV	0.513	0.326	0.098	0.054	0.008	0.000	0.000	0.000	0.000
CS	0.513	0.839	0.937	0.991	1	1	1	1	1

The influential combination of mixture proportions is introduced using Eq. (11) by PCA technique where presented in Table 1. It specified that consist of five predictors were adequate to prepare the higher explained variances. Table 1 presented the values of PCs and their variances of inputs. According to Table 1, it can be seen volumetric and weighted form of experimental variables that *CA*, *FA*, *SCM*, *W*, *B* based on PC1 are the most effective independent predictors variables that affecting on dependent outputs. Therefore, this combination of simulation variables along with age of specimens (*AS)* is used to construct proposed models to predict the mechanical characteristic of the concrete. The descriptive measures of the best combination of inputs for simulation of the mechanical characteristics of RCCP are presented in Table 2. Also, the correlation coefficients of selected independent variables for development

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0233.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

of proposed models have been captured on heatmap plot presented in Fig. 2. According to the matrix, there are not significant relationships between the developed matrixes of CS, TS and FS.

Table 2: Statistical measures of independent and dependent variables for CS, TS, and FS properties

Variables	Mean	Standard deviation	Median	Kurtosis	Skewness	Minimum	Maximum
CA	1014.9	184.2	1095	-0.68	-0.62	585	1325
FA	855.87	225.7	807	-0.13	-0.22	272.5	1263
SCM	86.26	72.23	90	-0.7	0.44	0	272.5
W	129.29	39.57	117	7.5	2.26	78	336.25
B	311.6	66.44	295	8.34	2.12	200	672.5
AS	35.54	42.55	28	2.25	1.6	1	180
CS	33.276	16.553	31.4	-0.46	0.38	1.88	83
TS	3.1828	1.2761	3.2	-0.25	0.08	0.14	6.4
FS	4.498	1.864	4.55	-0.47	0.07	0.4	8.9

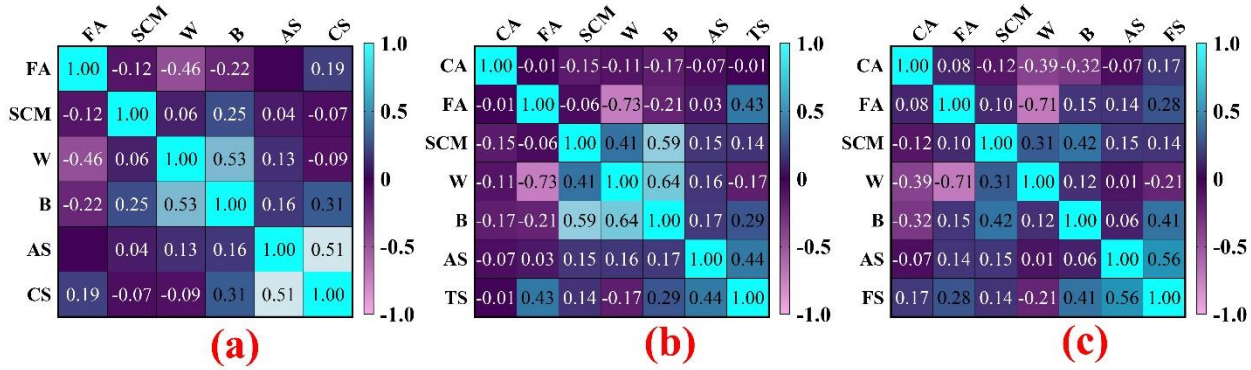

Fig. 2: Correlation matrix of inputs and outputs; (a): CS, (b): TS, (c): FS

3.2. *RCCP mechanical characteristics estimation using classification-based regression methods*

Application of Decision Tree classification System which is based on the artificial intelligence is a recent method proposed for solving engineering problems. The final properties of models are recaptured on the basis of the network calibrating. Then the network could generalize those learned in a similar condition [62]. In the present study, the modelling methods included are four classification-based regression methods namely the RF, CHAID, M5rule, and M5prime were explored for the prediction of characteristics of RCCP (the four methods are classified as ML-based models).

Definition of the matrix consisting of *CA, FA, SCM, W, B and AS* datasets indicated the independent variables, and the dependent variables, were CS, TS and FS that used in each decision tree-based regression models. RF, M5rule and M5p were performed using the *WEKA 3.9* and CHAID was implemented using *STATISTICA* software on an AMD A-12 9700, 10 computes core 2.5 GHz computer system.

To implement the *RF* model, the default of the Bagger algorithm was used in condition of bag size percent tuned at 200, the leaf number was set to eight and delta criterion set to 0.1007. It is noticeable, no generally mathematical formulation is utilized to fine the optimum number of trees. Commonly, a larger number of trees produces more precision results but increases computational cost.

The M5tree procedure for simulation of RCCP properties was generated using a set of tuning parameters to initialize of proposed model. A pruning factor 4.0 and smoothing option were selected to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0233.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

evaluate performance capability of M5 models toward proposing the mathematical linear formulations for RCCP. After classifying, the developed M5p consisting six input variables and three outputs was performed for simulation of CS, TS and FS of RCCP using 12, 18 and 3 rules. The proposed models have the optimum number of decision trees (linear models (LMs)) as this value achieve the lowest error in the training stage. These LMs (rules) on the basis of conditional sentences are illustrated in Fig. 3. Furthermore, Developed CS, TS and FS model are presented for AS smaller and greater than 10.5, 21 and 17.5 respectively, as in the M5p rule. Estimated coefficients for LMs based on proposed mathematical linear equation of RCCP properties (i.e., CS, TS and FS) are presented in Table A1-A3 in appendix. All inputs were participated in the simulation of the mechanical characteristics of RCCP; they are significant in implementation of the developed LMs.

