# Paper 1

## Using non-destructive tests for estimating uniaxial compressive strength and static Young’s modulus of carbonate rocks via some modeling techniques

* Predict relation between UCS(uniaxial compressive strength) and Young’s modulus(Es), dynamic poisson ratio(Vb) and dynamic young’s modulus(Ed).

**ML Models used**

1. Simple and Multivariate regression
2. Artificial neural network (ANN)
3. Support vector regression (SVR)

**Dataset:**

* Different intact limestone rock samples of Asmari formation (ranged from limestone to marl) were collected from five different dam sites located in the southwest of Iran.

**Parameters:**

Output : Uniaxial compressive strength (UCS) and the static modulus of elasticity (Es).

Inputs: Ultrasonic methods are used to get the data like density(p), primary wave velocity (Vp), and shear wave velocity (Vs), and these three parameters can be effective in predicting UCS and Es.

**Methods:**

* Multiple regression analyses

SRA provides a means of summarizing the relationship between two variables, whereas MRA is a statistical technique that uses several explanatory (predictor) variables to predict the outcome

* Artificial neural network

Multi-layer perceptron (MLP) is the best type of ANN and has the following three layers: (1) an input layer, which is applied to the present data in the network; it does not perform any computations, but serves only to feed the input data to the hidden layer.

(2) Hidden layer(s), which are applied to act as a collection of feature detectors.

(3) An output layer, which is applied to produce a suitable response to the given inputs.

ANN with one hidden layer had a lower performance than ANN with two hidden layers.

* SVM- support vector machine

Presenting the solution by means of small subset of training points, which gives very great computational advantages.

Idea of SVM is to map the training data from the input space into a higher dimensional feature space via a transfer function (/), and then to construct a separating hyper plane with maximum margin in the feature space.

* SVR- Support Vector Regression

Carries out linear regression in the high-dimension feature space by e-insensitive loss and, at the same time tries to minimize ||x||2 with the purpose of reducing model complexity.

The value of e influences the number of support vectors used for constructing the regression function. The bigger the value of e, the fewer support vectors are selected.

**Result:**

* SVR is fast comparetivvely to others.
* SVR more desirable

Results of simple regression analyses reported in these studies, it was concluded that no single relationship can be considered reliable for all rock types; they should be used with caution and only for specified rock types.

coefficient of determination (R2 ) and root mean square error (RMSE)

unconfined compressive strength and modulus of elasticity were correlated to the cited properties (Vd and Ed), using MRA, ANN and SVR.

* A comparison of SRA and MRA models indicates that the statistical index of coefficient of determination of equations, increased in going from one independent variable to two; however, RMSE values decreased
* To establish the most effective ANN structure, the number of hidden neurons in the ANN(1) model was selected as 6, with 9 and 8 selected for the ANN (2) model, The weights are adjusted to minimize differences (errors) between actual and desired output by using a sigmoid transfer function.
* SVR model is trained several times (at least 25 times) to select the best model. The most appropriate network structure provided the best training result.
* Scatter plot all 6 models

SVR and ANN are more powerful tools than regression methods in detecting and exploiting complex patterns on data by clustering, classifying, and ranking the data.

**Conclusion:**

SVR models surpassed the ANN and MLR models (with the higher R2 and lower RMSE).

# Paper 2

## Mechanical and Physical Based Artificial Neural Network Models for the Prediction of the Unconfined Compressive Strength of Rock

Basalt rocks as building stones is analysed

Porosity (g), dry density (c), Ultrasonic Pulse Velocity (VP), Point Load Index (Is(50)), Brazilian tensile strength (BTS), Schmidt Hammer Rebound hardness (SHR), slake durability index (SDI), shore Scleroscope hardness, and Brinell hardness tests, etc were recorded for specimens tested in the lab to develop indirect methods of estimating the rocks Unconfined Compressive Strength (UCS).

Simple regression (SR) analyses were performed to establish correlations between UCS and the results of each above-mentioned rock indices

Back Propagation-Artificial Neural Network (BP-ANN) approach was utilized to predict the USC of Basalt Rock.

**ML Models:**

Two ANN models were developed; one using the physical properties of rocks and the other one using the mechanical properties of rocks.

Artificial Neural Networks (ANNs), adaptive network-based fuzzy inference system (ANFIS), Genetic Programming (GP), and regression trees have been utilized in developing predictive models for complex problems

**Method:**

The performance of the ANN model in predicting UCS was compared to that of Multivariate Regression (MVR)

**Regression model:**

regression models in UCS prediction may result in several shortcomings. As such, these models predict the mean values only, and this will result in overproduction or under-prediction of the low and high UCS values.

