UNIVERSITY OF BRASILIA
FACULTY OF TECHNOLOGY
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

MACHINE LEARNING FOR GEOMEMBRANE-SAND INTERFACE ANALYSIS

ABENEZER TEFERA TANGA

SUPERVISOR: GREGÓRIO LUIS SILVA ARAÚJO, DSc.
CO-SUPERVISOR: FRANCISCO EVANGELISTA JUNIOR, PhD

A DISSERTATION SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE IN GEOTECHNICS
PUBLICATION: ---------

BRASILIA / DF: SEPTEMBER/2022

FACULTY OF TECHNOLOGY
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

MACHINE LEARNING FOR GEOMEMBRANE-SAND INTERFACE ANALYSIS

ABENEZER TEFERA TANGA

DISSERTATION OF MSc SUBMITTED TO THE DEPARTMENT OF CIVIL ENGINEERING OF THE UNIVERSITY OF BRASILIA AS PART OF THE NECESSARY REQUIREMENTS TO OBTAIN THE DEGREE OF MASTER OF SCIENCE.

APPROVED BY:

_____

GREGÓRIO LUIS SILVA ARAÚJO, DSc. (UnB)
(SUPERVISOR)

_____

FRANCISCO EVANGELISTA JUNIOR, PhD. (UnB)
(CO-SUPERVISOR)

_____

ENNIO MARQUES PALMEIRA, PhD. (UnB)
(INTERNAL EXAMINER)

_____

KRISHNA RAJIREDDIGARI REDDY, PhD. (UIC)
(EXTERNAL EXAMINER)

_____

BRASÍLIA/DF: SEPTEMBER/2022

CATALOGUING DATA

ABENEZER, TEFERA TANGA

Machine Learning for geomembrane-sand interface analysis. Distrito Federal, 2022.

LXXXIV, 84 p., 210x297 mm (ENC/FT/UnB, Masters of Science, Geotechnics, 2022).

Masters' dissertation – University of Brasilia, Faculty of Technology,

Department of Civil and Environmental Engineering.

1. Random Forest
2. Machine Learning
3. Geomembrane
4. Interface Shear
5. Friction angle

I. ENC/FT/UnB

II. Title (Series)

BIBLIOGRAPHIC REFERENCE

ABENEZER, T.T. (2022). Machine Learning for geomembrane-sand interface analysis. Dissertação do Mestrado, Publicação X.XX-XX/22, Departamento de Engenharia Civil e Ambiental, Universidade de Brasilia, Brasili, DF, 71 p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Abenezer Tefera Tanga

TITULO DA DISSERTAÇÃO DO MESTRADO: Machine Learning for geomembrne-sand interface analysis.

GRAU / ANO: Mestre / 2022.

Abenezer Tefera Tanga

SHIGS QI 7 CONJ.5 CASA 12

Bairro: Lago sul.

CEP: 70297 -400 – Brasília/DF - Brasil

DEDICATION

TO MY FATHER **TEFERA TANGA GURAE**
WHO LEFT THE LEGACY OF WISDOM AND THE
INSPIRATION OF HARD WORK.

ACKNOWLEDEGMENT

I would like to acknowledge and give my warmest thanks to my supervisor Prof. Gregório Luis Silva Araujo who made this work possible. His guidance, advice, devotion, encouragement and friendship carried me through all the stages of writing this dissertation. I would also like to thank my co-supervisor Prof. Francisco Evangelista Junior who has really helped me to walk through a new field of study and achieve the objectives of this project. Without my supervisors this study would not be done.

I would like to thank the current coordinator of the geotechnical engineering department, Prof. Rafael Cerqueira Silva and the previous two coordinators Prof. Michéle Dal Toé Casagrande and Prof. Juan Felix Rodriguez Rebolledo, for their tireless support in my post-graduate academic journey. And I thank the staff from the secretary's office for their continuous assistance.

I thank all the Professors who have moved me one step ahead in every discipline I had the opportunity to attend and for my colleagues who went extra mile for adapting me to my new language and culture experience.

I thank CAPES and the University of Brasilia for the financial support.

ABSTRACT

Evaluation of soil-geosynthetic interaction is important for analyzing the stability of the overall structure. This is because the interaction between the reinforcing geosynthetics and the reinforced soil can be a start for breakage which may cause a structural failure. Several researchers have been studying factors determining the interface shear strength in the laboratory and identified various components which affect the overall strength outcome. Machine Learning has a great potential for the analysis of parameters which are influenced by many variables. This dissertation brings a discussion about the use of Random Forest regression, which is a Machine Learning algorithm, for predicting geomembrane-sand interface friction angle.

The interfaces subjected for strength parameter investigation include geomembrane and cohesionless soil, and 495 interfaces from various literature were collected. The acquired interface data is utilized for the overall statistical and Machine Learning analysis. Fourteen parameters were recorded from the referred literatures as the factors determining the interface shear strength. The fourteen parameters are from three main interface components which are; laboratory test type, geomembrane properties and soil properties.

The presented data has been studied by using simple linear regression before initializing the Random Forest, to evaluate the interdependence between pairs of influencing parameters and their correlation with interface friction angle. The Pearson's correlation coefficient results are indicating the influence level between the interface components is mostly not strong. These correlation values imply the nonlinear distribution of the database and the importance of a multivariate and nonlinear algorithm for studying the referred types of interfaces.

After the data analysis, an inclusive Random Forest has been initialized to predict interface friction angle. It is observed only for 3% of the training set and 6% of the testing set that the friction angle estimation has exceeded ±5° from the laboratory records. The coefficient of determination measures shows strong coherence between friction angles from laboratory studies and Random Forest estimations by resulting $R^2 = 0.93$ and $R^2 = 0.92$; for the training and testing sets respectively. Thus, the Random Forest has forecasted interface friction angle adequately.

RESUMO

A avaliação da interação solo-geossintética é importante para analisar a estabilidade da estrutura geral. Isso ocorre porque a interação entre os geossintéticos de reforço e o solo reforçado pode ser um início de ruptura que pode causar uma falha estrutural. Vários pesquisadores têm estudado os fatores que determinam a resistência ao cisalhamento da interface em laboratório e identificaram vários componentes que afetam o resultado geral da resistência. O Machine Learning tem um grande potencial para a análise de parâmetros que são influenciados por muitas variáveis. Esta dissertação traz uma discussão sobre o uso da regressão Random Forest, que é um algoritmo de Machine Learning, para prever o ângulo de atrito da interface geomembrana-areia.

As interfaces submetidas à investigação de parâmetros de resistência incluem geomembrana e solo sem coesão, e 495 interfaces de várias literaturas foram coletadas. Os dados de interface adquiridos são utilizados para a análise estatística geral e de aprendizado de máquina. Quatorze parâmetros foram registrados da literatura referida como os fatores determinantes da resistência ao cisalhamento da interface. Os quatorze parâmetros são de três componentes de interface principais que são; tipo de teste de laboratório, propriedades da geomembrana e propriedades do solo. Os dados apresentados foram estudados usando regressão linear simples antes de inicializar o Random Forest, para avaliar a interdependência entre pares de parâmetros influenciadores e sua correlação com o ângulo de atrito da interface. Os resultados do coeficiente de correlação de Pearson estão indicando que o nível de influência entre os componentes da interface na maioria das vezes não é forte. Estes valores de correlação implicam a distribuição não linear da base de dados e a importância de um algoritmo multivariado e não linear para estudar os referidos tipos de interfaces.

Após a análise dos dados, uma Random Forest inclusiva foi inicializada para prever o ângulo de atrito da interface. Observa-se apenas para 3% do conjunto de treinamento e 6% do conjunto de teste que a estimativa do ângulo de atrito excedeu ±5° dos registros do laboratório. As medidas do coeficiente de determinação mostram forte coerência entre os ângulos de atrito dos estudos de laboratório e as estimativas da Random Forest resultando $R2 = 0,93$ e $R2 = 0,92$; para os conjuntos de treinamento e teste, respectivamente. Assim, a Random Forest previu o ângulo de atrito da interface adequadamente.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

γ_d: dry unit weight of soil

$\bar{\rho}$: Pearson's correlation coefficient

σ: Standard Deviation

$\Phi_{interface}$: Friction angle of geomembrane-sand interface

ϕ_{soil}: Friction angle of soil

ANN: Artificial Neural Network

ASTM: Standard Test Method for Measuring Asperity Height of Textured Geomembranes

C_c: Coefficient of curvature

CDS: Conventional direct shear test / equipment

CSPE: Chloro-sulphonated polyethylene

C_u: Coefficient of uniformity

D_{50}: Mean particle size

GPR: Gaussian Process Regression

HDPE: High density polyethylene geomembrane

IP: Inclined plane test: Test equipment

MDS: Medium size direct shear test/ equipment

LDPE: Low density polyethylene

Max: Maximum

Min: Minimum

ML: Machine Learning

MLR: Multiple Linear Regression

MSE: Mean Square Error

OOB: Out-of-bag technique

PVC: Polyvinyl chloride

R^2: Coefficient of determination

RFR: Random Forest Regression

RMSE: Root Mean Square Error

1. INTRODUCTION

Soil structures and waste disposal areas can be constructed by single or multilayer of geosynthetics for filtration, drainage containment or reinforcement. These synthetic products can be geomembrane, geotextile, geogrids, geonets, geosynthetic clay liners or geo-composites, among others and the interaction between geosynthetic layers and the reinforced soil is an important section for determining the overall structure stability. Laboratory experiments are the main practice to study geosynthetic interface strength and there are some test types to evaluate the interface shear strength between either geosynthetic-geosynthetic or geosynthetic-soil interface. Direct shear test, medium size direct shear test, inclined plane (ramp) test and pull-out test are the most practiced ones. From several experiment results by different authors interface shear strength can be affected by some critical factors like type of test equipment, applied normal stress, the strain rate, contact area between interface components, type of geosynthetics in the interface, thickness of geosynthetics, asperity height of geosynthetics, relative density of reinforced soil, specific gravity of soil, coefficient of curvature, coefficient of uniformity, D_{50} and friction angle of soil (O'Rourke et al. 1990; Koutsourais et al. 1991; Izgin et al. 1998; Lima 2000; Lopes et al. 2001; Mello 2001; Rebelo 2003; Viana 2007; Frost et al. 2012; Vieira et al. 2013; Bacas et al. 2015; Carbone et al. 2015; Faisal et al. 2015; Amanda et al. 2016; Choudhary et al. 2016; Vangla et al. 2017; Sánchez, 2018; Cen et al. 2018; Adeleke 2020; Lashkari & Jamali 2021; Pavanello, 2021; Araujo et al. 2022; and Pavanello, 2022). Based on the finding of these practical laboratory test results it is possible to build a prediction model by using Machine Learning algorithms. Machine Learning algorithms are set to learn from an available data and trained to make a projection of a certain parameter. Some of these algorithms are: Artificial Neural Network (ANN), Gaussian Process Regression (GPR), Support Vector Machine (SVM), M5 tree model and Random Forest Regression (RFR) and they have been applied in civil and geotechnical engineering parameters estimation and showed an adequate performance. In this study Random Forest regression, a powerful algorithm will be initialized to predict geomembrane-sand interface shear strength.

