

Boise State University

ScholarWorks

---

Civil Engineering Faculty Publications and  
Presentations

Department of Civil Engineering

---

3-2021

## Machine Learning Methods to Map Stabilizer Effectiveness Based on Common Soil Properties

Amit Gajurel

*Texas A&M University*

Bhaskar Chittoori

*Boise State University*

Partha Sarathi Mukherjee

*Indian Statistical Institute*

Mojtaba Sadegh

*Boise State University*

---

### Publication Information

Gajurel, Amit; Chittoori, Bhaskar; Mukherjee, Partha Sarathi; and Sadegh, Mojtaba. (2021). "Machine Learning Methods to Map Stabilizer Effectiveness Based on Common Soil Properties". *Transportation Geotechnics*, 27, 100506. <https://doi.org/10.1016/j.trgeo.2020.100506>

This is an author-produced, peer-reviewed version of this article. © 2021, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 4.0 International license. The final, definitive version of this document can be found online at *Transportation Geotechnics*, <https://doi.org/10.1016/j.trgeo.2020.100506>

# Machine Learning Methods to Map Stabilizer Effectiveness Based on Common Soil Properties

**Amit Gajurel**

Graduate Student

Department of Civil and Environmental Engineering  
Texas A&M University  
College Station, TX

**Bhaskar Chittoori**

Associate Professor

Department of Civil Engineering  
Boise State University  
Boise, ID

**Partha Sarathi Mukherjee**

Associate Professor

Indian Statistical Institute (ISI)  
Kolkata, West Bengal, India

**Mojtaba Sadegh**

Assistant Professor

Department of Civil Engineering  
Boise State University  
Boise, ID

## Abstract

Most chemical stabilization guidelines for subgrade/base use unconfined compressive strength (UCS) of treated soils as the primary acceptance criteria for selecting optimum stabilizer in laboratory testing. Establishing optimal additive content to augment UCS involves a resource-intensive trial-and-error procedure. Also, samples collected from discrete locations for laboratory trials may not be representative of the overall site. This study aims to minimize the number of laboratory trials and help strategize sampling locations by developing spatial maps of UCS at different treatment levels for lime and cement. These spatial maps were developed using machine-learning techniques, and using a database compiled from various reported studies on lime and cement stabilization of soils in the United States. Supervised learning methods under regression and classification categories were used to quantify and classify UCS values after treatments, respectively. Commonly available soil properties like Atterberg limits, gradation, and organic contents along with treatment type and amount were used as predictors and UCS values as the response. Median  $R^2$  for the best regression model was 0.75 for lime and 0.82 for cement, while the Correct Prediction Rate (CPR) for the best classification model was 92% for lime and 80% for cement. Results showed that satisfactory predictions could be made regarding stabilizer effectiveness using simple soil information commonly available. Best performing models for cement treatment were selected for generating the spatial maps for two counties in Montana. Soil samples collected from these counties were tested with different cement contents to verify the predictions. The results indicate that the Pearson's correlation coefficient for the regression model was 0.78 and CPR for the classification model was 92%. The authors hope that this study and future studies like these will increase data-driven-decision-making in geotechnical engineering practices.

**Keywords:** Machine Learning, Chemical Stabilization, Spatial Mapping, Strength Prediction, Classification, Regression

## 1. Introduction

Growing urbanization has led to construction of civil infrastructures on challenging and sometimes problematic ground. Problematic soils are often treated with chemical additives to modify and/or enhance their mechanical performance and make them suitable for construction. Soil properties alteration using calcium-based additives has been widely used in stabilization and/or modification of subgrade due to its established history of performance, wide range of application, ease of field mixing, and controlled alteration of soil properties to meet the specifications [1–3]. Substantial stabilization work for pavement subgrades has been performed using calcium-based additives such as lime and cement [4–7]. Standard guidelines for stabilization and/or modification of subgrade and base layers of pavement have been established by several agencies, such as the California Department of Transportation [8], the National Lime Association (NLA)[9], the Portland Cement Association (PCA) [10], the Texas Department of Transportation

(TxDOT)[11], the Department of the Army, the Navy, and the Air Force[12]. As per these guidelines, initial estimates of type and amount of stabilizers are based on Atterberg limits, soil gradation, organic contents, and soluble sulfates. Soil samples are prepared using the initial estimates and tested for unconfined compressive strengths (UCS). This process is iterated for different types and amounts of stabilizers until a target strength is achieved. For obtaining a single UCS value for a given type and amount of stabilizer, testing procedure generally takes 7-28 days excluding sample preparation time. Also, since randomly selected sampling locations are put through this process, the optimized stabilizer type and amount may not be representative of the entire site. As a solution to these problems, the current study is focused on developing machine learning models that could predict optimal stabilizer amounts using simple soil properties. This predicted amount can then be verified quickly in the laboratory, thereby the proposed machine learning based method minimizes the time and effort required for establishing optimal stabilizer amounts. Moreover, these models can generate spatial maps of stabilizer effectiveness that can help strategize sampling locations for laboratory testing.

Machine learning methods have gained a lot of attention in recent years in geotechnical engineering, and have shown great potential for solving real world large scale engineering problems[13–26]. Some of the widely used supervised learning methods in geotechnical engineering are support vector machines (SVM), classification and regression trees (CART), and artificial neural networks (ANN). These methods have been used for various applications including, soil classification using cone penetration test (CPT) data [13], assessing landslide susceptibility [21], predicting peak shear strength of fiber reinforced soil [14], and predicting undrained side resistance for drilled shafts [18].

In this study, a database of UCS values and corresponding treatment along with associated Atterberg limits, particle size distribution, and organic content of the soil was compiled from various research studies. This database was subsequently used to train and test different supervised machine learning models. The regression models studied herein include multiple linear regression (MLR), generalized additive models (GAM), K-nearest neighbors (KNN), and support vector machine (SVM), and classification models include logistic regression (LR), discriminant analysis, KNN, and SVM. The regression models generate quantitative predictions (actual value of UCS), whereas the classification models generate categorical predictions (pass/fail for a given threshold UCS value). Model performance was assessed using measures such as coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE) for regression models, and true positive rate (TPR), false positive rate (FPR), correct prediction rate (CPR), Cohen's Kappa (K), and area under the curve (AUC) for classification models. Both quantitative and categorical model responses were used to generate spatial distributions of UCS values and pass/fail potential at a specified threshold, respectively, for different types and amounts of treatments. Proposed models facilitate development of scientific sampling strategies, minimize laboratory testing, and provide insights into suitability of a stabilization (type and amount) without the need to collect samples in the preliminary stage.