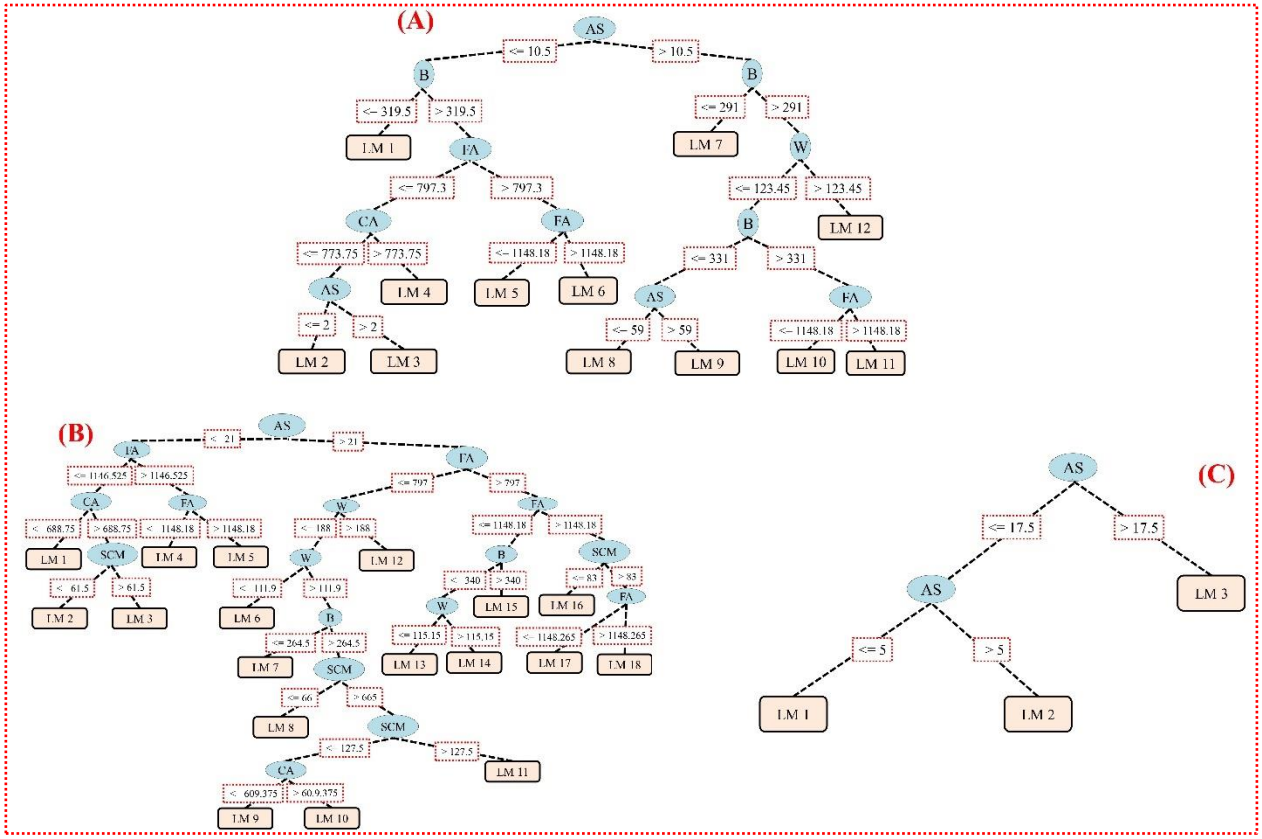

Fig. 3: Tree diagram of proposed M5p model; (A): CS, (B): TS, (C): FS

3.2.1. Compressive strength

The observed and simulated compressive strength values by the RF, M5rule, M5p and CHAID models for RCCP are illustrated in Fig. 4. As presented in Fig. 4, based on the closer the ratio to 1 (black and dotted line), there were better visual agreements between the observed CS and the simulated RF than other tree-based models. There were significant statistical correlations between observed and simulated CS values for the four models under study. To comparison of the proposed tree-based model's performances based on quantitative measures (i.e., NSE RSD, R and RMSE) table 3 is presented. The evaluation metrics over the training phase considered that RF simulated the CS with the highest precision (R = 0.986, NSE = 0.968 and minimum RSD = 0.561 MPa) in comparison with those resulted using other ensemble tree-based techniques such as CHAID (R = 0.925, NSE = 0.857 and RSD = 2.570 MPa). Moreover, M5rule model attained lower performance capability in terms of R (0.855), NSE (0.731), RSME (74.480

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0233.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

MPa), and RSD (5.460 MPa) versus M5p (R= 0.896, NSE= 0.797, RSME= 56.142 MPa, and RSD= 4.122 MPa), which M5p and M5rule provides permissible criteria.

In the testing phase, it is obvious that the *CS* simulated values modeled by RF indicated the best implementation with the highest NSE (0.931) and lowest RSD (1.181 MPa) values in comparison with other ML methods. Fig. 5 plotted the observed and simulated CS of RCCP and their relative error using tree-based techniques. it can trustworthily say that the estimated CS of RF and CHAID models were in robustness coherence with the observed data points. However, RF roughly could simulate extreme CS values.

Table 3: Predictive performance of the proposed models for CS prediction

Phase	Proposed models	Performance metrics			
		R	**NSE**	**RMSE**	**RSD**
Training	**RF**	**0.986**	**0.968**	**8.650**	**0.561**
	M5rule	0.855	0.731	74.480	5.460
	M5p	0.896	0.797	56.142	4.122
	CHAID	0.925	0.857	39.617	2.570
Testing	**RF**	**0.965**	**0.931**	**17.911**	**1.181**
	M5rule	0.828	0.680	83.507	6.499
	M5p	0.889	0.774	58.878	4.507
	CHAID	0.897	0.801	51.842	3.556

Bold text presented to the best performance

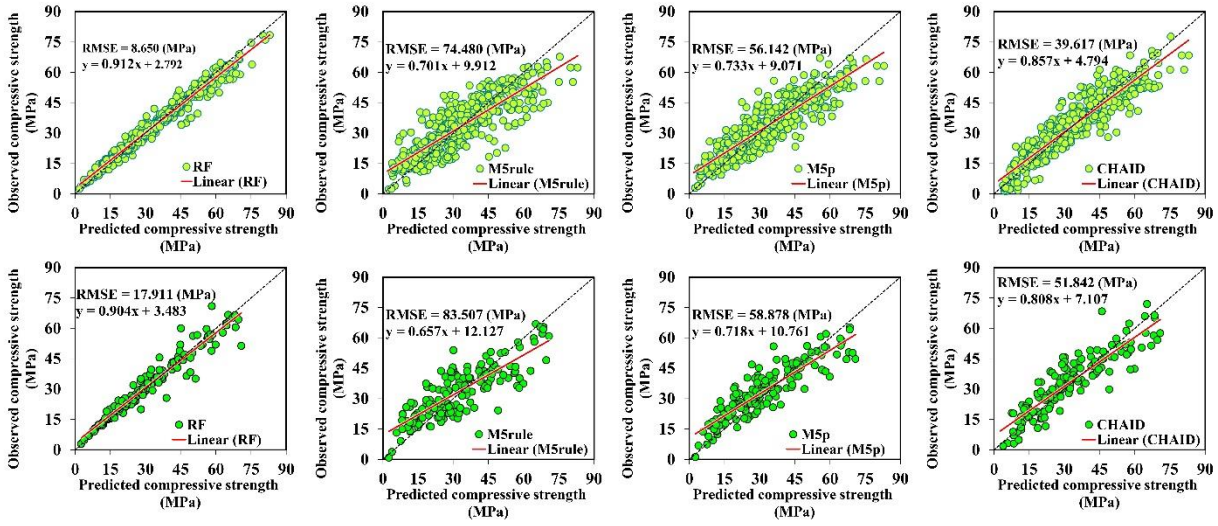

Fig. 4: Scatter plots of observed and simulated CS for training (light color) and testing (dark color) performances of the proposed models