1.1. OBJECTIVE OF THE STUDY

The general objective of this research is to study the applicability of Random Forest regression, a machine learning algorithm, for predicting geosynthetic interface friction angle. The studied geosynthetic interfaces are made from geomembrane and soil. In order to achieve the general objective, the following specific objectives are outlined:

- Evaluating performance of Random Forest regression for interface shear strength estimation in comparison with laboratory test results;
- Analyzing geomembrane interface data distribution and assessing its influence on the Random Forest algorithm;
- Evaluating a Random Forest application of feature importance ranking with in factors affecting interface shear strength;
- Inspecting the interdependence between geomembrane-sand interface parameters and their influence on the interface friction angle, based on the measured statistical values.

1.2. THESIS OUTLINE

This study is divided in five chapters; the first chapter deals with a general approach of the subject, portrayed how the geotechnical parameter and a machine learning algorithm are linked in this research, as well the general and specific objectives are addressed. The second chapter is the literature review on the subject matters. Both interface shear strength and the Random Forest algorithms are outlined according to referred literatures. Mainly the factors affecting interface shear strength are explained according to several laboratory investigations and some machine learning utilizing studies are referred to observe how artificial intelligence is practiced for civil engineering parameters projection. Chapter three discusses the applied tools and methodologies in the research. In this section data analysis, the tools utilized to make estimation and the mechanisms to evaluate the prediction performance are also explained. In chapter four interpretations from the data analysis and Random Forest projection are presented. The outcomes from the algorithm are compared and contrasted based on the laboratory investigation findings. The fifth and final chapter presents the conclusion obtained from this study and suggestions for future studies.

2. LITERATURE REVIEW

2.1. FACTORS AFFECTING GEOMEMBRANE-SAND INTERFACE

Soil-reinforcement interaction evaluation is of utmost importance for the stability of reinforced soil structures. The same applies to slopes of waste disposal areas where single or multiple layers of soils and geosynthetics are installed for filtration, drainage, containment, or reinforcement. The cover soil on the slope transfers shear stresses to the underlying geosynthetic layers; the interfaces have to resist the mobilized forces and stresses within safe margins (Palmeira et al. 2002). Studies utilizing direct shear, medium size direct shear and inclined plane equipment are the most practiced laboratory experiments. Each geosynthetics interface can be a combination of various materials. An interface may have a direct contact between soil and geosynthetics or can be an interaction between two geosynthetic products. Considered interfaces in this paper includes, geomembrane and cohesionless soil only. Table 2.1 shows three major groups of elements (type of laboratory test, geomembrane properties and soil properties) and fourteen individual strands listed in these main groups, affecting the overall shear strength of a geomembrane interface.

Table 2.1 factors affecting geomembrane interface shear strength.

Factors affecting interface strength		Interface shear parameters
2.1.1 Type of laboratory test	Type of test device	$\Phi_{interface}$
	Applied normal stress	
	Strain rate	
	Contact area	
2.1.2 Geomembrane type and property	Geomembrane thickness	
	Asperity height	
2.1.3 Soil properties	Relative density	
	C_c	
	C_u	
	D_{50}	
	ϕ_{soil}	

2.1.1. TYPE OF LABORATORY TESTS

In the study of geosynthetic reinforced structures, various authors simulate the actual onsite condition in a laboratory, for determining interface shear strength outcome. A safe and economic design of soil reinforcement requires a good understanding of interaction mechanisms that develop between the soil and the reinforcement. From number of standard and modified equipment utilized in various researches, the most common applications are conventional direct shear, medium size direct shear and inclined plane test. As shown in Figure 2.1, for determining the shear strength in a specific section of reinforced structure, it is important to consider appropriate apparatus which can closely simulate the onsite condition.

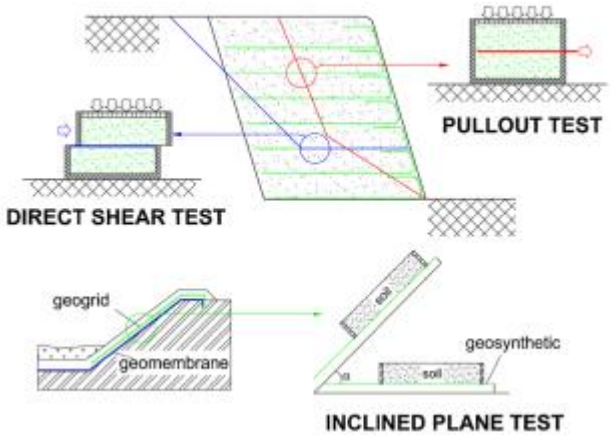


Figure 2.1. Soil–geosynthetic interaction mechanisms and test devices used to simulate the interface behavior (modified from Moraci et al. 2014).

Conventional direct shear (CDS) hardware is known for its simplicity and quick results. Figure 2.2.a shows CDS apparatus (Sánchez 2018). Numerous experimental studies have been conducted to assess the shear strength parameters of geosynthetic interfaces through CDS tests (Izgin & Wasti, 1998; Frost et.al 2011; Sánchez 2018; Lashkari & Jamali, 2021). Difference in the apparatus and boundary conditions can be observed according to each research. Palmeira, (2009) states that there are different procedures adopted to apply the normal stress such as a rigid and free top plate, a rigid top plate not allowed to rotate, a top plate fixed to the top halve of the cell or a flexible pressurized bag. The other difference among test arrangements is related to the

way geosynthetic is fixed between the top and the bottom half. CDS can have a circular or rectangular shear box and can accommodate smaller samples up to 100 cm^2.

(a) (b) (c)

Figure 2.2. Considered laboratory test equipment, (a) CDS apparatus, (b) MDS apparatus, (c) IP apparatus (Sánchez 2018).

Medium size direct shear (MDS) is the other widely utilized test instrument for geosynthetic interface studies. Figure 2.2.b shows MDS apparatus from Sánchez (2018). It is a direct shear test apparatus which allows larger samples and can accommodate a sample size bigger than 100 cm^2, which is an area commonly used in the CDS test. Several experimental studies have been conducted to assess the shear strength parameters of geosynthetic interfaces through MDS tests (Koutsourais et al.1991; Rebelo 2003; Faisal et al. 2015; Bacas et al. 2015; Choudhary and Krishna, 2016; Vangla and Gali 2016; Saeed 2017; Sánchez 2018; Cen et al. 2018) and they have shown that it is a suitable for studying larger contact area under higher applied normal stresses. In Cen et al. (2018) it is indicated the development of a large-scale shear apparatus can provide great convenience and save much cost for accurately and consistently simulating the shear behaviors of geomembrane and sand interfaces.

The third referred tool in this thesis is inclined plane (IP) or ramp test apparatus. IP test instruments are able to produce more realistic onsite conditions for landfills, slopes and canals as shown in Figure 2.2 (Sánchez 2018). Various works on IP test can be found in the literature (Izgin & Wasti, 1998; Lima 2000; Lopes 2001; Mello 2001; Palmeira et al. 2002; Viana 2007; Carbone et al. 2015; Sánchez 2018). IP test basically consists of increasing the inclination of the

ramp until a soil block slides on the geosynthetic layer fixed to the ramp surface. Similar test arrangements can be employed for tests on interfaces between different geosynthetics, such as geotextiles and geomembranes (Palmeira 2009). Coherently applied normal stress is the main feature of most geosynthetic interface studies.

The applied stress level is one of the main differences between the referred test types. Higher normal stresses can be applied for CDS and MDS apparatus and lower normal stresses should be applied for studies employing IP equipment (Palmeira et al. 2004; Sánchez 2018; Pavanello et al. 2022). The execution of direct shear tests using a standard apparatus under low normal stresses may yield to significant errors in the prediction of the interface friction angles (Palmeira et al. 2004). The effect of applied normal stress under similar boundary conditions on the same type of device is also studied by several authors (Izgin & Wasti, 1998; Palmeira et al. 2002; Bacas et al. 2015; Cen et al. 2018; Sánchez 2018; Lashkari 2021). In Sánchez (2018) it was utilized sand, three textured and one smooth geomembranes for CDS, MDS and IP tests under different stress levels. The author indicated that for soil geomembrane interfaces, CDS and MDS performs more accurately under normal stress greater than 25kPa and IP tests performs better for stress level lower than 25kPa. The author indicated that an increase in normal stress decreases the interface friction angle slightly. Table 2.2 summarizes the interface components, the apparatus type, the applied normal stress and relative interface friction angles used by Sánchez (2018). Similarly, Lashkari & Jamali (2021) also reflects that for sand and geomembrane interfaces peak interface friction angle increases when applied normal stress decreases.

Table 2.2. Interface friction angles under different stress levels (Sánchez 2018).

Interface types		Test types								
	Test type	CDS			MDS			IP		
	Normal stress	25	50	100	25	50	100	1.96	3.94	6.09
Sand/Textured geomembrane$_1$	Friction angles	34	33	32	31	30	31	32	31	30
Sand/Textured geomembrane$_2$		40	35	35	32	32	33	31	31	31
Sand/Textured geomembrane$_3$		38	35	32	34	30	33	32	32	32
Sand/Textured geomembrane$_4$		36	34	36	34	33	32	32	31	31
Sand/Smooth geomembrane		24	23	24	27	27	28	24	24	24

The influence of applied normal stress on the overall shear strength can be affected by other aspects of geosynthetic interface test. Carbone et al. (2015), Sánchez (2018) and Pavanello et al. (2021) indicated that the effect of applied normal stress is dependent on the sliding velocity of the IP upper box.

