

Investigating the Most Important Geomechanical Parameter of Unsaturated Soils using Different Artificial Intelligence Methods*

Reserch Articl

Hadi Fattahi¹ Fateme Jiryae²

1. Introduction

Today, engineers in most mining and civil projects deal with unsaturated soils. Unsaturated soil is a multi-phase porous medium composed of soil grains and the liquid and gas phase, which is usually air and water. Soil behavioral modeling and designs, regardless of unsaturated soil conditions, although simpler, reduce the safety factor and increase costs. The principle of effective stress is a basic principle in soil mechanics that transforms a porous environment consisting of several phases and multi-stresses into a continuous environment consisting of one phase and one stress to allow the application of the principles of solid mechanics. The determination of shear strength in unsaturated soils in the Bishop relation, in addition to the effective stress, depends on a parameter called the effective stress parameter (χ). This parameter determines the degree of suction participation in the effective stress. The value of this parameter is one in the saturated state and zero in the dry state. The value of this parameter is usually determined by thriaxial, shear, pressure plate, and filter paper tests. The relationship between the χ parameter and its related parameters is a nonlinear and complex relationship and the definition of this relationship is experimentally associated with many errors. Accurate predictions of χ can be achieved using advanced approaches such as intelligent methods and inferential systems. Among the intelligent methods, some are used for classification problems and some for regression problems. Some of them are able to solve both types of problems. Support Vector Machine (SVM) is one of the most widely used high performance methods. K-Nearest Neighbor (KNN) and Random Forest (RF) methods have also proven in many cases that they can work well.

In this research, to predict the χ parameter due to its effective parameters, artificial intelligence methods such as random forest, support vector machine and k-nearest neighbor are used for modeling. The modeling is performed in WEKA software and will use laboratory data collected from several articles. Inlet parameters in this model are soil-water characteristic curve fitting parameter, confining pressure, suction and volumetric water content and output variable, χ parameter.

2. The most important methods used in this research

2.1. Support vector machine. The support vector regression method consists of training algorithms that

work based on the classifier support vector machine. This method achieves a model for prediction by using statistical learning theory and minimizing structural risk. The SVR method was first proposed by Vepnik (1995). In this method, a nonlinear mapping is used to train the data and the data is transferred to a higher dimensional space and then the error function ϵ in Equation 1 is used to construct the regression shape.

$$L(y, f(x, a)) = |y - f(x, a)|_{\epsilon} \quad (1)$$

$$= \begin{cases} 0 & \text{if } |y - f(x, a)| \leq \epsilon \\ |y - f(x, a)| - \epsilon & \text{Otherwise} \end{cases}$$

Equation 1 shows that errors with values less than ϵ are not considered. Nonlinear relation, the input and output parameters are written as an optimal function according to Equation 2.

$$f(x) = w^T \varphi(x) + b \quad (2)$$

where x is the input variable vector, w is the weight vector, and b is the bias. This function is known as the ϵ -Insensitive Loss Function and its solution can be expressed as a quadratic problem in Equation 3.

$$\begin{cases} \text{Min } \frac{1}{2} w^T w + C \sum_{i=1}^L (\xi_i - \xi_i^*) \\ \text{s.t. } y_i - w \cdot \varphi(x_i) - b \leq \epsilon + \xi_i \\ w \cdot \varphi(x_i) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i \geq 0 \\ \xi_i^* \geq 0 \end{cases} \quad (3)$$

where C is the penalty parameter and ζ is the auxiliary variable. By introducing the Lagrangian coefficients and applying the kernel function instead of the internal multiplication of the operators, the above optimization problem is simplified and Equation 2 becomes Equation 4.

$$f(x) = \sum_{x_i \in SV} (a_i - a_i^*) K(x_i, x_j) + b \quad (4)$$

where a_i and a_i^* Lagrange coefficients and $k(x_i, x_j)$ are kernel functions.

2.2. K-Nearest Neighbor

The k-nearest neighbor method is a non-parametric statistical method used for classification and regression problems. This method is based on the concept of similarity, and predictions are made based on similar situations that have occurred in the past. In other words, by observing the independent variables in real time, the model searches for patterns similar to the current conditions in the historical series and selects possible

*Manuscript received: 17 December 2021; Revised, , Accepted, 16 May 2022 .

¹. Corresponding author. Associate Professor in Rock Mechanics, Faculty of Earth Sciences Engineering, Arak University of Technology, Arak, Iran. **Email:** H.fattahi@arakut.ac.ir

². MSc Student in Mining Engineering, Faculty of Earth Sciences Engineering, Arak University of Technology, Arak, Iran

options. Equation 5 shows The vector ($1 \times m$) of the predictive values of x_j at time t .

$$Pr_{jt} = (x_{jt}) \quad j = 1 \dots m \quad (5)$$

Also, the matrix ($n \times m$) of the value of predictor variables x_j in the historical time series is as Equation 6.

$$Pr_{j,(t-i)} = x_{j,(t-i)} \quad (6)$$

$$j = 1 \dots m \quad i = 1 \dots n$$

Using the distance meter function (Dist), in Equation 7, the distance between the vector $Pr_{j,t}$ and the rows of the matrix $Pr_{j,(t-i)}$ is calculated.

$$Dist(t-i) = f(w_j \cdot x_{j,(t-i)} \cdot x_{jt}) \quad (7)$$

where j is the index corresponding to the predictor variables and i is the index corresponding to the time step in the historical series. w_j are also the weights of the predictor variables. In order to improve the performance of the model, appropriate methods should be used to estimate the best neighbors and develop data transfer functions and distance measurement functions.

3. Case studies

The database included in the analyzes contains 120 data from the results of thriaxial, shear, pressure plate and filter paper tests, which were collected from several articles. These data include 7 variables: suction (s), bubbling pressure (hb), net confining pressure (p), residual water content (θ_r), saturated volumetric water content (θ_s), soil-water characteristic curve fitting parameter (λ), and the effective stress parameters (χ). For initial processing, the parameters s , hb , p , θ_r and θ_s are dimensionlessly entered into the model so that the dimensionless confining pressure parameter is obtained from the $\frac{P}{P_0}$ ratio (where the parameter is $P_0 = 101.325\text{KPa}$) and the suction parameter is dimensioned by the ratio $\frac{s}{hb}$. Also, from the $\frac{\theta_r}{\theta_s}$ ratio, the dimensionless volume water content parameter is obtained. Thus, in the modeling, the input parameters include $\frac{P}{P_0}$, $\frac{s}{hb}$, $\frac{\theta_r}{\theta_s}$ and λ , and the output parameter χ . Of this number of normalized data, 80% are used as training data and 20% as test data to evaluate the models.