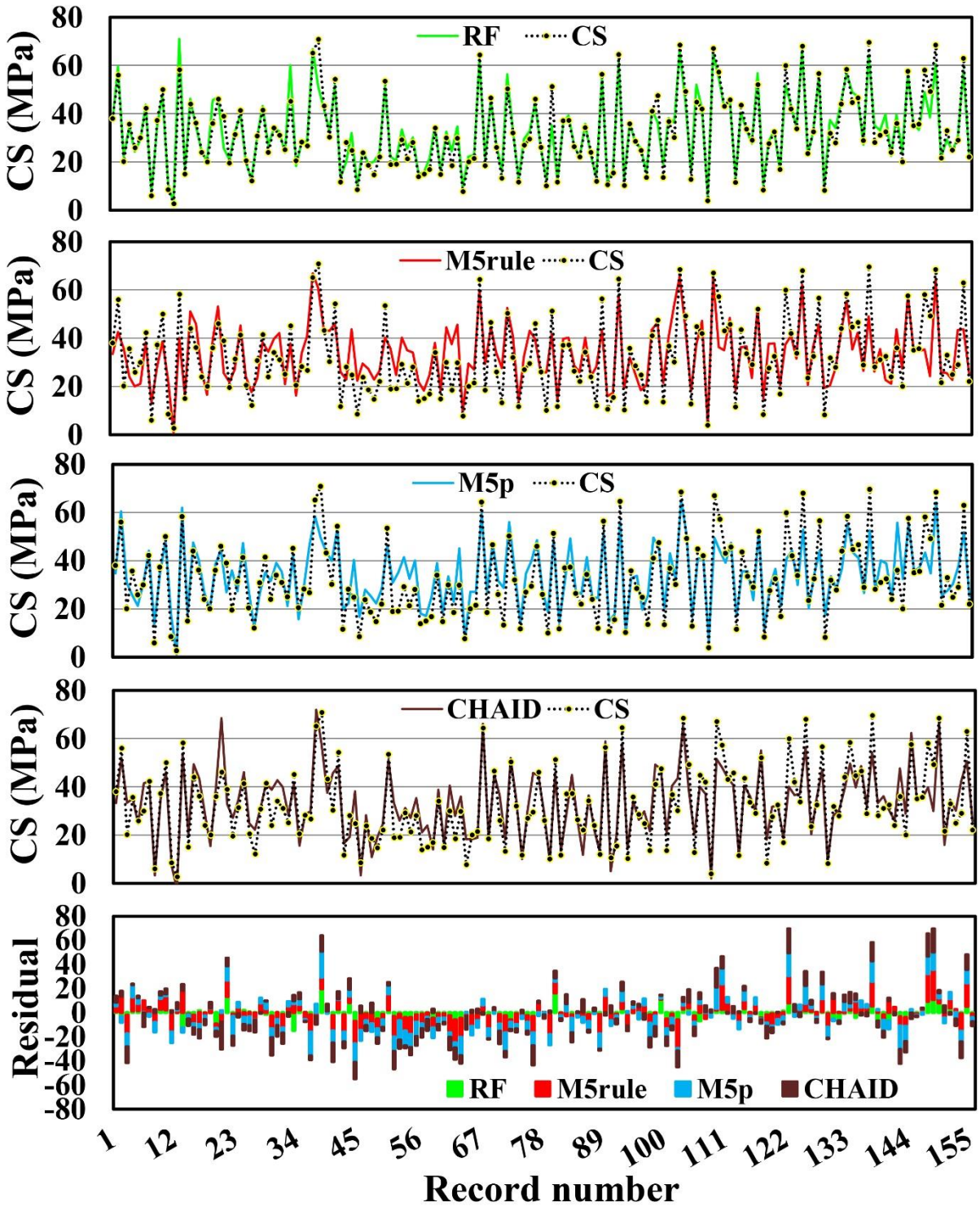

Fig. 5: Time series and residual plots of testing phase of the classification-based regression methods for CS estimation

3.2.2. Tensile strength

The performance indicators of the of the calibration and validation capability of developed tensile strength of RCCP using tree-based methods are reported in Table 4. According to Table 4, the RF model presented reliable and robustness performances in training and testing phases. The statistical assessment for the validation subset of the proposed RF, M5rule, M5p and CHAID techniques are (R= 0.984, NSE = 0.955 MPa, RMSE =0.070 MPa and RSD=0.062 MPa), (R= 0.850, NSE = 0.706 MPa, RMSE

=0.471 MPa and RSD=0.500 MPa), (R= 0.882, NSE = 0.776 MPa, RMSE =0.358 MPa and RSD=0.328 MPa), (R= 0.912, NSE = 0.817 MPa, RMSE =0.293 MPa and RSD=0.255 MPa), respectively. The graphical plots of subsets of the presented models are scattered in Figs. 6. It can be considered that presented tree-based models attained the perceptibly acceptable simulation result for TS of RCCP based on data correlated around ideal line (1:1 line). Although a few data points developed by M5p and M5rule around the TS of 2-5 MPa indicated some small divergence from 1:1 line, the results reported that all of the tree-based methods provided high accuracy to simulate of tensile strength. The time series and residual plots for tree-based simulation and actual TS presented in Fig. 7. It can be defined that the RF model generated the minimum RMSE and outperforms the M5rule, M5p and CHAID for estimation TS of RCCP.

Table 4: Predictive performance of the proposed models for TS prediction

Phase	Proposed models	Performance metrics			
		R	**NSE**	**RMSE**	**R**
Training	**RF**	**0.991**	**0.981**	**0.030**	**0.025**
	M5rule	0.892	0.791	0.338	0.320
	M5p	0.895	0.798	0.328	0.294
	CHAID	0.975	0.951	0.078	0.024
Testing	**RF**	**0.984**	**0.955**	**0.070**	**0.062**
	M5rule	0.850	0.706	0.471	0.500
	M5p	0.882	0.776	0.358	0.328
	CHAID	0.912	0.817	0.293	0.255

Bold text presented to the best performance

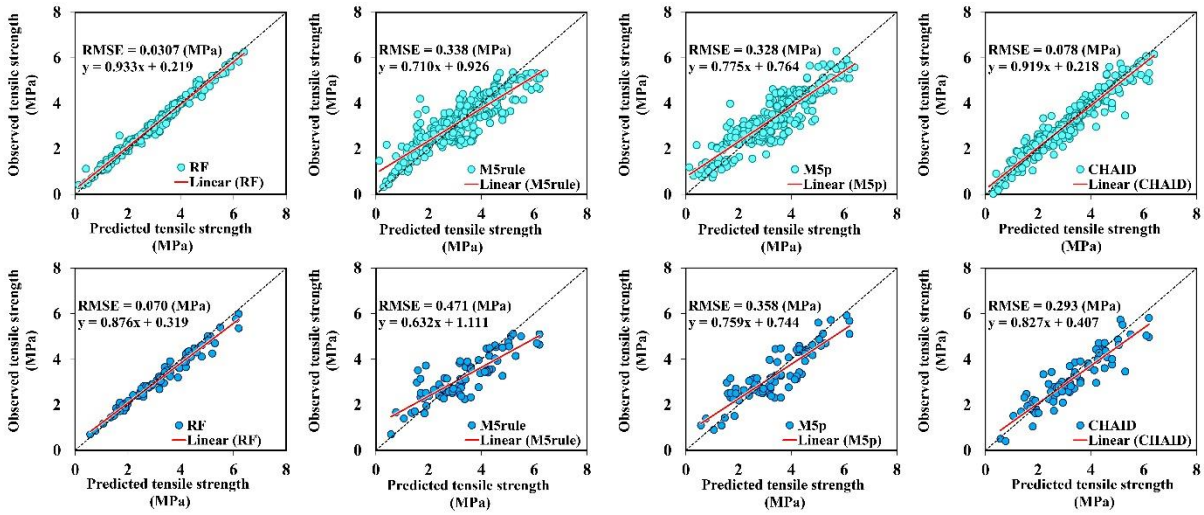

Fig. 6: Scatter plots of observed and simulated TS for training (light color) and testing (dark color) performances of the proposed models

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0233.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