The paper is organized as follows: It first presents a background on the current state-of-practice with respect to stabilizer selection, sampling location strategies, and the use of machine learning in geotechnical engineering. This section is followed by a discussion on compilation of the database used for model development in this study. Later, the various training algorithms for regression and classification methods along with their performance metrics, model selection, and model evaluation are briefly described. This section is followed by a discussion on the results with emphasis on model selection and evaluation. Finally, quantitative and categorical responses of the best performing models are compared with the laboratory data for three soil samples from Montana treated with different stabilizer types and amounts.

## **2. Background**

### **2.1. Current Practice for the Selection of Optimal Stabilizer Type and Amount**

Lime, cement, and fly ash, are the most common chemical stabilizers in practice [3,27]. When these stabilizers are mixed with soil in the presence of water, several processes like cation exchange, flocculation & agglomeration, cementitious hydration, and pozzolanic reactions occur [28–31], which in turn induce reduction of plasticity and increase in workability and compressive strength. Performance of these stabilized mixes is contingent on several factors such as plasticity of natural soil, type and amount of stabilizer, mixing and compaction methods, curing conditions, organic matter, gradation and pulverization, clay fraction, mineralogy and presence of soluble sulfates [30–35]. Considering these factors, several federal and state agencies have developed guidelines for improving engineering properties of soils[8–12,36], majority of which specify initial selection of stabilizer type based on the plasticity index (PI). For instance, TxDOT [11] recommends using cement as the first choice for soil with  $PI < 15$  and

lime for soils with  $PI \geq 15$ . Similarly, Caltrans [8] recommends cement for soils with  $PI < 15$ , cement or lime for  $15 \leq PI < 35$ , and lime for  $PI \geq 35$ . Initial amount of stabilizer is determined based on soil classification (in the case of cement) or Eades and Grim test (in the case of lime) [37]. For example, the Department of the Army, the Navy, and the Air Force [12] recommend initial cement content from 5% to 11% depending on the Unified Soil Classification System (USCS) classification of soil. A similar process is followed by other agencies like PCA[10], National Cooperative Highway Research Program (NCHRP) [36], and Federal Highway Administration (FHWA)[38]. After the initial selection, the Proctor test is conducted to determine Maximum Dry Unit Weight (MDUW) and Optimum Moisture Content (OMC) for the selected type and amount of stabilizer. Samples were prepared at the corresponding MDUW and OMC to determine UCS. If the test UCS values exceed a preset threshold, durability studies are conducted for this combination of treatment. Once both UCS and durability criteria are satisfied, the selected type and amount of stabilizer is termed as optimum. The process is repeated until the optimum values are determined for a given soil sample. Strength criterion in most of the guidelines is established based on UCS values. Hence, many researchers have used UCS as a measure of stabilization performance [30,33,39–45].

## **2.2. Current Practice in Selecting Sampling Locations for Laboratory Testing**

Selecting the right locations for in-situ testing and sampling is of utmost importance for any geotechnical investigation. The samples brought to the lab dictate many of the decisions made regarding the superstructure. For example, AASTHO (1988) reported that, for roadway and airfield pavements, subgrade conditions are critical as weak subgrades warrant thickened pavement section, removal, and replacement of poor subgrade, or soil stabilization/improvement. Therefore, identification of sampling locations must be performed carefully. These locations are typically selected by a preliminary desk study/literature review, during which useful information that may be present about site is gathered. Several studies have revealed that the common source of information for desk studies are a) historical records from prior site investigations, b) performance records from nearby facilities, c) geologic reports and publications, d) geologic maps (bedrock and surficial geology maps), soil survey maps, topographic and LiDAR maps, utility maps, insurance maps, etc., e) aerial photographs, satellite/remote sensing imagery, and f) review of nearby or adjacent projects[46–50]. Majority of geologic maps are published by United State Geological Survey (USGS), which has compiled information on depth of rock, location of rock outcrop, engineering properties of various soil types, geologic history and groundwater. Soil survey maps, which are compiled by the efforts of United States Department of Agriculture (USDA), provide data on physical/chemical properties and suitability/limitation for use - at each soil parcel of the mapped area. AASTHO (1988) reported that the soil survey data are most suitable in the preliminary planning of any geotechnical project. Furthermore, TxDOT's test procedure [51] for surveying and sampling soil for highways establishes scope of the soil survey to determine the extent and location of each type of soil, location of suitable fill, base, and aggregate material, and evaluate the need for stabilization of subgrade, subbase, and base material.

However, literature is sparse on including prior soils stabilization performance knowledge for reconnaissance of project area and delineation of sampling sites. Therefore, authors believe that a spatial map of stabilization performance, along with currently available spatial information, will aid in preliminary investigation process for projects considering stabilization of problematic soils.

## **2.3. Machine Learning in Geotechnical Engineering**

The complexity of environmental phenomena along with incomplete understanding of the underlying processes in geotechnical engineering have resulted in favor of statistically derived empirical and semi-empirical relations in contrast to constitutive models[16]. Use of experimental data to derive and/or update correlation is widely practiced since traditional geotechnical engineering started – which is the core idea in machine learning. For instance, the studies by Ching and Phoon [52,53] presented an excellent example of how the existing correlations between measurement and design parameters can be derived or updated using a global database. Machine learning is a set of tools for modeling and understanding complex datasets, which has been extensively used in geotechnical engineering. For instance, the study by Goh [17] assessed the possibility of using artificial neural networks (ANN) to predict liquefaction potential by modeling its complex relationship with soil parameters, such as standard penetration test (SPT) values, mean grain size, equivalent dynamic shear stress, total and effective stress, along with earthquake magnitude, and maximum horizontal acceleration at ground surface. This study argued that the prediction results of ANN are more reliable than the conventional methods. Lai et al. [20] compiled a database of liquefied and non-liquefied soil after several earthquakes, and demonstrated the efficacy of logistic regression for evaluating the liquefaction potential using only cone penetration test (CPT) data. Pal [54] showed a similar performance of SVM in predicting liquefaction

potential. Goh and Goh [55] did a similar study on 226 field records for liquefaction potential assessments using CPT data through SVM modeling and reported a classification success rate of 98%. A study for automated classification of soil using CPT data also showed that the predictive accuracy of decision trees (DT), ANN, and SVM was high (83%) [13]. A study on landslide susceptibility assessment based on various geological, morphological, and environmental parameters showed that SVM outperformed ANN and decision trees, as well as assessments made by experts [21]. A study on the application of SVM in settlement prediction of shallow foundation on cohesionless soil argued its superiority against the existing empirical methods [56]. The above finding is consistent with the study by Chou et al. [14] in which accuracy of various machine learning and meta-ensemble techniques was compared with theoretical and empirical models for predicting shear strength of fiber-reinforced soil. The results showed that the machine learning models outperformed the theoretical and empirical models.