Izgin and Wasti (1998) compared IP tests with MDS tests equipped with boxes of different dimensions (from 0.12 to 0.6 m^2) under a normal stress level ranging from 5 to 50 kPa. The tests were carried out on soil (Ottawa sand)-HDPE smooth and rough geomembrane interfaces considering the sliding angle as the main parameter of the test. They concluded that the direct shear test overestimates the interface shear strength angle and the authors noted a higher discrepancy if the small size box dimensions are considered.

2.1.2. GEOMEMBRANE PROPERTIES

There are various geosynthetic products manufactured in different forms according to their required purpose. They can be categorized as geomembrane, geotextile, geogrid, geo-composite, geofoam, geosynthetic clay liners, geocells ang geo-pipes. The geosynthetic products considered in this study are geomembranes with cohesionless sand in the interface. Geomembranes are often utilized as fluid barriers in geotechnical applications, such as landfills and water transportation canals. Due to their relative impermeability and chemical resistance characteristics, they are usually used alongside geotextiles or sand in landfills to constitute base, side-slope, and cover liner systems. Geomembranes can be manufactured by various polymers, such as; high density polyethylene (HDPE), low density polyethylene (LDPE), polyvinyl chloride (PVC) and chloro-sulphonated polyethylene (CSPE). On their surface finish, geomembranes can be smooth or textured. Figure 2.3 shows HDPE geomembranes with smooth and rough finishing.

(a) (b)

Figure 2.3. HDPE geomembranes: (a) Smooth, and (b) Textured (Cen et al. 2018).

Geomembranes can vary in thickness, unit weight, asperity height and asperity density. Numerous authors have investigated impact of geomembrane physical property on the interface shear strength results. The impact of geomembrane roughness on interface shear strength is studied and presented by several researchers. O'Rourke et al. (1990) analyzed the shear strength difference for interfaces consisting of two types of geomembrane polymers, HDPE and PVC. The interface friction angle for HDPE geomembranes tends to be smaller than an interface

consisting of PVC. This is due to the soil particles were sliding and created less interlocking bond with the rigid HDPE surface and for more compressible PVC surface the shear resistance showed greater strength.

Izgin et al. (1998) studied HDPE geomembranes of three different texture levels. As shown in Figure 2.4 shear strength was higher for interface consisting rough HDPE geomembranes.

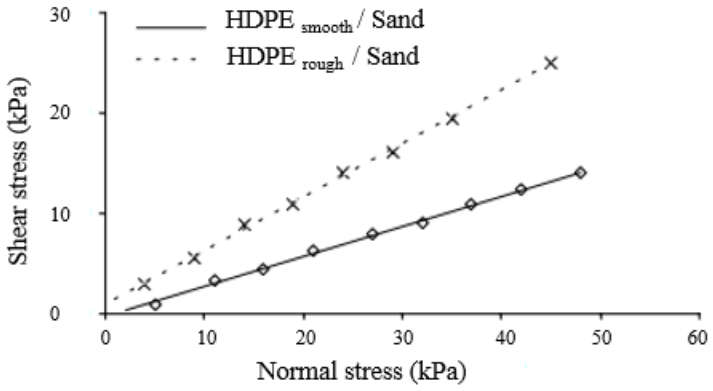


Figure 2.4. Influence of geomembrane roughness on interface strength (Izgin et al. 1998-modified).

Palmeira et al. (2004) studied HDPE geomembranes with different surface characteristics by using inclined plane test apparatus. Figure 2.6 shows the failure envelopes obtained for the tests with these geomembranes (smooth - GM $_{smooth}$, rough 1 - GM$_{rough1}$, rough 2 - GM $_{rough2}$) and coarse sand. As shown in Figure 2.5 the influence of surface roughness is exhibited by increasing the friction angle between soil and geomembrane approximately in 5 degrees.

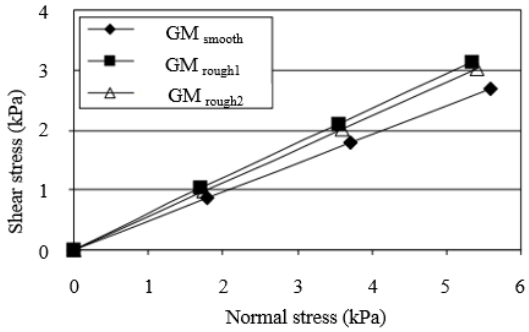
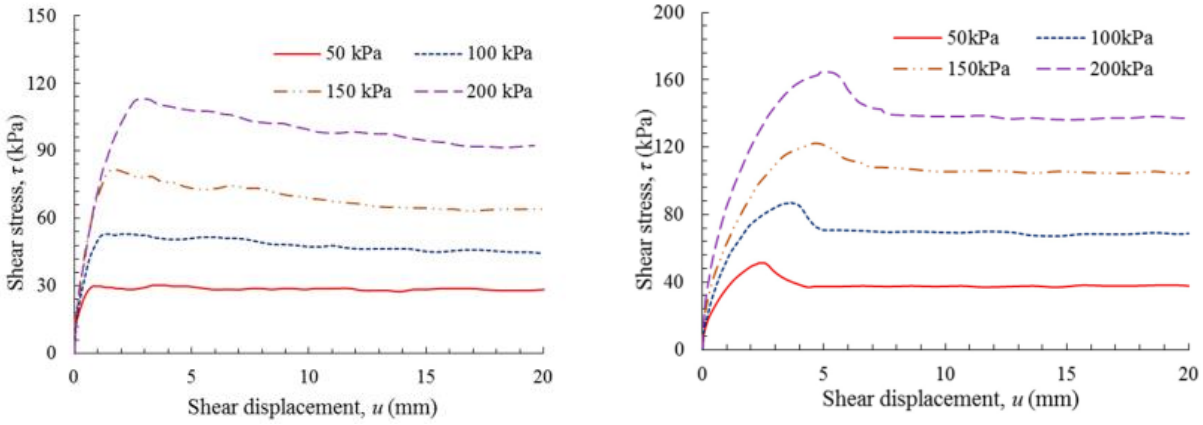


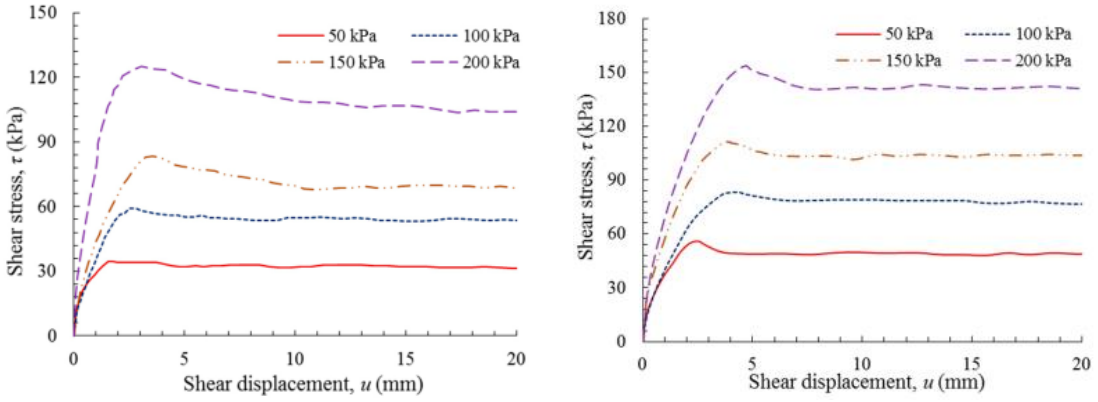
Figure 2.5. Failure envelopes for tests with HDPE geomembranes and coarse sand (Palmeira et al. 2004).

Cen et al. (2018) used smooth and textured HDPE geomembranes with a nominal thickness of 2 mm and density of 0.94 g/cm^3 in order to identify the effect of geomembrane texture. The interface was formed by using two sand types which are fine sand and sandy gravel soils. The test was done by utilizing medium size direct shear apparatus under four different normal stress levels (50kPa, 100kPa, 150kPa and 200kPa). As shown in Figure 2.6 and Fig 2.7 smooth HDPE interface shown a decrease in shear strength than interface composed by textured HDPE, in contact with fine sand or coarse sand and under all applied normal stresses. The authors found out higher peak shear stresses and corresponding shear displacements were observed for the textured geomembrane-soil interfaces. In the study, the friction angles of the textured geomembrane–soil interfaces are 12% to 15% higher than smooth geomembrane-soil interfaces.



(a) (b)

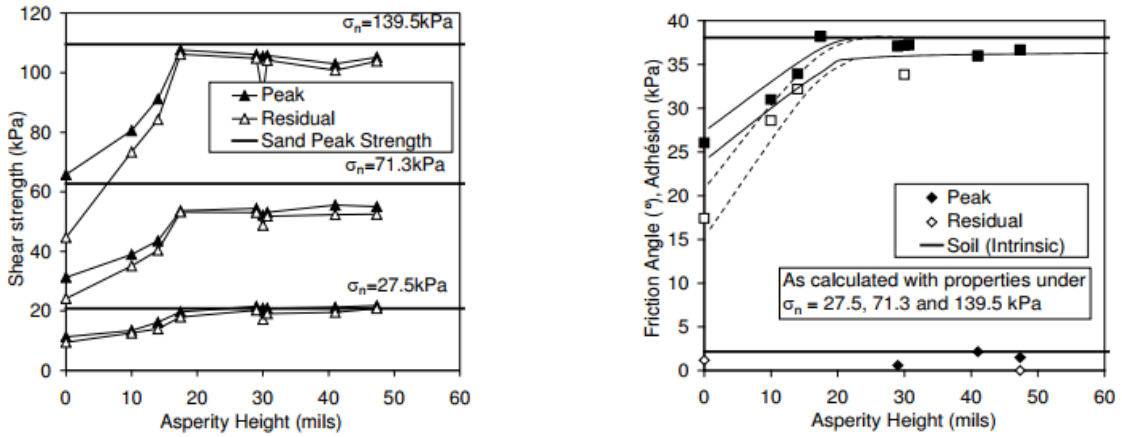
Figure 2.6. Shear stress versus shear displacement: (a) Smooth geomembrane-fine sand, (b) Textured geomembrane-fine sand (Cen et al. 2018).



(a) (b)

Figure 2.7. Shear stress versus shear displacement: (a) Smooth geomembrane-sandy gravel soil, (b) Textured geomembrane-sandy gravel soil (Cen et al. 2018).