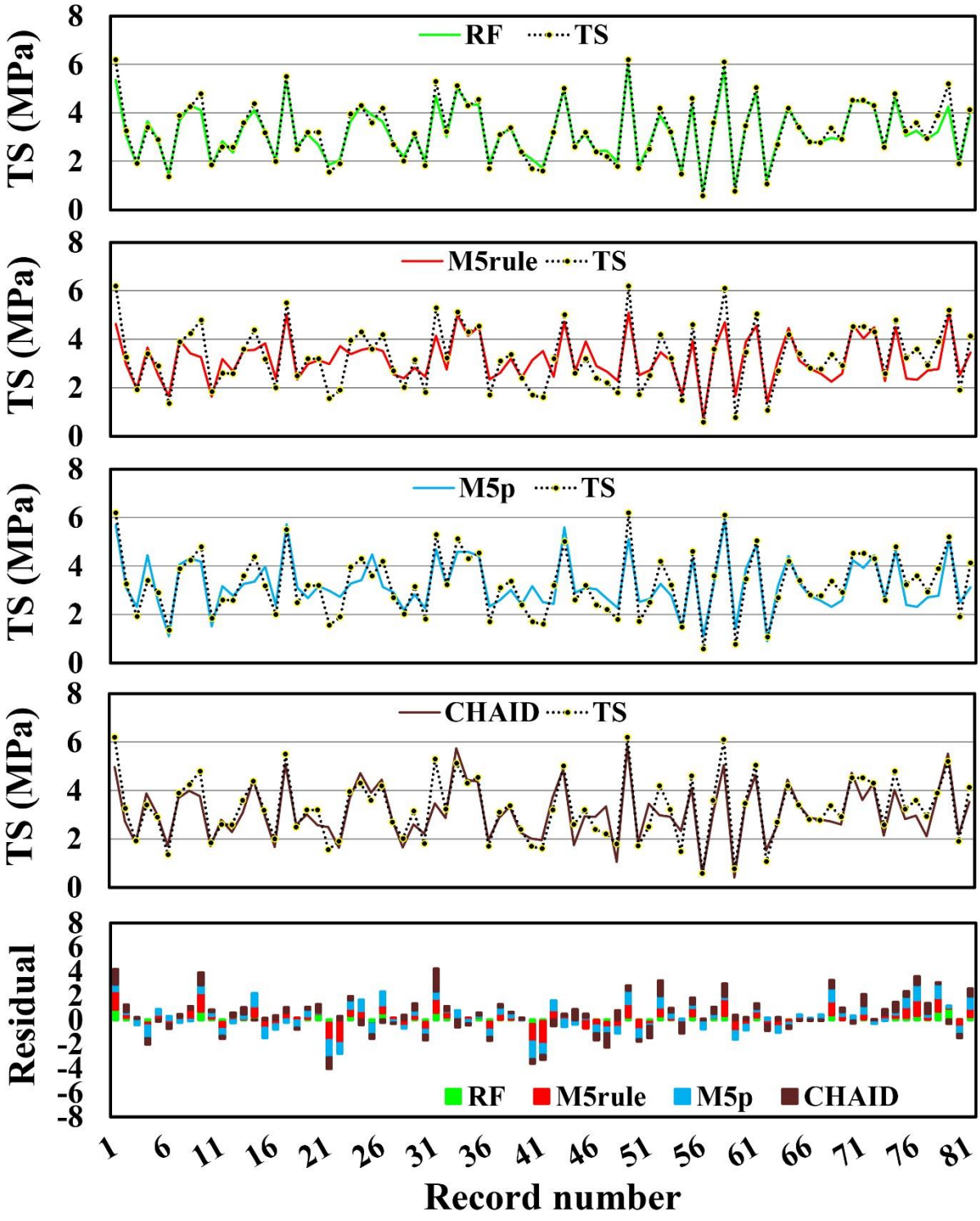

Fig. 7: Time series and residual plots of testing phase of the classification-based regression methods for TS estimation

3.2.3. Flexural strength

The applicability of tree-based models named RF, M5rule, M5p and CHAID was investigated for flexural strength of RCCP estimation. The statistical evaluation of the developed models in the simulation of FS considered in Table 5. In a 75-25% data splitting of this study, RF model (R= 0.988 and RSD=0.049 MPa), (R= 0.970 and RSD=0.108 MPa), outperformed than the other ML methods in both training and testing stages, respectively. RF has the lowest RMSE (0.197 MPa) and highest NSE (0.939); it enhanced

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0233.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

the precision of testing phase in terms of NSE of the M5rule, M5p and CHAID by 28%, 27% and 30%, respectively. Fig. 8 and 9 exhibited the plots for comparisons of the actual results with those of the four models inspired of tree-based regression methods. It can be shown in the aforementioned figures of the proposed models that the RF model has the highest accuracy in the simulation of FS during the train and test steps. It is also evident from this plots that the RF indicated a slightly higher precision in estimation of the local maximum and minimum of FS values comparing other ML methods.

Table 5: Predictive performance of the proposed models for FS prediction

Phase	Proposed models	Performance metrics			
		R	**NSE**	**RMSE**	**R**
Training	RF	**0.988**	**0.974**	**0.086**	**0.049**
	M5rule	0.937	0.878	0.416	0.234
	M5p	0.887	0.782	0.705	0.435
	CHAID	0.925	0.849	0.516	0.288
Testing	RF	**0.970**	**0.939**	**0.197**	**0.108**
	M5rule	0.853	0.673	1.068	0.689
	M5p	0.843	0.683	1.033	0.644
	CHAID	0.846	0.651	1.138	0.612

Bold text presented to the best performance

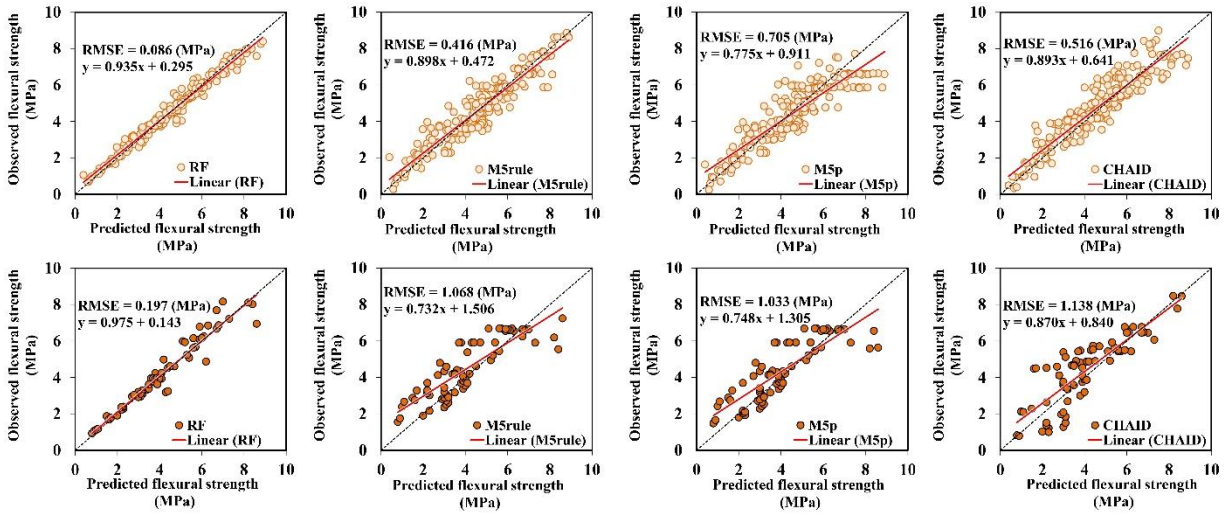

Fig. 8: Scatter plots of observed and simulated FS for training (light color) and testing (dark color) performances of the proposed models

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0233.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

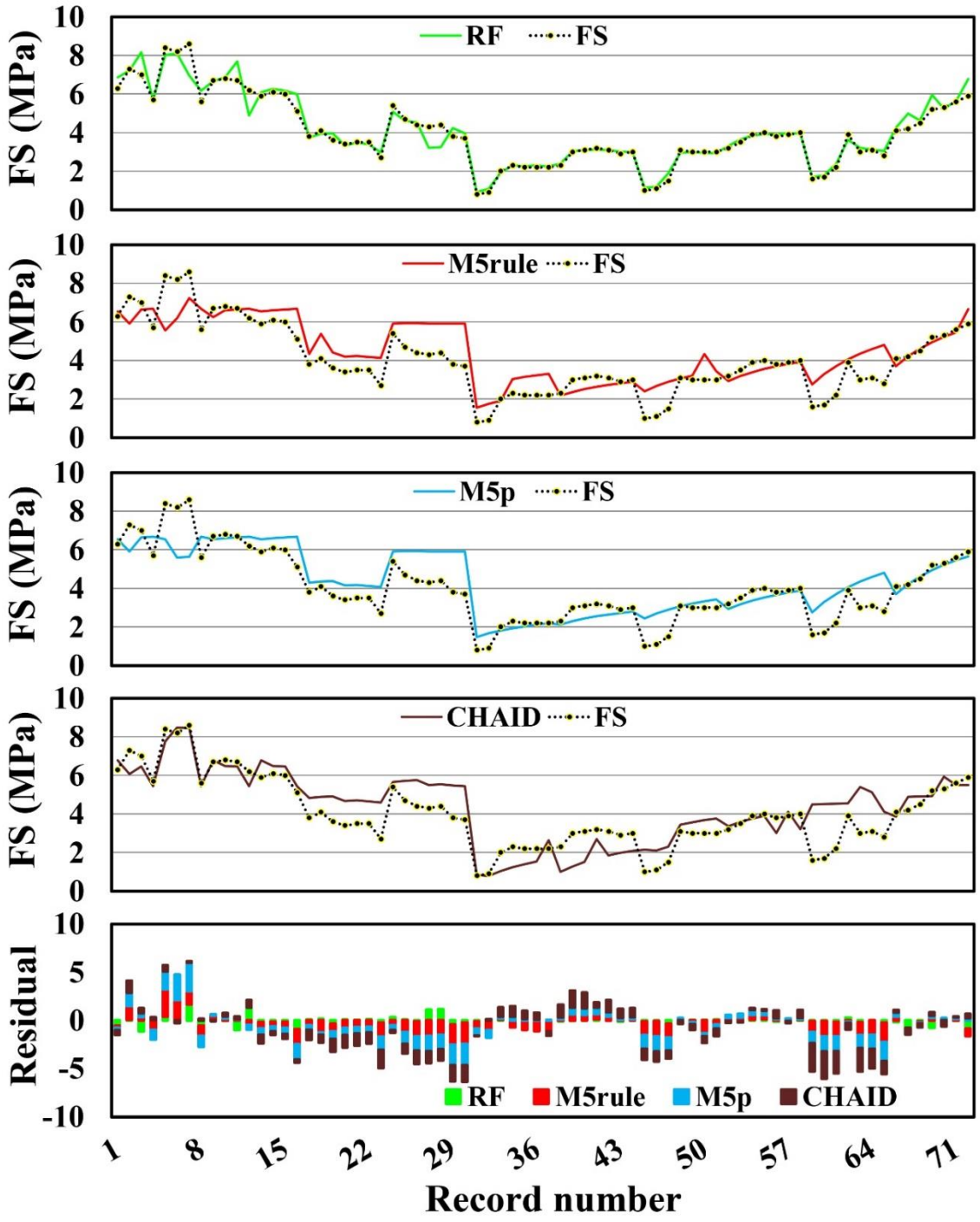

Fig. 9: Time series and residual plots of testing phase of the classification-based regression methods for FS estimation