Limited studies were done in developing statistical models that relate stabilization parameter such as UCS, MDUW, and OMC with soil and chemical additives. Alavi et al. [57] used ANN to predict the values of MDUW and OMC of soil-stabilizer mix based on liquid limit (LL), plastic limit (PL), linear shrinkage (LS), % sand, % fines, lime content, cement content, and asphalt content by training on the dataset compiled by Burroughs [58]. The reported result showed a mean absolute error (MAE) of 0.38 kN/m<sup>3</sup> (2.42 pcf) for MDUW and 0.717% for OMC. A similar study was conducted to predict the UCS value and MDUW of cement stabilized soil by Das et al. [19] using different classes of ANN and SVM based on the dataset compiled by Burroughs [58]. SVM models outperformed ANN with an RMSE of 1.26 MPa (182.74 psi) for UCS value and 0.80 kN/m<sup>3</sup> (5.10 pcf) for MDUW. Furthermore, Mozumder and Laskar [22] reported superiority of ANN over MLR in predicting UCS value of geopolymer stabilized clayey soil based on several predictors, including LL, plasticity index (PI), % stabilizers, molar concentration of alkali activator, ratio of alkali to binder, ratio of Na/Al, and ratio of Si/Al. In another study, Mozumder et al. [23] demonstrated success of SVM in predicting UCS value using similar parameters and data, as in [22], and reported RMSE of 0.75 MPa (108.77 psi) and MAPE of 4.5. In similar studies by Tinoco et al. [25,26], the applicability of various machine learning models in prediction of UCS value for jet grouting material was presented. Among multiple learners, SVM's prediction was superior with RMSE of 1.99 MPa (288.62 psi) and R<sup>2</sup> value of 0.51.

Although the machine learning models discussed above give a good insight into the sensitivity of the parameters to the UCS values, these models have some limitations. For example, the models proposed in Das et al. [19] requires parameters such as OMC along with other soil properties as model inputs. Estimating OMC requires laboratory testing, which means field sampling is a requirement to use this model. On the other hand, other studies [22,23] use parameters such as sodium to aluminum ratio among others, which are specific to alkali activation studies. To the best of authors' knowledge, the only study that highlighted this issue was Tinoco et al. [15]. However, in their model, they used % of clay and organic content as the only two input soil parameters for capturing the effect of soil in the soil-cement mix. Also, for most of these models, the RMSE reported in the literature ranges from 0.75 MPa (108.77 psi) to 1.99 MPa (288.62 psi). This value is comparatively high when dealing with subgrade modification/stabilization in which treated strength generally ranges from 0.34 MPa (50 psi) to 2.06 MPa (300 psi) [8,12,39].

### 3. Current Study Approach and Database Development

This study used soil properties in publicly available databases (such as soil survey geographic database - SSURGO) as input parameters to predict UCS values and generate a spatial distribution of UCS values at various treatment levels. First, a database comprising of soil properties, stabilizer type and amount, and the corresponding UCS values were compiled from the published literature. This database was then used to train and test various models developed using machine learning algorithms. The best performing models were used to generate spatial stabilization performance for various treatments. Results were then verified by conducting laboratory tests on samples collected from locations within the spatially mapped data. A systematic pictorial representation of this approach is summarized in Fig. 1.

#### 3.1. Compilation of Database of Stabilized US-Soils

The scarcity of existing database for physical and chemical properties and strength of chemically stabilized US soils was evident in the literature review process. Therefore, one of the primary objectives of this study was to create a database of various properties of chemically stabilized soils in the US. As shown in Fig. 1., this is the first step in the application of machine learning. Authors collected data from digitally available journals, thesis dissertations, and technical reports [39,42,43,59–63]. The major challenge faced in dataset generation was finding a consistent set of reported soil parameters, compaction energy for sample preparation, standard curing time and procedure, and sample aspect ratio. Considering all these constraints, the majority of the data in the database was from the laboratory works

at Iowa State University in the 1960's. All the samples in the database are cylindrical with a height and diameter of 50.8 mm (2 in.) and is compacted with standard proctor energy. The attributes in the collected data include name and location of soil, unified soil classification system (USCS), organic content, % sand, % silt, % clay, % gravel, liquid limit (LL), plasticity index (PI), % stabilizer (i.e. lime or cement), UCS values (7 days soaked/un-soaked and 28 days soaked/un-soaked). The summary statistics of each data source is presented in Table A.1 and Table A.2 in the appendix. Locations and number of different soil samples collected within the US is presented in Fig. A.1 (appendix). Graphical representation of Spearman co-relation matrix for input parameter for lime and cement treatment are shown in Fig. A.2 and Fig. A.3 (appendix), respectively.

### **3.2. Soil Survey Geographic Database (SSURGO)**

Most of the geotechnical manuals use the soil survey data by United States Department of Agriculture (USDA) in desk study phase for planning the preliminary in-situ and laboratory studies. This database (SSURGO) is also considered among the most useful available information for highway design[46–50]. SSURGO is a database of the digitally produced version of soil survey data over the course of 100 years in the US. Examples of available data in SSURGO include a) chemical soil properties – cation exchange capacities, gypsum, etc., b) soil health properties – available water capacity, organic matter, etc., c) soil physical properties – water capacity, bulk density, Atterberg's limits, gradation, etc., and d) soil qualities and features – USCS classification, parent material, etc.