The degrees of roughness of geomembranes are quantified by asperities. Asperities are individual projections of polymer that extend above the main surface of a textured geomembrane and characterized by their height, concentration, spacing and pattern (Yesiller 2005; ASTM D7466, 2015). In Blond and Eli (2006), it is shown that asperity height is a key factor to influence the shear strength for sand and geomembrane interfaces. The authors indicated a threshold height of 20 mils (0.5mm), in which beyond this value an increase in asperity height will not provide any additional shear strength to the interface. When the asperity height is close or beyond the observed threshold value of 20 mils (0.5mm), the peak shear strength is similar to the residual shear strength and shown in Figure 2.8. Figure 2.8 also shows residual shear strengths lower than peak shear strengths can be observed with small asperity heights.



(a) (b)

Figure 2.8. Asperity height influence on sand-geomembrane interface: (a) shear strength vs asperity height, (b) friction angle and adhesion vs asperity height (Blond and Eli 2006).

2.1.3. SOIL PROPERTIES

The soil physical property in the interface affects its interlock with the geomembrane and influences the overall shear strength. Many authors addressed the impact of soil property in geosynthetics interfaces. In O'Rourke et al. (1990) it was examined the influence of soil density on soil-geomembrane interaction. The study was carried out through friction tests of sand-geomembrane interfaces and it was found that the interface resistance is directly proportional to the soil density.

Lopes (2000) indicated well graded soil with a wide particle size range exhibited increase in interface shear strength either with smooth or rough geomembrane. The test utilized two types of soils, soil 1 with dimensions ranging between 0.074 mm and 9.54 mm and soil 2 with dimensions ranging between 0.074 mm and 2.00 mm. The author indicated there was an increase in the friction angle of the interface, 14.6% with smooth geomembrane, and 5.5% with rough geomembrane. Higher soil-geomembrane interface friction angles are measured when the geomembrane surface roughness allows the penetration of soil particles and the soil particle size has also an important influence on the soil-geomembrane interface friction angle. Broadly graded soils with large average soil particle sizes allow an increase in the interface resistance.

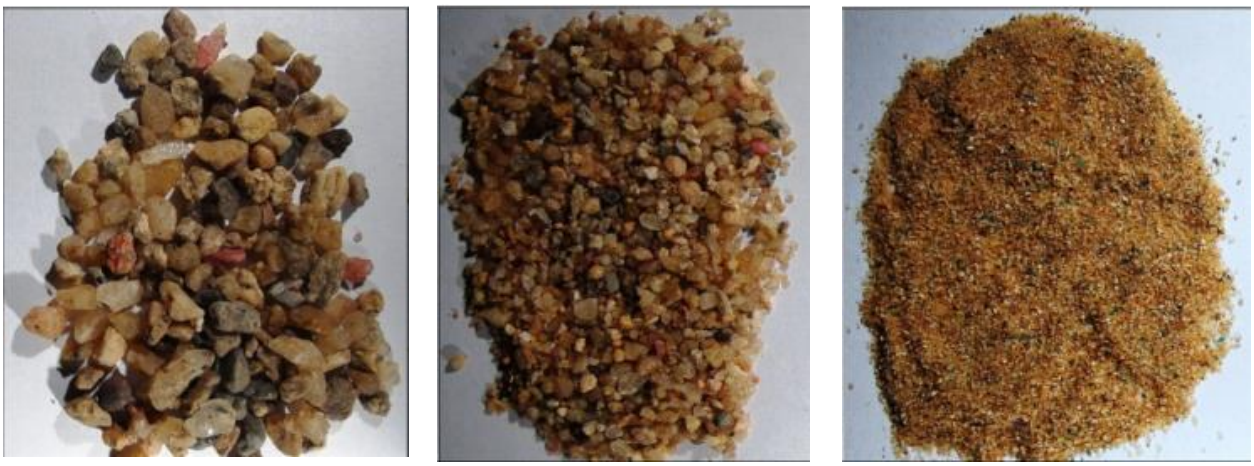
The influence of particle size and relative density of soil on the shear strength interface is indicated by Frost et al. (2002), Vangla and Latha (2015) and Choudhary and Krishna, (2016). In another research, Lopes et al. (2015) has studied the effect of particle size on soil-geomembrane interface by utilizing medium size direct shear test. The authors involved one smooth and one rough geomembrane with two different grains sized cohesionless soils; fine and coarse sand. The behavior of tested materials and the interface shear strength results are summarized on Table 2.3.

Table 2.3. Soil-geomembrane interface materials and formed interfaces strength.

	Name	GM 1	GM 2
	Behavior	Thickness = 2mm Smooth HDPE	Thickness = 2mm Rough HDPE
Name			
Soil 1	D_{50} = 0.43mm; C_u = 2.94 $\phi_{soil} = 36.2^o$	$\Phi_{interface} = 21.4^o$	$\Phi_{interface} = 31.2^o$
Soil 2	D_{50}= 1.3 mm; C_u = 3.64; $\phi_{soil} = 49.5$	$\Phi_{interface} = 24.5^o$	$\Phi_{interface} = 32.9^o$

As shown in Table 2.3, smooth and rough geomembranes exhibit different behaviors when in contact with either fine or coarse sands. The interface friction angle between GM1 and Soil2 has shown a 14.5% increase than the interface formed by GM1 and Soil1. This difference is due to the wider soil particle size distribution of Soil 2, which increases the soil-geomembrane contact surface. The authors concluded broadly graded soils with larger average soil particle sizes allow an increase in the soil-geomembrane interface resistance that is more significant to smooth geomembrane surfaces.

Vangla and Latha (2015) studied the effect of particle size on interface shear strength and the authors selected three sand types of different particle size. The considered coarse sand, medium sand and fine sand are shown on Figure 2.9. In order to examine the effect of particle size, the authors used sands with similar morphology and origin.



(a) (b) (c)

Figure 2.9. Photographs of sands: (a) coarse sand, (b) medium sand, (c) fine sand (Vangla and Latha 2015).

The test has been executed on modified medium size direct shear equipment under three applied normal stress levels and the interface shear strength for interfaces of the three sands are shown in Figure 2.10. It is seen that medium sand (D_{50} = 0.87mm, C_u = 1.96 and C_c = 0.97) is showing highest interfacial frictional resistance with the geomembrane compared to the coarse and fine sands. This is due to a greater number of effective contacts per unit area in the interface.

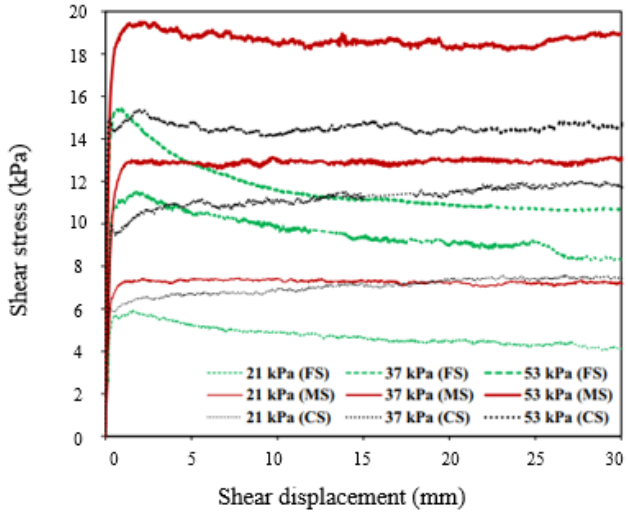


Figure 2.10. Shear stress versus shear displacement response of different sands with geomembrane interface direct shear tests. (Vangla and Latha 2015).

In Choudhary and Krishna (2016) three types of cohesionless soils were investigated to analyze the influence of soil particle size (D_{50}) on interface shear strength using geomembrane. The grain size distribution of the soils and the physical properties of soils in the study are shown in Figure 2.11 and Table 2.4, respectively. The specific gravity of sands was found to be 2.64 and the coarser sand, Soil 1 (D_{50} = 1.5 mm), has soil particle diameter values ranging from 1 to 2 mm. The finer sand, Soil 3 (D_{50} = 0.22 mm), has soil particle diameters from 0.09 to 0.5 mm and within this the soils are classified as poorly graded sands. The authors indicated that the interface friction angle from direct shear tests linearly increases with increase in D_{50} of soil.

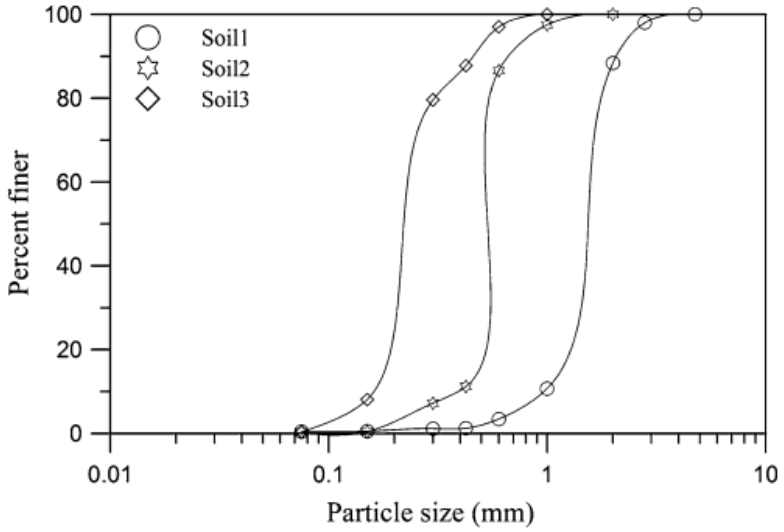


Figure 2.11. Grain size distribution curves of soils (Choudhary and Krishna 2016).

Table 2.4. Physical properties of three soils used in Choudhary and Krishna (2016).

Properties	Soil 1	Soil 2	Soil 3
G	2.64	2.64	2.64
D_{50} (mm)	1.5	0.5	0.22
C_u	1.6	1.27	1.56
C_c	0.9	1.03	1
γ_d (kN/m^3)	16.6	16.3	16.7

2.2. MACHINE LEARNING

Machine learning is the general term for when computers learn from data. It describes the intersection of computer science and statistics where algorithms are used to perform a specific task without being explicitly programmed; instead, they recognize patterns in the data and make predictions once new data arrives. The learning process of these algorithms can either be unsupervised or supervised, depending on the data being used to feed the algorithms. Unsupervised learning algorithms use only input data. This learning method is suitable for problems where there is little or no understanding what the expected results should be. Therefore, the goal of unsupervised learning is to gain knowledge and find structure in the data, not making predictions. Supervised learning makes use of a known relationship between input and output. The goal of the algorithm is to learn from the more correct answers in the training data and use the insights to make predictions when given new input. Supervised learning algorithms can perform two functions (i), Classification: which is identifying between labeled options and (ii), Regression delivering an estimation of continuous numerical value (IBM 2020). Regression is a method of modeling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables. Machine learning algorithm can be a simple linear regression or a multi-step nonlinear regression model. There are several civil engineering studies employing the linear and non-linear machine learning algorithms for data analysis and a parameter estimation.