4. Results and conclusion

Given that a significant percentage of the earth's surface is covered by unsaturated soils, it is obvious that these soils are encountered during operations in most mining and civil projects. The determination of shear strength in unsaturated soils in a Bishop relationship depends on a parameter called the χ parameter, which is a function of suction. In this research, models were developed and evaluated to predict the χ parameter. These models were performed using three intelligent methods of random forest, support vector machine and k-nearest neighbor in the form of WEKA software and in it, 120 laboratory samples collected from several papers were used. The results showed:

- Random forest model in both training dataset ($R^2 = 0.968$ and $RMSE = 0.051$) and test dataset ($R^2 = 0.869$ and $RMSE = 0.107$), with high accuracy can accurately predict the parameter χ ;
- Support vector machine and k-nearest neighbor models also have good performance and accuracy in training dataset;
- Sensitivity analysis performed on the parameters showed that the parameter of volumetric water content is very important in predicting the parameter χ , and in contrast, the parameter of the soil-water characteristic curve fitting is less important than other parameters.
- In general, it can be said that WEKA powerful and practical software can be used as a powerful tool in data mining and earth engineering sciences in estimating geomechanical parameters.

تخمین مهم ترین پارامتر ژئومکانیکی خاک های غیر اشباع با به کارگیری روش های مختلف هوش مصنوعی*

مقاله پژوهشی

هادی فتاحی^(۱)فاطمه جبریایی^(۲)

چکیده در پیش بینی رفتار سازه های ژئوتکنیکی که بر خاک های غیر اشباع احداث می شوند، دانستن خواص ژئومکانیکی به ویژه تنش مؤثر از اهمیت ویژه ای برخوردار است. تأثیر مکش در خاک های غیر اشباع موجب تغییر در رفتار حجمی و برشی خاک می شود. از این رو در تعیین مقاومت برشی در خاک های غیر اشباع، پارامتر تنش مؤثر (χ) به عنوان تابعی از مکش نقش اساسی دارد. تعیین این پارامتر نیازمند زمان و هزینه زیادی است که در آزمون های آزمایشگاهی صرف می شود. هدف از این تحقیق ارزیابی چند روش هوشمند برای ایجاد مدل هایی است که از طریق غیر مستقیم پارامتر χ را به طور دقیق تخمین بزنند. بدین منظور از ۱۲۰ داده (که از نتایج آزمایش های سه محوره، برشی، صفحه فشار و کاغذ فیلتر است) و روش های هوشمند جنگل تصادفی، ماشین بردار پشتیبان و نزدیک ترین همسایگی در نرم افزار WEKA استفاده شده است. نتایج نشان می دهد که مدل های توسعه یافته توسط سه روش هوشمند جنگل تصادفی، ماشین بردار پشتیبان و نزدیک ترین همسایگی، از عملکرد و دقت بسیار خوبی برخوردار هستند. اما مدل جنگل تصادفی در ارزیابی بر روی داده های آزمون با $R^2=0.918$ و $RMSE=0.079$ بهتر از دو مدل دیگر است. به علاوه در این تحقیق آنالیز حساسیت به منظور تعیین اهمیت پارامترهای مؤثر بر روی پارامتر χ انجام شد که از بین پارامترهای ورودی در مدل سازی، مشخص شد که پارامتر محتوای آب حجمی ($\frac{\theta_r}{\theta_s}$) بیشترین تأثیر را بر روی پارامتر χ دارد.

واژه های کلیدی تنش مؤثر، جنگل تصادفی، ماشین بردار پشتیبان، نزدیک ترین همسایگی، آنالیز حساسیت، نرم افزار WEKA.

مقدمه

رفتاری خاک و طراحی ها بدون در نظر گرفتن شرایط خاک غیر اشباع، گرچه ساده تر انجام می شود اما ضریب ایمنی را کاهش و هزینه ها را افزایش می دهد. اصل تنش مؤثر (Effective Stress) یک اصل اساسی در مکانیک خاک است که یک محیط متخلخل متشکل از چند فاز و چند تنشی را به یک محیط پیوسته تشکیل شده از یک فاز و یک تنش تبدیل می کند تا امکان استفاده از اصول مکانیک جامد فراهم شود. تعیین مقاومت برشی در خاک های غیر اشباع در رابطه بیشاپ (Bishop)، علاوه بر تنش مؤثر وابسته به پارامتری به نام پارامتر تنش مؤثر (χ) است. این پارامتر، میزان مشارکت مکش در تنش مؤثر را مشخص می کند. مقدار این پارامتر در حالت اشباع یک و در حالت خشک صفر است [3]. رابطه بین پارامتر χ و پارامترهای مرتبط آن یک رابطه غیر خطی و پیچیده است و تعریف این رابطه به طور تجربی با خطای زیادی همراه می شود. با استفاده از رویکردهای پیشرفته ای مانند روش های هوشمند و سیستم های استنباطی می توان به

امروزه مهندسان در اکثر پروژه های معدنی و عمرانی با خاک های غیر اشباع سر و کار دارند. پروژه هایی مانند احداث جاده ها، فضاهای زیرزمینی و تونل ها، ساخت سدهای خاکی و سنگ ریزه ای و فرودگاه ها، خاک ریزه های شیروانی ها و... از مواردی هستند که در حین عملیات با خاک های غیر اشباع برخورد می شود [1]. خاک غیر اشباع (Unsaturated Soil) یک محیط متخلخل چندفازی است که از دانه های خاک و فاز مایع و گاز که معمولاً آب و هواست، تشکیل می شود. از این رو، وجود نیروی کشش سطحی در مرز تماس آب و هوا و یا به عبارتی وجود پارامتر مکش (Suction)، فشار فاز سیال را متفاوت از حالت خاک اشباع می کند و منجر به تغییر در رفتار خاک می شود [2]. باران، تبخیر یا بالا آمدن سطح آب زیرزمینی، از عوامل زیست محیطی هستند که منجر به خشک شدگی یا ترشدگی و به عبارتی اشباع یا غیر اشباع بودن خاک می شوند. مدل سازی

* تاریخ دریافت مقاله ۱۴۰۰/۹/۱۸ و تاریخ پذیرش آن ۱۴۰۱/۲/۲۶ می باشد.

(۱) نویسنده مسئول، دانشیار، دانشکده مهندسی علوم زمین، دانشگاه صنعتی اراک.

(۲) دانشجو، دانشکده مهندسی علوم زمین، دانشگاه صنعتی اراک.

در این تحقیق برای پیش‌بینی پارامتر χ به‌واسطه پارامترهای مؤثر آن، از روش‌های هوش مصنوعی نظیر، جنگل تصادفی، ماشین بردار پشتیبان و نزدیکترین همسایگی، برای مدل‌سازی استفاده می‌شود. مدل‌سازی‌ها در نرم افزار هوشمند WEKA انجام می‌شود و در آن‌ها مجموعه داده‌های آزمایشگاهی جمع‌آوری شده از چندین مقاله به کار گرفته خواهد شد. پارامترهای ورودی در این مدل، پارامتر منحنی مشخصه آب-خاک (Soil-Water Characteristic Curve Fitting Parameter)، فشار محصور کننده، مکش و محتوای آب حجمی و متغیر خروجی، پارامتر χ است.