## **4. Model Development**

After the data compilation in the first stage, the second stage was to develop a predictive model using machine learning as shown in Fig. 1. Preprocessing of the data was conducted to select appropriate parameters and remove missing parameters before feeding the data for model development. Several parametric and non-parametric machine learning algorithms were implemented on the compiled database for regression and classification model development. The primary intent of the regression model is to predict the UCS of a treated soil characterized by its Atterberg limits, gradation, organic content, and type and amount of stabilizer. The primary goal of a classification model is to predict whether or not a soil, characterized by Atterberg limits, gradation, organic content, and type and amount of stabilizer, meets a certain threshold of strength. Implementation of the workflow was primarily performed using R language in RStudio[64,65].

### **4.1. Selection of Predictors**

Literature review of machine learning in stabilized soils' UCS prediction indicated that the primary focus of most studies leaned towards making inferences about the parameters – i.e., understanding relationships in the data. The focus of this study is more inclined towards predictive modeling – i.e., focus on accurate prediction. Therefore, for selecting the predictors, the authors considered a) availability of the parameters in the compiled database, b) availability of spatial distribution of parameters in SSURGO –spatial prediction of UCS is only possible if the spatially distributed parameters are available, c) parameters' effectiveness on enhancing stabilized soil strength from past studies[15,19,25,26,30,31], and d) multi-collinearity. Considering all these conditions, the soil parameters that qualified were LL, PI, % sand, % silt, % clay, organic content, and % stabilizer for UCS prediction. Separate models were developed for lime and cement treatments. For this particular study, 7-day soaked strength for cement treatment, and 28-day soaked strength for lime treatment were selected as UCS value to develop the model. Similar curing protocols were reported by different agencies [10,12]. The database has 167 complete examples for lime and 60 complete examples for cement treated soils. Interested readers can further utilize the generated database to develop model using UCS values for different curing protocols. Model parameters used in this study are summarized in Table A.3 (Appendix A) and data is presented in Appendix C.

### **4.2. Type of Machine Learning Models**

Several parametric – Multiple Linear Regression (MLR), Logistic Regression (LR), Discriminant Analysis (DA), semiparametric – Generalized Additive Model (GAM), and non-parametric – K – Nearest Neighbor (KNN), Support Vector Machine (SVM), machine learning models were chosen for this study. Such a wide range of models gives valuable insight into the advantage of increasing model's flexibility, at the expense of its interpretability, for improved predictive performance. Literature has pointed out that there is no specific algorithm which performs well for all data types[66–68]. A brief introduction to several machine learning algorithms used in the study for model development is presented in Appendix B. Details regarding these models can be found in published literatures[67–72].

#### **4.3. Prediction Accuracy of Regression Models**

The output of the regression model is quantitative. Some of the standard metrics used by authors for assessing the models' accuracy are discussed in the following section.

##### *Root Mean Square Error (RMSE)*

RMSE is calculated by taking the square root of the mean of squared residuals as given by Eq. (1). The unit of RMSE is the same as the unit of predictor and the value is generally interpreted as the measure of average distance between the actual and model predicted values. This value is more sensitive to large errors as the square operator gives greater weight to large errors.

$$RMSE = \sqrt{\frac{\sum_{i=0}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

where,  $y_i$  = actual value of response;  $\hat{y}_i$  = model predicted value of response;  $n$  = number of sample

##### *Mean Absolute Error (MAE)*

MAE is calculated by averaging the sum of absolute value of residuals as given by Eq.(2). The unit of MAE is the same as the unit of the predictor. The magnitude of MAE is less than or equal to RMSE and is less sensitive to large error.

$$MAE = \frac{\sum_{i=0}^n |y_i - \hat{y}_i|}{n} \quad (2)$$

##### *Coefficient of Determination ( $R^2$ )*

$R^2$  is a measure of percentage variance in the response explained by the model. This metric is the measure of the correlation between actual and predicted values, not the accuracy of the model. There are multiple formulations for  $R^2$ , but authors of the current study used the formulation in James et al [68] as given by Eq.(3).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

#### **4.4. Prediction Accuracy for Classification Models**

Classification models output discrete classes as well as a sort of continuous probability of class association for a given set of test data. Although, the formulation of regression and classification models are similar, assessment of models' performance is distinctly different between them. Some of the commonly used assessment metrics are discussed herein [69,73–76].

##### *Correct Prediction Rate (CPR)*

CPR is the most straightforward representation of classifiers' performance, as defined by the ratio of total correct predictions by the total number of samples given by Eq.(4). A confusion matrix is a cross-tabulation of predicted and actual classes for a given dataset. Sum of diagonal values of this matrix represents total correct class predictions. Further, computation of Cohen's kappa (K) provides information on whether the accuracy of CPR was due to chance (i.e., the relative frequency of each class) or efficacy of the classification model. Cohen's kappa is calculated by Eq. (5). The value of  $K=0$  refers to accuracy by chance, whereas  $K=1$  represents perfect agreement between the model's predictions and actual classes.

$$CPR = \frac{\text{sum of diagonal values of confusion matrix}}{\text{Total samples}} \quad (4)$$

$$Kappa(K) = \frac{O - E}{1 - E} \quad (5)$$

where, O = observed accuracy (from model's output i.e. CPR); E = expected accuracy (relative class frequency).

#### Area Under Curve (AUC)

AUC, the value of area of Receiver Operating Characteristics (ROC) curve, is a widely used measure of a classifier's performance [73,77]. Its values range from 0 to 1. For useful models, AUC shall exceed 0.5. The ROC curve for binary classification is a 2-D plot with a False Positive Rate (FPR) in the x-axis against a True Positive Rate (TPR) in the y-axis for every threshold. TPR and FPR are calculated, as per Eq. (6) and Eq.(7) respectively, by using the values from the confusion matrix. TPR represents the probability of actual true prediction when the model predicts true, which is also known as sensitivity. FPR represents probability of wrongly predicting the actual negative class as positive. AUC is particularly useful when comparing several classifiers without having to select a decision threshold.

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

where, TP = True Positive – Number of correctly predicted positive results; FN = False Negative – Number of wrongly predicted positive results

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

where, FP = False Positive – Number of wrongly predicted negative results; TN = True Negative – Number of correctly predicted negative results

#### **4.5. Model Selection and Model Assessment**

Model selection, herein, refers to finding the best performing model for a given machine learning algorithm by adjusting its hyperparameters. Model assessment refers to assessing the generalized performance of the selected model. The model assessment result is used to compare the performance of different machine learning algorithms. Model selection and model evaluations are commonly done using a k-fold cross validation. This technique involves randomly dividing the dataset into k folds. The first fold is stored as a validation set – i.e., independent test set, and the rest of the fold is used for model fitting. An appropriate model evaluation metric discussed in 4.3 and 4.4 is computed using the data on the validation set. This procedure is repeated 'k' times, where a different fold is treated as a validation set. The average value of the evaluation metric is reported as the k-fold cross validation estimate. For this study the value of k = 5 is used for cross-validation, as is recommended by several studies [14,67,78].