2.2.1. SIMPLE LINEAR REGRESSIONS AND CORRELATION MEASURES

Linear regressions are developed in the field of statistics and it is applied as a method for understanding the relationship between input and output numerical variables (Brownlee 2015). It is a linear relationship between the input variables (x) and the single output variable (y), that y can be calculated from a linear combination of the input variables (x). When there is a single input variable (x), the method is referred to as simple linear regression. Simple linear regressions are deployed in geotechnical studies such as Azzouz et al. (1976) for soil compressibility

analysis; Nair et al. (2018) for soil samples investigation; Linares-Unamunzaga et al. (2019) for prediction of soil-cement unconfined compression strength; and Mandhour, (2020) for predicting compression index of soil. The authors indicated that simple linear regressions have showed adequate prediction capability.

Correlation refers to the degree of relationship or dependency between two variables. Linear correlation refers to straight-line relationships between two variables. Linear correlations are widely utilized in Machine Learning investigations for civil and geotechnical engineering parameter analysis, by comparing a laboratory or field investigated result with the machine learning projections. It is an important step to observe the linear fit between the estimation from the algorithms and the practical reference data. There are numerous civil engineering parameter forecasting investigations which measured the performance of their developed model by using a linear correlation (Zhang et al. 2020; Jeremaiah et al. 2021; Ly et al. 2022). After a linear relationship developed between estimated and previously gathered experimental data, the performance evaluation which is the resemblance difference, is measured by error metrics. In several works of literature, the common correlation metrics measures are:

(i) Pearson's correlation coefficient (R).

The Pearson coefficient is a type of correlation coefficient that represents the linear relationship between two variables that are measured on the same interval or ratio scale. It is the covariance of the two variables divided by the product of their standard deviations and computed as:

$$\bar{\rho} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \tag{2.1}$$

where:

$\bar{\rho}$ = Pearson product-moment correlation coefficient

$Cov(x,y)$ = covariance of variables x and y

σ_x = standard deviation of x

σ_y= standard deviation of y

(ii) Coefficient of Determination (R^2)

R^2 measures how much variability in dependent variable can be explained by the model. It is the square of the Correlation Coefficient(R) and it is computed as:

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} \tag{2.2}$$

Where: $SS_{Residuals}$ is the sum of the square of residuals and,

SS_{Total} is the total sum of the squares

The functions of coefficient of determination are:

- It gives the ratio of the predicted and predicting variables
- It helps to find Explained variation / Total Variation
- It shows the strength of the linear association between the variables.
- It helps in determining the strength of association between different variables.

(iii) Mean square (MSE)

Mean squared error is a measure utilized to determine the performance of an estimator and it is necessary for relaying the concepts of precision, bias, and accuracy during statistical estimation. The measure of mean squared error needs a target of prediction or estimation along with a predictor or estimator, which is said to be the function of the given data. MSE is the average of squares of the "errors". The formula for mean squared error is given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2 \tag{2.3}$$

where; X_i is the vector denoting values of n number of predictions and a vector representing n number of true values.

In more generalist expression, if θ be some unknown parameter and $\theta_{obs,i}$ be the corresponding estimator, then the formula for mean square error of the given estimator is given by:

$$MSE(\theta_{obs,i}) = E[\theta_{obs,i}, -\theta)^2] \quad (2.4)$$

(iv) Root mean square (RMSE)

The Root mean square error (RMSE) of an estimator of a population parameter is the square root of the mean square error (MSE). The mean square error is defined as the expected value of the square of the difference between the estimator and the parameter. It is the sum of variance and squared bias.

2.2.2. MULTILPLE LINEAR REGRESSION (MLR)

Multiple linear regression is a regression model that estimates the relationship between a quantitative dependent variable and two or more independent variables using a straight line. MLR is a linear statistical technique that is beneficial for predicting the best relationship between dependent variable and several independent variables (Akan et al. 2015). It is based on least squares; the model is fit that the sum of squares of differences of observed and predicted values in minimized and a general MLR model can be formulated as:

$$Y = B_0 + B_1X_1 + \cdots + B_nX_n + \varepsilon \quad (2.6)$$

Where: Y indicates dependent variables, $X_i$indicates independent variables, $B_i$indicates predicted parameters and ε is the error term.

There are some studies which adopts this statistical algorithm for geotechnical engineering parameter analysis, such as Akan et al. (2015) for unconfined compressive strength prediction of jet grout column; Roy, (2016) for assessment of Soaked California Bearing Ratio and Williams and Ojuri, (2021) for modeling of soil hydraulic conductivity. In the studies it is shown that MLR functioned adequately on predicting the dependent variable.

2.3. ENSEMBLE LEARNING ALGORITHMS

An ensemble in the context of machine learning can be broadly defined as a machine learning system that is constructed with a set of individual models working in parallel and whose outputs are combined with a decision strategy to produce a single answer for a given problem. Ensemble methods combine multiple learning algorithms to obtain better predictive performance than the individual constitutive learning algorithms and it is a machine learning paradigm where multiple learners are trained to solve the same problem (Pintelas and Livieris, 2020). There are several ensemble learning algorithms in the studies employing Machine Learning as a tool. The most common Machine Learning techniques for civil and geotechnical parameter prediction are Artificial Neural Network, Gaussian Process Regression, Support Vector Machine and Random Forest Regression.

2.3.1. ARTIFICIAL NEURAL NETWORK (ANN)

Artificial Neural Networks (ANN) is one of the most popular AI techniques, which is based on the biological neural systems of human brain. It is more powerful in the case of unknow relationship among variables and its main feature is having an independent statistical distribution with self-learning and interdependent memory (Pham et. al. 2020b). ANN consists of a number of neurons (nodes) which construct its structure with three types of layers including input layers, hidden layers and output layers. ANN as a branch of artificial intelligence is simply an automated optimization system capable of learning the relationship and inter dependencies between multiple input variables of a given system and modelling such relations (trends and patterns) in the form of mathematical functions for easy prediction (Jeremaiah et al. 2021). Figure 2.12 shows the schematic illustration of typical ANN architecture.

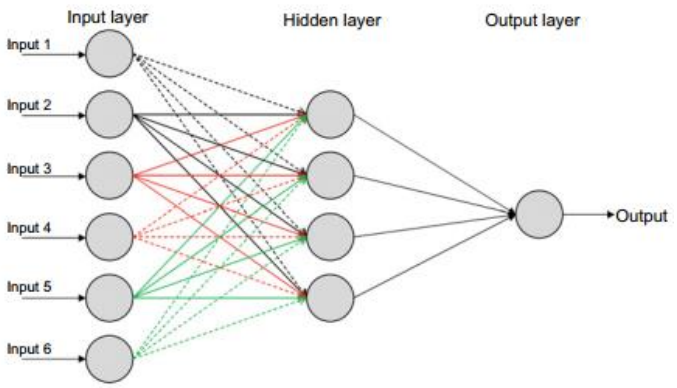


Figure 2.12. Schematic illustration of a typical ANN architecture (Pant and Ramana, 2021).

The performance and training efficiency of an ANN model depend largely upon four parameters, i.e., the number of neurons, the number of hidden layers, the regularization term, and the epoch size (Pierre et al. 2021). ANN one of the most applied Machine Learning algorithms in civil and geotechnical studies, and the technique is applied for analyzing several parameters, such as; surface settlement analysis (Suwansawat and Einstein, 2006), pipe failure rate (Akbar et al. 2014), compression coefficient of soft soil (Pham et al. 2020b), Geo-Mechanical Properties of Stabilized Clays (Jeremiah et al. 2021), and diameters of jet grouted columns (Piere et al. 2021). The authors indicated that ANN has estimated subjected parameters effectively.

2.3.2. GAUSSIAN PROCESS REGRESSION (GPR)

Gaussian process is a stochastic process, which provides a powerful tool for probabilistic inference directly governs the property of the functions and it has gained much attention in recent years (Rasmussen and Williams, 2006). A GPR is the generalization of a Gaussian distribution. While the latter is the distribution of a random variable, the Gaussian process describes a distribution over functions and its graphical illustration is shown in Figure 2.13. From function space view, the Gaussian process f(x) can be determined by the corresponding mean m(x) and covariance functions that are defined as:

$$m(x) = E(f(x)) \tag{2.7}$$

$$k(x, x') = E\big(\big(f(x) - m(x)\big)\big(f(x') - m(x')\big)\big) \tag{2.8}$$

Where: $k(x, x')$is the covariance (or kernel) function evaluated at x and x'
m(x) is the mean and E is the covariance of $f(x)$.

A Gaussian process $f(x)$ can be represented as:

$$f(x) \sim GP(m(x), k(x, x')) \tag{2.9}$$

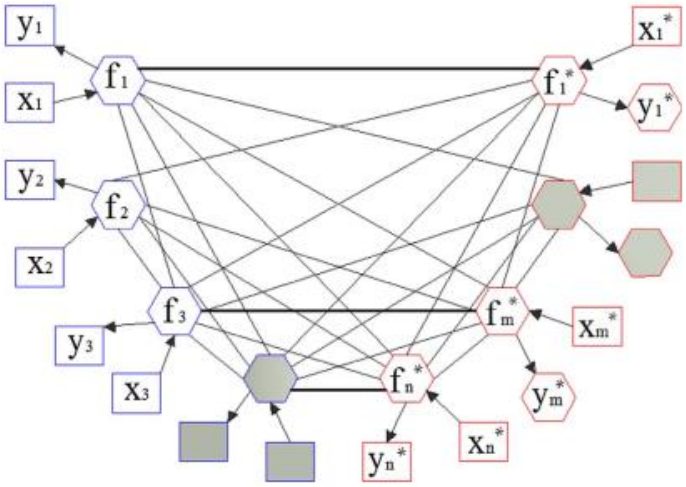


Figure 2.13. Graphical model for GPR (Li et al. 2017).