3.3. Models validity

External validation (EV) is used for comparison between the results of estimated and experimental event data. Golbraikh and Tropsha [74] have adopted the new external validation criteria to assess the estimation precision of the models according to the performance of validating data. EV means assessing the model performance with independent samples [75].

$$\sum_{i=1}^{n} \frac{t_{obs} \times t_{pre}}{t_{pre}^2} \qquad (12)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0233.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3707; doi:10.3390/app10113707

$\frac{t_{obs} \times t_{pre}}{t_{obs}^2}$ (13)

$t_{obs}$and t_{pre}represent the experimental and estimated target values, respectively.

$$m = (R^2 - R_0^2) / R^2 \quad < 0.1 \tag{14}$$

$$n = (R^2 - R_0'^2) / R^2 \quad < 0.1 \tag{15}$$

Furthermore, Roy and Roy used R_m (calculated by Eq.34) which is a stabilization criterion for external predictability of the models [76]. They found that an R_m value less than 0.5 shows an appropriate situation

$$R_m = R^2 \times \left(1 - \sqrt{|R^2 - R_0^2|}\right) > 0.5 \tag{16}$$

The determination coefficients passing through the source between the estimated and experimental values (R_0^2) and conversely ($R_0'^2$) are derived using the following equations:

$$R_0^2 = 1 - \sum_{i=1}^{n} t_{pre}^2 (1-k)^2 / \sum_{i=1}^{n} t_{pre} - \overline{t_{pre}})^2 \tag{17}$$

$$R_0'^2 = 1 - \sum_{i=1}^{n} t_{obs}^2 (1-k')^2 / \sum_{i=1}^{n} (t_{obs} - \bar{t}_{obs})^2 \tag{18}$$

The validation indicator and the related performance of CS, TS and FS prediction obtained by various models are presented in Table 6. According to this table, the RF models for CS, TS and FS which yielded Rm=0.691, Rm=0.834 and Rm=0716 values satisfy the conditions with best validation with respect to other used approaches such as CHAID, M5tree and CART models. In addition, the CART and M5tree values for CS (Rm =0.187), TS (Rm =0.195) and FS were less than the required value for Rm (Rm >0.5). Thus, it is seen that RF shows highest validity for predicting mechanical characteristics of RCCP and the computed correlations had not been accidentally.

Table 6: statistical measures of EV for all proposed models

	Model	*K*	*K'*	*m*	*n*	R_m
CS	RF	0.995	0.990	-0.071	-0.071	0.691
	M5rule	0.978	0.955	-0.452	-0.444	0.303
	M5p	0.971	0.984	-0.254	-0.261	0.436
	CHAID	0.986	0.983	-0.238	-0.240	0.552
TS	RF	1.035	0.961	-0.020	-0.019	0.834
	M5rule	1.038	0.927	-0.358	-0.328	0.354
	M5p	1.014	0.956	-0.282	-0.266	0.413
	CHAID	1.044	0.936	-0.181	-0.164	0.508
FS	RF	0.984	1.006	-0.060	-0.061	0.716
	M5rule	0.914	1.043	-0.278	-0.356	0.400
	M5p	0.934	1.018	-0.355	-0.402	0.353
	CHAID	0.910	1.043	-0.324	-0.380	0.370

Monte-Carlo Simulation (MCS) based uncertainty analysis is used for determining the randomness of model uncertainty. This method was first used by Ulam and Neman [77] in military projects for simulation of the probabilistic events. It is well known that CS, TS and FS contains various uncertainties such as uncertainty of input variables, uncertainty of model parameter, etc.

For this purpose, an investigation of quantitative uncertainty associated with outputs prediction rate (E) is performed using RF, M5rule, M5p and CHAID models. The MCS is performed for CS, TS and FS values. The individual error of prediction is calculated for all the datasets (Eq. 19). Equations 20 and 21 are utilized for calculation of the mean $(\bar{e})$ and standard deviation (S_e) of the estimation error, respectively [76]:

$$e_i = \log_{10}(t^{pre}{}_i) - \log_{10}(t^{exp}{}_i) \tag{19}$$

$$\bar{e} = \sum_{i=1}^{n} e_i \tag{20}$$

$$S_e = \sqrt{\sum_{i=1}^{n} \left(\frac{(e_i - \bar{e})^2}{n-1} \right)} \tag{21}$$

In the above equations n is the dataset length, t^{pre} and t^{obs} denote the estimated and experimental target values, respectively. A positive mean prediction denotes an overestimated prediction of targets value and a negative one denotes an underestimated value of target variable compared to the observed values. Thus, a confidence band could be drawn around the predicted error value through application of Wilson score approach [78]. Furthermore, $\pm 1.96\, S_e$ yields 95% confidence band around predicted P_i as follows:

$$\{P_i \times 10^{-\bar{e}-1.96S_e}, P_i \times 10^{-\bar{e}+1.96S_e}\} \tag{22}$$

The outputs of this analysis such as the uncertainty band width and Mean Absolute Deviation (MAD) are given in Table 7. According to this table, the positive mean prediction error indicates that the predicted values calculated by all these methods are higher than the experimental values. Also, it is seen that RF and CHAID methods for CS yielded the minimum (33.065% and 33.240) bandwidth uncertainties, respectively. Moreover, in other developed models, RF had lowest uncertainty and satisfied bandwidth criteria.

Table 7: Uncertainty quantification for all classification models

	Model	$\bar{e}$	S_e	*Median*	*MAD*	Uncertainty (%)
CS	RF	0.174	8.765	34.181	11.302	33.065
	M5rule	-0.017	3.313	32.539	12.530	38.509
	M5p	0.221	6.533	33.789	12.374	36.622
	CHAID	0.499	7.527	33.724	11.210	33.240
TS	RF	0.004	0.004	3.168	0.836	26.393
	M5rule	-0.038	0.361	3.091	0.973	31.507
	M5p	-0.012	0.201	3.108	0.943	30.346
	CHAID	0.048	0.578	3.116	0.883	28.348
FS	RF	-0.026	0.905	4.586	1.377	30.026
	M5rule	0.104	0.754	4.568	1.428	31.275
	M5p	0.008	0.338	4.563	1.455	31.896
	CHAID	0.188	0.798	4.806	1.447	30.119

3.4. Sensitivity analysis and variable importance

Sensitivity analysis (SA) of variables is a technique used to determine how different values of predictor variables will affected on output variable. For each independent variable, the SA% is as following [53]:

$$L_i = t_{max}(x_i) - t_{min}(x_i) \quad (23)$$

$$SA_i = \frac{L_i}{\sum_{j=1}^{M} L_i} \times 100 \quad (24)$$

where t_{max} and t_{min} = maximum and minimum of the estimated target over the i_{th} input domain, where other independent variable values are equal to their average values. The result of variable importance for the simulation of mechanical characteristics of RCCP is indicated in Fig. 10 based of RF model (best model). These figures presented that the most effective variable in CS, TS and FS of RCCP is the fine aggregate content.