For parametric models like MLR, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and LR, model selection is redundant since the structure of the model is already predefined and only requires calculation of model parameter from the training data. Unlike the parametric models, the semi-parametric and non-parametric models have specific parameters that have to be established prior to model assessment, called hyper-parameters or tuning parameters. For instance, the degree of freedom in case of splines in GAM is estimated by gradient descent in R using base R package [65] coupled with Latin hypercube sampling to minimize the RMSE obtained by 5-fold cross-validation. A range of degrees of freedom was established for each parameter and Latin-hypercube sampling was done to choose a random combination of the degree of freedom. In the case of KNN, the value of "k" was optimized using a grid search algorithm. A range of "k" was selected and a performance metric was optimized using 5-fold cross-validation. In the case of regression, the value of "k" with minimum RMSE was chosen. While for the classification, the value of "k" giving maximum CPR was chosen. SVM regression with  $\epsilon$ -insensitivity loss function was used for



this study.  $\epsilon$  was calculated as per Eq.(8)[79], while the regularization parameter (C), degree of polynomial, and kernel width ( $\gamma$ ) were estimated by a grid search algorithm using 5-fold cross validation. The values of the parameter that returned lowest RMSE was chosen for the model. In case of SVM classification, a similar grid search approach was followed by maximizing the correct prediction rate for calculation of model parameters.

$$\epsilon = 3\sigma \sqrt{\frac{\ln n}{n}} \quad (8)$$

where,  $\sigma$  is the square root of noise variance calculated from k-nearest neighbor regression.

After estimating hyper-parameters of models, selected models' generalized performance was assessed by 200 simulations of 5-fold cross-validation. The distribution of the average value of the evaluation metric for train and test sets was constructed for all regression and classification models. The best model among various learning algorithms was chosen based on the median performance in the majority of the metrics. The result of the model selection and assessment are presented in sections 5.1 and 5.2, respectively.

## 5. Results and Spatial Visualization

### 5.1. Model Selection

For GAM, using natural cubic splines, lack of sufficient data for fitting eight parameters resulted in a rank deficient matrix and results were removed from consideration. The model parameters for regression and classification are detailed in Tables 1 and 2, respectively. A threshold of 1.03 MPa (150 psi) for lime treated soil and 2.06 MPa (300 psi) for cement-treated soil is considered in classification effort.

### 5.2. Model Evaluation

Summary of model evaluation results is presented in Tables A.4, A.5, and A.6 (appendix). These results are calculated for train and test sets by 200 random simulations of 5-fold cross validation (2-fold cross validation for cement classification models). Interquartile percentiles Q1, Q2, and Q3 of evaluation metrics for various regression and classification models are presented in Fig. 2 – Fig. 5. The error bars in these figures represent the Q1 and Q3 range, while the marker represents the Q2 value. A star is given to each best performing metric based on the Q2 value. The model with the highest number of stars is selected as the best performing model.

Among regression models with lime treatment, the best performing model was SVM, closely followed by KNN. The performance of all other remaining models were identical. SVM with radial kernel had a median  $R^2$  of 0.83 (train) and 0.75 (test), a median MAE of 0.37 MPa (train) and 0.44 MPa (test), and a median RMSE of 0.42 MPa (train) and 0.50 MPa (test). In the case of regression models with cement treatment, the best performing model was MLR followed by GAM-SS and SVM-radial. MLR had a median  $R^2$  of 0.88 (train) and 0.82 (test), MAE of 0.36 MPa (train) and 0.45MPa (test), and RMSE of 0.50Mpa (train) and 0.53 MPa (train).

Similarly, for classification models with lime treatment and a cutoff strength of 1.034 MPa (150 psi), SVM with radial kernel had the best performance, which was closely followed by KNN and LR. SVM with radial kernel had a median correct prediction rate (CPR) of 0.95 (train) and 0.92 (test), a median Cohen's Kappa of 0.66 (train) and 0.60 (test), a median area under curve (AUC) of 0.99 (train) and 0.98 (test), a median true positive rate (TPR) of 0.97 (train) and 0.94 (test), and a median false positive rate (FPR) of 0.06 (train) and 0.10 (test). As for the classification model for cement treatment with a cutoff strength of 2.068 MPa (300 psi) LDA has the best overall stable performance, with a median CPR of 0.92 (train) and 0.80 (test), a median Cohen's Kappa of 0.8 (train) and 0.52 (test), a median AUC of 0.99 (train) and 0.90 (test), a median TPR of 0.94 (train) and 0.77 (test), and a median FPR of 0.05 (train) and 0.17 (test). Although SVMs performed well in every other metric, AUC values of the SVMs' showed a bi-modal distribution while a normal distribution was expected. The coefficients of parameters for LR became unstable and a lack of sufficient data hindered the execution of QDA.

### **5.3. Spatial Visualization and Comparison**

An example of application for the best performing models with their optimized hyperparameters is used for spatial visualization of quantitative and categorical results. For this purpose, Broadwater County in Montana was chosen. Soil parameters for the model were obtained from USDA's SSURGO database. The database was accessed by ArcGIS using the Soil Thematic Map Tool extension from USDA. The required parameters were imported from the database and individually rasterized. The rasters were then imported to R and were stacked as a raster brick using the "raster" package [80]. Regression and classification models were then applied to the raster brick at each cell – which resulted in a raster with predicted UCS value and a binary response for pass/fail, respectively, for a given type and amount of stabilizer. For instance, the regression result for application of 6% cement for Broadwater County is shown in Fig. 6 and classification result for the threshold of 2.06 MPa (300 psi) is shown in Fig. 7 using SVM-radial and LDA, respectively. For cement treatment in the laboratory, the authors also collected two samples from Broadwater County and one sample from Garfield County which are named NTF\_LP (Location 1), NTF\_HP (Location 2), and DC (Location 3), respectively. The treated UCS specimens in the laboratory were made using a 2:1 aspect ratio and were cured at 100% humidity for seven days. The predicted UCS from the model using the parameters from the SSURGO data at that sampling location was plotted against the laboratory strength in Fig. 8. The solid line in that figure represents the 45-degree line and the dotted line represents the 15% variation from the 45-degree line. Pearson's correlation coefficient between the actual and predicted value was 0.78. In addition to the regression results, the CPR for the classification model was 92% for a total of 13 samples.