پارامتر تنش مؤثر در روابط تجربی

در پروژه‌هایی که با خاک‌های غیراشباع سر و کار دارند، متغیر حالت تنش، نقش تعیین‌کننده‌ای در مدل‌سازی رفتاری خاک دارد. در خاک‌های اشباع تنش مؤثر به‌عنوان متغیر حالت تنش در نظر گرفته می‌شود، اما در خاک‌های غیراشباع متغیر حالت تنش، مطابق رابطه (۱) از دو متغیر حالت تنش مجزا تشکیل می‌شود.

$$\sigma' = (\sigma - u_a) + \chi(u_a - u_w) \quad (1)$$

که در آن χ پارامتر تنش مؤثر، u_a فشار آب حفره‌ای، $(\sigma - u_a)$ تنش خالص نرمال، $(u_a - u_w)$ نماینده مکش ارائه شده توسط خلیلی و خباز می‌باشد [11]. آن‌ها χ را به‌عنوان تابعی از نسبت مکش طبق رابطه (۲) معرفی کردند.

$$\chi = \begin{cases} \left(\frac{u_a - u_w}{u_e}\right)^{-0.55} & \text{for } u_a - u_w > u_e \\ 1 & \text{for } u_a - u_w \leq u_e \end{cases} \quad (2)$$

که در آن u_e مقدار ورودی هوا می‌باشد. این رابطه گذار از حالت اشباع، به حالت غیراشباع را نشان می‌دهد. تعیین پارامتر χ نیازمند زمان و هزینه زیادی است که در طی آزمون‌های آزمایشگاهی صرف می‌شود. طوری که مطالعات نشان می‌دهد این پارامتر بسیار غیرخطی بوده و روابط تجربی قادر به تخمین دقیق آن نخواهند بود [12]. بیشاپ رابطه (۳) را برای محاسبه مقاومت برشی خاک‌های غیراشباع ارائه داد [13].

$$\tau = c' + (\sigma - u_a)\tan\phi' + \chi s \tan\phi \quad (3)$$

پیش‌بینی‌های دقیقی از χ دست یافت. پژوهش‌هایی که در سال‌های اخیر در این راستا انجام شده‌است به شرح زیر است.

لی و همکاران (۲۰۰۳) [4] یک رابطه غیرخطی برای تخمین مقاومت برشی در خاک غیر اشباع ارائه دادند که در آن برای پیش‌بینی پارامتر چسبندگی C_{max} از روش شبکه عصبی مصنوعی (Artificial Neural Network) استفاده کردند.

کایادلن [5] با هدف بررسی قابلیت شبکه‌های عصبی مصنوعی برای مدل‌سازی رابطه پیچیده بین پارامترهای اساسی خاک، مکش و پارامتر χ ، پنج مدل شبکه عصبی مصنوعی را با پارامترهای ورودی متفاوت ایجاد کرد.

اجدری و همکاران [6] از مدل شبکه عصبی مصنوعی برای پیش‌بینی پارامتر تنش مؤثر استفاده کردند.

جوهری و همکاران [7] به توسعه یک مدل (Gene Expression Programming) به‌عنوان یک مدل شبکه عصبی مصنوعی برای پیش‌بینی پارامتر χ پرداختند. آن‌ها نشان دادند، پارامتر مکش مؤثرترین پارامتر در پیش‌بینی پارامتر χ است.

کانگو و همکاران [8] روش‌های شبکه عصبی مصنوعی و رگرسیون درختی (Classification and Regression Tree) را برای تخمین غیرمستقیم پارامترهای مقاومت برشی خاک غیر اشباع به‌کار بردند. آن‌ها برای ساخت مدل‌های مختلف از پارامترهای ورودی درصد تشکیل‌دهنده شن، ماسه، سیلت، خاک رس، چگالی خشک و شاخص پلاستیسیته استفاده کردند.

راهنما و همکاران [9] دو مدل سیستم استنتاج فازی (FCM- Fuzzy C-means Clustering method- Fuzzy) و (Subtractive clustering) FIS را برای پیش‌بینی کردند. آن‌ها نشان دادند پارامترهای فشار محصورکننده (Confining Pressure) و محتوای آب حجمی (Volumetric Water Content) مؤثرترین پارامترها در روند پیش‌بینی هستند.

تین بوی و همکاران [10] یک روش پیشنهادی برای پیش‌بینی مقاومت برشی خاک ارائه دادند و در آن دو روش حداقل مربعات ماشین بردار پشتیبان (Least Squares Support Vector Machine) و الگوریتم (Cuckoo Search) را تلفیق کردند. آن‌ها برای ارزیابی مدل ارائه شده از نمونه‌های جمع‌آوری شده از یک بزرگراه در ویتنام استفاده کردند.

مناسب می‌باشد. روش درخت تصمیم نوسانات بالایی را در نتایج خود نشان می‌دهد، به همین دلیل کار جنگل تصادفی کاهش نوسانات و واریانس برآوردها و در نتیجه بهبود نتایج درخت تصمیم است. جنگل تصادفی این کار را با ترکیب چند درخت تصمیم انجام می‌دهد، که در ساخت هر درخت تصمیم، تعدادی از متغیرها به‌طور تصادفی شرکت می‌کنند. مجموعه درخت‌ها با در نظر گرفتن n نمونه بوت استرپ از داده‌ها به‌عنوان داده‌های مشاهده‌ای مستقل ساخته می‌شوند. به این معنی که تعداد زیادی نمونه‌برداری همراه با جای‌گذاری انجام می‌شود. سپس بر روی هر نمونه یک درخت تصمیم تعریف می‌شود. اگر تعداد تمام متغیرهای مستقل M باشد، در هر شاخه از درخت به‌طور تصادفی m متغیر مستقل، برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیون نسبت $\frac{m}{M} = \frac{1}{3}$ و در حالت طبقه‌بندی $m = \sqrt{M}$ پیشنهاد می‌شود. سرانجام به تعداد درخت‌ها برای بردارهای ورودی، یک خروجی حاصل می‌شود و خروجی نهایی مدل از میانگین‌گیری این خروجی‌ها حاصل می‌شود [15]. شکل (۱) نمونه‌ای از جنگل تصادفی تشکیل شده از N درخت رویداد را نشان می‌دهد.

که در آن \hat{c} و $\hat{\phi}$ چسبندگی و زاویه اصطکاک داخلی مؤثر هستند. به‌طور کلی χ تابعی از درجه اشباع‌شدگی (sf) فرض می‌شود. رگرسیون‌های برازش‌شده بر روی برخی از داده‌های تجربی، تناسب خوبی را بین این دو پارامتر نشان می‌دهند. در رابطه (۴) χ تابعی از (sf) تعریف شده است [14].

