BE CIVIL ENGINEERING
PROJECT REPORT

Prediction of Swell Pressure of Expansive Soils Using Multi Expression Programming, An Artificial Intelligence Approach

Project submitted in partial fulfilment of the requirements for the degree of BE Civil Engineering

PROJECT ADVISOR
Associate Prof. Dr. Rai Waqas Azfar Khan

SUBMITTED BY

240857	Muhammad Umar Mujahid (Syndicate Leader)
240846	Mamoon Ajmal
240847	Ibtsam ur Rahman Khilji
240862	Muhammad Tariq Khan
216175	Zain Rasool
216661	Habib Ur Rehman

MILITARY COLLEGE OF ENGINEERING (MCE)

NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY (NUST)

RISALPUR CAMPUS, PAKISTAN

(2021)

This is to certify that the

BE Civil Engineering Project entitled

PREDICTION OF SWELL PRESSURE OF EXPANSIVE SOILS USING MULTI EXPRESSION PROGRAMMING, AN ARTIFICIAL INTELLIGENCE APPROACH

SUBMITTED BY

Muhammad Umar Mujahid (Syn. Leader)	**CMS-240857**
Mamoon Ajmal	**CMS-240846**
Ibtsam ur Rahman Khilji	**CMS-240847**
Muhammad Tariq Khan	**CMS-240862**
Zain Rasool	**CMS-216175**
Habib Ur Rehman	**CMS-216661**

Has been accepted towards the partial fulfilment of the requirement for

BE Civil Engineering Degree

Associate Prof. Dr. Rai Waqas Azfar Khan

Syndicate Advisor

Department of Construction Engineering and Management,

Military College of Engineering, Risalpur

National University of Sciences and Technology (NUST), Islamabad

This research work is dedicated to

Our beloved Parents and Teachers, who have been a source of inspiration for us.

Acknowledgements

All praise is due to Allah, Whose Divine guided us and kept us motivated to achieve the objective we set out to seek.

We want to convey our heartfelt gratitude to Associate Professor Dr. Rai Waqas Azfar khan under whose direction this research was led. His guidance was precious at each progression of this work. His outstanding output-oriented aptitudes helped us to get expertise on the topic rapidly. His help at every single phase has been significant.

Members of this syndicate are highly thankful to their cherishing parents, who helped us through each trouble of our life and yielded every comforts of their lives for our brilliant future. This is likewise a tribute to our teachers who guided us to go Up against the troubles with ingenuity and boldness & who made us what we are proud of today, Alhamdulillah.

The project could not be completed without the support of the people who stayed with us and gave us their support. We are thankful to Dr. Abubakar, Mr. Mohsin Ali Khan and Mr. Faisal Javed for supporting us all the way.

Last but not least, we want to thank our family and friends for their prayers, which helped us to accomplish this task.

Table of Contents

List of Figures

List of Tables/ Graphs

List of Abbreviations

Ps-ES	Swell Pressure of Expansive Soils
Ps	Swell Pressure
ES	Expansive Soils
GP	Genetic Programming
MEP	Multi-Expression Programming
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro Fuzzy Inference System
LGP	Linear Genetic Programming
GEP	Gene Expression Programming
RMSE	Relative Root Mean Square Error
MAE	Mean Absolute Error
RSE	Relative Squared Error
P	Performance Co-efficient
OF	Objective Function

Abstract

This project presents the development of a new empirical prediction model to evaluate swell pressure of expansive soils (Ps-ES). An extensive database comprising 168 Ps records was established after a comprehensive literature search. The performance of developed model was tested using mean absolute error (MAE), root squared error (RSE), root mean square error (RMSE), correlation coefficient (R), regression coefficient (R^2). The results in the increasing order of the contribution of each input parameter is in the order of OMC (28.27) > PI (27.59) > CF (14.59) > MDD (12.59) > SP (10.40) > silt (6.55).

The MEP model outperformed the other AI models found in literature for the prediction of swell pressure in terms of closeness of training, validation and testing data set with the ideal fit slope. The findings of this study can help researchers and designers to evaluate the swell characteristics of the expansive soils in pre-planning and pre-design phases of a construction project and for validation of the laboratory and field test results.

CHAPTER 1

INTRODUCTION

1.1 Importance of Geotechnical Surveys

A building with the weak foundation can never survive no matter how strong the superstructure is. Not only the strength of the foundation but the type of foundation and the treatment done before the foundation construction must be taken into account.

Since if a foundation type and the initial treatment of the soil on which the foundation is to be built are not given consideration, even the strongest of foundation material will fail when it encounters conditions for which it is not meant. The serious concern of selecting the best foundation type and of the specific pre-construction treatment of soil is based on the nature of the soil encountered on ground and to know the nature of the soil to make the above most important decision a series of a complex and extensive testing of soil is required.

Geotechnical engineering is one of the most complicated field of civil engineering and the major reason for this complexity are as follows.

- Soils are natural materials thus humans have no control over its quality and type.
- Soils are highly non-homogeneous in nature which makes them very difficult to generalize in terms of properties and responses to conditions.
- Most parameters of soil have highly multivariable dependencies.

- Apart from this, most of the testing procedures are very time-consuming, the testing machinery and operators required is very expensive and need skilled professionals to work on them.
- Owing to the above complexities, researchers have always tried to find ways to make correlations of simpler and easy to evaluate soil properties with the more complex, expensive, difficult to find and time-consuming parameters.

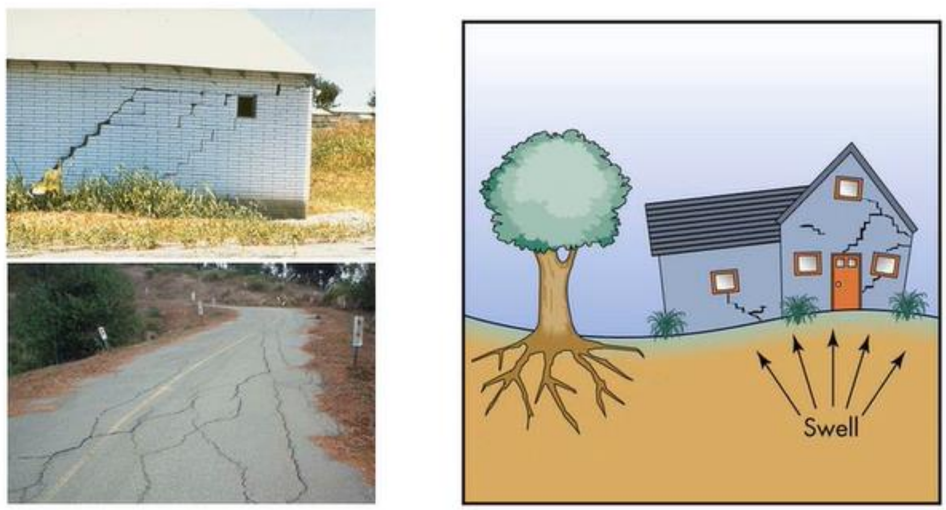


Figure 1.1 Swell behavior of Soil

One of the most devastating and difficult to handle parameter of soil is the swell pressure (Ps) that is caused when expansive soil encounters water as shown in Fig 1.1. This Figure shows the different types of damages that can be caused by swell pressure of expansive soils. Expansive soils are the soils that have a high clay fraction. Clay contains silicate minerals which are highly hydrophilic and makes strong hydrogen bonds with water. This increases the volume of the clay particles and thus they swell significantly. This swelling phenomena causes immense pressure on the foundations of buildings and the especially upon light structures such as roads, railways, pipelines etc.

Calculating the fundamental mechanical parameters of expansive/swelling soils is a prerequisite for a particular construction project.(Das et al., 2011; Huat et al., 2014; Ikeagwuani & Nwonu, 2019; Anand J. Puppala & Aravind Pedarla, 2017; Xu & Sun, 2001) Strong hydrophilic clay minerals, high plasticity, a higher fines fraction and a high void ratio are all characteristics of such soils.

It is critical to examine the swell parameters of these swelling soils in order to minimise post-construction failures of engineering structures caused by rising urbanisation and industrialisation associated with these swelling soils (Das et al., 2011; Ikeagwuani & Nwonu, 2019; Anand J Puppala & Aravind Pedarla, 2017).

Fig 1.2 and Fig 1.3 show the damages that have been caused by the swell pressure of soil to the lightweight structure like Railroads, Highway Roads etc.

Figure 1.2 Cracks in Lightweight Structures (Railroads)

Figure 1.3 Cracks in Lightweight Structures (Roads)

The swell behaviour is primarily defined by free swell (FS), swelling pressure (Ps) and swell under load, (Çanakcı et al., 2009; Erzin & Erol, 2007). The swelling behaviour of compressed hydrophilic clays have been extensively studied over the last decade in terms of the effects of particle gradation, plasticity, compaction, and so on the development of swell pressure of expansive soils (from here on referred to as Ps-ES) (Atemimi, 2020; Cherif et al., 2018; Klopp, 2019; Hong Li et al., 2019; Parastar et al., 2017; Pastor et al., 2019).

Building foundation damaged that displaced bricks and inward deflection of foundation can be observed in Fig 1.4.

Figure 1.4 Building damage

Desiccation Cracks in expansive soil can be seen in in Fig 1.5 caused by drying & swelling.

Figure 1. 5 Cracks in expansive soil

The swell pressure Ps (kPa) is usually measured in the laboratory using an oedometer system according to the widely established ASTM standard. Due to the lengthy nature of the test and the slower saturation, a variety of empirical equations for the prediction of Ps have been created before. These equations have been derived using classical methods which have a number of limitations that are discussed in sections to come. (Ashayeri & Yasrebi, 2009; ASTM, 2014; Erzin & Erol, 2004; Kayabali & Demir, 2011; Kumar et al., 2020).

Soft computing (SC) approaches are gaining popularity due to their greater predictive capability in comparison to older approaches. They are being developed to evaluate the complicated behaviours of a range of geotechnical engineering systems (Zhang et al., 2020). Additionally, with the fast development of machine learning methodologies, data mining for operations in chemistry, materials science, and civil engineering, in particular, has received widespread attention during the last two decades (Hao Li et al., 2019; Vyas et al., 2015).

Recent advances in artificial intelligence (AI) techniques, or soft computation (SC) methods like Bayesian neural network (BNN), multilayer perceptron neural network (MLPNN), the hybrid form of ANNs, i.e. adaptive neuro-fuzzy inference system (ANFIS), alternate decision trees, eXtreme gradient boosting (XGBoost), artificial neural networks (ANNs) (subtypes are; back-propagation neural network (BPNN), general regression neural network (GRNN), and k-nearest neighbour (KNN)¸ support vector machines (SVMs), multivariate adaptive regression splines (MARS), evolutionary algorithms (EA), multi expression programming (MEP), genetic expression programming (GEP), and Ensemble Random Forest regression, genetic algorithm (GA), so on have helped in the formation of models with common statistical methods like regressions (Alade et al., 2018; Çanakcı et al., 2009; Das, 2013; Gandomi & Roke, 2015; Iqbal et al., 2020; Ozbek et al., 2013; Pham et al., 2016; Sathyapriya et al., 2017; Shahin, 2015; Zhang et al., 2020).

Mechanistic learning is frequently used to assess estimating models in order to construct an intelligent structure(Das, 2013). Additionally, (Giustolisi et al., 2007) developed a color-coded categorization system for different mathematical models based on white, black and grey. The first kind, referred to as a white-box model, is one in which known variables and parameters are based on physical laws, resulting in correct physical connections and maximal transparency. However, (Shahin et al., 2009) suggested that because their underlying process is unknown, they are hard to characterize. Secondly, black-box models are built on regressive data-driven systems in which the functional form of the connections between the variables is unknown and must be approximated. Thirdly, grey-box models are those logical systems in which a mathematical framework effectively analyses the system's behaviour.

For example both ANNs and ANFISs are classified as 'black-box models' because of their (i) lack of transparency, (ii) inability to clearly describe the underlying physical processes, and (iii) inability to generate closed-form empirical equations (Mohammadzadeh S et al., 2019; Sun et al., 2015). Whereas GEP is categorised as a "grey box model" due to its simplified and symbolic representation of physical processes (Mehr, 2020; Naghadehi et al., 2018). Although Genetic programming based models are believed to perform better than neural network-based ANN and ANFIS models in geotechnical engineering, comparative studies using various AI tools are ongoing to gain additional insights (Çanakcı et al., 2009; Gandomi & Roke, 2015; Hanandeh et al., 2020).

Genetic Programming based models are stable and successful since no predefined association is assumed while constructing the model (Gandomi, Alavi, Mirzahosseini, et al., 2011; Giustolisi et al., 2007).

Earlier correlations for the Ps of high Expansive soils were determined mostly using traditional statistical studies, which had numerous disadvantages such as

(i) fewer data points, (ii) lesser correlation between commanding factors, and (iii) a lack of integrated comparative evaluation (Mohammadzadeh S et al., 2019). Furthermore the swell pressure test in the laboratory is both costly and time-consuming (Kumar et al., 2020). In the past, a number of studies have utilised simple geotechnical indices with AI approaches to estimate swell index (Chen et al., 2019; Das et al., 2011; Das et al., 2010; Mozumder & Laskar, 2015; Najjar et al., 1996; Salahudeen et al., 2020). This study attempts to create models that use simple and economically determinable essential geo-mechanical properties to precisely predict the swell behaviour of virgin expansive soils. A soft computing approach, Multi Expression Programming (MEP), were used in this work to generate prediction equations for Ps-ES. Nine soil properties i.e., clay fraction (CF), plasticity index (PI), specific gravity (Gs), maximum dry density (MDD), optimum moisture content (OMC), swell percent (SP), natural water content (w_n), percentage sand, and percentage silt were used as input variables. Based on previous research, these input parameters were taken from the Ps-ES database. The predicted variables, i.e., the output variables, was swell pressure (Ps). The major goals of the research were to (i) create MEP-based prediction model, (ii) obtain an empirical equation from the developed model. (iii) compare MEP model's performance with that of the GEP model for Ps prediction. To evaluate the suitability MEP model, the statistical performance measures such as mean absolute error (MAE), root squared error (RSE), root mean square error (RMSE), correlation coefficient (R), regression coefficient (R^2), and relative root mean square error (RRMSE) were used. Furthermore, a parametric research was carried out, and the output results were then analysed to categorise the majority of positive and negative input parameters using sensitivity analysis.

1.1 Problem Statement

As the importance of swell potential & their parameters is described earlier it is compulsory to calculate all these factors prior to any construction.

Although we have many other methods like ASTM lab protocols/ methods, but there are some limitations with these methods.

LAB Limitations

- The experimental procedure of these methods is really time consuming/demanding, and some of them may take up to week to complete the standard procedure of testing.
- If we're required to test a lot of samples for better reliability, we require a greater number of apparatus or we will have to wait for completion of sample that is under process of incrementally loading, obviously for more than one sample their will be time or cost restraint. Apart from that proper care of Soil Undisturbed sample, precautions of system & process incremental loadings should be taken into account.

- Secondly, for those tests, the testing machinery and operators required is very costly and need skilled professionals to work on them.
- Any mishap or un-intentional wrong handling may lead to severe results without any indication.

Literature Limitations

- While empirical relations are drawn by experts, they have to rely heavily on Assumptions, that makes corelations very limited.
- For the simplicity purpose, mathematical techniques used in these formulas are very limited, these practiced formulas do not give corelations with all other parameters OR (because of limited nature) does not cater for all those multi variable complex relations of input parameters, on which the output is relaying.
- They contain Very complex Graphical Interpretations, hence there may be a chance of human error.

Prev. AI techniques Limitations

- Although other AI methods have been applied but all of them fall under the category of black model, that don't give any information about what is happening behind the process & they do not give any empirical formula, that is one of our basic desire.
- Although in past many other AI methods have been applied to predict swell potential of soil, they do not give any empirical formula or they give complex and multi-functional mathematical equation which is really difficult to calculate every time & difficult to visualize them, While MEP gives simple equation where we can control complexity of our equation.

1.2 Project Objective

Going to the above complexities, Researchers have always tried to find ways to make correlations of all inputs & to evaluate soil properties with the more complex input parameters that are difficult to find and time-consuming parameters, while making the system as simple as possible.

So, SC is an emerging field in Geotechnical engineering that have potential to solve all above mentioned problems.

- Development of new reliable empirical prediction models to evaluate swell pressure of expansive soils (Ps-ES) using Multi Expression Programming MEP soft computing methods.
- Use of detailed information from datapoints about geotechnical indices alongside the swell pressure to make correlations and obtain higher degree of accuracy & deriving the simple mathematical equations for Ps-ES using MEP modelling. (for preplanning phase)
- To develop a timesaving & cost-effective system that can make complex relations of all input parameter without any prior assumption, on which output is relaying.

CHAPTER 2

LITERATURE REVIEW

2.1 Use of MEP in Geotechnical Engineering

2.1.1 Soil Classification Using Multi Expression Programming

Soil properties, such as LL, PL, % of gravel, sand, and fine particles and the soil colour are utilised as input factors for forecasting the classification of soil. (Alavi et al., 2010).

Following were the main observations of this research.

1. The results show that MEP-based formulations can anticipate target values with a high degree of accuracy. When compared to analytical and numerical results acquired by different researchers, the results obtained by MEP formulation are proven to be more reliable.
2. Using MEP and two separate function sets, two formulae for soil classification were developed.
3. Comparisons were made between the MEP formulations and the experimental findings & SLNN (an existing model in the literature,).
4. The results of this comparison show that the suggested MEP models can anticipate target values with a high degree of accuracy.
5. In addition to their high accuracy, MEP based equations are fairly short & straight-forward and are more practical for application than SLNN equations.
6. This analysis demonstrated that MEP is a highly promising strategy for the formulation of many civil engineering issues in the future.

2.1.2 Data Mining Approach for Complex Geotechnical Problems

The difficulty of geotechnical behaviour analysis stems from the multivariable interdependence of soil and rock reactions. Traditional kinds of the design solutions of engineering are made simpler to deal with this complicated behaviour. The simple solutions for the creation of conventional models may result in extremely huge mistakes. The aim of this study was to demonstrate the effectiveness of genetic programming (GP) variants such as multi-expression programming MEP, linear genetic programming LGP and gene expression programming GEP by using them for several complex geotechnical problems.