## **6. Discussions**

The results for this study indicate that for an end user who is interested in preliminary quantitative predictions, the regression model provides a decent estimation of UCS for any given type and amount of stabilization. Moreover, if the end user is more concerned about a treatment scenario passing a threshold or not, a classification model provides an excellent alternative. The performance results of these regression and classification models strongly support the argument that there isn't a single model that work best for all the datasets. Therefore, authors strongly suggest running a series of machine learning models before deciding on the specific model. Model performance and comparison of model predictions with laboratory values are discussed below.

### **6.1. Model Performance**

During regression model selection, lack of sufficient data to constrain the parameters of GAM model with natural cubic splines raised the issue of rank deficient matrix in both lime and cement models. Therefore, this model's results were excluded from further analysis. Although the development of the GAM model using smoothing splines was successful, its performance improvement against MLR was not significant (maximum of 6% in median values) for lime, and its performance was worse than MLR for cement treatment. This suggests that the residuals from MLR modeling are not due to its inability to model non-linear behavior and forcing non-linearity in to modeling effort – even with a model that has good regularization – doesn't increase the model prediction performance. Among non-parametric models, KNN for lime significantly outperformed MLR, i.e. a maximum of 47% increase in median performance. On the contrary, results for cement models using KNN was worse than MLR. Since KNN is an instance based learner, lack of sufficient data in case of cement treatment might have hindered its performance. Model selection results for SVM-linear and SVM-polynomial for lime and cement models revealed similar model performance. This suggests that projecting the dataset into kernel space of higher polynomial dimension wouldn't increase the model's prediction performance. The performance of SVM-linear and SVM-polynomial models for lime was similar to that of MLR. As expected, SVM-radial showed improvement in performance in comparison to MLR for lime models. But in case of the cement models, the performance of all the SVMs were worse than MLR. Although in this study, regularization parameter (C) and the degree of the polynomial (n) were optimized, the  $\epsilon$  – insensitivity loss was taken as a constant as suggested by Cherkassky and Ma [80]. Authors believe in the development of future models, tuning the  $\epsilon$  – insensitivity loss might improve the performance as SVM have been known to be very sensitive to hyper parameters [72,82]. Therefore, among all the regression models for lime, SVM-radial's performance was the best with a median test set MAE and RMSE of 0.44 MPa and 0.50 MPa, respectively. For cement model, MLR performance was superior with a median test set MAE and RMSE of 0.45 MPa and 0.53 MPa, respectively.

In the case of classification for lime with threshold of 1.03 MPa, the best performing model was SVM-radial, which was closely followed by KNN and LR. The " $\gamma$ " parameter for SVM-radial model was 0.02 and performance for all other models were comparatively similar. This suggests the presence of a somewhat linear separating decision

boundary between the two class of samples. In the case of cement samples, the coefficients for LR became unstable, AUC values for SVMs showed a bimodal distribution, and small sample size barred the development of QDA model. Therefore, LDA performance was optimum for cement treatment models. The performance of best model for lime treatment has a median CPR of 0.95 (train) and 0.80 (test), Cohen's kappa of 0.66 (train) and 0.61 (test), AUC of 0.99 (train) and 0.90 (test), TPR of 0.97 (train) and 0.94 (test), and FPR of 0.06 (train) and 0.10 (test). Similarly, the best performing cement treatment models has a median CPR of 0.92 (train) and 0.80 (test), Cohen's Kappa of 0.80 (train) and 0.52 (test), AUC of 0.99 (train) and 0.90 (test), TPR of 0.94 (train) and 0.77 (test), and FPR of 0.05 (train) and 0.11 (test).

The performance of regression model in this study using a limited set of parameters was comparable to those of previous studies, but provides a wide range of application possibility as the data needed to constrain parameters in the proposed model are readily available.

## **6.2. Comparison of Models' Predictions**

Prediction of stabilized strength using the parameters from SSURGO database at each of the three locations for the different amount of cement treatment showed good performance when compared to the laboratory data as shown in Fig. 8. Even though the aspect ratio for the soil in the laboratory and the aspect ratio of the sample used in the model development was different, most of the prediction was within/near the  $\pm 15\%$  of the 45-degree line with Pearson's correlation coefficient of 0.78. Moreover, the classification success rate for the threshold of 2.06 MPa (300 psi) was 92.31%, i.e. only one sample out of the 13 samples was mispredicted. The failing sample, which was predicted as pass instead of fail, had a UCS of 1.91 MPa. This value is very near to the cutoff point and the difference between the predicted and actual strength was 0.15 MPa, which is 7.24% of the cutoff value. Therefore, the results of regression as well as the classification models showed good prediction performance and further supports the application of these models for real-world applications.

## **7. Conclusion**

The purpose of this study was to assess applicability of data-driven models for prediction of UCS value for stabilized soils. Several models were developed using different machine learning algorithms and using only the parameter that are spatially available in databases like SSURGO. Such model would be very beneficial in the selection of optimum initial stabilizer in the laboratory works, aiding preliminary strength assessment for a given spatial location and delineation of laboratory sample collection, among others, without prior lab test. In addition, this model makes use of the vast amount of spatial soil data that is readily available. A database for the US soils was compiled and used for regression model (using MLR, GAM, KNN, SVMs) and classification model (using LR, DA, KNN, SVMs) development. The major findings from this study are:

- a. Median performance of regression models developed herein are similar or better than the performance reported in the literature for strength prediction of chemically stabilized soils.
- b. Performance of classification models developed was found to be satisfactory, with respect to the chosen threshold values.
- c. Comparison of the best performing regression and classification models' predictions with the observed laboratory strength further supports the applicability of these models for the US soils.
- d. Model prediction can be used to select an optimum stabilizer content, get an overview of stabilization performance for any spatial area within the US, and strategize the sampling operations for chemical stabilization projects.
- e. Analysis of the models' results strongly suggests that a single model cannot perform well for all data types. These results are in perfect agreement with research in the machine learning community.