**Result:**

results showed that the ANN model gave higher prediction performance compared to other models.

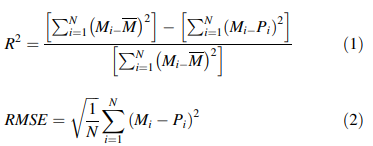
Both the fuzzy model and the MLR analyses gave a better prediction for the UCS than the SR

ANFIS model predicted the UCS with the highest accuracy among the other models

**SR – Simple Regression**

SR analyses were used to investigate the type of correlation between UCS (dependent variable) and each one of the index parameters

Different sorts of regression relationships such as linear, power, exponential, and logarithmic relationships between the UCS and rock indices were used to select the best relationship to fit the UCS of the rock. Both the coefficient of determination (R2 ) and Root Mean Square Error (RMSE) were calculated by this formula



where Mi is the measured UCS values, Pi is the predicted UCS values

To obtain higher R2 (lower RMSE), Multivariate Regression Analysis (MVR) could be used to establish a predictive model for predicting UCS amongst relevant rock properties.

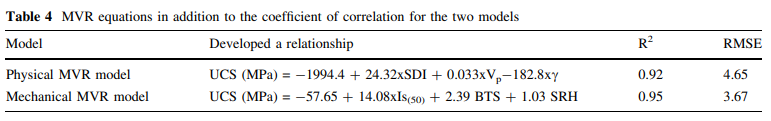
**Multivariate Regression Model (MVR)**

It is a higher version of SR, by considering multiple independent variables to obtain the best-fit equation with the highest R2 and lowest RMSE values between the input and output variables.

The statistical package for Social Science (SPSS) program was used to develop the MVR models.

For regression models, the value of R2 can be improved by increasing the number of parameters, fitting too many variables could result in overfitting problems.

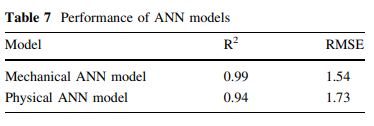
ance but add nothing to the model. In this study, two separate predictive models were developed. One model was developed to predict the UCS based on rock physical indices (c, Vp, and SDI), and the other model was developed based on rock’s mechanical strength indices (BTS, SHR, and Is(50))



**ANN Model**

Two ANN models were developed to predict the UCS. The first model was based on experimental results of the physical properties, including (c, VP, and SDI). The second model was based on the mechanical properties of the rock, namely (Is(50), BTS, and SHR).

The back-propagation learning (BP) algorithm was used for supervised learning to train the weights of the multilayer feed-forward neural network using gradient descent. BP algorithm searches for the minimum value of the error function by adjusting the weights.

****

**Conclusion:**

Both Multivariate Regression analyses and the ANN exhibited better predictive performances than SR.

Higher advantage for the ANN models over the MVR models.

Both physical and mechanical properties basedANN models can be used to predict the UCS value of Basalt rock with high accuracy.

Mechanical based-ANN model showed slightly better performance over the physical based-ANN model.

# Paper 3

## An ensemble tree-based machine learning model for predicting the uniaxial compressive strength of travertine rocks

Three standalone tree-based machine learning models (random forest (RF), M5 model tree, and multivariate adaptive regression splines (MARS)) for the prediction of UCS in travertine rocks from the Azarshahr area of northwestern Iran. Additionally, an ensemble committee-based artificial neural network (ANN) model was developed to integrate the advantages of the three standalone models and obtain further accuracy in UCS prediction.

Rock test data including p-wave velocity (Vp (Km/s)), Schmidt Hammer (Rn), porosity (n%), point load index (Is (MPa)), and UCS (MPa) were acquired from 93 travertine core samples.

Data was divided in ratio of 70:15:15 (train: validate: test)

**Data Set:**

Travertine, which is characterized by its high porosity, fine grain, and banded structure, is a particular form of carbonate deposit (also referred to as a type of limeston)

Travertine samples, for modeling purposes, were obtained from ten quarries in four different provinces in the area. Overall, 30 travertine blocks were collected.

Porosity (n%), P-wave velocity [Vp (Km/s)], Schmidt rebound hardness (Rn), Is (MPa), and UCS (MPa) were the primary physical and mechanical properties measured by the laboratory tests.

**Multiple linear and nonlinear regression models:**

To identify the relationships between multiple independent or explanatory variables (Xi) and a dependent variable (Yi), multiple linear regression (MLR) and multiple nonlinear regression (MNLR) model was used.