$$\chi = S_F^k = \left(\frac{\theta}{\theta_s}\right)^k \quad (4)$$

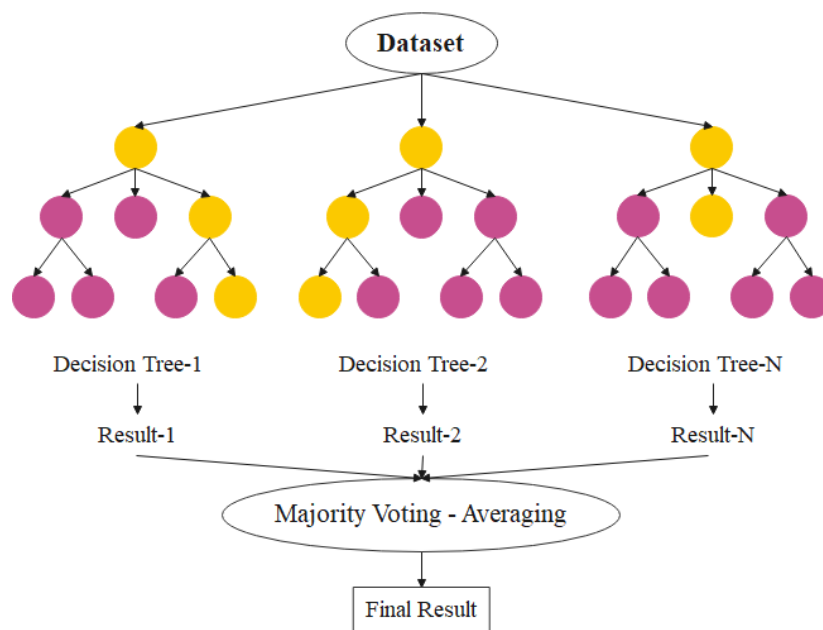
که در آن θ محتوای آب حجمی، θ_s محتوای آب اشباع و k یک پارامتر بهینه شده است که از طریق بهترین برازش بین داده‌های اندازه‌گیری شده و پیش‌بینی شده به دست می‌آید.

معرفی روش‌های هوشمند

در این بخش به توصیف برخی از تکنیک‌های داده‌کاوی که در این تحقیق بکارگرفته شده‌اند، پرداخته خواهد شد.

جنگل تصادفی (Random Forest) RF

جنگل تصادفی روشی برگرفته از روش درخت تصمیم است و برای اهداف طبقه‌بندی و رگرسیون داده‌های گسسته و پیوسته



شکل ۱. نمایی از جنگل تصادفی متشکل از N درخت رویداد

شکل رابطه (۸) در می‌آید.

$$f(x) = \sum_{x_i \in SV} (a_i - a_i^*)K(x_i, x_j) + b \quad (8)$$

که در آن a_i و a_i^* ضرایب لاگرانژ و $k(x_i, x_j)$ تابع کرنل است [17].

نزدیک‌ترین همسایگی (KNN)

روش نزدیک‌ترین همسایگی یک روش آماری ناپارامتری است که برای مسائل طبقه‌بندی و رگرسیون مورد استفاده قرار می‌گیرد. این روش مبتنی بر مفهوم تشابه بوده و پیش‌بینی‌ها بر اساس موقعیت‌های مشابهی که در گذشته اتفاق افتاده، انجام می‌شود. به عبارتی با مشاهده متغیرهای مستقل در زمان واقعی، مدل به جستجوی الگوهای مشابه شرایط فعلی در سری تاریخی می‌پردازد و گزینه‌های محتمل را انتخاب می‌کند [18]. رابطه ۹ بردار $1 \times m$ مقادیر پیش‌بینی‌کننده x_j در زمان t را نشان می‌دهد.

$$Pr_{jt} = (x_{jt}) \quad j = 1 \dots m \quad (9)$$

هم‌چنین ماتریس $n \times m$ مقادیر متغیرهای پیش‌بینی‌کننده x_j در سری زمانی تاریخی به صورت رابطه (۱۰) است.

$$Pr_{j,(t-i)} = x_{j,(t-i)} \quad (10)$$

$$j = 1 \dots m, i = 1 \dots n$$

با استفاده از تابع فاصله‌سنج $Dist$ ، در رابطه (۱۱)، فاصله بین بردار $Pr_{j,t}$ با سطرهای ماتریس $Pr_{j,(t-i)}$ محاسبه می‌شود.

$$Dist(t-i) = f(w_j \cdot x_{j,(t-i)} \cdot x_{jt}) \quad (11)$$

که در آن j اندیس مربوط به متغیرهای پیش‌بینی‌کننده و i اندیس مربوط به گام زمانی در سری تاریخی است. w_j نیز وزن‌های متغیرهای پیش‌بینی‌کننده می‌باشد. در ادامه می‌بایست به منظور بهبود عملکرد مدل، از روش‌های مناسبی برای تخمین بهترین همسایه‌ها، توسعه توابع انتقال اطلاعات و توابع فاصله‌سنجی استفاده کرد [19].

معرفی اجمالی نرم‌افزار هوشمند WEKA

تا به امروز نرم‌افزارهای فراوانی در حوزه داده‌کاوی عرضه

ماشین بردار پشتیبان (Support Vector Machine) SVM

ماشین بردار پشتیبان یک الگوریتم نظارت‌شده یادگیری ماشین است که از نوع مدل طبقه‌بندی ماشین بردار پشتیبان برای کلاس‌بندی داده‌ها و از نوع مدل رگرسیون بردار پشتیبان، برای حل مسائل پیش‌بینی استفاده می‌شود. روش رگرسیون بردار پشتیبان از الگوریتم‌های آموزشی تشکیل شده که بر اساس ماشین بردار پشتیبان طبقه‌بندی‌کننده کار می‌کند. این روش با بهره‌گیری از تئوری یادگیری آماری و مینیم کردن ریسک ساختاری به مدلی برای پیش‌بینی می‌رسد. روش (Support Vector Regression) SVR (نخستین بار توسط وینیک ارائه شد [16]). در این روش برای آموزش داده‌ها از یک نگاشت غیرخطی استفاده می‌شود و داده‌ها به فضایی با ابعاد بالاتر منتقل می‌شوند و سپس برای ساخت شکل رگرسیون از تابع خطای ϵ در رابطه (۵) استفاده می‌شود.

$$L(y, f(x, a)) = |y - f(x, a)|_{\epsilon} = \begin{cases} 0 & \text{if } |y - f(x, a)| \leq \epsilon \\ |y - f(x, a)| - \epsilon & \text{Otherwise} \end{cases} \quad (5)$$

رابطه (۵) نشان می‌دهد که خطاهایی با مقادیر کمتر از ϵ در نظر گرفته نمی‌شوند. رابطه غیرخطی پارامترهای ورودی و خروجی مطابق رابطه (۶) به شکل یک تابع بهینه نوشته می‌شود.

$$f(x) = w^T \varphi(x) + b \quad (6)$$

که در آن x بردار متغیر ورودی، w بردار وزن و b بایاس است. این تابع به عنوان تابع زیان ϵ -غیرحساس (ϵ -Insensitive Loss Function) شناخته می‌شود و حل آن می‌تواند به شکل مسئله درجه دوم در رابطه (۷) بیان شود.

$$\begin{cases} \text{Min } \frac{1}{2} w^T w + C \sum_{i=1}^L (\xi_i - \xi_i^*) \\ \text{s.t. } y_i - w \cdot \varphi(x_i) - b \leq \epsilon + \xi_i \\ w \cdot \varphi(x_i) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i \geq 0 \\ \xi_i^* \geq 0 \end{cases} \quad (7)$$

که در آن C پارامتر جریمه و ξ متغیر کمکی است. با معرفی ضرایب لاگرانژ و کاربرد تابع کرنل به جای ضرب داخلی عملگرها، مسئله بهینه‌سازی فوق ساده‌تر شده و رابطه (۶) به