GEP, LGP, and MEP are the Genetic Programming variants that distinguish between an individual's genotype and phenotype. GEP, MEP and LGP approaches are more compatible with computer architecture than the classical GP. This significantly accelerates their execution. These approaches excel at directly capturing the information inherent in experimental data without making assumptions about the fundamental principles guiding the system. This is one of their most significant benefits over most classic constitutive modelling approaches. (Alavi & Gandomi, 2011).

The following conclusions were made based on the results obtained from this research.

1. GEP, MEP and LGP modelling capabilities were demonstrated by predicting slope stability, settlement around tunnels, relative settlement of rockfill dams and the soil liquefaction. Findings were found to be more accurate when compared to those produced by other models provided in the literature. The basic models created using GEP, MEP and LGP gave excellent analytic tools for geotechnical engineers.

2. These ideas address the drawbacks of prior methods for analysing geotechnical engineering systems provided in the literature. Unlike ANN and many other soft computing techniques, GP approaches give equations that may can be used for ordinary design practise.
3. MEP is unique in its capacity to encode several computer programmes for a single problem on a single chromosome. Based on numerical trials, the LGP, GEP, and MEP methods surpass related strategies substantially.

2.1.3 Soil Deformation Moduli (Secant Modulus and Reloading Modulus)

The goal of this research was to use the MEP technique for creating new constitutive models for predicting the deformation moduli of soil. (Alavi et al., 2012).

The researchers gave the following conclusions from their research.

1. The suggested models provide accurate estimates of soil deformation moduli.
2. The Es and Er formulation outperform various empirical models observed in the literature.
3. The validation stages show the models' efficacy for broad application to soil moduli determination. In general, the models developed are appropriate for fine-grained soils.

2.1.4 Compression Index of Fine-Grained Soils

Multi-expression programming, MEP was used to develop a non-linear model for predicting the compression index of fine soils. This model connects the soil compression index, Cc to its LL, PL and void ratio. 108 consolidation experiments performed on various soils were used to create the model. (Mohammadzadeh et al., 2014).

The researchers gave the following conclusions from their research.

1. The results show that the MEP technique describes the soil compression index properly, resulting in very high prediction performance.
2. The derived model outperforms the current empirical formulae for the soil compression index substantially.
3. Despite ANN's somewhat better performance for analysed problem, a significant benefit of MEP over ANN is that it produces simpler equations that can be easily calculated by hand.
4. By integrating the data of various soil types, the performance of this model can be enhanced to produce more accurate results for a larger range of database.
5. In general, the models developed using this technique should be utilised for pre-planning and pre-design purposes, as well as to assess the general validity of laboratory or field test findings.
6. Furthermore, when testing is not possible, these models are effective options for determining Cc.

2.1.5 Consolidation Assessment

Considering the complicated nature of consolidation, the Multi Expression Programming, MEP was used to generate several models that are accurate and simple for calculating the degree of consolidation.(Sharifi et al., 2020).

The major findings of this research are given below.

1. The solutions consist of two groups of equations, one for calculating the average degree of consolidation and the other for calculating the time factor.
2. The results of the consolidation tests performed on four different clays obtained from the literature demonstrated that the proposed models performed well.

2.1.6 Prediction of Soil Compaction Parameters

Using multi expression programming (MEP), this work provides the construction of a new prediction model for soil compaction parameters (i.e. optimal water content and maximum dry density).(Wang & Yin, 2020).

Following were the main conclusions.

1. For training dataset, an R^2 greater than 0.87 suggested that the model can predict the maximum dry density and the optimal water content with good accuracy.
2. The model is also accurate in predicting these parameters (MDD and OWC) based on testing data with $R^2 > 0.85$.
3. In comparison to prior research, the current model is trustworthy and more appropriate for a broader range of soil types, including the coarse grained and high plasticity soils.
4. Because the model requires only 4 physical parameters of soil and the compaction energy, it is suitable for geotechnical applications. For example, if we assume a certain value of the compaction energy, the suggested model can automatically determine the optimal water content (OWC) and the maximum dry density (MDD) of soil with the four fundamental physical attributes known.

Compaction energy may thus be simply back determined based on the design necessary compaction parameter values.

2.2 Prediction of Swell Pressure Using Different Methods

Because of the surface properties of various clay types, expansive clays expand in volume when they come into contact with water. Swelling Pressure of Soil refers to the pressure exerted by expansive soil if it is not permitted to expand or if the volume change of the soil is halted.

2.2.1 Conventional Laboratory Techniques

These are the earliest known methods for predicting Ps. Since the study of expansive soils began, many tests such as the free swell, zero swell, loaded swell, limited swell, constant volume, swell-consolidation, and double oedometer test have been utilized; all of these tests are basically adaptations to the simple oedometer test. Among the tests described, the zero swell, constant volume, free swell, and swell consolidation tests are the most widely utilized. (ASTM, 1996; Petry et al., 1992)

2.2.2 Swell Pressure Prediction by Suction Methods

Suction methods were used in this study to predict swell pressure using the thermocouple psychrometer technique. It was found that the log soil suction and swell pressure had linear relationship. Using multiple regression analyses, soil suction was related to plasticity index PI, water content and the dry density. The correlations revealed a simple regression equation for predicting soil suctions quickly based on easily determined soil properties.

The results of the standard constant-volume (ASTM, 1996) swell tests were used to investigate soil suction versus swell pressure behavior. (Erzin & Erol, 2007)

The swell pressure for a clay soil can be predicted if the plasticity index, dry density & the initial soil suction are known. These parameters can be determined in a short period of time using simpler techniques than the oedometer test.

Major findings of this research are discussed below.

1. Suction methods for categorization of swell behavior in expansive soils outperform traditional oedometer tests.
2. These tests take less time than oedometer tests & provide data that is relevant to a variety of moisture content circumstances.
3. A linear relationship was established among log initial suction & swell pressure.
4. It is possible to conclude that the initial soil suction is the most important condition of suction that describes prospective swell pressures.

Erzin also considered the use of ANN (Artificial Neural Network) to analyze swell pressure vs soil suction. (Erzin, 2007)

2.2.3 Direct Method versus Indirect Methods

The direct approach was adopted in this study & the swelling pressure was determined using a load cell under constant volume. The direct method's findings were compared to indirect approaches such as the limited swell, zero swell, swell consolidation & double oedometer tests. (Kayabali & Demir, 2011)

Following were the major observations of this research.

1. The restricted swell test understates the swell pressure. This test requires many identical specimens for a single test which can be challenging at times.
2. The swell-consolidation test exaggerates the swell pressure.
3. No correlation was seen between swell pressure recorded with measured by the double oedometer test & the swell pressure measured by direct method.
4. The swell pressure measured by zero swell test overestimates the swell.
5. There was a significant correlation between swell pressure measured directly & free swell. It has been suggested that swell pressure may be anticipated using free swell.
6. The equipment used for this study to directly measure swell pressure is simple, produces reliable and reproducible findings. It is offered as an easy-to-use method for determining the swelling pressure of expansive soils.

Because of the stiffness of the load cell, swelling pressure measured directly is thought to be somewhat less than the real amount. Higher the stiffness value of the load cell, smaller will be the degree of inaccuracy in swelling pressure measurements.

2.2.4 Statistical Regression Models

This study employed ten distinct soil samples to create 2 linear regression models called as M1 & M2, which predict swelling pressure based on index properties and initial placement conditions. (James et al., 2013)

The researchers gave the following observations about their research.

1. A dataset of ten distinct soil samples was employed in this investigation. The two models were calibrated using eight of these specimens. The last two soil specimens were used to validate the models. This was done by calculating absolute error by taking the difference between the predicted swelling pressure by the new models & the actual swelling pressure.
2. When the models were ranked according to their absolute inaccuracy, M1 performed the best.
3. M1 produced consistent forecasts of swelling pressure with varying initial water content & initial dry unit weight amounts. Because it is a linear model, M1 is simple & the most user friendly of the two models.

2.2.5 Use of Artificial Intelligence Approaches: ANN, ANFIS & GEP

New empirical prediction models were created for swell pressure using the three soft computing methods i.e., artificial neural networks (ANNs), gene expression programming (GEP) & adaptive neuro fuzzy inference system (ANFIS). Nine geotechnical parameters were chosen as the predictor variables based on their relevance & their ease of determining. The network was trained, tested and the predictions were compared to the observed outcomes. (Jalal et al., 2021)

Following were the major conclusions.

1. The GEP & ANN based models can predict the Ps with high accuracy and without having any need for the prior assumptions.

2. The mathematical equations derived by GEP are much simpler than the ANFIS & ANN models.
3. The GEP model developed can be applied in normal design practices for the prediction of Ps.

2.3 Summary

The objectives and outcomes of the research explained above are given in summarised form along with proper references in the upcoming Table 2.1 and Table 2.2.

Table 2. 1 Use Of MEP In Geotechnical Engineering

USE OF MEP IN GEOTECHNICAL ENGINEERING		
Reference	**Objective**	**Outcomes**
Alavi et al., 2010	Formulation of Soil Classification	➢ Higher degree of accuracy ➢ Short and very simple MEP-based prediction equations
Alavi et al., 2012	Formulation of soil deformation moduli i.e. Es & Er	➢ Precise estimations of the soil deformation moduli ➢ The formulation outperforms several empirical models found in literature ➢ Suitable for fine-grained soils
Mohammadzadeh et al., 2014	Formulation of compression index of fine-grained soils	➢ Significantly better performance than the existing empirical equations ➢ MEP provided simplified equations for Cc that can be readily used via hand calculating ➢ The model is suggested to be used for pre-planning and pre-design purposes ➢ Good alternative to determine Cc when testing is not possible
Sharifi et al., 2020	Consolidation assessment	➢ Several models developed for different depths of soil ➢ Proper performance of the proposed models on comparison with results of the consolidation test carried out on four different clays attained from the literature
Wang & Yin, 2020	Prediction of soil compaction parameters	➢ Compared to the previous studies, the present model is more applicable and reliable for a wider range of soil classifications, even for the soils with high plasticity and coarse-grained soils from the database ➢ The model requires only four physical properties of soil and compaction energy therefore it can be recommended for geotechnical application

Table 2. 2 Prediction Of Swell Pressure Using Different Methods

PREDICTION OF SWELL PRESSURE USING DIFFERENT METHODS		
Reference	**Objective**	**Outcomes**
ASTM, 1996	Conventional laboratory methods	➢ Earliest known techniques for the prediction of Ps ➢ Free swell, zero swell, and swell-consolidation tests ➢ Time consuming and costly
Erzin & Erol, 2007	Swell pressure prediction by suction methods	➢ Thermocouple psychrometer technique used ➢ Linear relationship between soil suction and Ps ➢ Soil suction correlated to W.C, P.I and D.D using multiple regression analyses ➢ advantages over conventional oedometer tests
Kayabali & Demir, 2011	Measurement of swell pressure: direct method versus indirect methods	➢ The results of Direct and inDirect methods were compared ➢ Considerably high correlation between swell pressure measured using the direct method and free swell ➢ The apparatus devised measuring swell pressure by direct method is simple, robust, and gives reasonably reliable and repeatable results
James et al., 2013	Prediction of swell pressure using the Statistical regression models	➢ Ranking the models by the absolute error of each model, M1 preformed most accurately ➢ M1 resulted in steady predictions of swelling pressure across fluctuating initial dry unit weight and initial water content values ➢ M1 is the most simple and user friendly of the new models because it is a linear model and is easier to calibrate and estimate the independent variables in the laboratory
Jalal et al., 2021	Prediction of swell pressure using AI approaches: ANN, ANFIS and GEP	➢ GEP and ANN-based formulated models can accurately predict the Ps with high accuracy and without any prior assumption ➢ The mathematical equations derived by GEP are much simpler than the ANN and ANFIS models

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction to Artificial Intelligence

Before diving into the research methodology, one should have a clear understanding of Artificial Intelligence and how it’s implemented in various problems. A simplistic answer to the question "What Artificial Intelligence means?" is that it varies depending on whom you ask. A person slightly familiar with technology might think of it as robots or some kind of advanced computer systems. They'd describe AI as a terminator-like figure that can take decisions & act freely. If you ask about AI from a researcher, he/she will tell you that it is a set of computer programs that can produce results without being explicitly instructed to do so. And they'd all be correct. To summarise, Artificial intelligence (AI) is in fact a replication of human intellect in robots that are built to think like people & copy their behaviour.

Artificial intelligence is based on the idea that human intellect may be characterised in such a manner that a computer could simply duplicate it and complete tasks ranging from the most basic to the most complicated. Artificial intelligence has a goal of imitating human cognitive processes. To the extent that they can be described concretely, researchers & developers in the area are making unexpectedly quick progress in replicating tasks like as learning, reasoning, and perception. Some predict that inventors may soon have a capability to create systems that could beat humans' ability to study or reason about any topic. Others, however, remain suspicious, owing to the fact that all cognitive activity is laced with value judgements that are vulnerable to human experience.

As technology makes further advancements, earlier artificial intelligence criteria become obsolete. Computers/devices that calculate fundamental calculations or detect text with the use of optical character recognition, for example, are no longer regarded to possess artificial intelligence since these abilities are now assumed to be inherent in computers. AI is constantly developing to help a wide range of sectors. Machines are wired utilising a multidisciplinary method that incorporates mathematics, computer science, linguistics, psychology & other disciplines.

3.2 Collection of Data

A database was created from a comprehensive literature review of globally published research publications by collecting 168 Ps entries for the final creation of accurate predictive model employing MEP approach. Between 1995 and 2020, data was collected for the typical expansive soils ranging from low to extremely high nature, with more than 95 percent of datapoints collected after 2000. It was important to choose datapoints that provided precise information regarding geotechnical indices as well as swell pressure parameters. The following Table 3.1 shows the dataset that we're using in our research.

Table 3.1 Dataset obtained from different international publications used in this project

Swell pressure, P_s												
Serial No.	**Reference**	**Clay fraction (CF)**	**Liquid Limit (LL)**	**Plasticity Index (PI)**	**Specific Gravity (Gs)**	**Max. Dry Density (MDD)**	**Optimum Moisture Content (OMC)**	**Swell Potential (SP)**	**Natural Water Content (w_n)**	**Sand**	**Silt**	**Swell Pressure (Ps)**
		%	%	%		kN/m^3	%	%	%	%	%	KPa
1	(Abdalqadir et al., 2020)	34	45.2	19.83	2.72	18.34	15	5.96	2.4	23.2	36.3	142
2	(Benyahia et al., 2020)	68	73.3	54.3	2.6	16.97	16.4	7.81	16.4	4	24.5	271
3		64	74.8	41.5	2.64	16.19	22.09	6.6	22.1	4.2	32.7	258
4		66	71.7	52.68	2.6	15.99	19.39	6.57	19.4	2	31	236
5	(Kowalska and Ptaszek, 2019)	33	71	39	2.77	15.6	22.5	7	2.8	5.4	55.3	120
6	(Al-Rawas et al., 2005)	65	148	116	2.68	14.9	23	9.3	3.7	0	35	170
7	(Soundara and Selvakumar, 2020)	72.5	74	35	2.66	17.55	19	5.43	2.2	4	23.5	270
8	(Taher et al., 2020)	70	31	12.9	2.7	16.8	18.4	14.9	12.1	7	15	139
9	(Shi et al., 2002)	26	58.5	33	2.7	16.5	19	2.7	19	22	27.5	75
10		58	63	35.5	2.7	14.5	26	1.3	26	4.7	36	89
11		32	59.5	30.5	2.7	14.5	30	1.82	30	22	27.5	51
12		41	64	35.5	2.7	15	28.5	4.49	28.5	18	26	59
13		19	43	21.5	2.68	14	25	3.4	25	13	67	29
14		43	57.5	28.5	2.69	15.5	26.5	5.05	26.5	15	29	125
15	(Phani Kumar and Sharma, 2004)	69	80	52	2.72	13.75	40	10.8	14	7	24	90
16	(Sabtan, 2005)	27	52	20	2.66	17.36	9.2	2.4	9.2	34.3	31.8	215
17		26	58	24	2.66	17.36	8.8	2.5	8.8	34.3	31.8	144
18		61	46	25	2.66	17.36	12.4	5.6	12.4	4.7	36	283
19		32	67	30	2.66	17.36	8.1	4.8	8.1	28	20	204
20		73	58	35	2.66	17.36	4.4	6.5	4.4	2	24	302
21		34	36	24	2.66	17.36	5.5	3.8	5.5	22.2	34.2	237
22		71	72	40	2.66	17.36	3.7	7.8	3.7	2.5	22.5	341