**Random Forest:**

The RF, known as an ensemble method, produces a set of repeated predictions of the same phenomenon by combining multiple decision tree algorithms.

Decision trees can be grouped into classification or regression types.

For each internal node of a rule (of the tree), data partitioning is performed repeatedly until a previously-specified stop condition is reached. Each terminal node, or leaf, has a simple regression model attachment applied specifically to that node section.

**M5 Model Tree**

This model, with linear regression functions at the terminal (leaf) nodes, has been used for continuous-class learning purposes and more recently for engineering problems.

The M5 model tree is a type of binary decision tree, which is generally applied to categorical datasets. The algorithm can be applied to quantitative data, which is an advantage in comparison with other tree base.

The M5 model tree is developed in two steps.

In the first step, the input-target data are divided into subcategories, and a decision tree is created. The division of the data is carried out based on two factors;

* first, the treatment of the standard deviation of the class values, and
* second, the calculation of the expected decrease in this error as a consequence of testing each attribute at that node

The standard deviation reduction (SDR) is computed.



M5 model tree selects nodes with the highest expected error reduction after scanning all of the possible divisions in the resulting tree structure.

**Multivariate Adaptive Regression Splines**

MARS is a multivariate nonparametric technique used to predict continuous numeric results. MARS is a flexible technique for organizing variables.

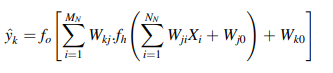
The MARS model aims to divide the solution space (i.e., the input-target matrix) into various intervals that indicate the feature space of the indicator variables. The individual splines are then fit to each interval .

each data interval, a unique mathematical regression equation is determined.

This process can be carried out in two stepwise methods, forward and backward.

**Ensemble model: the ANN-committee-based model**

A feed-forward multilayer perceptron (MLP) was employed to construct the ANN-committee model. The MLP was organized into three layers, including an input, one or more hidden layers, and an output layer.



where fh is the activation function of the hidden neuron, fo is the activation function of the output neuron, y^k are the computed output variables, NN is the number of neurons in the input layer.

**Method:**

Seventy percent of the original dataset was randomly selected for the training phase, and the remainder of the dataset was partitioned for the validation (15%) and testing (15%) phases.

Before developing the machine learning models, all variables were normalized to a value between zero and one by a scaling factor to guarantee that all inputtarget variables received equal attention during the training phase.

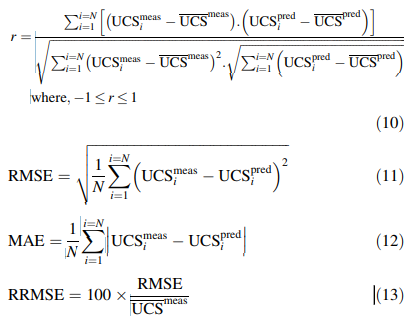
the initial number of weak learners (i.e., regression trees) was set to 800, and the initial number of leaves in each tree was set to five.

The M5 model tree was constructed using a set of tuning parameters for model initialization. A minimum tree split value of five, a smoothing value of 15, and a split threshold value of 0.05 were selected.

the MARS modeling process consisted of two stages: the forward and backward stages. In the forward stage, the reflected pair(s) of the BFs were added and the potential knots were identified to obtain the greatest decrease in the training error (RMSE).

construction and evaluation of the standalone (i.e., MARS, RF, M5 tree) models, integration of the predicted UCS values from the three developed machine learning models was performed to improve the prediction accuracy of the UCS data, and for subsequent use in an ANN model

These metrics are comprised of the correlation coefficient (r), root mean square error (RMSE), mean absolute error (MAE), and their normalized equivalents expressed in percentages (RRMSE and RMAE)





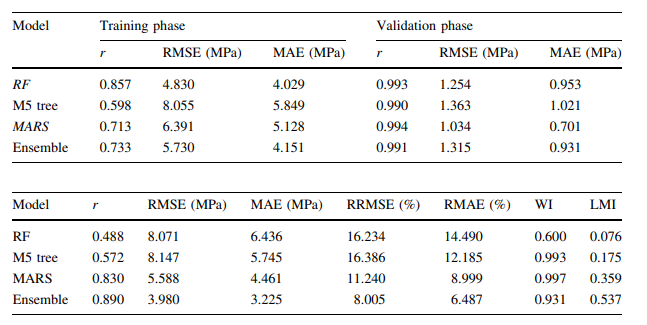
**Result:**

Simple regression analysis was used to determine the relationships between the predicted (i.e., UCS) and predictor variables (i.e., Vp, Rn, n% and Is).

comparison of the statistical performances of the RF, M5 tree and MARS models, as well as the ANN-based ensemble model, in predicting UCS in travertine rocks. Statistical performance metrics (r, RMSE, and MAE) are provided for both the training and validation phases of each model.