همچنین از نسبت $\frac{\theta_r}{\theta_s}$ پارامتر محتوای آب حجمی بدون بعد به دست می آید. بدین صورت در مدل سازی ها پارامترهای ورودی شامل $\frac{\theta_r}{\theta_s}$ ، $\frac{s}{h_b}$ ، $\frac{p}{p_0}$ و λ و پارامتر خروجی χ می شوند. جدول (۱) مشخصات آماری داده های مربوط به این پارامترها را نشان می دهد. از این تعداد داده نرمالایز شده، ۸۰ درصد به عنوان داده های آموزش و ۲۰ درصد، به عنوان داده های آزمون جهت ارزیابی مدل ها به کار گرفته می شوند.

$$R^2 = 1 - \frac{\sum_{k=1}^n (Y_{mea} - Y_{pre})^2}{\sum_{k=1}^n Y_{mea}^2 - \frac{\sum_{k=1}^n Y_{pre}^2}{n}} \quad (12)$$

$$MAE = \frac{1}{n} \sum_{k=1}^n \left| \frac{Y_{mea} - Y_{pre}}{Y_{mea}} \right| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (Y_{mea} - Y_{pre})^2} \quad (14)$$

مدل سازی و تحلیل نتایج

به منظور ارزیابی عملکرد و دقت مدل های هوشمند چندین شاخص و معیار شامل ضریب هم بستگی مربع (R^2)، میانگین درصد خطای مطلق (MAE) و جذر میانگین خطای مربع (RMSE) معرفی شده اند. در تمامی این معادلات (معادلات ۱۲ تا ۱۴) Y_{pre} و Y_{mea} به ترتیب مقادیر اندازه گیری شده و پیش بینی شده هستند و n تعداد نمونه ها می باشند.

شده اند که هر یک متناسب با نوع و فرمت داده هایی که پردازش می کنند، شامل الگوریتم های مختلفی می شوند. نرم افزار WEKA از جمله ابزارهای قدرتمند داده کاوی است و شامل الگوریتم های روز یادگیری ماشین و ابزارهای پیش پردازش داده ها می باشد. در این نرم افزار می توان روش های مختلف داده کاوی را بر روی داده ها آزمایش کرد و با ایجاد چندین کار، نتایج مدل سازی برای روش های مختلف یا داده های متفاوت را به طور هم زمان مقایسه کرد. مجموعه روش ها برای مسائلی مانند رگرسیون، طبقه بندی، خوشه بندی، کاوش قواعد انجمنی و انتخاب ویژگی، در این نرم افزار گرد آمده اند. همه روش ها، ورودی های خود را با فرمت ARFF دریافت می کنند که می توان آن را به وسیله یک درخواست از پایگاه داده تولید کرد. به علاوه کتابی جامع و مرتبط با این نرم افزار [20] برای آشنایی بیشتر معرفی می شود.

توصیف پایگاه داده برای مدل سازی با روش های

هوشمند

پایگاه داده شرکت داده شده در آنالیزها، شامل ۱۲۰ داده از نتایج آزمایش های سه محوره، برشی، صفحه فشار و کاغذ فیلتر است که از چندین مقاله جمع آوری شده است [21-29]. این داده ها شامل ۷ متغیر شامل: مکش (s)، فشار حبابی (h_b)، فشار خالص محصور کننده (p)، محتوای آب باقی مانده (θ_r)، محتوای آب حجمی اشباع شده (θ_s)، پارامتر منحنی مشخصه آب-خاک (λ) و پارامتر تنش مؤثر (χ) می باشند. به منظور پردازش اولیه، پارامترهای θ_s و θ_r ، p ، h_b ، s ، λ و χ را به صورت بی بعد وارد مدل سازی می شوند بدین صورت که از نسبت $\frac{p}{p_0}$ پارامتر فشار محصور کننده بدون بعد حاصل می شود (که در آن پارامتر $P_0=101.325\text{KPa}$ است) و پارامتر مکش نیز توسط نسبت $\frac{s}{h_b}$ بدون بعد می شود.

جدول ۱ مشخصات آماری پارامترهای ورودی و خروجی مدل ها

پارامترها	مینیمم	ماکزیمم	میانگین	انحراف معیار
فشار محصورکننده ($\frac{p}{p_0}$)	۰	۳/۹۴۸	۱/۲۴۳	۱/۰۳۴
مکش ($\frac{s}{h_b}$)	۰	۳۰۰	۲۴/۲۶۲	۴۶/۴۱۵
محتوای آب حجمی ($\frac{\theta_r}{\theta_s}$)	۰	۰/۷۱۴	۰/۲۴۹	۰/۲۳۱
منحنی مشخصه آب-خاک (λ)	۰/۱۹	۱۱/۸۲	۱/۲۹۴	۲/۴۱۹
پارامتر تنش مؤثر (χ)	۰/۰۹۱	۱	۰/۶۸۹	۰/۲۷۴

همسایگی (KNN) را برای مجموعه داده آموزش و آزمون مورد ارزیابی قرار می‌دهد. در بخش داده‌های آموزش مدل نزدیک‌ترین همسایگی به‌طور نسبی دارای مقدار R^2 بالا و RMSE پایین‌تری می‌باشد و از این نظر مدل جنگل تصادفی پس از مدل نزدیک‌ترین همسایگی قرار می‌گیرد اما با نظر به نمودار بخش داده‌های آزمون، مدل جنگل تصادفی دارای مقدار R^2 بالاتر و RMSE پایین‌تری است. بنابراین مدل جنگل تصادفی ضمن عملکرد و دقت خوب در داده‌های آموزش، همان عملکرد را در داده‌های آزمون نیز حفظ کرده‌است.

هوشمند برای داده‌های آزمون

به‌علاوه در شکل‌های (۴) تا (۶) نیز می‌توان عملکرد مدل‌های هوشمند در داده‌های آموزش و آزمون را در کنار هم مشاهده کرد. در این نمودارها نقاط مقادیر اندازه‌گیری شده و پیش‌بینی شده پارامتر γ رسم شده‌است. در شکل مربوط به مدل جنگل تصادفی در هر دو نمودار، نقاط نزدیک به خط، قطری و دارای برآزش مناسبی هستند اما در مدل نزدیک‌ترین همسایگی، نمودار داده‌های آموزش بسیار منظم و روی خط قطری اما در داده‌های آزمون بسیار پراکنده هستند. بنابراین مدل جنگل تصادفی در هر دو مجموعه دارای عملکرد بهتر و دقت بالاتری است و در مجموع مدل کارآمدی می‌باشد.