For civil and geotechnical engineering studies GPR has been applied for landslide displacement (Rohmer and Foerster, 2011), surface settlement (Ocak and Seker, 2013), compressive strength of high-performance concrete (Hoang et al. 2016), water inflow prediction in tunnel construction (Li et al. 2017) and early strength of high-performance concrete (Ly et al. 2022).

In Ocak and Seker, (2013) short-term surface settlements caused by earth pressure boring machines are predicted for twin tunnels, which will be excavated in the Esenler and Kirazli stations of the Istanbul Metro line. A total of eighteen input parameters are used for predicting surface settlements and the data source has been iteratively processed during the parameter selection phase. The parameter selection phase tries to understand the effect of each parameter on the results. The parameter with the least effect on the result is eliminated; the remaining

parameters are processed for the next steps and the dataset is divided 90% for training and 10% for testing the developed GPR model. The authors indicated GPR predicted surface settlement with less error margin.

2.3.3. SUPPORT VECTOR MACHINE (SVM)

SVM was developed by Cortes and Vapnik, (1995). It is a supervised learning algorithm for solving regression and classification problems. For a classification problem, the basis of SVM is to find a hyperplane so that all data points $x_i \in R^n$ in the same tag are on the same side of the plane and the distance between the nearest point of each tag to the hyperplane is the largest. In other words, the best hyperplane is found by maximizing the width of the margin. Support Vector Regression is similar to Linear Regression in that the line equation is $y = wx + b$. In Support Vector Regression, this straight line is referred to as a hyperplane. The data points on either side of the hyperplane closest to the hyperplane are called Support Vectors which are used to plot the boundary line. Unlike other Regression models that operate to minimize the error between the real and predicted value, the Support Vector Regression aims to fit the best line within a threshold value (Distance between the hyperplane and boundary line), a. Thus, it is observed that Support Vector Regression is a model intended to satisfy the condition $-a < y - wx + b < a$. It used the points with this boundary to predict the value. This basic idea is presented in Figure 2.14 for SVM training set.

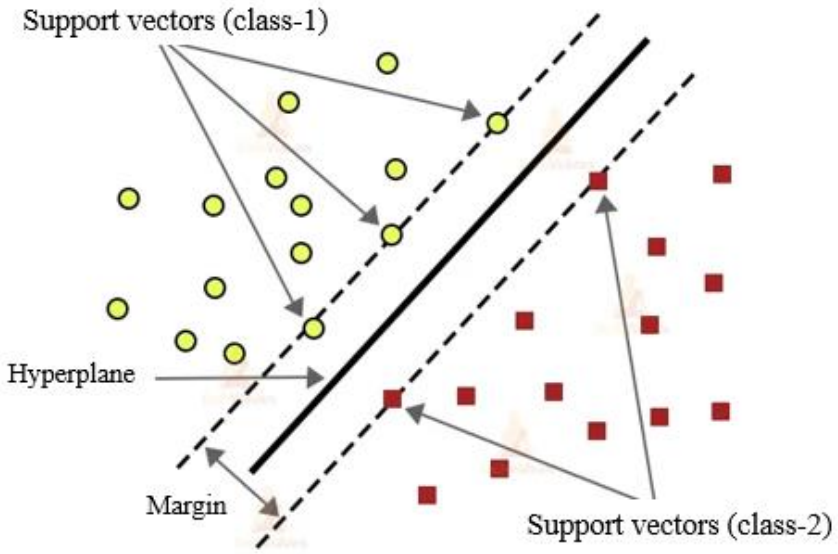


Figure 2.14. A hyperplane for SVM training set (modified from Truong and Pham, 2021).

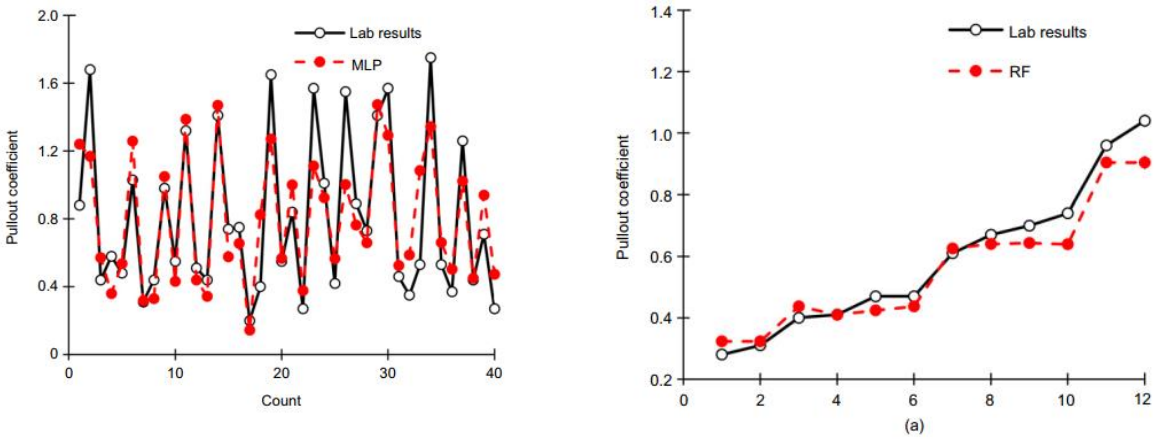
For civil and geotechnical engineering studies a SVM technique is implemented for soil moisture prediction (Gill et al. 2006); predict slope stability (Ramya and Vinodhkumar, 2017); soil classification (Vijay et al. 2020); prediction of tunnel boring machine penetration (Afradi, et al. 2020); regression for ultimate strength of trusses (Truong and Pham, 2021) and the authors indicated SVM as an efficient Machine learning algorithm for subjected parameter analysis. For example, in Ramya and Vinodhkumar, (2017) the authors stated that the use of SVMs is very advantageous for the prediction of slope stability because it can perform nonlinear regression efficiently for high-dimensional data sets.

2.4. RANDOM FOREST (RF)

RF is developed by Breiman, (2001) and it is a supervised machine learning algorithm which is constructed from decision trees and used to solve regression problems. RF is implemented based on bagging decision trees by employing random split selection and it extracts multiple sub-sample sets from the original sample set, each of them independently forms a decision tree $(h(x, \theta_k), k = 1,2, \dots, K)$(Liu et al. 2020). Decision trees randomly select features to split the internal nodes and eventually form a random forest. For the regression algorithm, the prediction result is the average of all decision tree output results (Zhang et al. 2021). Number of trees,

maximum depth of trees and minimum number of samples to split a node are called hyperparameters of a RF. A proper optimization of hyperparameters leads to a more accurate prediction of the algorithm (Zhou et al.2019). Technical investigations implementing RF for geosynthetic interface analysis are rare and the algorithm is utilized more on other geotechnical and civil engineering parameters analysis with datasets showing non-linearity. Random Forest easily adapts to nonlinearity found in data and tends to make a better prediction (Schonlau et al. 2020).

Pant & Ramana (2021) investigated the algorithm efficiency for estimation of pullout coefficient of geogrid. There are other studies employing RF technique for geotechnical parameter prediction, such as; ultimate bearing capacity of strip footing (Dutta et al. 2019), assessment of pile drivability (Zhang et al. 2019), shear strength of rockfill material (Zhou et al. 2019), undrained shear strength of soil (Pham et al. 2020a; Zhang et al. 2021), soil unconfined compressive strength (Ly et al. 2020), and factor of safety against basal heave for braced excavations in anisotropic clays (Zhang et al. 2022). From a broader perspective, RF algorithm has been employed in the analysis of civil engineering parameters, such as; the strength of rubberized concrete (Sun et al. 2019), compressive strength of high-performance concrete (Han et al. 2019; Farooq et al. 2020; Liu et al. 2020) and permeability of pervious concrete (Huang et al. 2020). The authors indicated RF as an important tool for adequate parameter prediction.
In Pant & Ramana (2021) RF is applied for inspecting the pullout behavior of geogrids based on 198 laboratory test results. For each recorded pullout coefficient of geogrid, 7 strands influencing this parameter are also collected, namely; normal stress, D_{50} of the soil, fines content, specimen length (L), space between longitudinal members (S_L), space between transverse members (S_T) and ultimate tensile strength of geogrid (T_{ult}). The authors used five-fold cross-validation which is 80% of the data has been trained the model algorithm and then the accuracy of the model is tested on 20% of the remaining dataset. Figure 2.15 shows scatter plot between the observed and the predicted geogrid coefficient values. It is indicated that the RF has projected the geogrid coefficient close to the target values by resulting $R^2 = 0.98$ for training and $R^2 = 0.99$ for testing set.



(a) (b)

Figure 2.15. Prediction capability of RF algorithm for geogrid coefficient, (a), for the training and, (b) for the testing datasets (Pant & Ramana 2021).

In Dutta et al. (2019) the paper presented the prediction of the ultimate bearing capacity of a strip footing resting on layered soil using Random Forest regression. The problem statement for the footing resting on layered soil to predict the ultimate bearing capacity is shown in Figure 2.16. In the study, 181 ultimate bearing capacity (q_{ult}) results and 7 influencing elements for each q_{ult} result have been gathered from laboratory and finite element method approaches. The considered 7 influencing elements are friction angle of the dense sand layer (ϕ_1), friction angle of the loose sand layer (ϕ_2), unit weight of the dense sand layer in kN/m^3 (γ_1), unit weight of the loose sand layer in kN/m^3 (γ_2), ratio of the depth of the footing to the width of the footing (D/B), ratio of the thickness of the dense sand layer below base of the footing to the width of footing (H/B) and (H+D)/B.

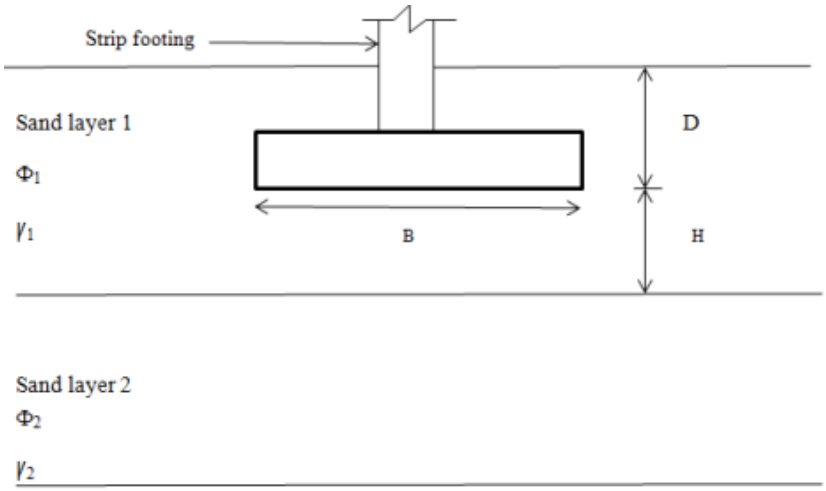


Figure 2.16. Stirp footing on a layered soil study for predicting ultimate bearing capacity by using RF (Dutta et al. 2019).