23		61	87	45	2.66	17.36	6.2	6.6	6.2	2.7	38	247
24		21	32	8	2.75	17.36	8.7	1.7	8.7	14	31	123
25		93	94	70	2.75	17.36	2.7	10.2	2.7	1	6	480
26		91	105	79	2.75	17.36	2.3	10.8	2.3	1	8	521
27		41	72	33	2.75	17.36	9.1	4.3	9.1	19	24	223
28		33	39	21	2.75	17.36	7.2	3.1	7.2	14	27.5	223
29		84	80	63	2.75	17.36	4.9	9.2	4.9	4	14	425
30		77	85	50	2.75	17.36	4.3	7.8	4.3	4	19	263
31		53	82	55	2.75	17.36	5.3	7.2	5.3	12	30	257
32		58	67	52	2.77	17.36	2.1	8.4	2.1	7	33	377
33		84	79	53	2.77	17.36	2.3	9.7	2.3	5.5	13	383
34		71	62	46	2.77	17.36	1.3	8.5	1.3	4	23	421
35		44	57	41	2.77	17.36	5.5	5.4	5.5	12	28	249
36		79	93	56	2.77	17.36	2.4	10.5	2.4	6	15	430
37		48	58	28	2.77	17.36	7.3	4.1	7.3	8	40	182
38		49	62	31	2.77	17.36	5.1	5.7	5.1	28	20	244
39		91	89	58	2.77	17.36	3.8	8.2	3.8	1	7	453
40	(Puppala et al., 2006)	55	44	22	2.46	16.26	17.49	8.1	3.2	3	22	116
41	(Seda et al., 2007)	45	52	34	2.81	16.5	21.5	8.2	21.6	5.4	37.4	125
42		27	77.2	42.5	2.73	17.2	18	4.09	18	22	27.5	210
43	(Yan and Wu, 2009)	33	56	26	2.69	16.5	20	3.34	20	8	40	170
44		31	55.4	23.9	2.7	16.8	22	2.97	22	22	27.5	142
45		47	116	75	2.66	14.9	31.4	5.68	19.1	5	50.1	125
46	(Zheng et al., 2009)	48	132	80	2.67	15	29.2	5.09	20.1	5.4	55.3	150
47		28.2	108	61	2.64	18	19.9	4.91	17.6	24	35.4	150
48		51	59	27	2.68	16.2	26.2	2.3	0.9	8	40	75
49	(Lin and Cerato, 2011)	62	54	34	2.78	16.7	20.6	5.7	2.3	2	37	141
50		50	70	49	2.77	15.5	24.2	9.3	3.7	33	17	230
51		23	78	34	2.77	13.5	32	0	0	2	75	65
52		45	71	28	2.75	11.7	40	0	0	19	24	107
53	(Al-Rawas, 1999)	55	94	54	2.78	15.1	28	0.6	0.2	28	20	73
54		48	64	40	2.77	16.6	17	0.9	0.4	14	27.5	193
55		48	59	37	2.75	19	7	13.3	5.3	28	20	221
56	(Sabat and Nanda, 2011)	56	60	28	2.61	16.1	21	4.78	1.9	18	26	128
57	(Al-Mukhtar et al., 2012)	55	95	70	2.65	12.94	32	6.6	2.6	5.4	37.4	150

58	(Gandhi, 2013)	60	62	37	2.7	13.1	22.1	5.4	2.2	0	40	216
59	(Rashid et al., 2013)	43	55	32	2.69	15.3	22	4.5	7	2	58	85
60	(Sabat, 2013)	60	61	30	2.67	16.7	20.7	5.2	2.1	12	28	132
61	(Malekzadeh and Bilsel, 2014)	52	57	29	2.56	14.68	24	4.4	1.8	8	40	200
62	(Radhakrishnan et al., 2014)	70	85.2	52.13	2.61	15.21	24.7	20.6	8.2	2	28	295
63	(Reddy et al., 2015)	73	98	62	2.56	16.2	26	12	4.8	25	73	192
64		72	76	46	2.6	16	25	15	6	26	72	230
65		71	64	38	2.58	16.4	23	17	6.8	26	71	280
66	(Zumrawi, 2015a)	66	69	36	2.69	14.21	25.38	10.1	4	18	16	105
67		56	61	31	2.76	15.22	23.4	9.5	3.8	25	19	93
68		70	72	40	2.72	13.8	25.6	15	6	10	20	130
69		57	68	37	2.7	14.33	24	8.7	3.5	19	24	95
70		52	59	30	2.66	14.5	22.7	7	2.8	28	20	90
71		62	74	39	2.7	13.2	26.7	10.5	4.2	20	18	122
72	(Ameta et al., 2007)	27	100	50	2.69	16.18	16	9.9	4	7	66	104
73		35	50	27	2.69	16.18	20	5.8	2.3	12	53	61.3
74	(Dang et al., 2016; Hasan et al., 2016)	65	86	49	2.64	12.65	36.5	9.8	30.8	18.3	16	80
75	(Shalabi et al., 2017)	22.5	51.9	24	2.71	18.02	15.6	5.44	2.2	10.3	64.1	103
76	(Zumrawi and Mohammed, 2017)	63	77	49	2.72	14.51	26.5	32.7	13	5	33	250
77		34	54	31	2.76	15.08	20	8.5	3.4	21.8	36.3	123
78		53	76	52	2.64	14.64	26	18.5	7.4	4.7	36	210
79	(Zumrawi et al., 2017b)	62	63	45	2.74	13	17.45	5.5	2.2	8	30	124
80	(Akgün et al., 2018)	17	50	25	2.71	14.62	22	3.2	1.3	14	75	42
81		22	39	12	2.69	16	23.4	0.2	22.5	5	66	18.4
82		37	70	34	2.69	12.1	28.2	0.5	25.3	1	51	12.5
83		48	62	31	2.67	13.1	26.05	1.8	27.4	3	32	52.1
84	(Lin and Cerato, 2014)	51	59	27	2.8	15.9	26.2	2.3	0.9	4	33	75
85	(Dayioglu et al., 2017)	45	57	28	2.67	17.1	17.6	4.32	1.7	0	55	230

86	(Zumrawi, 2015b)	56	61	31	2.69	14.21	25.38	9.5	3.8	25	19	90
87		70	72	40	2.76	15.22	23.4	15	6	10	20	130
88		52	54	29	2.72	13.8	25.6	7	2.8	28	20	95
89		62	70	42	2.7	14.33	24	10	4	20	18	122
90		14	37	10	2.66	14.5	22.7	1.6	0.6	64	22	28
91		21	45	17	2.7	13.2	26.7	4	1.6	61	18	50
92	(Eyo et al., 2019)	26	58	28	2.6	15	17	13	5.2	11	74	90
93		30	85	48	2.65	13.9	21	17	6.8	5.4	70	102
94		35	130	82	2.69	13.5	23	28	11.2	2.1	65	160
95	(Syed et al., 2020)	50	52	32	2.59	16.5	24.6	4.4	1.8	27	22	77
96	(Gupta and Sharma, 2016)	43	62.7	32.3	2.29	15	22.8	4	1.6	12	28	78
97	(Parik and Patra, 2020)	35.03	50.3	26.5	2.65	15.6	21.1	3.4	1.4	2.1	62.8	43.5
98		14.69	45.8	20.5	2.75	16	20.8	1.85	0.7	3.1	81.2	32
99	(Zumrawi et al., 2017a)	46	57	38	2.66	14.5	22.7	4.7	1.9	3	34	55
100	(Chittoori et al., 2019)	30	43	19	2.6	15.65	21.5	2.58	1	22	27.5	70
101	(Gheris and Hamrouni, 2020)	65.2	52.2	47.97	2.7	14.22	31.7	5.04	31.7	2	33	100
102	(Phanikumar and Singla, 2016)	60	79	53	2.73	12	22.7	21.78	8.7	14	26	152
103	(He et al., 2018)	40	76	58	2.84	13.47	31	7.1	2.8	18	26	93
104	(Rosenbalm & Zapata, 2017)	32.2	48	27	2.72	17.15	18	3.46	1.4	3	32	120
105		48.6	65	42	2.71	16.48	19	5.32	2.1	11	34	230
106	(Ramesh et al., 2012)	40	64	37	2.75	17.3	18	14.5	5.8	18	40	135
107	(Basma et al., 1998)	20	56.8	21.9	2.7	17.5	21	8.36	8.9	14	40	182
108	(Kumar et al., 2020)	58	59	31	2.68	15.89	21	5.9	30.9	12	30	202
109		54	88	53	2.72	16.19	21	6.75	30.2	22	24	260
110	(Ozer et al., 2012)	35.2	65	30	2.7	13.7	29.1	2.6	29.1	23.2	36.3	19.8
111		35.1	68	35	2.7	12.9	23.1	7.7	23.1	21.8	36.3	51.1

112		35.5	67	34	2.7	12.6	32.7	5.3	32.7	22.2	34.2	57.4
113		23.2	51	23	2.7	16.7	17	4.7	17	39.3	26.7	71.6
114		21.8	50	23	2.7	14.9	23.2	2.6	23.2	38.5	31.8	18.8
115		26.5	54	26	2.7	15.9	19.6	5.5	19.6	34.3	31.8	72.2
116	(Yenes et al., 2012)	33.2	115	70.57	2.63	14.51	27.92	8.35	27.9	8	12.7	133
117		33.1	100	64.35	2.63	12.55	29.76	6.09	35.9	5.4	10	83.4
118	(Baby et al., 2016)	50	60.2	28.2	2.67	15	22.8	4.3	1.7	26.7	23.2	124
119	(Mirzababaei et al., 2017)	93.3	74	47	2.71	16.2	16.8	11.5	4.6	3.6	3.1	219
120		27.5	35	13.8	2.69	18.62	16.5	3	1.2	7	66	120
121	(Kaczyński & Grabowska-Olszewska, 1997)	43	56.5	28	2.73	19.32	23.2	13.2	5.3	15	29	131
122		35	72	30.5	2.73	19.26	16.5	1.8	0.7	18	26	165
123		47.5	80	47.5	2.74	20.75	17.5	2.5	1	28	20	182
124		59	54	32	2.74	19.47	19.5	2.2	0.9	7	33	169
125	(Azzam, 2012)	40	52	30	2.61	14.8	13	30	13	5	55	258
126	(Carraro et al., 2010)	35	42	26	2.64	16.3	18.6	1.6	19.4	18	72	96
127	(Yazdandoust & Yasrobi, 2010)	30	65	41	2.59	12.72	34.4	11	4.4	24	46	65
128		53	71	48	2.59	11.95	36.9	16.4	6.5	28	19	84.4
129		73	96	67	2.59	11.1	31	22	8.8	21	6	137
130	(Trouzine et al., 2012)	55.3	45	22	2.55	15.25	20	4.6	1.8	16	17.6	67
131		61.2	133	83	2.61	11.5	20	15.9	6.3	20	18.8	148
132		23	53	24	2.69	15.1	16.5	4.3	16.5	2	75	64
133		23	51	28	2.69	16.1	23.2	6	23.2	2	75	95
134		23	58	28	2.69	15.4	17.2	5.4	17.2	2	75	86
135		23	51	25	2.69	16.2	16.5	5.7	16.5	2	75	102
136		24	51	26	2.67	16.3	17.5	6.9	17.5	2	74	107
137		24	50	26	2.67	17.9	19.5	7.6	19.5	2	74	134
138		24	54	27	2.67	16.2	14.5	6.5	14.5	2	74	95
139		20	42	22	2.67	16.5	18	6.6	18	2	74	93
140	(Rashid, 2015)	20	51	24	2.67	14.8	16	4.2	16	2	74	43
141		20	51	28	2.67	16.1	20.5	6.8	20.5	2	74	102
142		20	56	28	2.67	14.9	10.5	5.4	10.5	2	74	68
143		20	54	26	2.67	16.3	17.5	6.7	17.5	2	74	105
144		22	51	23	2.67	16	18.3	5.2	18.3	2	74	90
145		21	42	21	2.65	16.8	11.4	5.8	11.4	1	78	98
146		22	50	25	2.65	15.5	17.9	5.8	17.9	1	78	76
147		22	42	21	2.65	16.5	18.6	5.4	18.6	1	78	86
148		22	58	28	2.65	16.4	15	7.5	15	1	78	100

149		22	55	27	2.65	14.5	18.2	3.9	18.2	1	78	62
150		22	52	25	2.65	14.6	15	4.9	15	1	78	55
151		22	56	28	2.65	16.1	18	6.6	18	1	78	105
152		21	52	26	2.65	16.2	22.7	6.7	22.7	1	78	91
153		25	55	28	2.7	14.9	10.1	4.7	10.1	1	74	71
154		21	46	23	2.7	16.2	18	5.5	18	1	74	83
155		21	37	21	2.7	16.8	22.3	6.7	22.3	1	74	94
156		21	50	21	2.7	16.8	11.5	6.5	11.5	1	74	105
157		42	50	25	2.7	15.6	25.3	5	25.3	1	74	81
158		42	50	23	2.7	15.2	22.1	4.2	22.1	1	74	67
159		42	51	24	2.7	15.1	16.5	5.1	16.5	1	74	60
160	(Mujtaba et al., 2018)	23	55	30	2.65	18.3	15	7.6	11	11	74	154
161		22	40	20	2.62	17.9	13.5	5	9.7	11	74	128
162	(She et al., 2020)	60.4	63	38.06	2.7	18	16.2	19	7.6	3.9	35.7	183
163	(Mumtaz et al., 2020)	60	56	31	2.74	18.3	14	6	2.4	3	37	180
164	(Khennouf & Baheddi, 2020)	71	72.3	43.08	2.74	17.5	14.1	5.4	14.1	4.5	24.5	190
165	(Pedarla et al., 2019)	22	55	37	2.69	16.33	19	8.8	3.5	34.3	31.8	183
166		25	63	42	2.69	16.49	19	12	4.8	7	66	194
167		20	46	26	2.66	16.93	19	6.2	2.5	11	65	88.3
168		11	24	12	2.61	14.62	27	9.1	3.6	13	75	158

3.3 Data Division and Pre-Processing

After generating the basic mathematical equations for Ps utilising MEP modelling, the goal was to avoid utilising correlations & achieve a greater degree of accuracy. The specific soil parameters used to calculate Ps may be found in Tables 3.1 given previously (Supporting Material). As previously stated, the input variables were the grain size distribution, Atterberg limits, compaction characteristics & expansivity (i.e., Gs, MDD, CF, PI, OMC, Wn, SP, Sand, and Silt). Based on a recent literature analysis (Akan & Keskin, 2019; Berrah et al., 2020; Elbadry, 2017; Kumar et al., 2017; Mawlood & Hummadi, 2020; Saputra & Putra, 2020; Soleimani et al., 2018), the most critical elements influencing the behaviour of expansive soil swelling were identified.

Table 3.2 shows the description of statistical distribution for whole set of parameters addressed here. It provides information on the typical geotechnical indices of expansive soils that influence their swell properties. The Ps of the expanding soil are seen to vary from 12.5 to 521 KPa. These values (as shown in Table 3.2) are advised to be utilised while calculating the Ps utilising the proposed AI model in this work. The effectiveness of generated model is heavily influenced by the distribution of datapoints (Gandomi & Roke, 2015). Furthermore, the accuracy of prediction of the AI model for a certain prediction aim is mostly influenced by the data feature, data size, and the inherent link between input and output variables (Maeda, 2018).

Table 3.2 Statistical description of the entire dataset

	count	mean	std	min	25%	50%	75%	max
CF	168.0	43.766190	19.463599	11.00	25.000	42.50	59.250	93.30
LL	168.0	64.029762	19.962508	24.00	51.675	59.00	72.000	148.00
PI	168.0	36.187321	16.083106	8.00	26.000	31.00	43.560	116.00
GS	168.0	2.686250	0.063827	2.29	2.660	2.69	2.720	2.84
MDD	168.0	15.783571	1.721913	11.10	14.635	16.10	17.225	20.75
OMC	168.0	19.449464	7.867586	1.30	16.350	20.00	24.000	40.00
SP	168.0	7.231726	5.263008	0.00	4.300	5.80	8.550	32.70
wn	168.0	10.438690	9.083966	0.00	2.675	6.80	17.925	35.90
Sand	168.0	11.841667	11.447592	0.00	2.075	7.50	19.250	64.00
Silt	168.0	41.497619	22.450499	3.10	24.000	33.50	66.000	81.20
Ps	168.0	147.129762	94.800793	12.50	84.150	123.00	193.250	521.00

Previous research has shown that combining too many inputs with very low correlation with the desired output has a detrimental impact on model performance and increases its complexity (Abunama et al., 2019; Javed et al., 2020). The term "complex" refers to a transitory period resulting from a reversal, an adventure, or their combination. The computational complexity of a certain problem refers to the computing difficulty in evaluating a certain problem (Papadimitriou, 2020b). The greater the number of various classes within a surface, space, or spatial object, the greater its spatial complexity, which may be described as the degree of complexity required to reduce the structure of a two-dimensional or higher-dimensional item (Papadimitriou, 2020a). As a result, the necessity of geographic data complexity for processing & maintaining huge soil data sets is critical, because the more spatially complex an area is, the more time consuming and precise the environmental management plan will be (Papadimitriou, 2009) Finally, nine parameters were chosen as predictors of the dependent variables for the MEP model construction based on the evidence presented (Ps). The Ps is clearly controlled by all factors, notably CF, OMC & PI. Following the acquisition of data points, it is common practice to divide the available data into three subsets like training, testing, and validation (Iqbal et al., 2020; Maeda, 2018). The dataset was divided into 3 sets using python's library sklearn.

3.4 Modelling Parameters

As previously stated, numerous MEP fitting parameters must be established before modelling in order to construct an efficient and generalised model. The input parameters are chosen using earlier recommendations and a trial-and-error method (Mousavi et al., 2010). The population size determines the number of programmes that will emerge in the population. A model having large population would be complicated, highly accurate, and may take long time period to converge. Moreover, once the size of the model is raised above a certain limit, the issue of overfitting of the model may develop. The procedure began by assuming a total of ten subpopulations.

Table 3.3 Fitting Parameters for our model

Number of subpopulation	50
Size of subpopulation	250
Code length	40
Crossover probability	1
Mathematical operators	+, −, ×, ÷,sqrt,exp,Power,ln,
Tournament size	4
Operators	0.5
Variables	0.5
Number of generations	500

Table 3.3 shows the parameters that were chosen for the two models produced in the study. For the sake of simplicity in the final expressions, the function set considers the four basic mathematical operators i.e subtraction, addition, multiplication, and division. The number of generations specifies the level of accuracy that the algorithm should reach before being terminated. A run with high number of generations would result in a model with the fewest statistical errors. Similarly, the rate of crossover and mutation is a measure of the likelihood of the off springs undergoing these genetic processes. The cross over rate ranges between 50% and 95%. Several combinations of these parameters were tested on the data, and the optimum combination was chosen based on the model's overall performance qualities, as shown in Table 3.3. Overfitting of the data is one of the challenges with AI-based modelling. A model works well on the original data, but its efficiency suffers dramatically when applied to previously unseen data. To prevent this problem, it has been suggested that the trained model be tested on an unknown or testing dataset (Pyo et al., 2020; Qiu et al., 2020). As a result, the whole database was randomly partitioned into training, validation, and testing sets. While modelling, the training and validation data were processed. The validated model is next evaluated on the third dataset, which was not utilised during model building. It was assured that data distribution is consistent across all three datasets. In the current study, 70%, 15%, and 15% of the data were used for training, validation, and testing, respectively as shown in fig 3.1. The resulting models outperformed the competition on all three datasets. MEPX 2021.05.18.0 a commercially available computer tool, was used to implement the MEP algorithm.