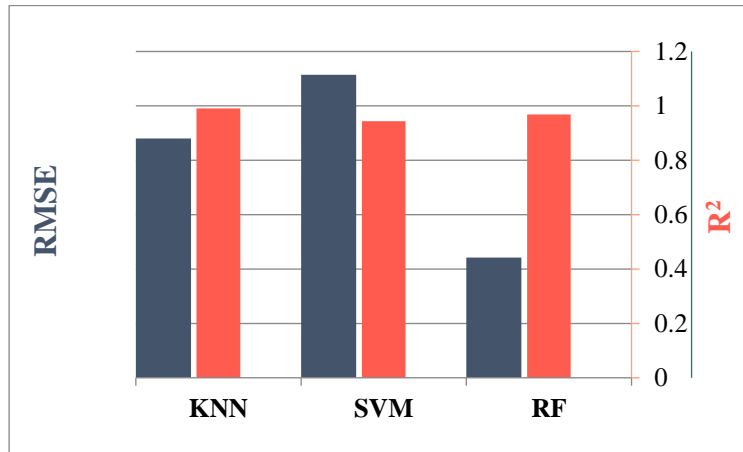
مطابق با روابط فوق، هر قدر مقادیر MAE و RMSE به صفر نزدیک‌تر و مقدار R^2 به سمت عدد ۱ نزدیک‌تر باشد، مدل پیش‌بینی شده به مقدار واقعی نزدیک‌تر شده و به‌عنوان یک مدل ایدئال پذیرفته می‌شود.

در ادامه، پس از آماده‌سازی داده‌ها نوبت به ساخت مدل‌هایی برای پیش‌بینی/تخمین پارامتر γ با استفاده از روش‌های هوشمند در نرم‌افزار WEKA می‌رسد. از جمله روش‌های هوشمند موجود در نرم‌افزار، سه روش هوشمند جنگل تصادفی، ماشین بردار پشتیبان و نزدیک‌ترین همسایگی که دارای عملکرد و دقت بهتری بوده‌اند، برای ساخت مدل‌های مدنظر مورد استفاده قرار گرفته‌اند. همان‌طور که قبلاً اشاره شد، برای ارزیابی عملکردها مدل‌ها از سه شاخص آماری استفاده می‌شود که در جدول (۲) مقادیر R^2 ، RMSE و MAE مربوط به هر یک از مدل‌ها، برای مجموعه داده‌های آموزش و آزمون به‌طور مجزا ارائه شده‌است. همان‌طور که مشخص است، ضریب R^2 برای هر سه مدل هوشمند در بخش آموزش داده‌ها، دارای مقادیر بالا و قابل توجهی است. البته مقادیر همین ضریب در بخش آزمون مدل‌های هوشمند، شرایط متفاوتی دارد. می‌توان گفت مدل جنگل تصادفی، در این بخش از شاخص‌های آماری مطلوبی برخوردار است.

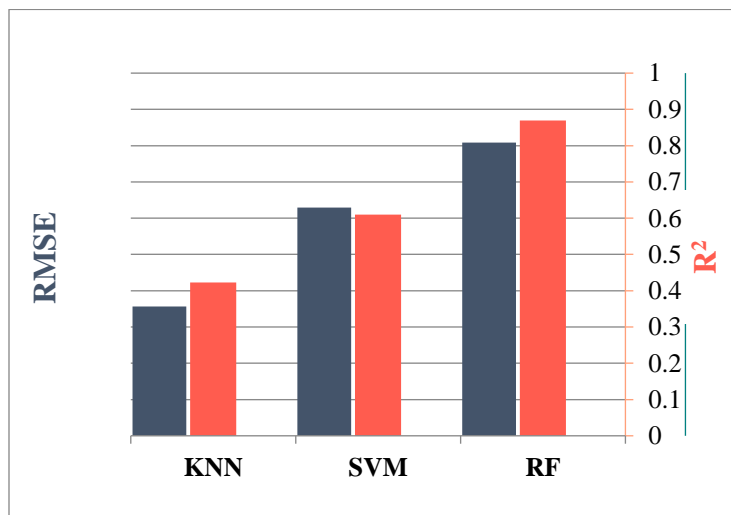
هم‌چنین نمودارهای نشان داده‌شده در شکل (۲) و (۳) به‌صورت گرافیکی مقادیر R^2 و RMSE سه مدل هوشمند جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM) و نزدیک‌ترین

جدول ۲ مقدار شاخص‌های ارزیابی مدل‌ها در داده‌های آموزش و آزمون

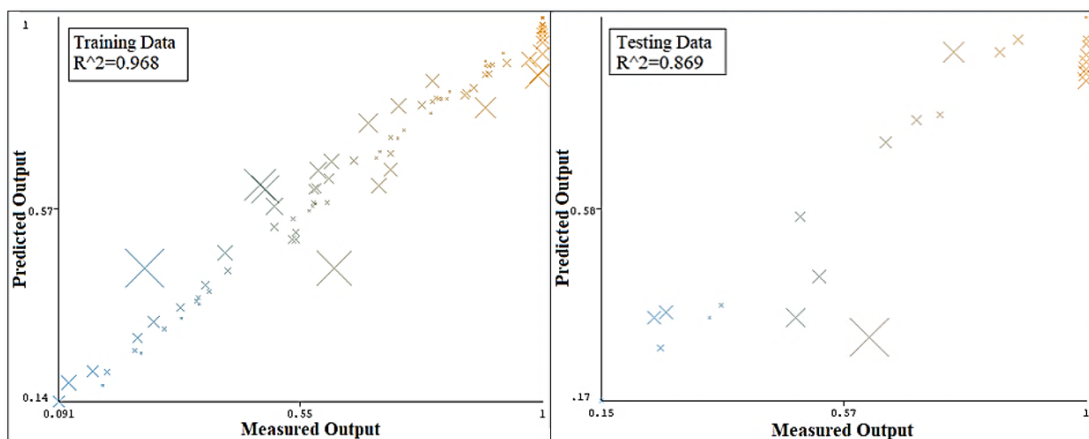
مدل‌ها			MAE	RMSE	R^2
جنگل تصادفی	برای داده‌های آموزش	۰/۹۶۷۹	۰/۰۵۱۳	۰/۰۳۴۳	
	برای داده‌های آزمون	۰/۸۶۹۲	۰/۱۰۶۹	۰/۰۸۶۶	
ماشین بردار پشتیبان	برای داده‌های آموزش	۰/۹۴۳۰	۰/۰۶۵	۰/۱۰۶	
	برای داده‌های آزمون	۰/۶۱۰۱	۰/۱۸۸۹	۰/۱۵۷۸	
نزدیک‌ترین همسایگی	برای داده‌های آموزش	۰/۹۹۰۸	۰/۰۲۵۸	۰/۰۰۳۷	
	برای داده‌های آزمون	۰/۴۲۳۰	۰/۲۴۲۵	۰/۱۹۱	



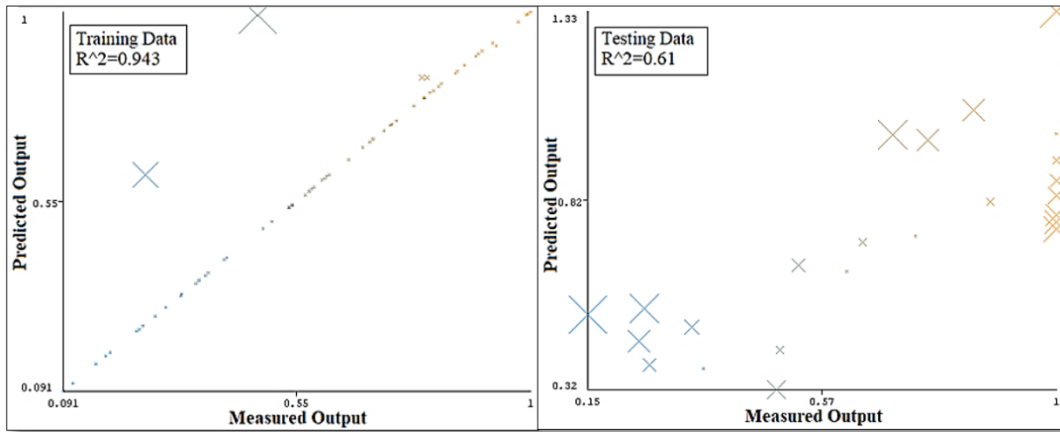
شکل ۲ نمودار ستونی مقادیر RMSE و R² سه مدل هوشمند برای داده‌های آموزش