The authors trained the algorithm by randomly selected 70% of the data and tested the prediction ability by the remaining 30%. The performance of the algorithm has been measured by different metrics regression methods to identify how far the estimation varied from the targeted ultimate bearing capacity values. Figure 2.17 shows the coefficient of determination (R^2) measurements between RF estimation and targeted ultimate bearing capacity. It is indicated that the RF has projected the ultimate bearing capacity close to the target values by resulting R^2 = 0.98 for training and R^2 = 0.96 for testing set.

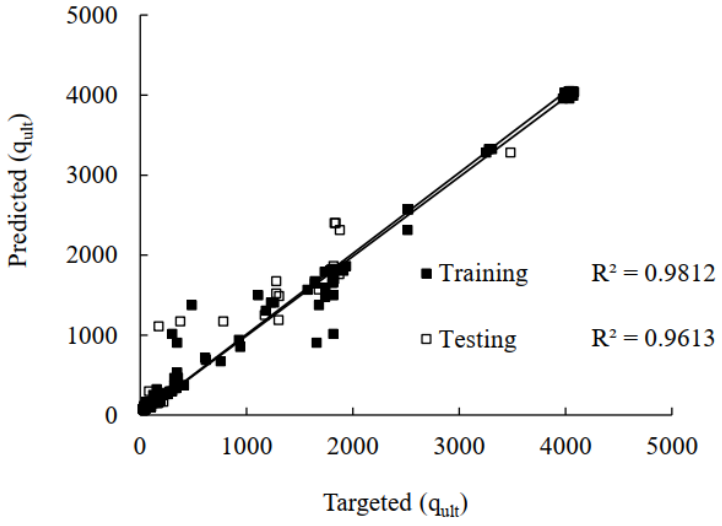
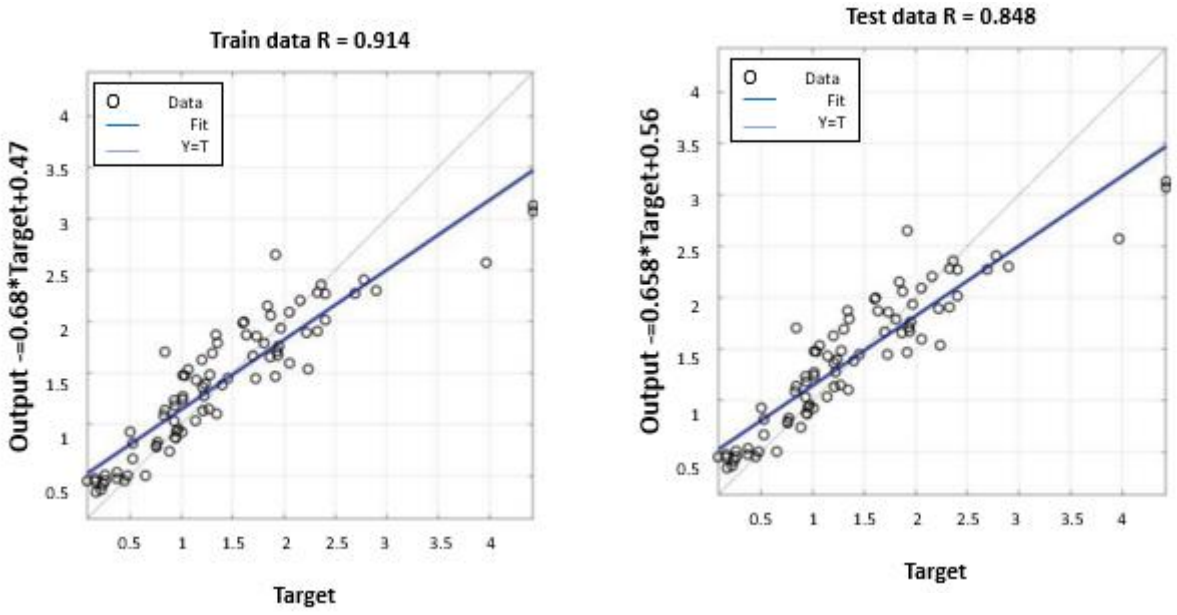


Figure 2.17. Variation of targeted with the predicted ultimate bearing capacity of the footing resting on layered soil using RFR (Dutta et al. 2019).

In Ly et al. (2020) the technique is applied to predict soil unconfined compressive strength based on a laboratory investigation data from a Long Phu 1 power plant in Vietnam. In the study 118 samples were collected to determine the soil properties and to generate the study data as well. There were 6 soil properties for each examined sample which are considered as elements affecting unconfined compressive strength of soil (q_u) such as; the clay content (%), void ratio, liquid limit (%), moisture content (%), plastic limit (%), and specific gravity. The data set which consists the influencing elements and the q_u results is utilized by 70% for training the RF algorithm and 30% for testing the estimation capability. Figure 2.18 shows the correlation coefficient (R) measurements between RF estimation and targeted unconfined compressive strength. It is indicated that the RF has projected q_u close to the target values by resulting $\bar{\rho}$ = 0.914 for training and $\bar{\rho}$ = 0.848 for testing set. Recruiting the factors influencing the predicted parameter is the common features of RF utilizing studies. The overall data is always divided in to training and testing sets and the final prediction is measured by its variation from the previously collected experimental values of the analyzed parameter.



(a) (b)

Figure 2.18. Prediction capability of RF algorithm for the UCS in a regression form, (a), for the training and, (b) for the testing datasets (Ly et al. 2020).

RF can pair up with other Machine Learning algorithms for a better functionality and accuracy and this due to hyperparameters of RF can be optimized by the functions of other algorithms. In Pham et al. (2020a) for a study analyzing undrained shear strength of soil, the hyperparameters of RF were optimized by an algorithm called Particle Swarm Optimization. It is a computational method which is a form of evolution algorithm in which each particle, in a random population of particles is considered as a given approach to seek a solution for solving the problem. The Particle Swarm Optimization is more focused on the position of each particle and this position is upgraded considering its current position and velocity (Eberhart and Kennedy, 1995). Five hyperparameters of RF were optimized by particle swarm optimization in the study and the authors indicated a hybrid approach has resulted a better prediction performance compared to an individual RF application. A Pearson's correlation coefficient value increased from 0.86 to 0.87 for the training set and $\bar{\rho}$ value increased from 0.87 to 0.89 for the testing set after the hyperparameters optimization. From the works of literature investigating civil engineering parameter, Sun et al. (2019) and Huang et al. (2020) also adopted an optimizing algorithm (Beetle Antenna Search) and indicated RF performed better with a collaboration of other Machine Learning algorithm. In this paper for geomembrane and sand interface analysis, to optimize RF hyperparameters, Differential Evolution algorithm, which is a powerful minimizer of stochastic functions is adopted (Pierre et al. 2021).

Naturally, the RF algorithm allows evaluating the feature importance. Feature importance refers to a technique that calculate a score for all the input features for a given model and the scores simply represent the importance of each feature. A higher score means that the specific feature will have a larger effect on the model that is being used to predict a parameter (Shin, 2021). In Ly et al. (2020) the estimation of predictor importance values was conducted by summing changes in the risk due to splits on every predictor and dividing the sum by the number of branch nodes. For unconfined compressive strength determination, specific gravity of soil has been identified as a most influential element among the other 5 factors and the rank agreed with actual laboratory investigations. In Zhou et al. (2019) for a study estimating shear strength of rock fill materials, 13 influencing factors were considered and during feature importance analysis the RF model excluded two of the input strands as non-influential factors. Similarly, in Dutta et al. (2019) the study utilized 7 factors affecting ultimate bearing capacity of strip footing. For the

sensitivity analysis different combination of the input parameters was used. For each of the combination, one of the input parameters was removed and the RF was carried out in order to check the influence of this omitted input parameter on the output. Further, for each of the combination of the input parameters, the performance measures were calculated. The omission of two variables (friction angle and unit weight of loose sand layer) has decreased the developed model performance significantly and the authors indicated that this feature importance score agrees with actual experiments. Therefore, in this research the feature importance analysis for factors affecting geomembrane and sand interface will be presented and compared to the works of literature to evaluate the developed RF model.

2.5. DIFFERENTIAL EVOLUTION (DE)

Differential Evolution (DE) is a population-based approach for the global optimization of nonlinear and non-differentiable continuous space functions (Storn and Price 1997). DE can handle nonlinear and non-differentiable multi-dimensional objective functions, while requiring very few control parameters to steer the minimization and these characteristics make the algorithm more practical to find solution for optimization problems. DE is designed to be a stochastic direct search method. Direct search methods also have the advantage of being easily applied to experimental minimization where the cost value is derived from a physical experiment rather than a computer simulation. The DE directly converts the solution to the optimization problem, into a D-dimensional solution vector, and each solution vector is the basic individual of the evolution. DE is designed to be a stochastic direct search method. Direct search methods also have the advantage of being easily applied to experimental minimization where the cost value is derived from a physical experiment rather than a computer simulation.

DE has been employed for some civil and geotechnical engineering parameters analysis, such as: soil water retention parameters (Ou, 2015); soil stabilization optimization (Kang et al. 2020); assessment of slope stability (Vu, 2022) and pipe embedding parameters estimation (Lu et al. 2022). In the other studies DE is utilized to optimize hyperparameters of Machine Learning algorithms, such as: for predicting a rain water induced soil erosion by support vector machine (Dinh et al. 2021); identification of the surrounding rock parameters in the tunnel construction

process by Gaussian Process Regression (Jiang et al. 2021); tunnel displacement prediction by Gaussian Process Regression (Zheng et al. 2021) and predicting diameters of jet grouted columns by Artificial Neural Network (Piere et al. 2021). The optimization methodology of DE for one of machine learning algorithms (GPR) is shown on Fig 2.19.