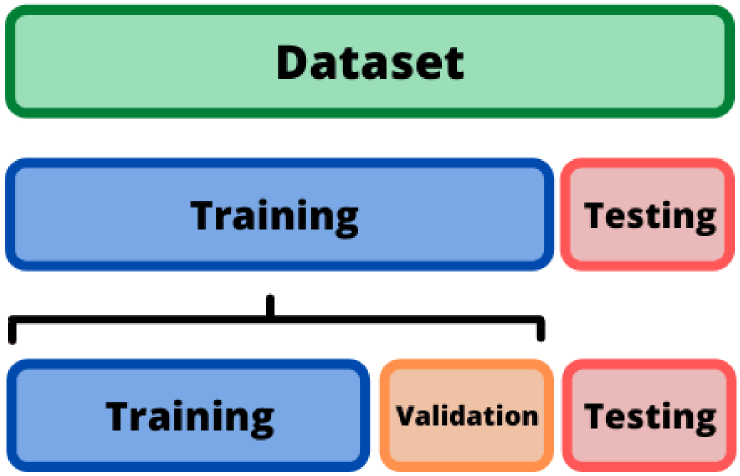


Figure 3. 1 Division of dataset

The algorithm begins by producing a population of viable solutions. The process is iterative, with each generation bringing us closer to a solution. Within the solution population, the fitness of each generation is evaluated. The MEP algorithm evolves until there is no change in the pre-specified fitness function, i.e., R or root mean squared error (RMSE). The objective function (OF) is also assessed for each trained model in this study to assess overall efficiency since it takes into account the effects of R, RMSE, and the number of data points. If the model findings for the three datasets (training, validation, and testing) are not correct, the procedure is repeated by gradually increasing the number and size of subpopulations. The final model is then chosen based on the lowest OF. However, it was discovered that the performance of certain models on the training set was superior to the testing set, indicating that the model was over-fitted, which should be avoided. It should be mentioned that the evolutionary period of the number of generations has an effect on the correctness of the produced model. A model would continue to evolve endlessly with these types of algorithms owing to the introduction of additional variables into the system. However, in the current work, the model was halted either after 500 generations or when the change in fitness function was less than 0.1 percent. Furthermore, an ideal model should fulfil many performance indicators, as explained in the following discussion.

The models' effectiveness is assessed by estimating numerous statistical error metrics. R, RMSE, mean absolute error (MAE), relative root mean square error (RRMSE), relative squared error (RSE), and performance index (ρ) are among them. Another approach to prevent model overfitting is to choose the optimal model by minimising the objective function OF, as recommended by (Azim et al., 2020; Gandomi & Roke, 2015). This method was used in the current investigation, and the OF is known as the fitness function. These statistical measures have the following expressions: Eqs. (I) – (VII).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (e_i - m_i)^2}{n}} \qquad \text{Eq I}$$

$$MAE = \frac{\sum_{i=1}^{n} |e_i - m_i|}{n} \qquad \text{Eq II}$$

$$RSE = \frac{\sum_{i=1}^{n} (m_i - e_i)^2}{\sum_{i=1}^{n} (\bar{e} - e_i)^2} \qquad \text{Eq III}$$

$$RRMSE = \frac{1}{|\bar{e}|} \sqrt{\frac{\sum_{i=1}^{n} (e_i - m_i)^2}{n}} \qquad \text{Eq IV}$$

$$R = \frac{\sum_{i=1}^{n} (e_i - \bar{e}_i)(m_i - \overline{m}_i)}{\sqrt{\sum_{i=1}^{n} (e_i - \bar{e}_i)^2 \sum_{i=1}^{n} (m_i - \overline{m}_i)^2}} \qquad \text{Eq V}$$

$$\rho = \frac{\mathrm{RRMSE}}{1 + R} \qquad \text{Eq VI}$$

$$OF = \left(\frac{n_T - n_{TE}}{n}\right)\rho_T + 2\left(\frac{n_{TE}}{n}\right)\rho_{TE} \qquad \text{Eq VII}$$

Fitness Functions for model evaluation

Where e_i, m_i, e_i, and mi represent the ith experimental, predicted, mean experimental, and mean predicted values, respectively, and n represents the total number of data points utilized for modelling.

A high R value and minimal statistical errors imply that the model is accurate. R has been proposed by researchers to assess linear dependence between input and output (Nguyen et al., 2019), and a value greater than 0.8 denotes a high relationship between predicted and experimental values (Gandomi, Alavi, Mirzahosseini, et al., 2011). However, because it is insensitive to multiplication or division of output with a constant, it cannot be used alone as a measure of overall model efficiency. The average magnitude of the mistakes is quantified by the RMSE and MAE. However, each metric has its own relevance. Errors are squared before to estimating the average in RMSE, so larger errors are given more weight. A high RMSE score shows that the number of predictions with large errors is significantly larger than desired and should be reduced. MAE, on the other hand, gives minimal weight to significant errors and is always less than RMSE. Similarly, Despotovic et al. (2016) proposed that a model is excellent if the RRMSE value is between 0 and 0.10, and good if the value is between 0.11 and 0.20. The values of ρ and OF range from 0 to infinity. If the value of ρ and OF is < 0.2, the model is regarded satisfactory (Gandomi & Roke, 2015). It should be noted that OF considers three elements at the same time: R, RRMSE, and the relative fraction of data in separate sets, namely training and testing. As a result, a low score signifies higher overall performance of a specific model. As previously stated, numerous trial runs were undertaken, and the model with the lowest OF is presented in this study.

A high R value and minimal statistical errors imply that the model is accurate. R has been proposed by researchers to assess linear dependency between input and output (Nguyen et al., 2019), and a value larger than 0.8 denotes a high relationship between anticipated and experimental values (Gandomi, Alavi, Mirzahosseini, et al., 2011; Gandomi, Alavi, & Yun, 2011). However, because it is insensitive to multiplication or division of output with a constant, it cannot be used alone as a measure of overall model efficiency. The average size of the errors is quantified by the RMSE & MAE. Both parameters, however, have their own significance. Mistakes are squared before to estimating the average in RMSE, thus greater errors are given more weight. A high RMSE score shows that the number of predictions with large errors is significantly larger than expected and should be reduced. MAE, on the other hand, gives minimal weight to significant mistakes and is always smaller than RMSE. Similarly, Despotovic et al. (2016) proposed that a model is excellent if the RRMSE value is between 0 & 0.10, and good if the value is between 0.11 & 0.20. The values of and OF range from 0 to infinity. If the value of p and OF is 0.2, the model is regarded satisfactory (Gandomi & Roke, 2015). It should be mentioned that OF considers three elements at the same time, namely R, RRMSE, and the relative percentage of data in separate sets, namely training and testing. As a result, a low number signifies a model's greater overall performance. As previously stated, numerous trial runs were carried out, and the model with the lowest OF is given in this study.

CHAPTER 4

ANALYSIS AND RESULTS

4.1 Formulation of the Empirical Equations

The output given by the Multi Expression Programming model has been decoded to get the empirical equation for the swell pressure of expansive soils. The equation thus derived is given below in (Equation-a).

$$Ps = 2 * (A + SP) - (OMC + \exp(B)) - A - \big((B * MDD) + OMC\big) + \big(MDD * sqrt(PI + CF + \tan(PI + CF))\big) - \tan\big((2 * \mathrm{B} * MDD) + silt\big)$$

Where,

$$A = \frac{PI + CF + \tan(PI + CF)}{OMC + \exp(\sin(MDD))}$$

$$\mathrm{B} = \sin(MDD)$$

4.2 Evaluation Through Scatter Plots

The experimental and the predicted results are compared using scatter plots shown in the figures below. It can be seen from the scatter plots that the model is well fitted across all the three sets i.e., training, validation and testing. The slopes are as follows.

- Training: y = 0.9039x
- Validation: y = 0.9174x
- Testing: y = 0.9806x

The fact that the model is well fitted and that the model possesses a strong correlation between the experimental and the predicted values is evident by the slopes of the regression lines given above. It can be seen that the slopes are very close to 1 and thus approaching the condition of the ideal fit i.e., a slope of 1:1 between the experimental and the predicted values. The scatter plots for the individual sets and also for all the three sets combined has been shown below (Fig. 4.1-4.4) for better visualization.

Further it can be seen that the residual values of all the sets are close to the best fitted line and are following a general trend in terms of the residuals which is also another indication of a good and generalized model.

We can also see from the combined scatter plot that the residual values of all the sets are almost equally spread out in the plot which means that all the three sets taken were statistically consistent which is a basic requirement of creating a good model and for the ideal training of the model and has been met in our research.

As Fig 4.1 shows the Comparison of actual and predicted results for training test data. A graph with the independent variable on the horizontal axis and the dependent variable on the vertical axis is made & comparison has been conducted.

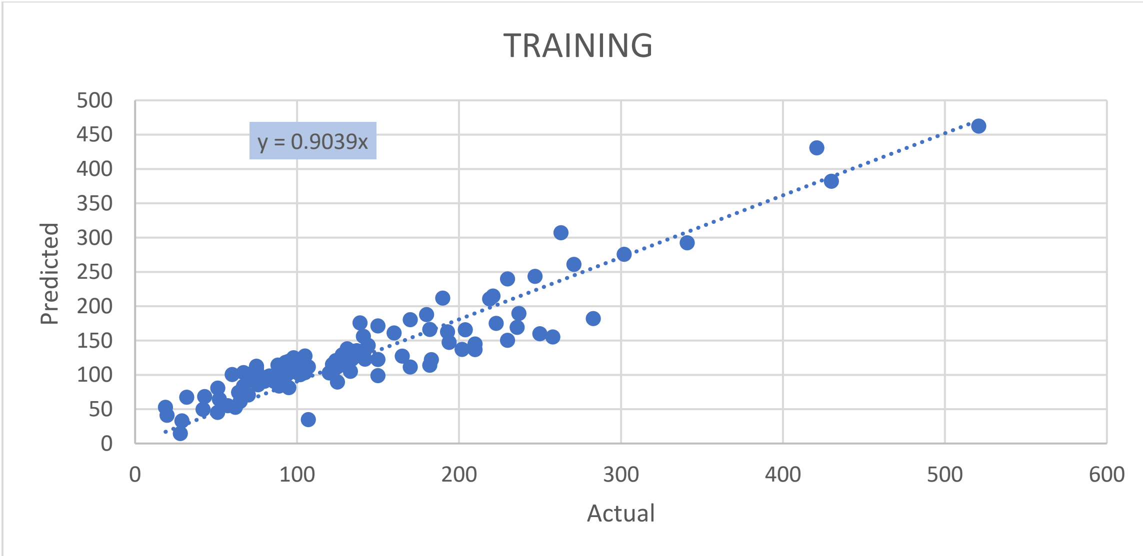


Figure 4.1 Scatter Graph for Training Set

In the following Fig 4.2, the Comparison of actual and predicted results for validation data has done. It was 15% of our total dataset for this model.

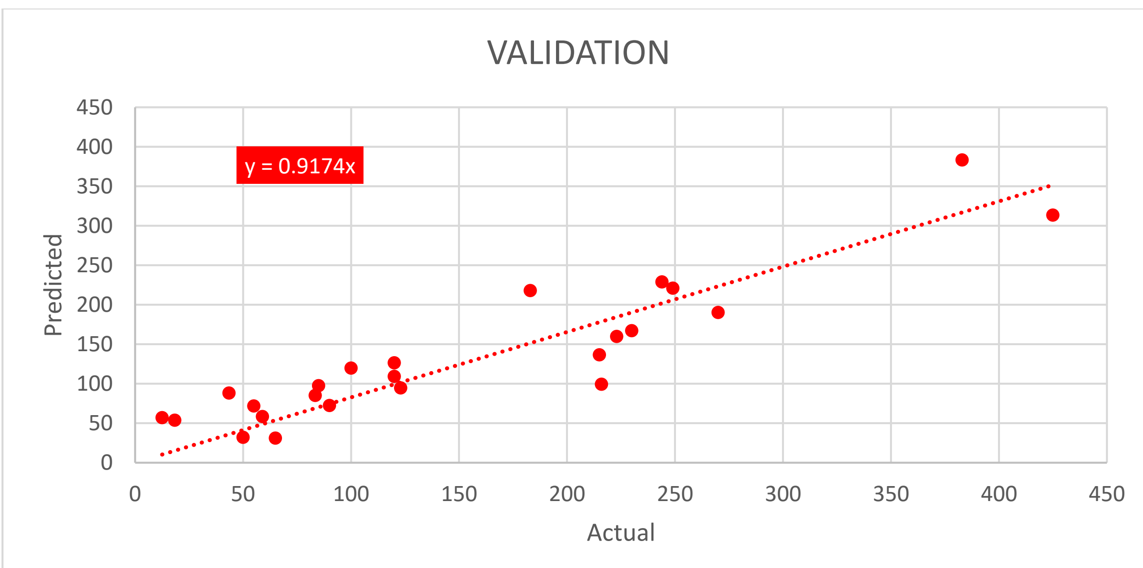


Figure 4.2 Scatter Graph for Validation Set

Now in Fig 4.3, Comparison of actual and predicted results for validation data has done. It was also 15% of our total dataset for this model.

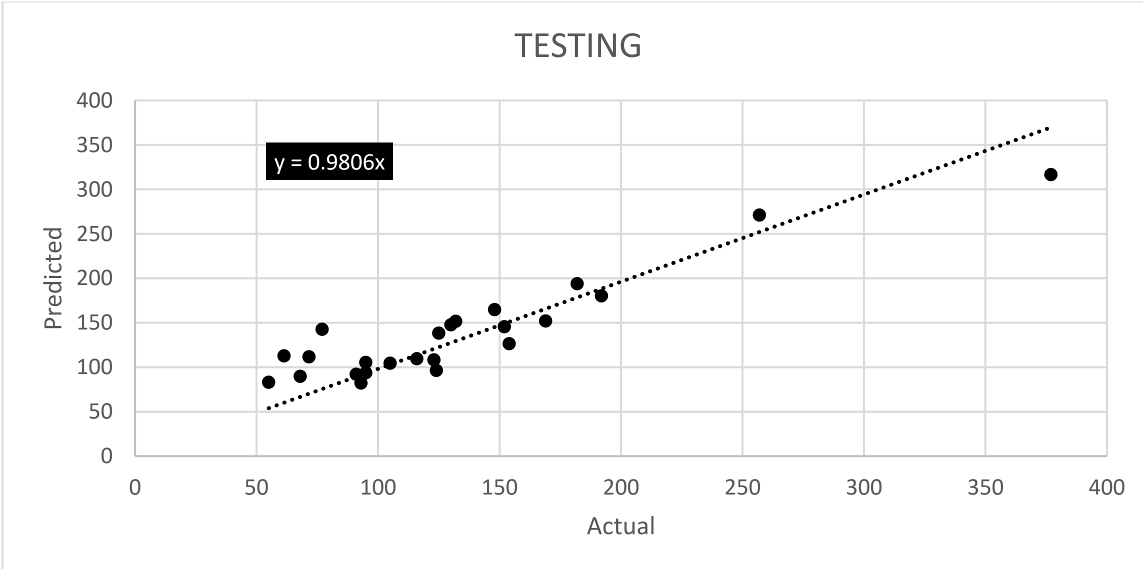


Figure 4.3 Scatter Graph for Testing Set

Now combing all previous divisions of data set, combined scatter plot is made showing all data points for training, validation & testing data as shown in Fig 4.4

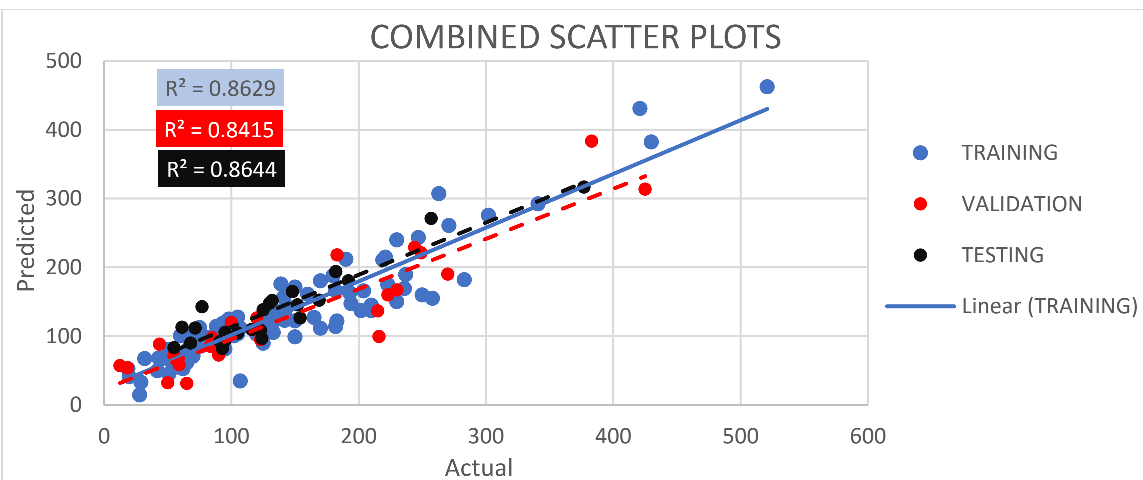


Figure 4.4 Scatter Graph for Combined Results

4.3 Evaluation Through Series Plot

The experimental and the predicted values are plotted on the series plot to have further a visualization of the model. Following is the series plot.