Simple regression analysis demonstrated a meaningful relationship between UCS and Is and a relatively weak relationship between USC and the other measured parameters.

The MARS model (r = 0.830, RMSE = 5.588 MPa, MAE = 4.461 MPa, WI = 0.997, and LMI = 0.359) showed superior performance in the testing phase for prediction of UCS, followed by the M5 tree and RF models.

****

# Paper 4

## Prediction of the strength and elasticity modulus of gypsum using multiple regression, ANN, and ANFIS models

Strength and deformability characteristics of the rock are also essential for determining their suitability for various construction purposes. Elastic and failure properties are mainly used in numerical and analytical methods in design approaches.

aims to predict the UCS and tangent Young’ s modulus (Et) of the gypsum based on its water content, porosity, sonic velocity, Schmidt hammer rebound number, and point load index. This is accomplished by using multiple regression (MR), ANN, adaptive neuro-fuzzy inference system (ANFIS) models.

**Dataset:**

Gypsum samples have been collected from various locations of Sivas basin and tested. The tests included the determination of water content, porosity, sonic velocity, Schmidt hammer rebound number, point load index, UCS and tangent Young’s modulus. Index test results were first correlated with UCS and E.

121 sample sets were used in the analyses. However, we have the data of 250 or more samples of gypsum. UCS, elasticity modulus, porosity, water content, point load Schmidt hammer index and sonic velocity tests were conducted on each rock sample.

**ML Model**

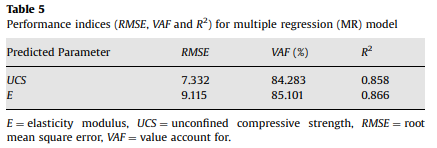
**Multiple regression models:**

The MR equation takes the form y ¼ b1x1+b2x2+?+bnxn+c, where {b1,b2,y,bn} are the regression coefficients. The parameter c is a constant representing the value of y when all the independent variables are 0.

The major conceptual limitation of all regression techniques is that one can only ascertain relationships, but never be sure about underlying causal mechanism.

Two MR analyses were carried out to correlate the measured UCS and modulus of elasticity to five rock properties selected, namely, effective porosity, point load index, Schmidt hammer rebound number, water content, and sonic velocity.

values for VAF and root mean square error (RMSE) indices were also calculated to control the performance of the prediction capacity of predictive models developed.

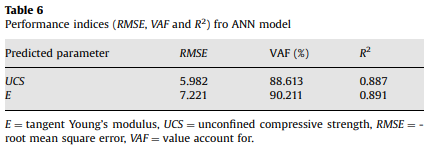


**Artificial neural networks (ANNs) models:**

A trained neural network can be thought of as an ‘‘expert’’ in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer ‘‘what if’’ questions.

Neural networks may be used as a direct substitute for auto correlation, multivariable regression, linear regression, trigonometric, and other statistical analysis and techniques.

Cross-correlation between predicted and observed values indicated that the ANN model constructed is highly acceptable for prediction of UCS and E. RMSE, VAF and R2 values are



**Adaptive neuro-fuzzy inference system (ANFIS) models**

In ANFIS, both of the learning capabilities of a neural network and reasoning capabilities of fuzzy logic were combined in order to give enhanced prediction capabilities, as compared to using a single methodology alone.

The fuzzy inference system (FIS) is a knowledge representation where each fuzzy rule describes a local behaviour of the system.

A hybrid intelligent system called ANFIS (the adaptive neuro-fuzzy inference system) for predicting UCS and E was also applied.

According to the RMSE, VAF, R2 values and crosscorrelation between predicted and observed values , ANFIS model constructed to predict UCS and E has a high prediction performance.

**Result:**

Use of MR, ANN and ANFIS models, for the prediction of UCS and elasticity modulus of gypsum, was described and compared.

Results of simple regression analyses, there are statistically meaningful relationships between elasticity modulus and UCS with Schmidt hammer rebound number, point load index, porosity, water content, and sonic velocity.

MR, ANN, ANFIS for the prediction of the elasticity modulus and UCS were then constructed using five inputs and two outputs.

t is shown that the constructed ANFIS model exhibits a high performance for predicting UCS and E. The performance comparison also showed that the ANFIS is a good approach for minimizing the uncertainties in the rock engineering projects.

# Paper 5

## Uniaxial compressive strength prediction through a new technique based on gene expression programming

Feasibility of gene expression programming (GEP) model in indirect determination of UCS values of sandstone rock samples is examined.