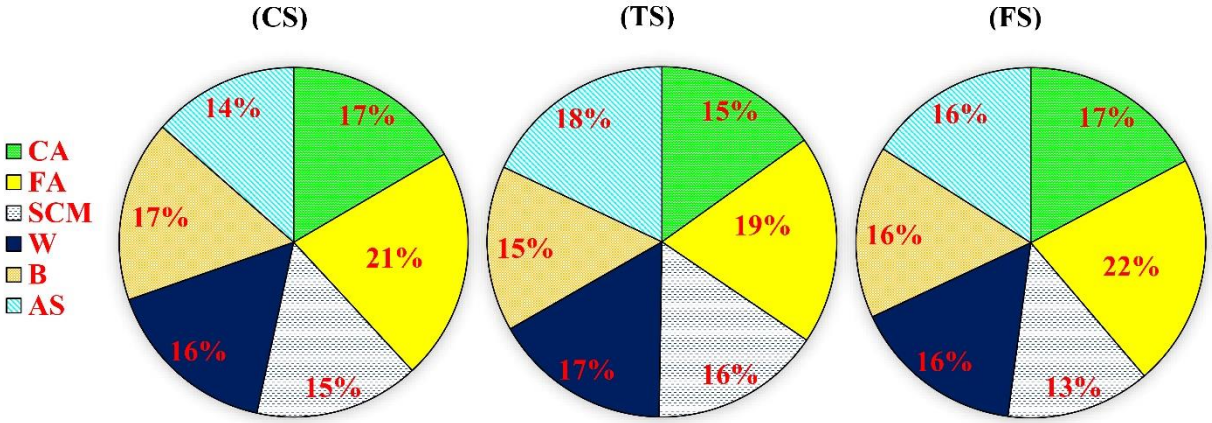

Fig. 10: Sensitivity analysis of variable importance

4. Conclusions

In this research, classification-based regression methods based on RF, M5rule, M5p and CHAID were applied as a ML tools to propose new predictive models of the mechanical characteristics of RCCP. The models were constructed using comprehensive datasets of RCCP design codes. According to results of this research, the following point of view can be outlined:

- Developing RF, M5rule, M5p and CHAID revealed that the CS, TS and FS of RCCP are mainly related to six inputs including CA, FA, SCM, W, B and AS that selected based on PCA technique.

- Presented CS, TS and FS models indicate that the RF method presented more precision simulation in comprising with the other three tree-based techniques, with respect to R, NSE, RMSE and RSD measures for training and testing phase.

- The proposed RF and CHAID models as novel ML method confirmed all of the required criterion of the external validation statistical condition, which satisfy their validity.

- The Monte-Carlo uncertainty investigation for implemented tree-based methods was carried out. The robustness of the proposed ML methods was verified in this paper. Moreover, sensitivity analysis of variable importance considered the highest importance of predictors factor influenced on the mechanical characteristics of RCCP to be the fine aggregate content.

Appendix A

Table A1: Inputs coefficient of developed M5p for CS formulation

Linear Model	Coefficient						
	CA	FA	SCM	W	B	AS	X
LM1	0.000	0.005	- 0.043	- 0.067	0.113	1.543	- 11.878
LM2	0.010	0.002	- 0.039	- 0.078	0.082	1.681	- 8.495
LM3	0.010	0.002	- 0.042	- 0.078	0.082	1.6815	- 7.7941
LM4	0.010	- 0.001	- 0.024	- 0.084	0.0819	0.7478	- 2.6946
LM5	0.005	- 0.013	- 0.048	- 0.220	0.265	1.068	- 23.049
LM6	0.112	- 0.023	- 0.081	- 0.149	0.148	1.317	- 68.118
LM7	-0.016	0.001	- 0.072	- 0.100	0.251	0.104	- 4.059
LM8	0.002	- 0.004	- 0.003	0.345	0.066	0.059	- 15.986
LM9	0.002	- 0.004	- 0.003	0.293	0.066	0.058	- 5.490
LM10	0.007	- 0.006	- 0.003	0.055	0.086	0.114	16.015
LM11	0.009	- 31.824	- 0.003	0.055	0.086	0.074	36.545
LM12	-0.012	- 0.0003	0.049	- 0.236	0.083	0.116	50.6158

X denoted the M5p intercept coefficient in LM

Table A2: Inputs coefficient of developed M5p for TS formulation

Linear Model	Coefficient						
	CA	FA	SCM	W	B	AS	X
LM1	-0.004	0.0005	- 0.002	- 0.002	- 0.029	0.213	7.607
LM2	0.001	0.0003	- 0.006	- 0.0008	0.002	0.015	0.004
LM3	0.0004	0.0003	0.003	0.0003	0.002	0.174	- 0.0899
LM4	0.000	- 0.006	- 0.004	- 0.003	0.003	0.125	9.550
LM5	0.000	- 0.008	- 0.007	- 0.003	0.003	0.107	12.874
LM6	-0.0001	0.000	- 0.003	- 0.016	0.004	0.002	3.923
LM7	-0.0002	0.000	- 0.004	- 0.010	0.005	0.002	2.984
LM8	0.0001	- 0.001	- 0.003	- 0.016	0.004	0.002	5.166
LM9	-0.0005	0.000	- 0.004	- 0.011	0.004	0.002	4.006
LM10	-0.0005	0.000	- 0.004	- 0.011	0.004	0.002	3.982
LM11	-0.0004	0.000	- 0.004	- 0.011	0.004	0.001	3.902
LM12	0.0004	- 0.0003	- 0.002	- 0.002	0.003	0.004	2.258
LM13	0.001	0.001	0.0009	0.011	0.006	0.005	- 1.933
LM14	0.001	0.001	0.0009	0.002	0.006	0.010	- 1.058
LM15	0.003	0.001	0.001	0.002	0.007	0.003	- 1.188
LM16	0.001	0.0004	0.001	0.0005	0.004	0.004	1.367
LM17	0.001	0.0006	- 0.001	0.0005	0.004	0.003	1.295
LM18	0.001	0.0006	- 0.001	0.0005	0.004	0.003	1.195

X denoted the M5p intercept coefficient in LM

Table A3: Inputs coefficient of developed M5p for FS formulation

Linear Model	Coefficient						
	CA	FA	SCM	W	B	AS	X
LM1	0.001	0.0001	- 0.009	- 0.003	0.011	0.283	- 1.899
LM2	0.002	0.0001	- 0.001	- 0.013	0.015	0.037	- 1.207
LM3	0.003	0.0001	- 0.0003	- 0.0008	0.017	0.012	- 3.863

X denoted the M5p intercept coefficient in LM

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER, grant number XXX" and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding, any errors may affect your future funding.

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hashemi, M., Shafigh, P., Karim, M. R. B., & Atis, C. D. (2018). The effect of coarse to fine aggregate ratio on the fresh and hardened properties of roller-compacted concrete pavement. *Construction and Building Materials*, *169*, 553-566.
2. Modarres, A., Hesami, S., Soltaninejad, M., & Madani, H. (2018). Application of coal waste in sustainable roller compacted concrete pavement-environmental and technical assessment. *International Journal of Pavement Engineering*, *19*(8), 748-761.
3. Lam, M. N. T., Le, D. H., & Jaritngam, S. (2018). Compressive strength and durability properties of roller-compacted concrete pavement containing electric arc furnace slag aggregate and fly ash. *Construction and Building Materials*, *191*, 912-922.
4. Chhorn, C., Kim, Y. K., Hong, S. J., & Lee, S. W. (2019). Evaluation on compactibility and workability of roller-compacted concrete for pavement. *International Journal of Pavement Engineering*, *20*(8), 905-910.
5. Adamu, M., Mohammed, B. S., Shafiq, N., & Liew, M. S. (2018). Durability performance of high volume fly ash roller compacted concrete pavement containing crumb rubber and nano silica. *International Journal of Pavement Engineering*, 1-8.
6. Adamu, M., Mohammed, B. S., & Liew, M. S. (2018). Mechanical properties and performance of high volume fly ash roller compacted concrete containing crumb rubber and nano silica. *Construction and Building Materials*, *171*, 521-538.