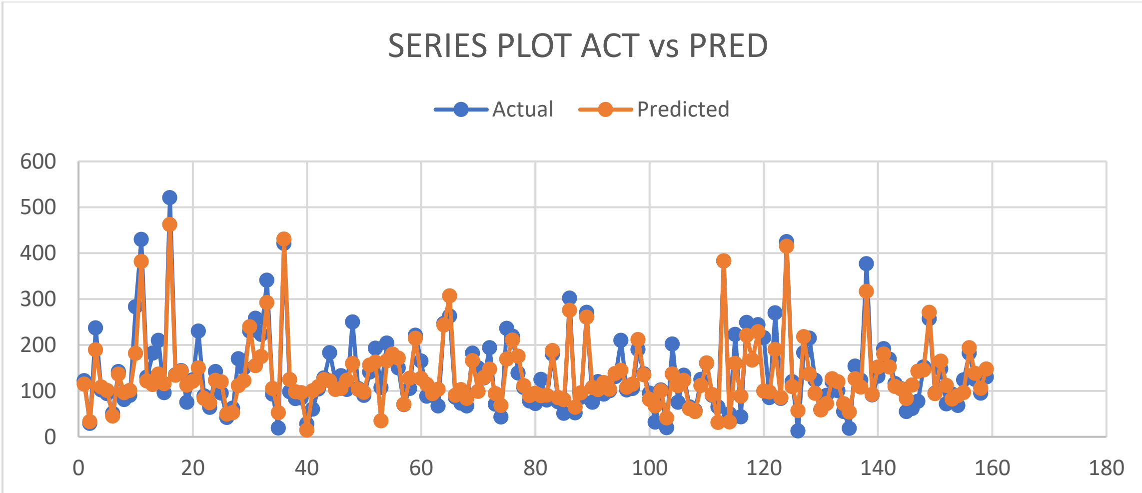


Figure 4.5 Series Plot Actual vs. Predicted

It can be clearly seen that the actual and the predicted values are very close together and the predictive capability of our model is very efficient, in Fig 4.5.

4.4 Evaluation Through Fitness Functions

Table 4.1 Statistical Evaluation Metrics

Data	RMSE	MAE	RSE	R^2	R	P	OF
TRAINING	35.85713	24.944	0.16269	0.8629	0.928924	0.132391	0.110731
VALIDATION	32.47242	25.54723	0.217598	0.8415	0.917333	0.131484	
TESTING	34.94825	27.3584	0.136267	0.8644	0.929731	0.126839	

Fitness functions (as mentioned in Table 4.1) are the statistical metrics that define the goodness of fit of any model. They include.

- Gain or reward functions
- Loss functions
- Performance indicators

For the evaluation of models, it is necessary that the models are evaluated on the bases of all three kinds of metrics i.e., loss functions, reward functions and the performance indicators. Lower the loss function value and greater the gain function value the better is the model. Apart from this the evaluation based on the performance indicators is done according to the set standard limits of the literature. All this has been done step wise in the coming sections.

4.4.1 Based on Gain functions

The two most common gain functions used for the evaluation of predictive models are.

4.4.1.1 Coefficient of Correlation (R)

This quantifies the linear dependence between the inputs and the outputs. (Nguyen et al., 2019) A Pearson coefficient of correlation greater than 0.8 means a strong correlation between predicted and experimental values.(Gandomi, Alavi, & Yun, 2011; Gandomi & Roke, 2015) As shown in the statistical evaluation metrics Table 4.1, the "R"s value of our model is greater than 0.8 and goes above 0.9 which shows that the model has an excellent correlation between the predicted and the experimental values. However, this is not the only metric to be considered since it assumes linear relationships between the quantities and thus it must be used in combination with other metrics.

4.4.1.2 Coefficient of Determination (R squared)

Another metric used is the coefficient of determination or R squared. This metric shows the ratio between the difference of the squared residuals of the best fitted line and the squared residuals of the mean line to the squared residuals of the mean line. This gives us an indication of how well our model is fitted on the data w.r.t. the mean line. Its value ranges from 0 to 1 and the closer it is to one the better fitted is the model. The value above 0.8 is considered a very good coefficient of determination. The values obtained for our model are.

- R training: 0.863
- R validation: 0.8415
- R testing: 0.864

This value is a very big improvement to the model since the previous model developed using Gene Expression Programming had low values of R squared i.e., 0.81 for training and 0.78 for testing. This shows the robustness of MEP over GEP. Moreover, it must be noted that the values of R squared for training and testing of our MEP model unlike the previous GEP model are almost equal with a difference of only 0.001 which shows that the model is very generalized and performs equally well on unknown (external) data. This in turn means that the problem of overfitting has been dealt with properly. This is a great improvement of the MEP model over the previous GEP model.

4.4.2 Based on Loss functions

The loss functions used for the evaluation of the model are.

4.4.2.1 Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Root Squared Error (RSE)

The value of RMSE, MAE and RSE of our model for all three datasets are low and very close for all the sets which is an indicator of high accuracy and generalization capacity. This is also shown by the histogram given below (in Fig 4.6). The histogram shows that the frequency of small errors is more than the larger errors and the larger errors are very few from among the residuals which shows the reliability of the model. Moreover it

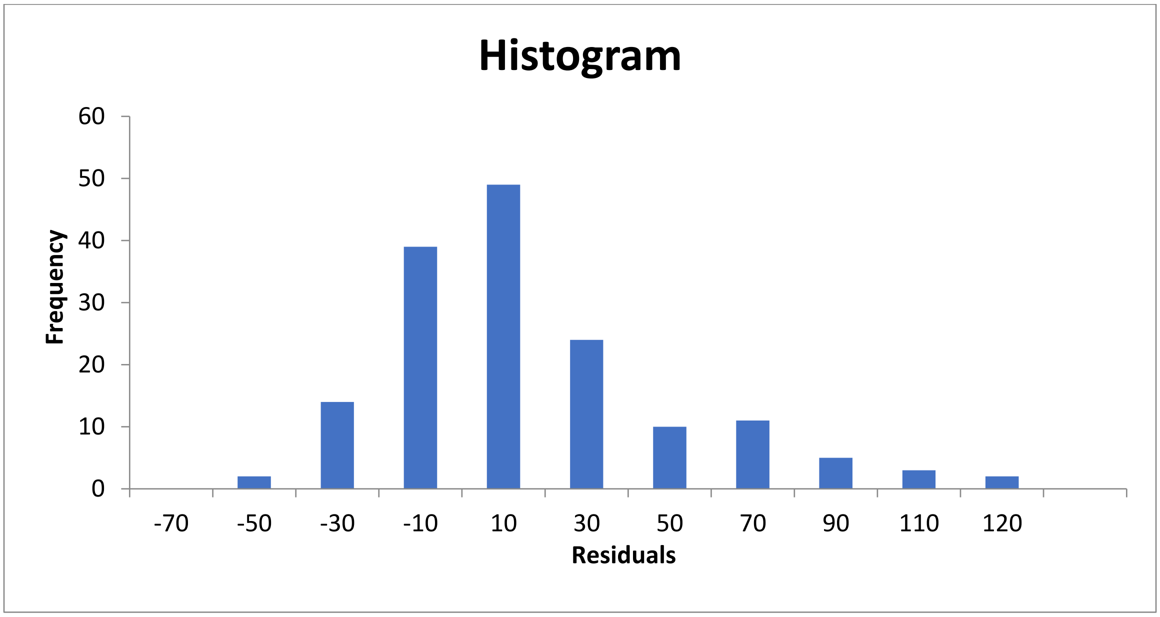


Figure 4.6 Graph for Histogram

4.4.3 Based on Performance Indicators

Two performance indicators have been used to analyse the model i.e.

- Performance coefficient
- Objective function

4.4.3.1 Based On Performance Coefficient

The value of performance coefficient shows the reliability of the model and it takes into account both the Relative root mean squared error and the coefficient of correlation thus giving us a comparison to get an idea of how much the model is reliable. The value of performance coefficient ranges from 0 to infinity and a model with a value lesser than 0.2

is considered a good one.(Gandomi & Roke, 2015) Our model's values of performance coefficient are significantly lesser than 2 thus our model is a very good and reliable one.

4.4.3.2 Based On Objective Function

Objective function is another performance indicator and the speciality of the objective function is that rather than taking into account each dataset separately it takes the whole database and gives us a visualization of the performance by taking into account relative root mean squared error (RRMSE) and the correlation coefficient along with the relative percentage of the training and the testing datasets. This gives the overall performance of the model.

A good model has the value of objective function lesser than 0.2.(Gandomi & Roke, 2015) our objective function value is significantly lower than 0.2 i.e. 0.11 thus this value validates our overall model performance rather than the performance of each set. The value lesser than 0.2 tells that the issue of overfitting of the model has been eliminated effectively.

4.5 Parametric Analysis

Parametric analysis of the model was carried out to check the sensitivity of the input parameters with the output. The graphs through Fig 4.7 - Fig 4.12 shows the trend between the inputs and the outputs. The parametric analysis of the model was created by taking the mean of all the parameter inputs and then keeping all the inputs equal to the mean and except one input at a time and then drawing the graphs between the obtained output versus the inputs. The resulting graphs are given below from Fig 4.7 - Fig 4.12. the trends shown by the graphs ID in accordance with the intuitive relation that the inputs have with the outputs.

As shown in Fig 4.7 CF was having almost linear relation with Sell pressure of soil.

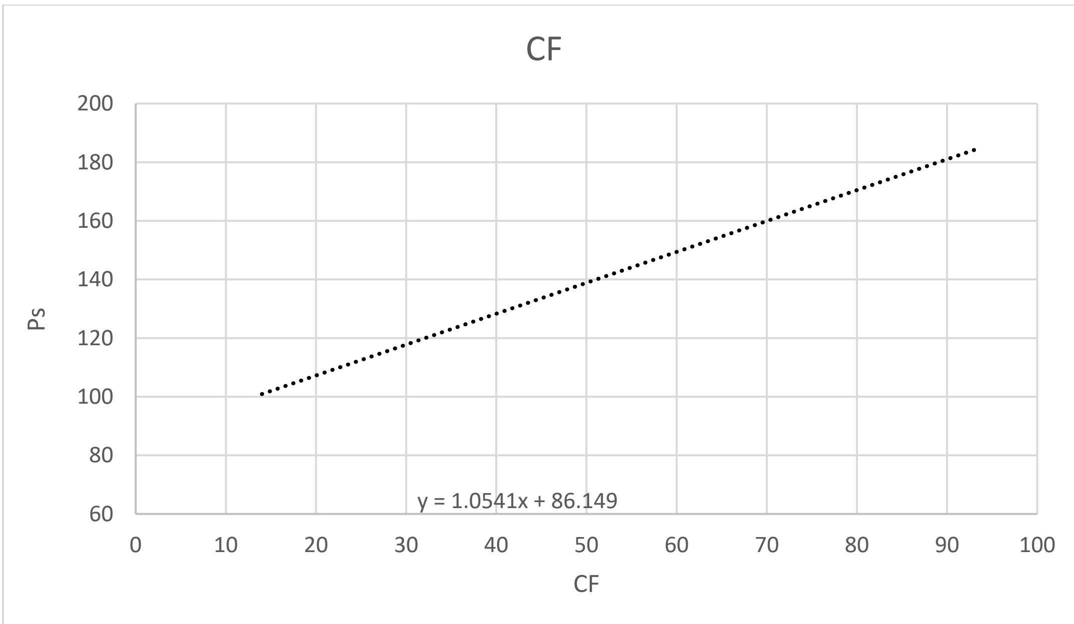


Figure 4.7 Graph for Variation in Ps with change in CF

As shown in following Fig 4.8 relation of PI is given with Sell pressure of soil. A quadratic function or a parabolic curve can be seen. Quadratic relation is also given with PI on X-Axis.

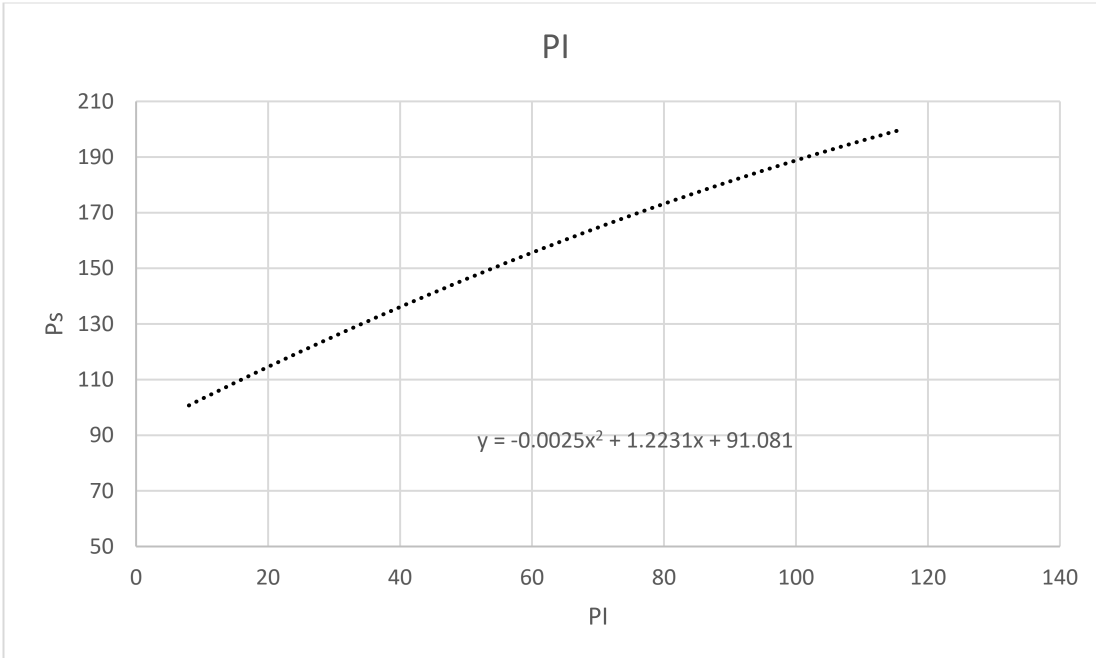


Figure 4.8 Graph for Variation in Ps with change in PI

As shown in following Fig 4.9 relation of MDD is given with Sell pressure of soil. In equation of Ps the parameter B is a sinusoidal function of MDD i.e., B=Sin (MDD) that can be observed in graph

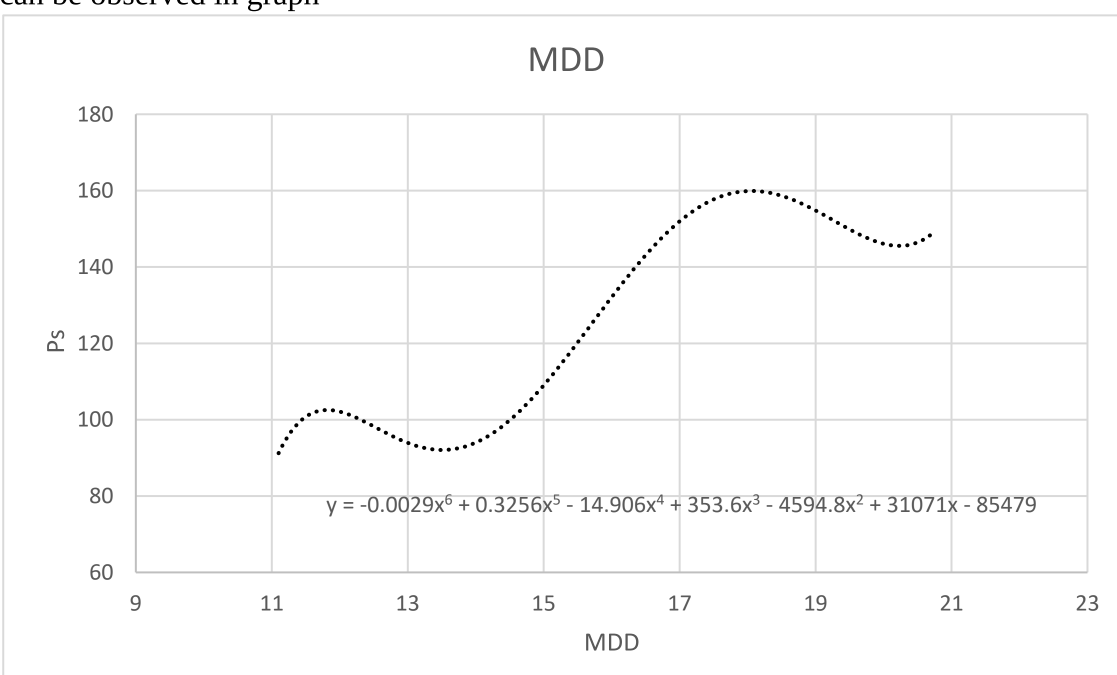


Figure 4.9 Graph for Variation in Ps with change in MDD

In the following Fig 4.10 relation of PI is given with OMC of soil. A cubic 3rd degree function can be seen, with equation given.

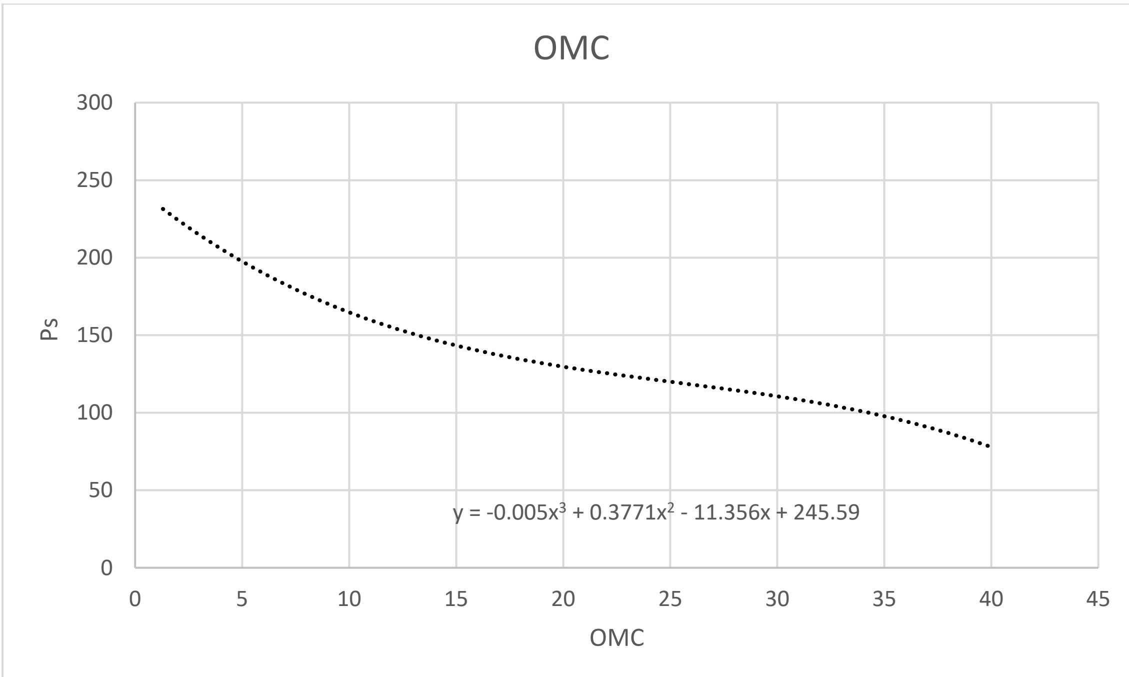


Figure 4.10 Graph for Variation in Ps with change in OMC

As shown in Fig 4.11 SP was having almost linear relation with Swell potential of soil. Greater the potential of swell in the soil, greater will be the swell pressure exerted by soil.