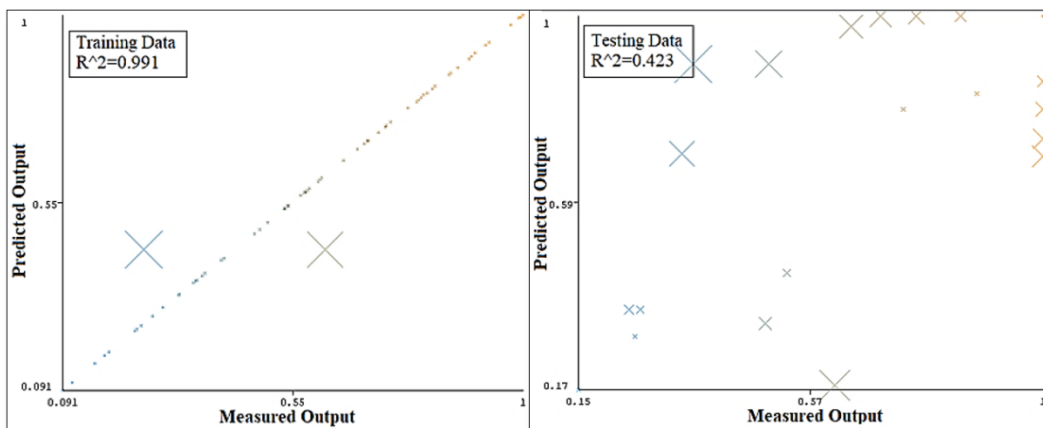
شکل ۳ نمودار ستونی مقادیر RMSE و R² سه مدل



شکل ۴ نمودار مقادیر اندازه‌گیری شده و مقادیر پیش بینی شده آن در مدل جنگل تصادفی در نرم‌افزار WEKA



شکل ۵ نمودار مقادیر اندازه‌گیری شده λ و مقادیر پیش‌بینی شده آن در مدل ماشین بردار پشتیبان در نرم‌افزار WEKA



شکل ۶ نمودار مقادیر اندازه‌گیری شده λ و مقادیر پیش‌بینی شده آن در مدل نزدیک‌ترین همسایگی در نرم‌افزار WEKA

x_2, x_3 و x_4 به ترتیب بیانگر پارامترهای $\frac{P}{P_0}$ ، $\frac{s}{h_b}$ و $\frac{\theta_r}{\theta_s}$ و λ و y بیانگر پارامتر λ است.

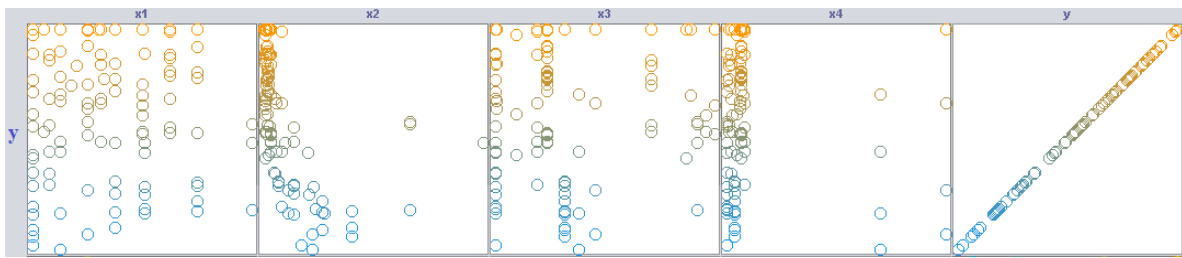
جدول ۳ آنالیز حساسیت برای تأثیر پارامترهای ورودی مدل‌های هوشمند

در نرم‌افزار WEKA

پارامترها	رتبه	مقدار شایستگی متوسط
$\frac{\theta_r}{\theta_s}$	۱	۰/۱۵۷
$\frac{P}{P_0}$	۲	-۰/۰۵۶
$\frac{S}{h_b}$	۳	-۰/۱۶
λ	۴	-۰/۵۲۵

مدل‌سازی‌های انجام شده در قسمت‌های بالا، پارامترهای ورودی، هر یک با تأثیر متفاوتی در پیش‌بینی پارامتر λ مؤثر بوده‌اند. به منظور تشخیص تأثیرگذارترین پارامترها در تخمین، آنالیز حساسیتی در نرم‌افزار WEKA انجام شده است. ارزش و تأثیرگذاری هر پارامتر از طریق اندازه‌گیری هم‌بستگی پارامتر مورد نظر با پارامتر هدف تعیین می‌شود. در جدول (۴) با توجه به مقدار شایستگی متوسط، رتبه هر یک از پارامترها از لحاظ اهمیت مشخص شده است، که مطابق آن، $\frac{\theta_r}{\theta_s}$ مؤثرترین پارامتر و λ بی‌ارزش‌ترین پارامتر گزارش شده است.

هم‌چنین با رسم نمودار پراکندگی دو به دوی پارامترها می‌توان از چگونگی رابطه و رفتار آن‌ها نسبت به یکدیگر مطلع شد. شکل (۷) نمودار پراکندگی همه پارامترها را نسبت به پارامتر هدف یعنی پارامتر تنش مؤثر نشان می‌دهد. به این صورت که x_1



شکل ۷ نمودار پراکندگی پارامترها نسبت به پارامتر خروجی در نرم افزار WEKA

نتیجه گیری

- مدل جنگل تصادفی هم در مجموعه داده های آموزش (RMSE=0.051 و $R^2=0.968$) و هم در داده های آزمون (RMSE=0.107 و $R^2=0.869$)، با دقت بالا قابلیت انجام پیش بینی دقیق پارامتر χ را دارد.
 - مدل های ماشین بردار پشتیبان و نزدیک ترین همسایگی نیز در خصوص مجموعه داده های آموزش، عملکرد و دقت خوبی دارند.
 - آنالیز حساسیت انجام شده بر روی پارامترها نشان داد، پارامتر محتوای آب حجمی اهمیت بالایی در پیش بینی پارامتر χ دارد و در مقابل، پارامتر منحنی مشخصه آب-خاک، نسبت به سایر پارامترها اهمیت کمتری دارد.
 - به طور کل می توان گفت نرم افزار قدرتمند و کاربردی WEKA می تواند به عنوان یک ابزار قوی در علم داده کاوی و علوم مهندسی زمین در تخمین پارامترهای ژئومکانیکی به کار گرفته شود.
- با توجه به این که درصد قابل توجهی از سطح زمین را خاک های غیراشباع دربر گرفته اند، بدیهی است در حین عملیات در اکثر پروژه های معدنی و عمرانی با این خاک ها برخورد می شود. تعیین مقاومت برشی در خاک های غیراشباع در رابطه بیشاپ وابسته به پارامتری به نام پارامتر χ است، که تابعی از مکش می باشد. در این تحقیق مدلهایی برای پیش بینی پارامتر χ ایجاد و ارزیابی شد. این مدل ها به کمک سه روش هوشمند جنگل تصادفی، ماشین بردار پشتیبان و نزدیک ترین همسایگی در قالب نرم افزار WEKA انجام شد و در آن از ۱۲۰ نمونه آزمایشگاهی جمع آوری شده از چندین مقاله استفاده شد. پارامترهای ورودی شامل پارامتر منحنی مشخصه آب- خاک (soil-water characteristic curve fitting parameter)، فشار محصور کننده، مکش و محتوای آب حجمی و پارامتر خروجی، پارامتر تنش مؤثر است. نتایج نشان داد:

مراجع

1. Fredlund, D.G., "Unsaturated soil mechanics in engineering practice". *Journal of geotechnical geoenvironmental engineering*, Vol. 132, No. 3, pp. 286-321, (2006).
2. Aitchison, G.D., Russam, K., and Richards, B., "Engineering concepts of moisture equilibria and moisture changes in soils". *Ministry of Transport*, Vol., (1966).
3. Sheng, D., "Constitutive modelling of unsaturated soils: Discussion of fundamental principles". *Unsaturated soils*, Vol. 1, No. pp. 91-112, (2011).
4. Lee, S., Lee, S.R., and Kim, Y., "An approach to estimate unsaturated shear strength using artificial neural network and hyperbolic formulation". *Computers Geotechnics*, Vol. 30, No. 6, pp. 489-503, (2003).
5. Kayadelen, C., "Estimation of effective stress parameter of unsaturated soils by using artificial neural networks". *International journal for numerical analytical methods in geomechanics*, Vol. 32, No. 9, pp. 1087-1106, (2008).

6. Ajdari, M., Habibagahi, G., and Ghahramani, A., "Predicting effective stress parameter of unsaturated soils using neural networks". *Computers and Geotechnics*, Vol. 40, No. pp. 89-96, (2012).
7. Johari, A., Habibagahi, G., and Nakhaee, M., "Prediction of unsaturated soils effective stress parameter using gene expression programming". *Scientia Iranica*, Vol. 20, No. 5, pp. 1433-1444, (2013).
8. Kanungo, D., Sharma, S., and Pain, A., "Artificial Neural Network (ANN) and Regression Tree (CART) applications for the indirect estimation of unsaturated soil shear strength parameters". *Frontiers of earth science*, Vol. 8, No. 3, pp. 439-456, (2014).
9. Rahnema, H., Hashemi Jokar, M., and Khabbaz, H., "Predicting the effective stress parameter of unsaturated soils using adaptive neuro-fuzzy inference system". *Scientia Iranica*, Vol. 26, No. 6, pp. 3140-3158, (2019).
10. Tien Bui, D., Hoang, N.-D., and Nhu, V.-H., "A swarm intelligence-based machine learning approach for predicting soil shear strength for road construction: a case study at Trung Luong National Expressway Project (Vietnam)". *Engineering with Computers*, Vol. 35, No. 3, pp. 955-965, (2019).
11. Khalili, N. and Khabbaz, M., "A unique relationship for χ for the determination of the shear strength of unsaturated soils". *Geotechnique*, Vol. 48, No. 5, pp. 681-687, (1998).
12. Cui, Y.J., et al., "Unsaturated soil mechanics". *International Society of Soil Mechanics and Geotechnical Engineering*, Vol., (2004).
13. Bishop, A.W., "The principle of effective stress". *Teknisk ukeblad*, Vol. 39, No. pp. 859-863, (1959).
14. Vanapalli, S., et al., "Model for the prediction of shear strength with respect to soil suction". *Canadian geotechnical journal*, Vol. 33, No. 3, pp. 379-392, (1996).
15. vladimir, et al., "A classification and regression tool for compound classification and qsar modeling svetnik". *Journal of Chemical Information Computer Sciences*, Vol. 43, No. 6, pp. 1947-1958, (2003).
16. Cortes, C. and Vapnik, V., "Support vector machine". *Machine learning*, Vol. 20, No. 3, pp. 273-297, (1995).
17. Zhang, D., et al., "Parameter optimization for support vector regression based on genetic algorithm with simplex crossover operator". *Journal of Information Computational Science*, Vol. 8, No. 6, pp. 911-920, (2011).
18. Shahrabi, J. and Hejazi, T., *Data mining 2, Tehran, Industrial university of amirkabir*. 2013, Jahad daneshgahi Press.(In Persian).
19. Azmi, M. and Araghinejad, S., "Development of K-Nearest Neighbour Regression Method in Forecasting River Stream Flow". *Journal of Water and Wastewater*, Vol. 23, No. 2, pp. 108-119, (2012).
20. Witten, I.H., et al. Practical machine learning tools and techniques. in DATA MINING. 2005.
21. Thu, T.M., Rahardjo, H., and Leong, E.-C., Effects of hysteresis on shear strength envelopes from constant water content and consolidated drained triaxial tests, in Unsaturated soils 2 . . . p. 1212-1222.
22. Russell, A.R. and Khalili, N., "A bounding surface plasticity model for sands exhibiting particle crushing". *Canadian Geotechnical Journal*, Vol. 41, No. 6, pp. 1179-1192, (2004).
23. Russell, A. and Khalili, N., "A unified bounding surface plasticity model for unsaturated soils". *International Journal for Numerical Analytical Methods in Geomechanics*, Vol. 30, No. 3, pp. 181-212, (2006).

24. Rampino, C., Mancuso, C., and Vinale, F., "Experimental behaviour and modelling of an unsaturated compacted soil". *Canadian Geotechnical Journal*, Vol. 37, No. 4, pp. 748-763, (2000).
25. Rahardjo, H., Heng, O.B., and Choon, L.E., "Shear strength of a compacted residual soil from consolidated drained and constant water content triaxial tests". *Canadian Geotechnical Journal*, Vol. 41, No. 3, pp. 421-436, (2004).
26. Miao, L., Liu, S., and Lai, Y., "Research of soil-water characteristics and shear strength features of Nanyang expansive soil". *Engineering Geology*, Vol. 65, No. 4, pp. 261-267, (2002).
27. Lee, I.-M., Sung, S.-G., and Cho, G.-C., "Effect of stress state on the unsaturated shear strength of a weathered granite". *Canadian Geotechnical Journal*, Vol. 42, No. 2, pp. 624-631, (2005).
28. Khalili, N., Geiser, F., and Blight, G., "Effective stress in unsaturated soils: Review with new evidence". *International journal of Geomechanics*, Vol. 4, No. 2, pp. 115-126, (2004).
29. Bishop, A.W. and Blight, G., "Some aspects of effective stress in saturated and partly saturated soils". *Geotechnique*, Vol. 13, No. 3, pp. 177-197, (1963).