## **8. Appendices**

Appendices are attached in the upload.

## 9. Data Availability Statements

Models or code that support the findings of this study are available from the corresponding author upon reasonable request. The data used in the analysis can be found in TC304 database open sharing project 304dB.

## 10. Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. However, the authors would like to heartily thank Montana Department of Transportation (MDT) for supporting this research by providing soil samples that were tested for comparison with the model predictions.

## References

- [1] Chittoori S, Pedarla A, Puppala AJ, Hoyos LR, Nazarian S, Saride S. Leachate Studies on Lime and Portland Cement Treated Expansive Clays. *Geo-Frontiers 2011 Adv. Geotech. Eng.*, vol. 41165, 2011, p. 4479–88.
- [2] Nelson J, Miller DJ. *Expansive soils: Problems and Practice in Foundation and Pavement Engineering*. John Wiley & Sons; 1997.
- [3] Puppala AJ. Advances in ground modification with chemical additives: From theory to practice. *Transp Geotech* 2016;9:123–38.
- [4] Berry KM, Hasan A, Cooling TL. Geotechnical Aspects of the St. Louis-Lambert International Airport Expansions. *Aviat. A World Growth*, 2007, p. 197–204.
- [5] Cole LW, Cepco C. PENNDOT'S Efforts to Improve Pavement Sub-grade During Highway Reconstruction. *Airf. Highw. Pavements Spec. Conf.*, Atlanta, Georgia: American Society of Civil Engineers; 2006.
- [6] Petry TM, Little DN. Review of Stabilization of Clays and Expansive Soils in Pavements and Lightly Loaded Structures—History, Practice, and Future. *J Mater Civ Eng* 2002;14:447–60.
- [7] Thompson MR. Deep-plow lime stabilization for pavement construction. *Transp Eng J ASCE* 1972;98:311–23.
- [8] Jones D, Rahim A, Saadeh S, Harvey JT. *Guidelines for the Stabilization of Subgrade Soils In California*, University of California Pavement Research Center; 2011.
- [9] National Lime Association. *Mixture Design and Testing Procedures for Lime Stabilized Soil*. Technical Brief; 2006.
- [10] Portland Cement Association. *Soil Cement Laboratory Handbook*. 1992.
- [11] TxDOT. *Guidelines for Modification and Stabilization of Soils and Base for Use in Pavement Structures*. Texas Department Transp 2005.
- [12] U.S. Army TM 5-882-14/AFM 32-1019. *Soil Stabilization for Pavements*. 1994.
- [13] Bhattacharya B, Solomatine DP. Machine learning in soil classification. *Neural Networks* 2006;19:186–95.
- [14] Chou J-S, Yang K-H, Lin J-Y. Peak Shear Strength of Discrete Fiber-Reinforced Soils Computed by Machine Learning and Metaensemble Methods. *J Comput Civ Eng* 2016;30:04016036.
- [15] Tinoco J, António Alberto Santos C, Da Venda P, Correia AG, Lemos L. A Data-driven Approach for qu Prediction of Laboratory Soil-cement Mixtures. *Procedia Eng* 2016;143:566–73.
- [16] Das SK. Artificial Neural Networks in Geotechnical Engineering: Modeling and Application Issues. *Metaheuristics Water, Geotech. Transp. Eng.*, 2013, p. 231–67.
- [17] Goh ATC. Seismic Liquefaction Potential Assessed by Neural Networks. *J Geotech Eng* 1994;120:1467–80.
- [18] Goh ATC, Kulhawy FH, Chua CG. Bayesian neural network analysis of undrained side resistance of drilled shafts. *J Geotech Geoenvironmental Eng* 2005;131:84–93.
- [19] Das SK, Samui P, Sabat AK. Application of Artificial Intelligence to Maximum Dry Density and Unconfined Compressive Strength of Cement Stabilized Soil. *Geotech Geol Eng* 2011;29:329–42.
- [20] Lai S, Chang W, Lin P. Logistic Regression Model for Evaluating Soil Liquefaction Probability Using CPT data. *J Geotech Geoenvironmental Eng* 2006;132:694–704.
- [21] Marjanović M, Kovačević M, Bajat B, Voženilek V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng Geol* 2011;123:225–34.
- [22] Mozumder RA, Laskar AI. Prediction of unconfined compressive strength of geopolymer stabilized clayey soil using Artificial Neural Network. *Comput Geotech* 2015;69:291–300.