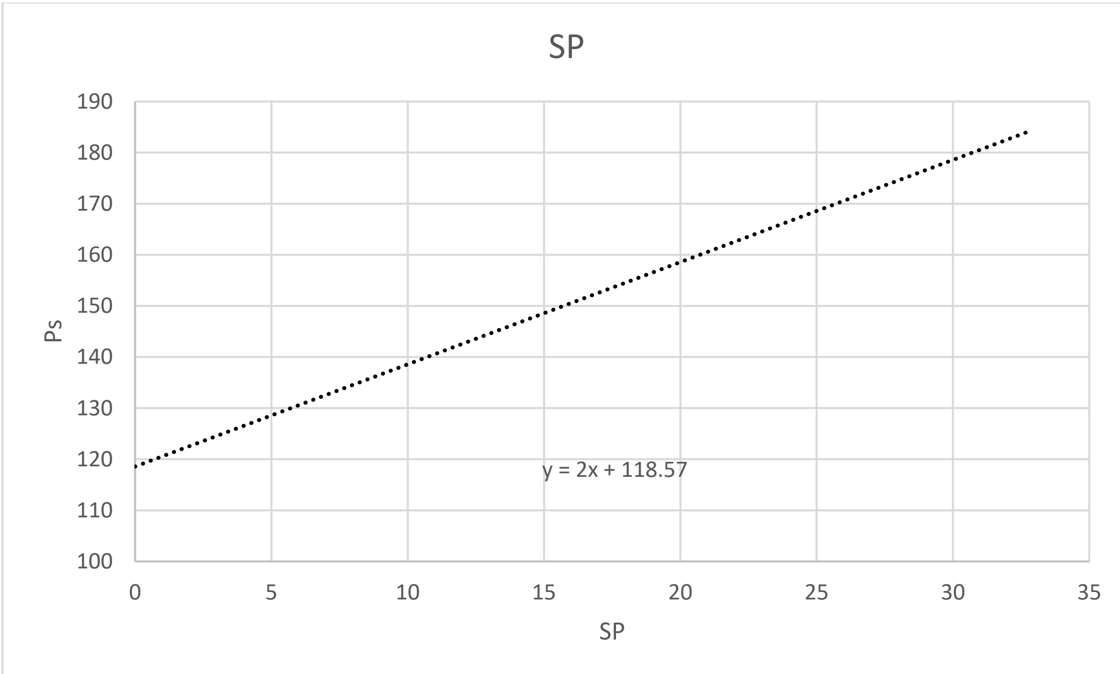


Figure 4.11 Graph for Variation in Ps with change in SP

As shown in Fig 4.12 Ps is having linear relation with Silt percentage in the soil.

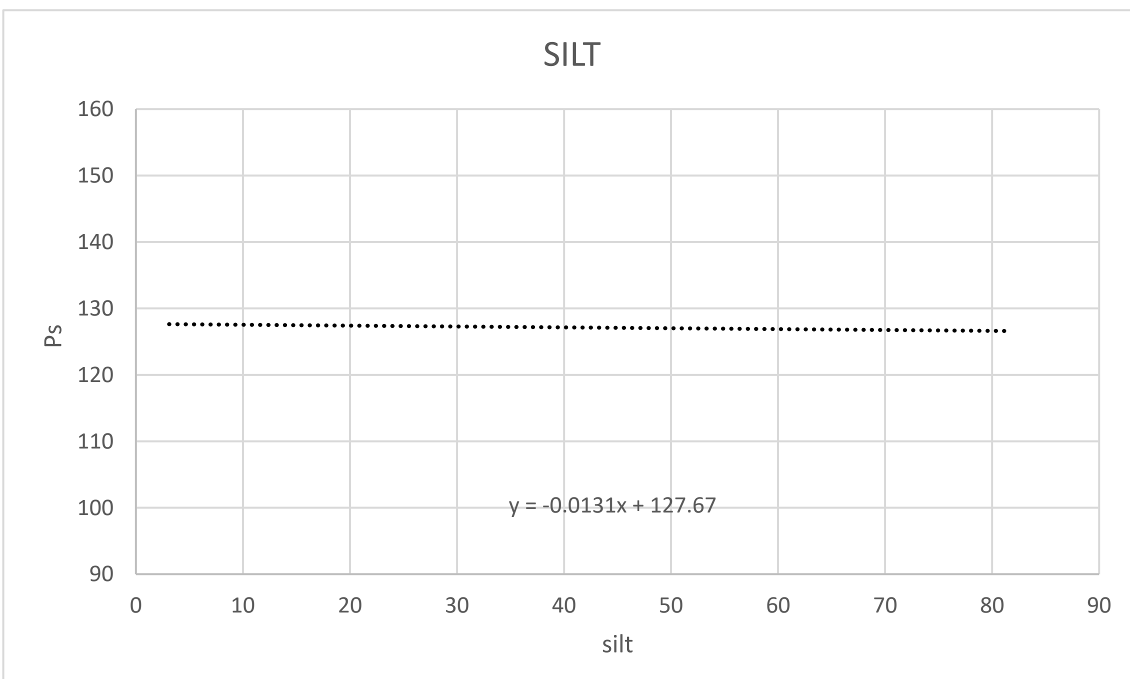


Figure 4.12 Graph for Variation in Ps with change in Silt

4.6 Sensitivity Analysis

Sensitivity analysis was carried out for the model and the relative contribution of each input variable that the model has incorporated has been formulated the results of the sensitivity analysis are as follows in Table 4.2.

Table 4.2 Sensitivity percentages of the input variables

Parameters	Percent sensitivity
CF	14.59
PI	27.59
MDD	12.59
OMC	28.27
SP	10.40
Silt	6.55
Total percentage	100.00

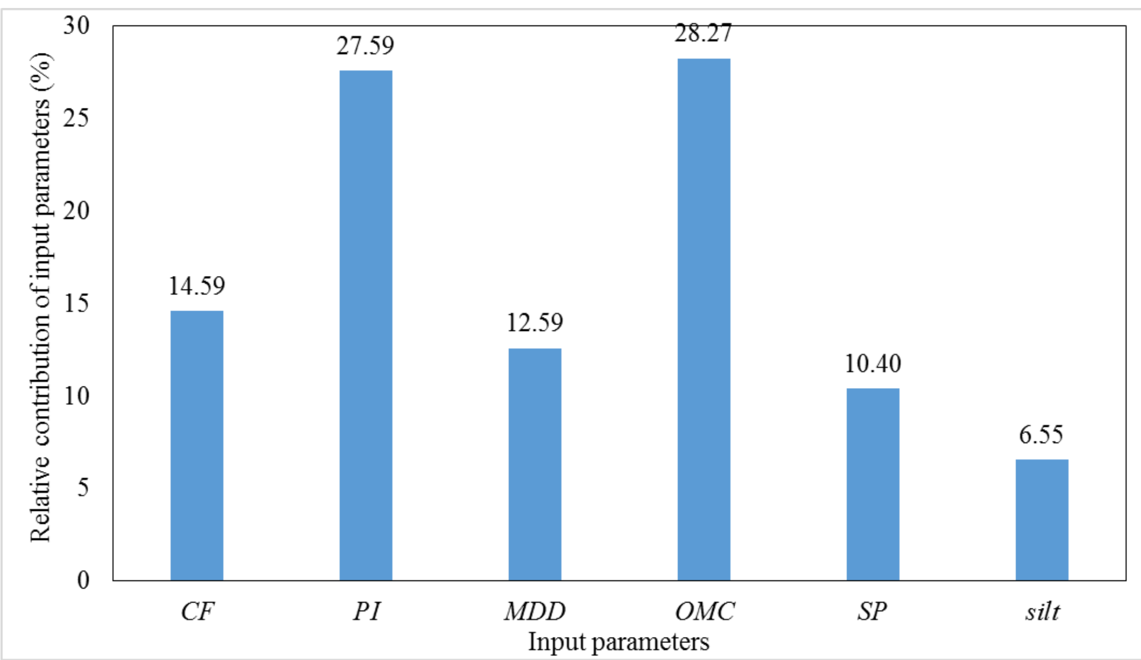


Figure 4.13 Relative contribution of the input parameters (results of sensitivity analysis)

This Graph (Fig 4.13) shows the Results of sensitivity analysis, in form of Relative contribution of the input parameters. The results in the increasing order of the contribution of each input parameter is in the order of OMC (28.27) > PI (27.59) > CF (14.59) > MDD (12.59) > SP (10.40) > Silt (6.55).

Model with output of Ps and the relative contribution of CF parameter is given in this Fig 4.14 following an accuracy of 0.95.

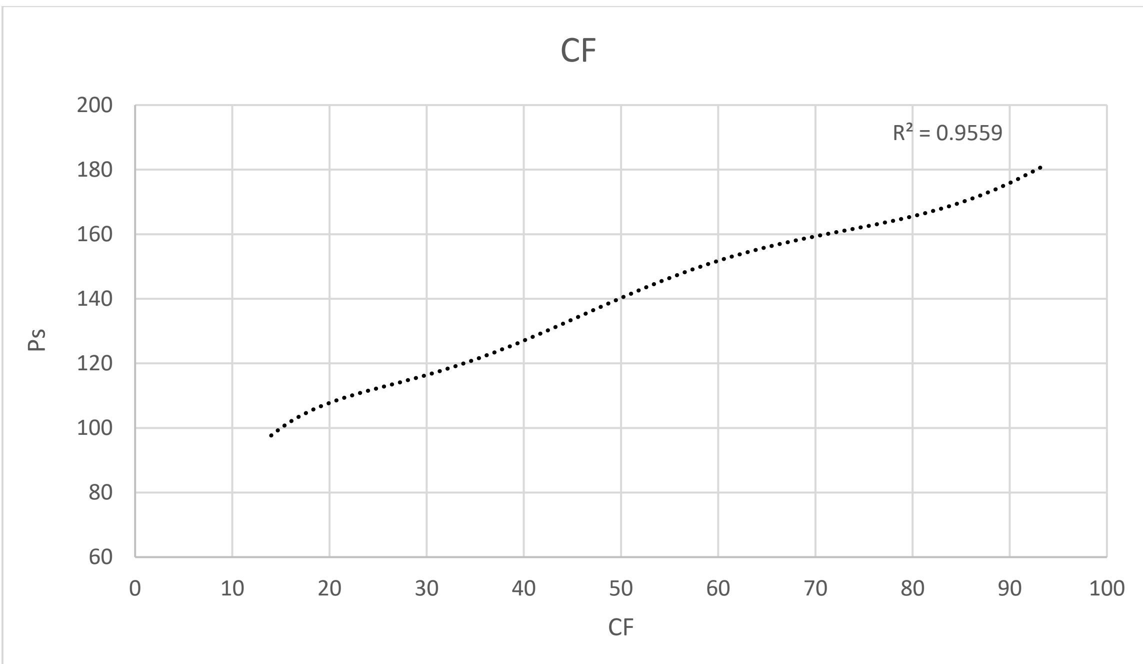


Figure 4.14 Relative contribution of CF on Ps

As shown in following Fig 4.15 sensitivity analysis of PI is given with Sell pressure of soil. For this relative contribution value of R is 0.9.

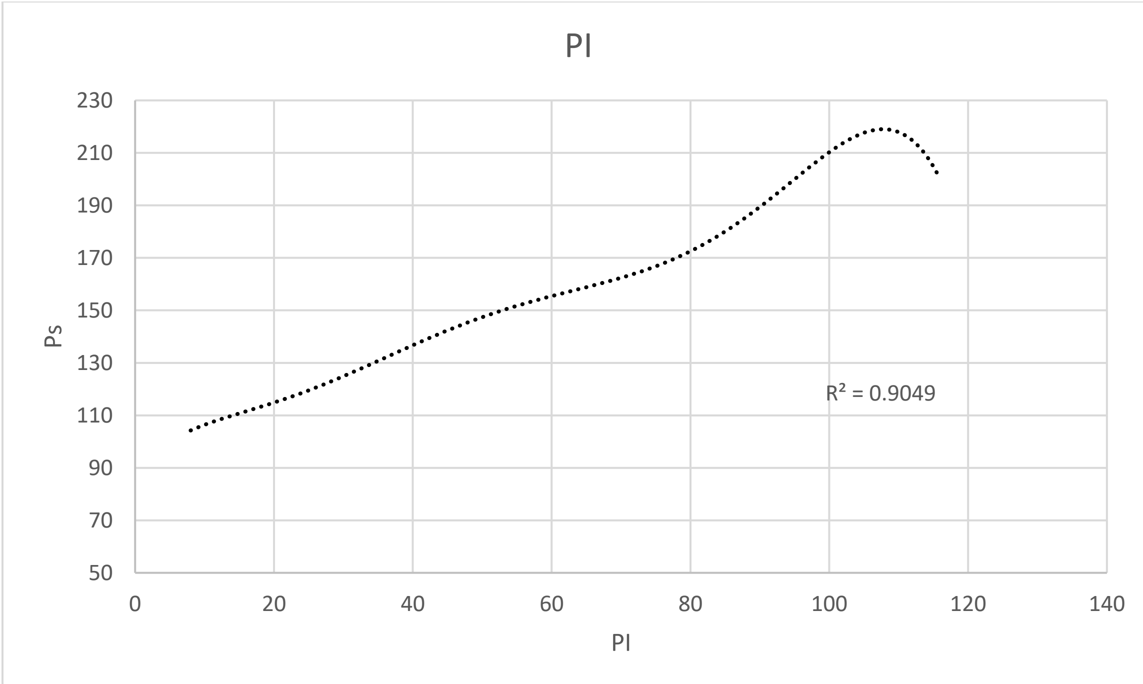


Figure 4.15 Relative contribution of PI on Ps

As shown in following Fig 4.16 sensitivity analysis of MDD is given with Sell pressure of soil. For this relative contribution value of R is 0.96. This analysis is created to understand the impact range of variables on outcome Ps.

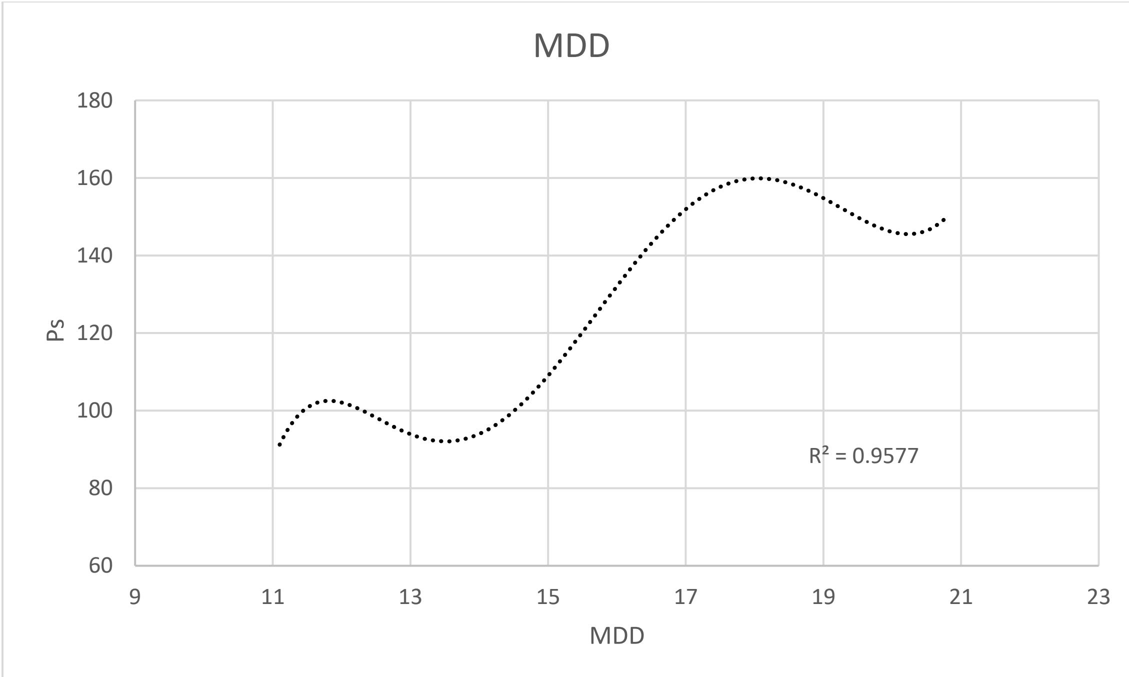


Figure 4.16 Relative contribution of MDD on Ps

Analysis are done to understand the impact of a range of variables on a given outcome.