Several GEP models were constructed to estimate UCS of the rock and finally, the best GEP model was selected.

In order to indicate capability of the proposed GEP model, linear multiple regression (LMR) was also performed.

**DataSet:**

several laboratory tests including Brazilian test, density test, slake durability test and UCS test were conducted on 47 samples of sandstone which were collected from the Dengkil, Malaysia.

More than 40 sandstone block samples were taken from the Dengkil site and then transferred to the laboratory.

Tests conducted were uniaxial compressive strength (UCS) test, Brazilian test and slake durability test. Density of the samples was also determined for the correlation purpose.

**Gene expression programming:**

Genetic algorithm (GA) is one of the most popular evolutionary algorithms (EAs) which have two new modified versions of genetic programming (GP) and gene expression programming (GEP).

In GEP, solutions are encoded strings that follow Karva language called chromosome and are able to express as trees (ET).

only GP and GEP can develop mathematical equation for independent variable of problem, which this feature can be beneficial and practical tool for contractors and other engineers at the working field.

In GEP, by randomly combining the input variables and arithmetic functions (e.g+,-,/,x; Sin; Cos; Ln; Sqrt) as terminal set and function set, respectively, the solutions of the problem are made.

**UCS prediction:**

UCS prediction In this section, applications of two different models namely linear multiple regression (LMR) and GEP in estimating UCS were described in detail.

In UCS modeling, input parameters including DD, Id2 and BTS were utilized to construct the developed LMR and GEP models.

**LMR model:**

LMR model LMR method was used to develop a new multiple equation for UCS prediction using results of laboratory study.

Coefficient of determination (R2 ) values of 0.881 and 0.930 were achieved for training and testing of the proposed LMR equation.

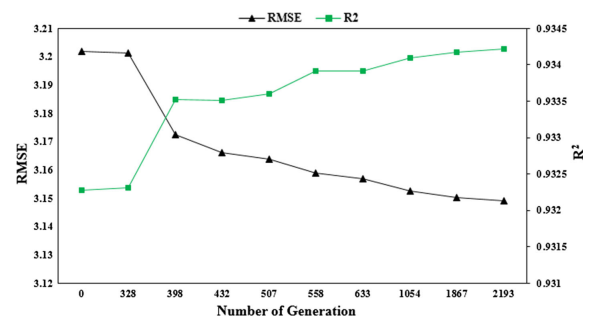
**GEP model:**

a GEP-based predictive model is developed to estimate the UCS using the same dataset of the LMR part and GeneXpro Tools 4.0 software.

The process of GEP modeling can be summarized in the following:

1. A fitness function was selected as a criterion for appraisal the merit of each chromosome. RMSE is the prevalent fitness function that used in the GEP modeling process.
2. Assigning a set of terminals (T) and functions (F) to the generation of chromosomes by combination of them.
3. GEP architecture parameters (i.e., head size, a number of genes and the number of chromosomes) should be assigned. The number of genes is the parameter that can determine the sub-ETs in each chromosome.
4. considering the suggested values by previous researchers, some other GEP models were constructed using trial-and error procedure.
5. to link the generated genes, there is a need to define a linking function. There are different linking functions such as subtraction (-), addition (+), division (/) and multiplication (x).

. The variation of the R2 and RMSE values during the GEP training phase is illustrated

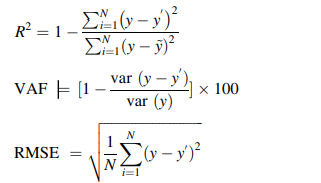


**Method:**

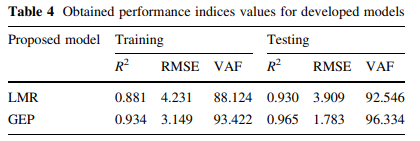
whole datasets were divided into two sections: training and testing. Based on several investigators (e.g. 38 datasets (80%) were utilized randomly for UCS model developments, whereas the remained 9 datasets (20%) were chosen to test the model.

**Result:**

R2 , variance account for (VAF) and RMSE were calculated to control the performance prediction of all proposed techniques:

****

where y and y0 are the measured and predicted values, respectively, y~ is the mean of the y values and N is the total number of data.

****

In order to predict UCS with high level of accuracy, LMR and GEP were performed and developed.

It was found that performance indices results of GEP equation are higher than LMR equation. The R2 equal to 0.965 for testing dataset recommends capability of the GEP method in estimating UCS, while value of 0.930 is obtained for LMR predictive model.