- [23] Mozumder RA, Laskar AI, Hussain M. Empirical approach for strength prediction of geopolymer stabilized clayey soil using support vector machines. *Constr Build Mater* 2017;132:412–24.
- [24] Suman S, Mahamaya M, Das SK. Prediction of Maximum Dry Density and Unconfined Compressive Strength of Cement Stabilised Soil Using Artificial Intelligence Techniques. *Int J Geosynth Gr Eng* 2016;2:1–11.
- [25] Tinoco J, Gomes Correia A, Cortez P. Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns. *Comput Geotech* 2014;55:132–40.
- [26] Tinoco J, Correia AG, Cortez P. Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time. *Constr Build Mater* 2011;25:1257–62.
- [27] Tastan EO, Edil TB, Benson CH, Aydilek AH. Stabilization of organic soils with fly ash. *J Geotech Geoenvironmental Eng* 2011;137:819–33.
- [28] Mitchell JK, Soga K. Fundamentals of soil behavior. vol. 3. New York: John Wiley & Sons; 2005.
- [29] Chittoori B, Puppala AJ. Quantitative Estimation of Clay Mineralogy in Fine-Grained Soils. *J Geotech Geoenvironmental Eng* 2011;137:997–1008.
- [30] Thompson MR. Factors influencing the plasticity and strength of lime-soil mixtures. University of Illinois at Urbana Champaign, College of Engineering. Engineering Experiment Station.; 1967.
- [31] Prusinski J, Bhattacharja S. Effectiveness of Portland cement and lime in stabilizing clay soils. *Transp Res Rec J Transp Res Board* 1999;215–27.
- [32] Croft JB. The influence of soil mineralogical composition on cement stabilization. *Geotechnique* 1967;17:119–35.
- [33] Pedarla A, Chittoori S, Puppala AJ. Influence of mineralogy and plasticity index on the stabilization effectiveness of expansive clays. *Transp Res Rec* 2011;2212:91–9.
- [34] Tremblay H, Duchesne J, Locat J, Leroueil S. Influence of the nature of organic compounds on fine soil stabilization with cement. *Can Geotech J* 2002;39:535–46.
- [35] Hampton MB, Edil TB. Strength gain of organic ground with cement-type binders, *Soil Improvement for Big Digs*. Soil Improv. big digs, ASCE; 1998, p. 135–48.
- [36] Little DN, Nair S. Recommended practice for stabilization of subgrade soils and base material. NCHRP Project 20-07 Final report. *Transp Res Board* 2009.
- [37] Eades JL, Grim RE. A quick test to determine lime requirements for lime stabilization. *Highw Res Rec* 1966.
- [38] Carpenter SH, Croveti MR, Smith KL, Rmeili E, Wilson T. Soil And Base Stabilization And Associated Drainage Considerations. Volume I: Pavement Design And Construction Considerations. Final Report. Federal Highway Administration. Office of Technology Applications; 1992.
- [39] Veisi M, Chittoori B, Celaya M, Nazarian S, Puppala AJ, Solis C. Accelerated Stabilization Design of Subgrade Soils. Center for Transportation Infrastructure Systems, The University of Texas at El Paso; 2010.
- [40] Little N. Evaluation of structural properties of lime stabilized soils and aggregates. vol. 1. National Lime Association; 1999.
- [41] Wen H, Muhunthan B, Wang J, Li X, Edil T, Tinjum JM. Characterization of Cementitiously Stabilized Layers for Use in Pavement Design and Analysis. *Transportation Research Board*; 2014.
- [42] Sariosseiri F, Muhunthan B. Effect of cement treatment on geotechnical properties of some Washington State soils. *Eng Geol* 2009;104:119–25.
- [43] Davidson DT, George PL, Mateos M, George KP. Moisture-Density, Moisture-Strength and Compaction Characteristics of Cement-Treated Soil Mixtures. *Bull* 353 1962.
- [44] Ladd CC, Moh ZC, Lambe TW. Recent soil-lime research at the Massachusetts Institute of Technology. *Highw Res Board Bull* 1960.
- [45] Chittoori BCS, Puppala AJ, Wejrungsikul T, Hoyos LR. Experimental Studies on Stabilized Clays at Various Leaching Cycles. *J Geotech Geoenvironmental Eng* 2013;139:1665–75.
- [46] AASTHO. Manual on Subsurface Investigations. Washington, D.C.: 1988.
- [47] Leohr JE, Lutenege A, Rosenblad B, Boeckmann A. Geotechnical Site Characterization. Washington, D.C.: 2016.
- [48] WSDOT. Geotechnical design manual. 2010.
- [49] CDOT. Geotechnical Design Manual. 2017.
- [50] IDOT. Geotechnical Manual. 2015.
- [51] Tex-100-E. Surveying and Sampling Soils for Highways. 1999.

- [52] Ching J, Phoon K-K. Transformations and correlations among some clay parameters—the global database. *Can Geotech J* 2014;51:663–85.
- [53] Ching J, Phoon K-K. Correlations among some clay parameters—the multivariate distribution. *Can Geotech J* 2014;51:686–704.
- [54] Pal M. Support vector machines-based modelling of seismic liquefaction potential. *Int J Numer Anal Methods Geomech* 2006;30:983–96.
- [55] Goh ATC, Goh SH. Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data. *Comput Geotech* 2007;34:410–21.
- [56] Samui P. Support vector machine applied to settlement of shallow foundations on cohesionless soils. *Comput Geotech* 2008;35:419–27.
- [57] Alavi AH, Gandomi AH, Gandomi M, Sadat Hosseini SS. Prediction of maximum dry density and optimum moisture content of stabilised soil using RBF neural networks. *IES J Part A Civ Struct Eng* 2009;2:98–106.
- [58] Burroughs VS. Quantitative criteria for the selection and stabilisation of soils for rammed earth wall construction. University of New South Wales, 2001.
- [59] Pietsch PE, Davidson DT. Effects of lime on plasticity and compressive strength of representative Iowa soils. Iowa State University, 1961.
- [60] Davidson DT, Bruns EW. Comparison of type I and type III Portland cements for soil stabilization. *Highw Res Board Bull* 1960.
- [61] Remus MD, Davidson DT. Relation of Strength to composition and density of lime-treated clayey soils. *Highw Res Board Bull* 1961.
- [62] Whitehurst EA. Stabilization of Tennessee Gravel and Chert Bases. *Highw Res Board Bull* 1955.
- [63] Davidson DT, Mateos M, Barnes HF. Improvement of Lime Stabilization of Montmorillonitic Clay Soils with Chemical Additives. *Highw Res Board Bull* 1960.
- [64] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2017.
- [65] RStudio Team. RStudio: Integrated Development for R 2016.
- [66] Friedman J. Introduction to computational learning and statistical prediction. Technical report, Department of Statistics, Stanford University; 1995.
- [67] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. vol. 1. Springer series in statistics New York; 2001.
- [68] James G, Witten D, Hastie T, Tibshirani R. Introduction to Statistical Learning. Springer; 2013.
- [69] Kuhn M, Johnson K. Applied predictive modeling. vol. 26. Springer; 2013.
- [70] Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20:273–97.
- [71] Vapnik V. Statistical learning theory. 1998. vol. 3. Wiley, New York; 1998.
- [72] Vapnik V. The nature of statistical learning theory. Springer science & business media; 2013.
- [73] Majnik M, Bosni Z. ROC analysis of classifiers in machine learning : A survey 2013;17:531–58.
- [74] Ferri C, Hernández-Orallo J, Modroiu R. An experimental comparison of performance measures for classification. *Pattern Recognit Lett* 2009;30:27–38.
- [75] Tofallis C. A better measure of relative prediction accuracy for model selection and model estimation. *J Oper Res Soc* 2015;66:1352–62.
- [76] Gneiting T. Making and evaluating point forecasts. *J Am Stat Assoc* 2011;106:746–62.
- [77] Bradley A. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 1997;30:1145–59.
- [78] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int. Jt. Conf. Artif. Intell.*, vol. 14, Montreal, Canada; 1995, p. 1137–45.
- [79] Cherkassky V, Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 2004;17:113–26.
- [80] Hijmans RJ, Etten J van. raster: Geographic analysis and modeling with raster data 2017.