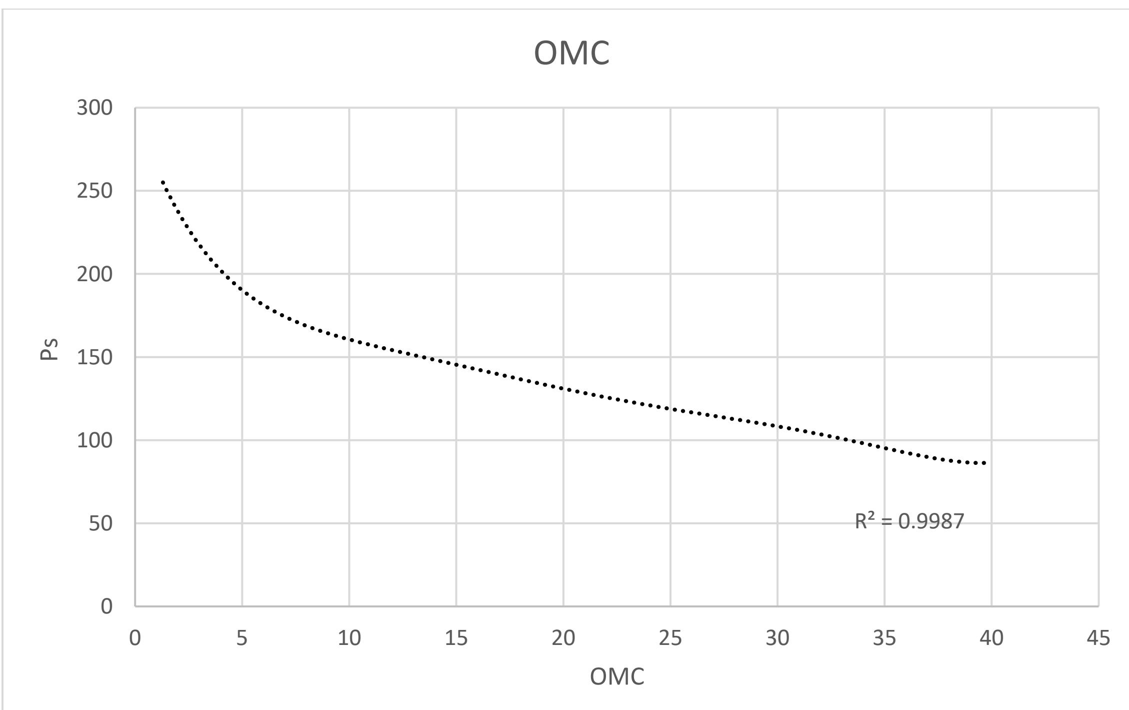


Figure 4.17 Relative contribution of OMC on Ps

As shown in following Fig 4.18 sensitivity analysis of Swell potential is given with Sell pressure of soil. This analysis is created to understand the impact range of variables on outcome Ps. For this relative contribution value of R is 1 that shows strong impact.

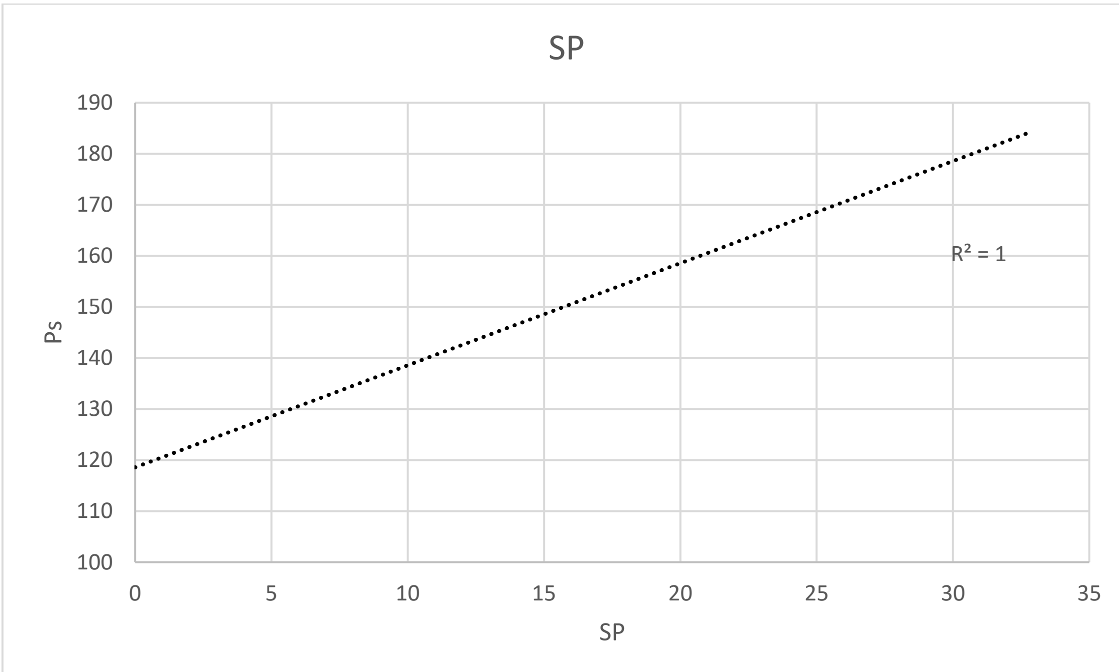


Figure 4.18   Relative contribution of SP on Ps

In upcoming Fig 4.19 it shows relative contribution of silt on output Swell pressure of soil. For this relative contribution less value of R shows less weak impact.

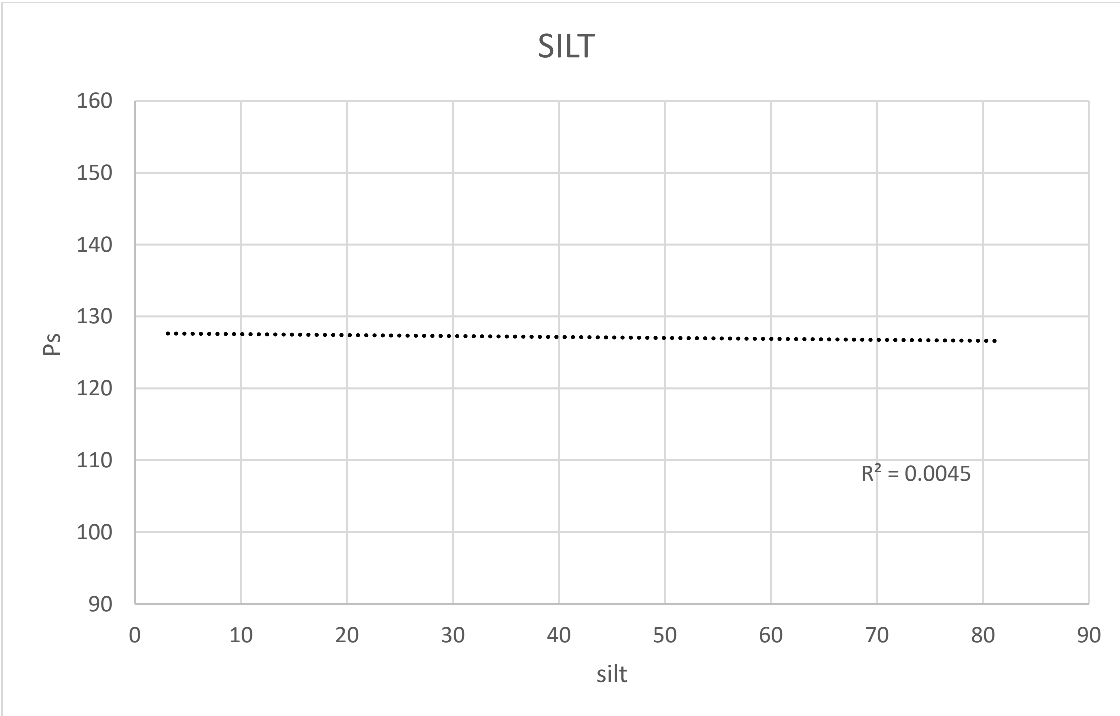


Figure 4.14 Relative contribution of Silt on Ps

These analysis were carried out for the model and the relative contribution of each input variable that the model has incorporated has been formulated the results of the sensitivity analysis & used as an indicator.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The type of artificial intelligence (AI) technique which we used to determine the swell pressure (Ps) of environmentally vulnerable expansive soils is Multi expression programming (MEP). Two databases were created based on an exhaustive review of 145 globally published research publications, with 168 values for Ps.

When the influence of input factors on the swell pressure (Ps) was investigated, it was determined that the Ps fluctuates linearly and increases with MDD, CF, SP, and PI but decreases with OMC and Gs.

Formulated models consist of MEP predict the Ps with great accuracy and without any previous assumptions. Additionally, MEP's prediction of swell pressure features is superior to those of ANN. The MEP technique simplifies the derivation of the swell characteristics while maintaining a reasonable degree of agreement between simulated and experimental data. This demonstrates the versatility of the MEP technique, since it handles both linear and nonlinear data.

Numerous stages, including data processing and division, model simplification, sensitivity analysis, and a parametric investigation, are necessary to avoid over-fitting the respective model generated using MEP and have been covered in length in the work. According to the sensitivity analysis, the increasing order of input significance for Ps was as follows: OMC (28.27) > PI (27.59) > CF (14.59) > MDD (12.59) > SP (10.40) > silt (6.55).

Additionally, a parametric analysis was conducted, and the resulting trends were found to be consistent with previous research findings.

All models are usually evaluated using RMSE, NSE, MAE, RSE, R, RRMSE & In general, the comparison results indicate that MEP is effective and dependable strategies for Ps prediction.

The mathematical equations obtained through MEP are significantly easy and clear than those suggested by ANN and ANFIS. On the other hand, the later strategies have the disadvantages of data overfitting, neural network limitations, and network structural confusion. It is advised that the created MEP model be used in normal design practise. Additionally, the innovative MEP technique evaluates relevant linkages between portrayed physical processes and does not require prior solution, which distinguishes it from others. It should be noted that the suggested mathematical expressions from the MEP methodology may usually be estimated within the data range of input parameters utilised for formulation. As more data points become available, these mathematical equations based on MEP may be improved to find the swell pressure across a bigger range.

The existing models may be used successfully for future applications to estimate the expansive soil swell pressure (Ps) utilising basic geotechnical indicators that are efficient, timesaving, economically viable and dependable in dealing with sensitive expansive clay problems. Consequently, it can lead to the sustainable building of structures sitting on or in expansive soils, resulting in less energy utilization and lower construction costs, i.e., the sustainable use of environmental resources.

The swell pressure is modelled by utilizing the distinguishing features of innovative artificial intelligence technique i.e., MEP. The empirical formulations provided are based on experimental data from the literature. The generated models produce findings that are in good agreement with the experimental results and function similarly well on unseen data. The created models' accuracy and dependability were evaluated using several performance indicators such as RMSE, RSE, R, MAE, and RRMSE. Furthermore, the evaluation of OF and ρ revealed that the constructed models are highly generalizable, and the issue of overfitting has been properly handled. The R value is 0.863 for training, 0.8415 for validation and 0.864 for testing.

5.2 Recommendations

Finally, based on the present research findings, this is vital to note that AI approaches are pretty strong and useful instruments for solving issues with complicated processes, notably in geo-environmental engineering. Simple mathematical formulas can be intelligently generalised to previously unknown facts. Furthermore, it is recommended that the results of this study be validated with more recent data, and that other AI methods be studied, such as Support vector machines (SVMs), Ensemble Random Forest (RF) regression, and Gradient boosted (GB) trees, among others. Majority of soft computing approaches continue to face criticism due to fundamental flaws such as model interpretability, knowledge extraction, and model uncertainty. For good understanding of the learning process, extra emphasis must be made to gaining previous information about the hidden physical process by engineering judgement along with strong basics of applied statistics.

– – –

REFRENCES

Abunama, T., Othman, F., Ansari, M., & El-Shafie, A. (2019). Leachate generation rate modeling using artificial intelligence algorithms aided by input optimization method for an MSW landfill. *Environmental Science and Pollution Research*, *26*(4), 3368-3381. https://link.springer.com/article/10.1007/s11356-018-3749-5

Akan, R., & Keskin, S. N. (2019). The effect of data size of ANFIS and MLR models on prediction of unconfined compression strength of clayey soils. *SN Applied Sciences*, *1*(8), 1-11. https://doi.org/https://doi.org/10.1007/s42452-019-0883-8

Alade, I. O., Bagudu, A., Oyehan, T. A., Abd Rahman, M. A., Saleh, T. A., & Olatunji, S. O. (2018). Estimating the refractive index of oxygenated and deoxygenated hemoglobin using genetic algorithm–support vector regression model. *Computer methods and programs in biomedicine*, *163*, 135-142.

Alavi, A. H., & Gandomi, A. H. (2011). A robust data mining approach for formulation of geotechnical engineering systems. *Engineering Computations*.

Alavi, A. H., Gandomi, A. H., Sahab, M. G., & Gandomi, M. (2010). Multi expression programming: a new approach to formulation of soil classification. *Engineering with computers*, *26*(2), 111-118.

Alavi, A. H., Mollahasani, A., Gandomi, A. H., & Bazaz, J. B. (2012). Formulation of secant and reloading soil deformation moduli using multi expression programming. *Engineering Computations*.

Ashayeri, I., & Yasrebi, S. (2009). Free-swell and swelling pressure of unsaturated compacted clays; experiments and neural networks modeling. *Geotechnical and Geological Engineering*, *27*(1), 137.

ASTM, D. (1996). 4546 Standard Test Methods for One-Dimensional Swell or Settlement Potential of Cohesive Soils. *ASTM International, West Conshohocken, PA, USA*.

ASTM, D. (2014). 4546-14: Standard test methods for one-dimensional swell or collapse of soils. *West Conshohocken, PA, USA*.

Atemimi, Y. K. (2020). Effect of the Grain Size of Sand on Expansive Soil. Key Engineering Materials,

Azim, I., Yang, J., Javed, M. F., Iqbal, M. F., Mahmood, Z., Wang, F., & Liu, Q.-f. (2020). Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming. Structures, Berrah, Y., Boumezbeur, A., Charef, N., & Brahmi, S. (2020). Design of Experiments (DOE) techniques to Predict Swelling Pressure of Expansive Soils In Tebessa (Algeria). In *CIGOS 2019, Innovation for Sustainable Infrastructure* (pp. 823-828). Springer.

Çanakcı, H., Baykasoğlu, A., & Güllü, H. (2009). Prediction of compressive and tensile strength of Gaziantep basalts via neural networks and gene expression programming. *Neural Computing and Applications*, *18*(8), 1031-1041. https://doi.org/https://doi.org/10.1007/s00521-008-0208-0

Chen, X., Shi, X., Zhou, J., Du, X., Chen, Q., & Qiu, X. (2019). Effect of overflow tailings properties on cemented paste backfill. *Journal of environmental management*, *235*, 133-144.

Cherif, M. M., Amal, M., & Ramdane, B. (2018). Effect of swelling mineral on geotechnical characteristics of clay soil. MATEC Web of Conferences,

Das, S. K. (2013). 10 Artificial neural networks in geotechnical engineering: modeling and application issues. *Metaheuristics in Water Geotech Transp Eng*, *45*, 231-267.

Das, S. K., Samui, P., & Sabat, A. K. (2011). Application of artificial intelligence to maximum dry density and unconfined compressive strength of cement stabilized soil. *Geotechnical and Geological Engineering*, *29*(3), 329-342.

Das, S. K., Samui, P., Sabat, A. K., & Sitharam, T. (2010). Prediction of swelling pressure of soil using artificial intelligence techniques. *Environmental Earth Sciences*, *61*(2), 393-403.

Elbadry, H. (2017). Simplified reliable prediction method for determining the volume change of expansive soils based on simply physical tests. *HBRC journal*, *13*(3), 353-360. https://doi.org/https://doi.org/10.1016/j.hbrcj.2015.10.001

Erzin, Y. (2007). Artificial neural networks approach for swell pressure versus soil suction behaviour. *Canadian Geotechnical Journal*, *44*(10), 1215-1223.

Erzin, Y., & Erol, O. (2004). Correlations for quick prediction of swell pressures. *Electronic Journal of Geotechnical Engineering*, *9*(1).

Erzin, Y., & Erol, O. (2007). Swell pressure prediction by suction methods. *Engineering Geology*, *92*(3-4), 133-145.

Gandomi, A. H., Alavi, A. H., Mirzahosseini, M. R., & Nejad, F. M. (2011). Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. *Journal of materials in civil engineering*, *23*(3), 248-263.

Gandomi, A. H., Alavi, A. H., & Yun, G. J. (2011). Formulation of uplift capacity of suction caissons using multi expression programming. *KSCE Journal of Civil Engineering*, *15*(2), 363-373. https://doi.org/https://doi.org/10.1007/s12205-011-1117-9

Gandomi, A. H., & Roke, D. A. (2015). Assessment of artificial neural network and genetic programming as predictive tools. *Advances in Engineering Software*, *88*, 63-72.

Giustolisi, O., Doglioni, A., Savic, D. A., & Webb, B. (2007). A multi-model approach to analysis of environmental phenomena. *Environmental Modelling & Software*, *22*(5), 674-682.

Hanandeh, S., Ardah, A., & Abu-Farsakh, M. (2020). Using artificial neural network and genetics algorithm to estimate the resilient modulus for stabilized subgrade and propose new empirical formula. *Transportation Geotechnics*, *24*, 100358.

Huat, B. B. K., Prasad, A., Asadi, A., & Kazemian, S. (2014). *Geotechnics of Organic Soils and Peat* (Vol. (1st ed.)). CRC Press. https://doi.org/https://doi.org/10.1201/b15627

Ikeagwuani, C. C., & Nwonu, D. C. (2019). Emerging trends in expansive soil stabilisation: A review. *Journal of Rock Mechanics and Geotechnical Engineering*, *11*(2), 423-440. https://doi.org/https://doi.org/10.1016/j.jrmge.2018.08.013

Iqbal, M. F., Liu, Q.-f., Azim, I., Zhu, X., Yang, J., Javed, M. F., & Rauf, M. (2020). Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. *Journal of hazardous materials*, *384*, 121322.

Jalal, F. E., Xu, Y., Iqbal, M., Javed, M. F., & Jamhiri, B. (2021). Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP. *Journal of environmental management*, *289*, 112420.