7. Ashrafian, A., Gandomi, A. H., Rezaie-Balf, M., & Emadi, M. (2019). An Evolutionary Approach to Formulate the Compressive Strength of Roller Compacted Concrete Pavement. *Measurement*, 107309.
8. TAHERI AMIRI, M. J., Ashrafian, A., Haghighi, F. R., & Javaheri Barforooshi, M. (2019). Prediction of the Compressive Strength of Self-compacting Concrete containing Rice Husk Ash using Data Driven Models. *Modares Civil Engineering journal*, *19*(1), 196-206.
9. Yaseen, Z. M., Deo, R. C., Hilal, A., Abd, A. M., Bueno, L. C., Salcedo-Sanz, S., & Nehdi, M. L. (2018). Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Advances in Engineering Software*, *115*, 112-125.
10. Yaseen, Z. M., Tran, M. T., Kim, S., Bakhshpoori, T., & Deo, R. C. (2018). Shear strength prediction of steel fiber reinforced concrete beam using hybrid intelligence models: a new approach. Engineering Structures, 177, 244-255.
11. Amlashi, A. T., Abdollahi, S. M., Goodarzi, S., & Ghanizadeh, A. R. (2019). Soft computing based formulations for slump, compressive strength, and elastic modulus of bentonite plastic concrete. *Journal of Cleaner Production*, *230*, 1197-1216.
12. Gholampour, A., Gandomi, A. H., & Ozbakkaloglu, T. (2017). New formulations for mechanical properties of recycled aggregate concrete using gene expression programming. *Construction and Building Materials*, *130*, 122-145.
13. Asteris, P. G., Armaghani, D. J., Hatzigeorgiou, G. D., Karayannis, C. G., & Pilakoutas, K. (2019). Predicting the shear strength of reinforced concrete beams using Artificial Neural Networks. *Computers and Concrete*, *24*(5), 469-488.
14. Ly, H. B., Pham, B. T., Dao, D. V., Le, V. M., Le, L. M., & Le, T. T. (2019). Improvement of ANFIS Model for Prediction of Compressive Strength of Manufactured Sand Concrete. *Applied Sciences*, *9*(18), 3841.
15. Sun, J., Zhang, J., Gu, Y., Huang, Y., Sun, Y., & Ma, G. (2019). Prediction of permeability and unconfined compressive strength of pervious concrete using evolved support vector regression. *Construction and Building Materials*, *207*, 440-449.
16. Ashrafian, A., Shokri, F., Amiri, M. J. T., Yaseen, Z. M., & Rezaie-Balf, M. (2020). Compressive strength of Foamed Cellular Lightweight Concrete simulation: New development of hybrid artificial intelligence model. *Construction and Building Materials*, *230*, 117048.
17. Deng, F., He, Y., Zhou, S., Yu, Y., Cheng, H., & Wu, X. (2018). Compressive strength prediction of recycled concrete based on deep learning. *Construction and Building Materials*, *175*, 562-569.
18. Sun, J., Zhang, J., Gu, Y., Huang, Y., Sun, Y., & Ma, G. (2019). Prediction of permeability and unconfined compressive strength of pervious concrete using evolved support vector regression. *Construction and Building Materials*, *207*, 440-449.
19. Shahmansouri, A. A., Bengar, H. A., & Jahani, E. (2019). Predicting compressive strength and electrical resistivity of eco-friendly concrete containing natural zeolite via GEP algorithm. *Construction and Building Materials*, *229*, 116883.
20. Feng, D. C., Liu, Z. T., Wang, X. D., Chen, Y., Chang, J. Q., Wei, D. F., & Jiang, Z. M. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*, *230*, 117000.

21. Iqbal, M. F., Liu, Q. F., Azim, I., Zhu, X., Yang, J., Javed, M. F., & Rauf, M. (2020). Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. *Journal of hazardous materials*, *384*, 121322.
22. Asteris, P. G., Ashrafian, A., & Rezaie-Balf, M. (2019). Prediction of the compressive strength of self-compacting concrete using surrogate models. *Comput. Concr*, *24*, 137-150.
23. Hassan, M. A., Khalil, A., Kaseb, S., & Kassem, M. A. (2017). Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Applied Energy*, *203*, 897-916.
24. Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., ... & Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and forest meteorology*, *263*, 225-241.
25. Behnood, A., Olek, J., & Glinicki, M. A. (2015). Predicting modulus elasticity of recycled aggregate concrete using M5′ model tree algorithm. Construction and Building Materials, 94, 137-147.
26. Behnood, A., Behnood, V., Gharehveran, M. M., & Alyamac, K. E. (2017). Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Construction and Building Materials*, *142*, 199-207.
27. Han, Q., Gui, C., Xu, J., & Lacidogna, G. (2019). A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. *Construction and Building Materials*, *226*, 734-742.
28. Mohamed, O. A., Ati, M., & Najm, O. F. (2017). Predicting Compressive Strength of Sustainable Self-Consolidating Concrete Using Random Forest. In *Key Engineering Materials* (Vol. 744, pp. 141-145). Trans Tech Publications.
29. Ashrafian, A., Amiri, M. J. T., Rezaie-Balf, M., Ozbakkaloglu, T., & Lotfi-Omran, O. (2018). Prediction of compressive strength and ultrasonic pulse velocity of fiber reinforced concrete incorporating nano silica using heuristic regression methods. *Construction and Building Materials*, *190*, 479-494.
30. Gholampour, A., Mansouri, I., Kisi, O., & Ozbakkaloglu, T. (2018). Evaluation of mechanical properties of concretes containing coarse recycled concrete aggregates using multivariate adaptive regression splines (MARS), M5 model tree (M5Tree), and least squares support vector regression (LSSVR) models. *Neural Computing and Applications*, 1-14.
31. AzariJafari, H., Amiri, M. J. T., Ashrafian, A., Rasekh, H., Barforooshi, M. J., & Berenjian, J. (2019). Ternary blended cement: An eco-friendly alternative to improve resistivity of high-performance self-consolidating concrete against elevated temperature. *Journal of Cleaner Production*, *223*, 575-586.
32. Ramezanianpour, A. A., Mohammadi, A., Dehkordi, E. R., & Chenar, Q. B. (2017). Mechanical properties and durability of roller compacted concrete pavements in cold regions. *Construction and Building Materials*, *146*, 260-266.
33. ACI 325-10R-95, State-of-the-Art Report on Roller-compacted Concrete Pavements, 2001, 32 pp.
34. Rao, S. K., Sravana, P., & Rao, T. C. (2015). Strength and compaction characteristics of fly ash roller compacted concrete. *International Journal of Scientific Research in Knowledge*, *3*(10), 260-269.

35. Mardani-Aghabaglou, A., & Ramyar, K. (2013). Mechanical properties of high-volume fly ash roller compacted concrete designed by maximum density method. *Construction and Building Materials*, *38*, 356-364.
36. Pavan, S., & Rao, S. K. (2014). Effect of fly ash on strength characteristics of roller compacted concrete pavement. *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, *11*(6), 04-08.
37. Atiş, C. D., Sevim, U. K., Özcan, F., Bilim, C., Karahan, O., Tanrikulu, A. H., & Ekşi, A. (2004). Strength properties of roller compacted concrete containing a non-standard high calcium fly ash. *Materials Letters*, *58*(9), 1446-1450.
38. Tangtermsirikul, S., Kaewkhluab, T., & Jitvutikrai, P. (2004). A compressive strength model for roller-compacted concrete with fly ash. *Magazine of concrete Research*, *56*(1), 35-44.
39. Rao, S. K., Sravana, P., & Rao, T. C. (2016). Experimental studies in Ultrasonic Pulse Velocity of roller compacted concrete pavement containing fly ash and M-sand. *International Journal of Pavement Research and Technology*, *9*(4), 289-301.
40. Cao, C., Sun, W., & Qin, H. (2000). The analysis on strength and fly ash effect of roller-compacted concrete with high volume fly ash. *Cement and concrete research*, *30*(1), 71-75.
41. Rao, S. K., Sravana, P., & Rao, T. C. (2015). Investigation on pozzolanic effect of Fly ash in Roller Compacted Concrete pavement. *IRACST-Engineering Science and Technology: An International Journal (ESTIJ)*, *5*(2), 202-206.
42. Ghahari, S. A., Mohammadi, A., & Ramezanianpour, A. A. (2017). Performance assessment of natural pozzolan roller compacted concrete pavements. *Case studies in construction materials*, *7*, 82-90.
43. Mohammed, B. S., & Adamu, M. (2018). Mechanical performance of roller compacted concrete pavement containing crumb rubber and nano silica. *Construction and Building Materials*, *159*, 234-251.
44. Debbarma, S., Ransinchung, G. D., & Singh, S. (2019). Feasibility of roller compacted concrete pavement containing different fractions of reclaimed asphalt pavement. *Construction and Building Materials*, *199*, 508-525.
45. Hashemi, M., Shafigh, P., Karim, M. R. B., & Atis, C. D. (2018). The effect of coarse to fine aggregate ratio on the fresh and hardened properties of roller-compacted concrete pavement. *Construction and Building Materials*, *169*, 553-566.
46. Lam, M. N. T., Le, D. H., & Jaritngam, S. (2018). Compressive strength and durability properties of roller-compacted concrete pavement containing electric arc furnace slag aggregate and fly ash. *Construction and Building Materials*, *191*, 912-922.
47. Vahidi, E. K., Malekabadi, M. M., Rezaei, A., Roshani, M. M., & Roshani, G. H. (2017). Modeling of mechanical properties of roller compacted concrete containing RHA using ANFIS. *Computers and Concrete*, *19*(4), 435-442.
48. Shamsaei, M., Aghayan, I., & Kazemi, K. A. (2017). Experimental investigation of using cross-linked polyethylene waste as aggregate in roller compacted concrete pavement. *Journal of cleaner production*, *165*, 290-297.
49. Hesami, S., Modarres, A., Soltaninejad, M., & Madani, H. (2016). Mechanical properties of roller compacted concrete pavement containing coal waste and limestone powder as partial replacements of cement. *Construction and Building Materials*, *111*, 625-636.

50. Rao, S. K., Sravana, P., & Rao, T. C. (2015). Strength and compaction characteristics of fly ash roller compacted concrete. *International Journal of Scientific Research in Knowledge*, *3*(10), 260-269.
51. Rao, S. K., Sravana, P., & Rao, T. C. (2016). Abrasion resistance and mechanical properties of Roller Compacted Concrete with GGBS. *Construction and Building Materials*, *114*, 925-933.
52. Hesami, S., Modarres, A., Soltaninejad, M., & Madani, H. (2016). Mechanical properties of roller compacted concrete pavement containing coal waste and limestone powder as partial replacements of cement. *Construction and Building Materials*, *111*, 625-636.
53. Rao, S. K., Sravana, P., & Rao, T. C. (2016). Experimental studies in Ultrasonic Pulse Velocity of roller compacted concrete pavement containing fly ash and M-sand. *International Journal of Pavement Research and Technology*, *9*(4), 289-301.
54. Modarres, A., Hesami, S., Soltaninejad, M., & Madani, H. (2018). Application of coal waste in sustainable roller compacted concrete pavement-environmental and technical assessment. *International Journal of Pavement Engineering*, *19*(8), 748-761.
55. Rashad, A. M. (2013). A preliminary study on the effect of fine aggregate replacement with metakaolin on strength and abrasion resistance of concrete. *Construction and Building Materials*, *44*, 487-495.
56. Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.
57. Adusumilli, S., Bhatt, D., Wang, H., Bhattacharya, P., & Devabhaktuni, V. (2013). A low-cost INS/GPS integration methodology based on random forest regression. *Expert Systems with Applications*, *40*(11), 4653-4659.
58. Zhou, J., Shi, X., Du, K., Qiu, X., Li, X., & Mitri, H. S. (2016). Feasibility of random-forest approach for prediction of ground settlements induced by the construction of a shield-driven tunnel. *International Journal of Geomechanics*, *17*(6), 04016129.
59. Zhou, J., Li, X., & Mitri, H. S. (2015). Comparative performance of six supervised learning methods for the development of models of hard rock pillar stability prediction. *Natural Hazards*, *79*(1), 291-316.
60. Troncoso, A., Salcedo-Sanz, S., Casanova-Mateo, C., Riquelme, J. C., & Prieto, L. (2015). Local models-based regression trees for very short-term wind speed prediction. *Renewable Energy*, *81*, 589-598.
61. Arnett, F. C., Edworthy, S. M., Bloch, D. A., Mcshane, D. J., Fries, J. F., Cooper, N. S., ... & Medsger Jr, T. A. (1988). The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, *31*(3), 315-324.
62. Quinlan, J. R. (1992, November). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence* (Vol. 92, pp. 343-348).
63. Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*, *42*(11).
64. Abdelkader, S. S., Grolinger, K., & Capretz, M. A. (2015, December). Predicting energy demand peak using M5 model trees. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 509-514). IEEE.
65. Attar, N. F., Pham, Q. B., Nowbandegani, S. F., Rezaie-Balf, M., Fai, C. M., Ahmed, A. N., ... & El-Shafie, A. (2020). Enhancing the Prediction Accuracy of Data-Driven Models for Monthly Streamflow in

Urmia Lake Basin Based upon the Autoregressive Conditionally Heteroskedastic Time-Series Model. Applied Sciences, 10(2), 571.

66. Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *29*(2), 119-127.
67. Kamber, M., & Pei, J. (2006). *Data Mining*. Morgan kaufmann.
68. Sharp, A. (1998). *The performance of segmentation variables: A comparative study* (Doctoral dissertation, Department Marketing University of Otago).
69. Gallagher, C. A., Monroe, H. M., & Fish, J. L. (2000). An Iterative Approach to Classification Analysis. *Journal of Applied Statistics*, *29*, 256-266.
70. Lungu, C., Ersali, S., Szefler, B., Pirvan-Moldovan, A., Basak, S., & Diudea, M. V. (2017). DIMENSIONALITY OF BIG DATA SETS EXPLORED BY CLUJ DESCRIPTORS. *Studia Universitatis Babes-Bolyai, Chemia*, *62*(3).
71. Jolliffe, I.T. Principal Component Analysis; Springer Series in Statistics, 2nd ed.; Springer: New York, NY, USA, 2002; ISBN 978-0-387-95442-4.
72. Gosav, S., Praisler, M., & Birsa, M. L. (2011). Principal component analysis coupled with artificial neural networks—A combined technique classifying small molecular structures using a concatenated spectral database. *International journal of molecular sciences*, *12*(10), 6668-6684.
73. Defernez, M., & Kemsley, E. K. (1999). Avoiding overfitting in the analysis of high-dimensional data with artificial neural networks (ANNs). *Analyst*, *124*(11), 1675-1681.
74. A. Golbraikh, A. Tropsha, Beware of q2!, Journal of Molecular Graphics and Modelling. 20 (2002) 269–276. doi:10.1016/s1093-3263(01)00123-1.
75. Sattar, A. M. (2013). Gene expression models for the prediction of longitudinal dispersion coefficients in transitional and turbulent pipe flow. *Journal of Pipeline Systems Engineering and Practice*, *5*(1), 04013011.
76. P.P. Roy, K. Roy, On Some Aspects of Variable Selection for Partial Least Squares Regression Models, QSAR & Combinatorial Science. 27 (2008) 302–313. doi:10.1002/qsar.200710043.
77. K. Binder, D.M. Ceperley, J.-P. Hansen, M.H. Kalos, D.P. Landau, D. Levesque, H. Mueller-Krumbhaar, D. Stauffer, J.-J. Weis, Monte Carlo methods in statistical physics, Springer Science & Business Media, 2012.
78. R.G. Newcombe, Two-sided confidence intervals for the single proportion: comparison of seven methods, Statistics in Medicine. 17 (1998) 857–872. doi:10.1002/(sici)1097-0258(19980430)17:8<857::aid-sim777>3.0.co;2-e.