James, R., Unnikrishnan, A., & Sharma, R. S. (2013). Statistical regression models for predicting the swelling pressure of compacted expansive soils. *International Journal of Geotechnical Engineering*, *7*(4), 431-437.

Javed, M. F., Amin, M. N., Shah, M. I., Khan, K., Iftikhar, B., Farooq, F., Aslam, F., Alyousef, R., & Alabduljabbar, H. (2020). Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete. *Crystals*, *10*(9), 737.

Kayabali, K., & Demir, S. (2011). Measurement of swelling pressure: direct method versus indirect methods. *Canadian Geotechnical Journal*, *48*(3), 354-364.

Klopp, H. W. (2019). *Effect of Soil ESP and Salinity on Soil Atterberg Limits, Swelling and Hydrologic Properties*. The University of Wisconsin-Madison.

Kumar, R., Bhargava, K., & Choudhury, D. (2017). Correlations of uniaxial compressive strength of rock mass with conventional strength properties through random number generation. *International Journal of Geomechanics*, *17*(2), 06016021.

Kumar, T. A., Raheena, M., Robinson, R., & Thyagaraj, T. (2020). A Rapid Method of Determination of Swell Potential and Swell Pressure of Expansive Soils Using Constant Rate of Strain Apparatus. *Geotechnical Testing Journal*, *43*(6).

Li, H., Wu, A., & Wang, H. (2019). Evaluation of short-term strength development of cemented backfill with varying sulphide contents and the use of additives. *Journal of environmental management*, *239*, 279-286.

Li, H., Zhang, Z., & Zhao, Z.-Z. (2019). Data-mining for processes in chemistry, materials, and engineering. *Processes*, *7*(3), 151.

Maeda, T. (2018). *How to rationally compare the performances of different machine learning models?* (2167-9843).

Mawlood, Y. I., & Hummadi, R. A. (2020). Swelling pressures and size effect correlations of expansive soils. *Journal of the Chinese Institute of Engineers*, *43*(7), 657-665.

Mehr, A. D. (2020). An ensemble genetic programming model for seasonal precipitation forecasting. *SN Applied Sciences*, *2*(11), 1-14.

Mohammadzadeh, D., Bazaz, J. B., & Alavi, A. H. (2014). An evolutionary computational approach for formulation of compression index of fine-grained soils. *Engineering Applications of Artificial Intelligence*, *33*, 58-68.

Mohammadzadeh S, D., Kazemi, S.-F., Mosavi, A., Nasseralshariati, E., & Tah, J. H. (2019). Prediction of compression index of fine-grained soils using a gene expression programming model. *Infrastructures*, *4*(2), 26.

Mousavi, S., Alavi, A., Gandomi, A., Arab Esmaeili, M., & Gandomi, M. (2010). A data mining approach to compressive strength of CFRP-confined concrete cylinders. *Structural engineering & mechanics*, *36*(6), 759.

Mozumder, R. A., & Laskar, A. I. (2015). Prediction of unconfined compressive strength of geopolymer stabilized clayey soil using artificial neural network. *Computers and Geotechnics*, *69*, 291-300.

Naghadehi, M. Z., Samaei, M., Ranjbarnia, M., & Nourani, V. (2018). State-of-the-art predictive modeling of TBM performance in changing geological conditions through gene expression programming. *Measurement*, *126*, 46-57.

Najjar, Y. M., Basheer, I. A., & Mcreynolds, R. (1996). Neural modeling of Kansas soil swelling. *Transportation Research Record*, *1526*(1), 14-19.

Nguyen, T., Kashani, A., Ngo, T., & Bordas, S. (2019). Deep neural network with high-order neuron for the prediction of foamed concrete strength. *Computer-Aided Civil and Infrastructure Engineering*, *34*(4), 316-332.

Ozbek, A., Unsal, M., & Dikec, A. (2013). Estimating uniaxial compressive strength of rocks using genetic expression programming. *Journal of Rock Mechanics and Geotechnical Engineering*, *5*(4), 325-329.

Papadimitriou, F. (2009). Modelling spatial landscape complexity using the Levenshtein algorithm. *Ecological Informatics*, *4*(1), 48-55.

Papadimitriou, F. (2020a). The Probabilistic Basis of Spatial Complexity. In *Spatial Complexity* (pp. 51-61). Springer.

Papadimitriou, F. (2020b). What is Spatial Complexity? In *Spatial Complexity* (pp. 3-18). Springer.

Parastar, F., Hejazi, S. M., Sheikhzadeh, M., & Alirezazadeh, A. (2017). A parametric study on hydraulic conductivity and self-healing properties of geotextile clay liners used in landfills. *Journal of environmental management*, *202*, 29-37. https://doi.org/https://doi.org/10.1016/j.jenvman.2017.07.013

Pastor, J. L., Tomás, R., Cano, M., Riquelme, A., & Gutiérrez, E. (2019). Evaluation of the improvement effect of limestone powder waste in the stabilization of swelling clayey soil. *Sustainability*, *11*(3), 679.

Petry, T., Sheen, J.-S., & Armstrong, J. (1992). Effects of pre-test stress environments on swell. International conference on expansive soils,

Pham, B. T., Bui, D. T., Dholakia, M., Prakash, I., & Pham, H. V. (2016). A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. *Geotechnical and Geological Engineering*, *34*(6), 1807-1824. https://doi.org/https://doi.org/10.1007/s10706-016-9990-0

Puppala, A. J., & Pedarla, A. (2017). Innovative ground improvement techniques for expansive soils. *Innovative Infrastructure Solutions*, *2*(1), 1-15.

Puppala, A. J., & Pedarla, A. (2017). Innovative ground improvement techniques for expansive soils. *Innovative Infrastructure Solutions*, *2*(1), 24. https://doi.org/10.1007/s41062-017-0079-2

Pyo, J., Hong, S. M., Kwon, Y. S., Kim, M. S., & Cho, K. H. (2020). Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. *Science of The Total Environment*, *741*, 140162.

Qiu, R., Wang, Y., Wang, D., Qiu, W., Wu, J., & Tao, Y. (2020). Water temperature forecasting based on modified artificial neural network methods: Two cases of the Yangtze River. *Science of The Total Environment*, *737*, 139729.

Salahudeen, A., Sadeeq, J., Badamasi, A., & Onyelowe, K. (2020). PREDICTION OF UNCONFINED COMPRESSIVE STRENGTH OF TREATED EXPANSIVE CLAY USING BACK-PROPAGATION ARTIFICIAL NEURAL NETWORKS. *Nigerian Journal of Engineering Vol*, *27*(1).

Saputra, N. A., & Putra, R. (2020). The Correlation Between CBR (California Bearing Ratio) and UCS (Unconfined Compression Strength) Laterite Soils in Palangka Raya as Heap Material. IOP Conference Series: Earth and Environmental Science,

Sathyapriya, S., Arumairaj, P., & Ranjini, D. (2017). Prediction of unconfined compressive strength of a stabilised expansive clay soil using ANN and regression analysis (SPSS). *Asian Journal of Research in Social Sciences and Humanities*, *7*(2), 109-123.

Shahin, M. A. (2015). Genetic programming for modelling of geotechnical engineering systems. In *Handbook of Genetic Programming Applications* (pp. 37-57). Springer.

Shahin, M. A., Jaksa, M. B., & Maier, H. R. (2009). Recent Advances and Future Challenges for Artificial Neural Systems in Geotechnical Engineering Applications. *Advances in Artificial Neural Systems*, *2009*, 308239. https://doi.org/10.1155/2009/308239

Sharifi, S., Abrishami, S., & Gandomi, A. H. (2020). Consolidation assessment using multi expression programming. *Applied Soft Computing*, *86*, 105842.

Soleimani, S., Rajaei, S., Jiao, P., Sabz, A., & Soheilinia, S. (2018). New prediction models for unconfined compressive strength of geopolymer stabilized soil using multi-gen genetic programming. *Measurement*, *113*, 99-107.

Sun, W., Hu, P., Lei, F., Zhu, N., & Jiang, Z. (2015). Case study of performance evaluation of ground source heat pump system based on ANN and ANFIS models. *Applied Thermal Engineering*, *87*, 586-594.

Vyas, R., Goel, P., & Tambe, S. S. (2015). Genetic programming applications in chemical sciences and engineering. In *Handbook of genetic programming applications* (pp. 99-140). Springer.

Wang, H.-L., & Yin, Z.-Y. (2020). High performance prediction of soil compaction parameters using multi expression programming. *Engineering Geology*, *276*, 105758.

Xu, Y., & Sun, D. A. (2001). DETERMINATION OF EXPANSIVE SOIL STRENGTH USING A FRACTAL MODEL. *Fractals*, *09*(01), 51-60. https://doi.org/10.1142/S0218348X01000506

Zhang, W., Zhang, R., Wu, C., Goh, A. T. C., Lacasse, S., Liu, Z., & Liu, H. (2020). State-of-the-art review of soft computing applications in underground excavations. *Geoscience Frontiers*, *11*(4), 1095-1106.

Keywords

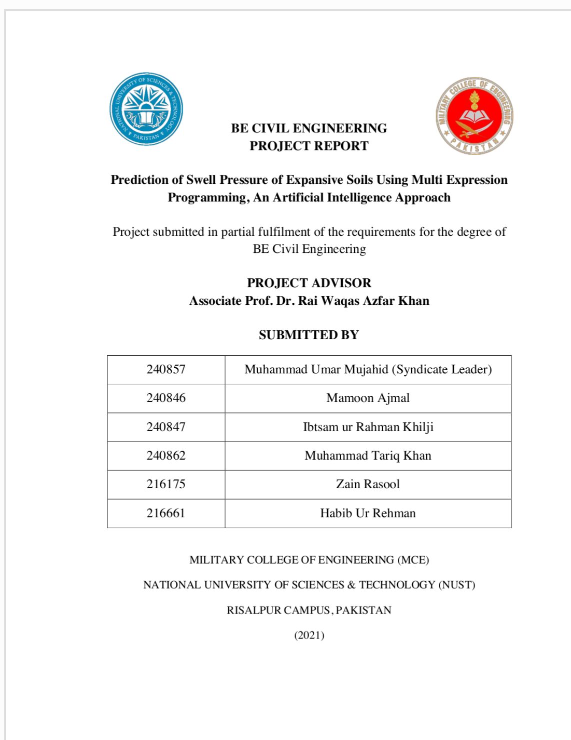
Expansive Soils, Artificial Intelligence Approaches, Swell Pressure, Sensitivity and Parametric Analyses, Multi Expression Programming, Soft Computing Techniques

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Muhammad Umar Mujahid
Assignment title:	Plagiarism detection (June 2021) Part 2 (Moodle TT)
Submission title:	Umar FYP
File name:	97060_Muhammad_Umar_Mujahid_Umar_FYP_715833_84067...
File size:	8.69M
Page count:	75
Word count:	12,345
Character count:	61,016
Submission date:	04-Jul-2021 04:58AM (UTC-0700)
Submission ID:	1600549143



BE CIVIL ENGINEERING
PROJECT REPORT

Prediction of Swell Pressure of Expansive Soils Using Multi Expression Programming, An Artificial Intelligence Approach

Project submitted in partial fulfilment of the requirements for the degree of BE Civil Engineering

PROJECT ADVISOR
Associate Prof. Dr. Rai Waqas Azfar Khan

SUBMITTED BY

240857	Muhammad Umar Mujahid (Syndicate Leader)
240846	Mamoon Ajmal
240847	Ibtsam ur Rahman Khilji
240862	Muhammad Tariq Khan
216175	Zain Rasool
216661	Habib Ur Rehman

MILITARY COLLEGE OF ENGINEERING (MCE)

NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY (NUST)

RISALPUR CAMPUS, PAKISTAN

(2021)

Umar FYP

by Muhammad Umar Mujahid

Umar FYP

ORIGINALITY REPORT

17%	5%	15%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Fazal E. Jalal, Yongfu Xu, Mudassir Iqbal, Muhammad Faisal Javed, Babak Jamhiri. "Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP", Journal of Environmental Management, 2021 Publication	5%
2	Muhammad Farjad Iqbal, Muhammad Faisal Javed, Momina Rauf, Iftikhar Azim, Muhammad Ashraf, Jian Yang, Qing-feng Liu. "Sustainable utilization of foundry waste: Forecasting mechanical properties of foundry sand based concrete using multi-expression programming", Science of The Total Environment, 2021 Publication	4%
3	doctorpenguin.com Internet Source	1%
4	Kamil Kayabali, Saniye Demir. "Measurement of swelling pressure: direct method versus	1%

indirect methods", Canadian Geotechnical Journal, 2011
Publication

5	www.cs.bham.ac.uk Internet Source	1 %
6	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Student Paper	<1 %
7	Rachel James, Avinash Unnikrishnan, Radhey S. Sharma. "Statistical regression models for predicting the swelling pressure of compacted expansive soils", International Journal of Geotechnical Engineering, 2013 Publication	<1 %
8	Engineering Computations, Volume 28, Issue 3 (2011-03-20) Publication	<1 %
9	Muhammad Izhar Shah, Shazim Ali Memon, Muhammad Sohaib Khan Niazi, Muhammad Nasir Amin, Fahid Aslam, Muhammad Faisal Javed. "Machine Learning-Based Modeling with Optimization Algorithm for Predicting Mechanical Properties of Sustainable Concrete", Advances in Civil Engineering, 2021 Publication	<1 %
10	Han-Lin Wang, Zhen-Yu Yin. "High performance prediction of soil compaction	<1 %

	parameters using multi expression programming", Engineering Geology, 2020 Publication	
11	Yusuf Erzin, Orhan Erol. "Swell pressure prediction by suction methods", Engineering Geology, 2007 Publication	<1%
12	Danial Mohammadzadeh S., Jafar Bolouri Bazaz, Amir H. Alavi. "An evolutionary computational approach for formulation of compression index of fine-grained soils", Engineering Applications of Artificial Intelligence, 2014 Publication	<1%
13	Fazal E. Jalal, Yongfu Xu, Mudassir Iqbal, Babak Jamhiri, Muhammad Faisal Javed. "Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms", Transportation Geotechnics, 2021 Publication	<1%
14	link.springer.com Internet Source	<1%
15	trap.ncirl.ie Internet Source	<1%
16	Submitted to Higher Education Commission Pakistan Student Paper	<1%

17	idoc.pub Internet Source	<1 %
18	Erzin, Y.. "Swell pressure prediction by suction methods", Engineering Geology, 20070710 Publication	<1 %
19	asq.org Internet Source	<1 %
20	www.hindawi.com Internet Source	<1 %
21	www.margaretfarren.net Internet Source	<1 %
22	Nasrin Zare-Junaghani, Hamid Mehrnahad, Mehdi Torabi-Kaveh. "Assessing Dispersivity and Expansivity of Clay Soils in the South-East of Yazd with Aim of Investigating Correlation between Them", Periodica Polytechnica Civil Engineering, 2019 Publication	<1 %
23	Submitted to University of Melbourne Student Paper	<1 %
24	Submitted to University of California, Los Angeles Student Paper	<1 %
25	tind-customer-agecon.s3.amazonaws.com Internet Source	<1 %

26	www.ofcoursecme.nl Internet Source	<1 %
27	Submitted to University of Science and Technology Student Paper	<1 %
28	Submitted to VHS Virtual High School Student Paper	<1 %
29	Yusuf Erzin. "Artificial neural networks approach for swell pressure versus soil suction behaviour", Canadian Geotechnical Journal, 2007 Publication	<1 %
30	Amir Hossein Alavi, Amir Hossein Gandomi, Hadi Chahkandi Nejad, Ali Mollahasani, Azadeh Rashed. "Design equations for prediction of pressuremeter soil deformation moduli utilizing expression programming systems", Neural Computing and Applications, 2012 Publication	<1 %
31	Seyyed Mohammad Mousavi, Amir Hossein Alavi, Ali Mollahasani, Amir Hossein Gandomi. "A hybrid computational approach to formulate soil deformation moduli obtained from PLT", Engineering Geology, 2011 Publication	<1 %
	repository.up.ac.za	

32	Internet Source	<1 %
33	epublications.marquette.edu Internet Source	<1 %
34	kt.ijs.si Internet Source	<1 %
35	slidelegend.com Internet Source	<1 %
36	hdl.handle.net Internet Source	<1 %
37	soda.swedishict.se Internet Source	<1 %
38	Hiram Ponce, Paulo V. de Campos Souza, Augusto Junio Guimarães, Guillermo González-Mora. "Stochastic parallel extreme artificial hydrocarbon networks: An implementation for fast and robust supervised machine learning in high-dimensional data", Engineering Applications of Artificial Intelligence, 2020 Publication	<1 %
39	Muhammad Izhar Shah, Muhammad Faisal Javed, Abdulaziz Alqahtani, Ali Aldrees. "Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning	<1 %

models based on consistent big data", Process Safety and Environmental Protection, 2021
Publication

---

40 Rao, Sudhakar. "Identification and classification of expansive soils", Expansive Soils Recent Advances in Characterization and Treatment, 2006.
Publication
<1%

---

41 Unsaturated Soils Research and Applications, 2012.
Publication
<1%

---

42 epubs.surrey.ac.uk
Internet Source
<1%

---

43 opac.wbut.ac.in
Internet Source
<1%

---

44 "Advances in Characterization and Analysis of Expansive Soils and Rocks", Springer Science and Business Media LLC, 2018
Publication
<1%

---

45 "Advances in Emerging Trends and Technologies", Springer Science and Business Media LLC, 2020
Publication
<1%

---

46 L. D. Suits, T. C. Sheahan, Sonal Singhal, Sandra L. Houston, William N. Houston. "Effects of Testing Procedures on the
<1%

Laboratory Determination of Swell Pressure of Expansive Soils", Geotechnical Testing Journal, 2011
Publication

47	eprints.utar.edu.my Internet Source	<1 %
48	Z. Wang, A.J. Sobey, Y. Wang. "Corrosion prediction for bulk carrier via data fusion of survey and experimental measurements", Materials & Design, 2021 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		