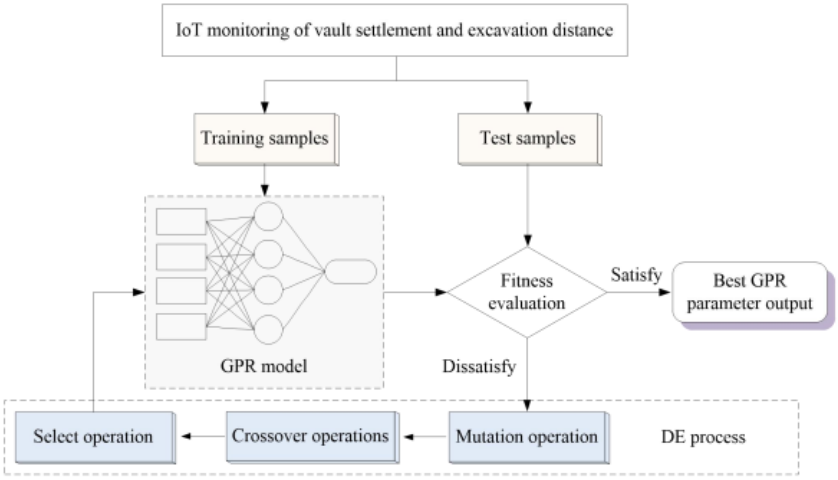


Figure 2.19. DE process to searching the optimal parameters of GPR (Zheng et al. 2021).

Similarly, for sand-geomembrane interface friction angle prediction, the four steps of DE are used to optimize RF hyperparameters.

3. METHODOLOGY

3.1. DATA ANALYSIS

Machine Learning applications depend heavily on the collected data in which each algorithm will learn through the gathered information and with this, training the model for regression purposes will be possible. As it is indicated in the literature review a ML parameter prediction requires the data incorporating the target variable and factors affecting this output element. In this study interface friction angle is the desired output of a prediction by the RF model for interface consisting sand and geomembrane. As the input parameters, 11 interface shear components and the 3 considered test types are recorded and organized together. The total number of 495 interface friction angles are gathered from previously executed laboratory test results and for each interface friction angle, the following inputs are gathered as influencing factors; the strain rate (mm/min or ^{o}/min), applied normal stress (kPa), a contact area between interface components (cm^2), thickness of geomembrane (mm), asperity height of geomembrane (mm), relative density of reinforced soil (%), the specific gravity of soil (g/cm^3), coefficient of curvature (C_c), coefficient of uniformity (C_u), D_{50} (mm), friction angle of soil (^{o}) and the type of test equipment (CDS, MDS or IP). Thus, 14 affecting parameters are utilized throughout the analysis and Figure 3.1 shows the registered versus missing elements of these components for the 495 investigated interfaces.

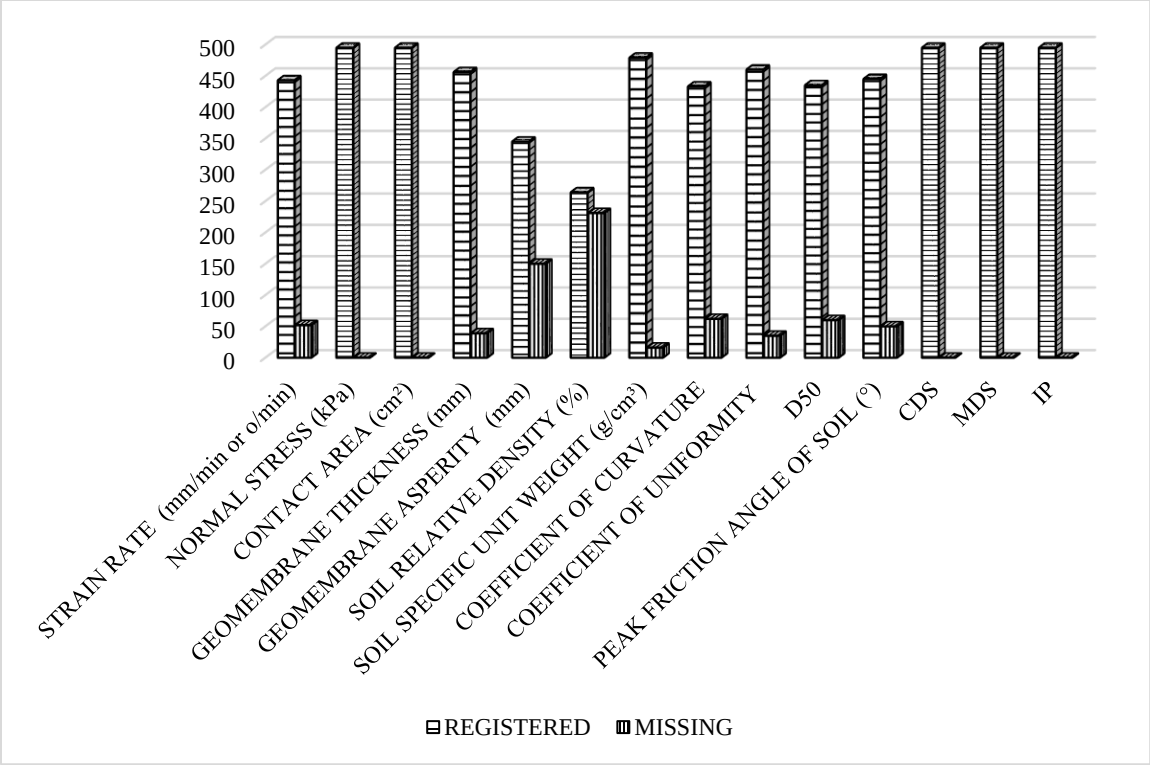


Figure 3.1. Collected and missing elements of the dataset.

In this research 495 interface friction angles and 14 influencing factors for each interface are recorded. The data is accumulated from 16 different research papers, with investigations from 6 different countries. The variety in the type of test equipment and boundary conditions from different laboratories is aimed to keep randomness for the gathered input - output values. The pondered studies and their related information are listed in Table 3.1.

Table 3.1. Data providing literature.

No.	Study theme	Author (s)	Φ Samples	CDS	MDS	IP	Country
1	Shear strength characteristics of sand-polymer interfaces	O'Rourke et al. (1990)	12	✓			USA
2	Interfacial Friction Study of Cap and Liner Components for Landfill design	Koutsourais & Sprague (1991)	20		✓		USA
3	Geomembrane – sand interface friction	Izgin & Wasti 1998	195	✓			Turkey
4	Study of soil-geosynthetic interaction in environmental protection works with the use of inclined plane equipment	Lima Junior (2000)	15				Brazil
5	Shear behavior of geosynthetics in the inclined plane test-influence of soil particle size and geosynthetics structure	Lopes et al. (2001)	14				Portugal
6	Study of soil-geosynthetic interaction in slopes of waste disposal works	Mello (2001)	12				Brazil
7	Interaction between soils and geosynthetic layers in large-scale ramp tests	Palmeira et al. (2004)	15				Brazil
8	Interface strength between geomembranes and soils through the Ring Shear test	Rebelo (2003)	15		✓		Brazil
9	Study of the stability and hydraulic conductivity of conventional and alternative capping systems for waste disposal works (translation)	Viana (2007)	16				Brazil
10	Microscale Geomembrane-Granular Material Interactions	Frost et al. (2011)	17	✓			USA
11	Shear behavior of sand-smooth geomembrane interfaces through micro-topographical analysis	Vangla & Gali (2016)	15		✓		India
12	Effect of time on soil-geomembrane interface shear strength	Alzahrani (2017)	18		✓		USA
13	Laboratory Investigation of Shear Behavior of High-Density Polyethylene Geomembrane Interfaces	Cen et al. (2018)	16		✓		China
14	Study of some aspects that influence the adhesion between geosynthetics and different materials	Sánchez (2018)	31	✓	✓		Brazil

15	Global and local sand–geosynthetic interface behavior	Lashkari & Jamali (2021)	49	✓			Iran
16	Researches from Nortène geosynthetic factory		35		✓		Brazil

Initially, the characteristics of the data has been statistically measured before exported to a RF algorithm. The observation was through six statistical measures, namely: minimum (Min) value, maximum (Max) value, mean, median, standard deviation (σ), and coefficient of variation were used and their relevant weight is shown in Table 3.2. The minimum and maximum values are distinguished to measure the range of each presented parameters of test type, geomembrane and soil properties. For analysis on the test type, as shown in Table 3.2, the strain rate has a range between 0.1mm/min or °/min to 3mm/min or °/min or. The least strain rate was observed for a CDS and the highest strain rate was recorded from an IP test. On the other hand, the applied normal stress of the referred studies has a minimum amount of 1kPa, over an IP equipment and the maximum amount of 300kPa on a CDS test. Similarly, the shear box size in CDS apparatus resulted the least contact area of 36cm² and a 192cm x 47cm ramp for IP test resulted the highest contact area of 9024 cm². In the investigation of the information about geomembrane properties, Table 3.2 shows 3mm for a smooth HDPE and 0.5mm for a PVC geomembrane are the maximum and minimum thicknesses. The range for geomembrane asperity height is between 0mm and 1.71mm. The smooth geomembranes have the no asperities measures and the textured geomembranes exhibited the highest asperity height. The sand physical property in the data as well can be highlighted by the minimum and maximum values. In the referred actual laboratory studies, a 20.8% relative density for a very loose soil and a 98% relative density for a very dense soil is recorded. In the data the investigated soils have cohesionless property and exhibited smallest specific unit weight of 1.45 g/cm^3 and highest unit weight of 2.94 g/cm^3. The grain size distribution peak measures are also marked and a well-graded sandy- gravel has a highest C_c = 1.68 and C_u = 46.68. Whereas, the smallest Cc and Cu measures are for fine sands, resulting Cc = 0.80 and C_u = 1.11. The particle sizes of the soil throughout the data ranges from 0.17mm to 3.08mm and the friction angle lays between 25° to 49.5° for the sand collected from the literature. The minimum and maximum values of the reported interface components highlight the scope of the studies contemplated in the data assembly.

Table 3.2. Basic statistics for test results.

Prediction and output variables	Statistics of the overall data set					
	Mean	Median	Min	Max	σ	Coefficient of variation
Strain rate (mm/min or ^{o}/min)	1.11	1	0.10	3	0.67	0.60
Normal stress (kPa)	43.59	31.65	1	300	48.77	1.12
Contact area (cm²)	1521	100	36	9024	2994.94	1.97
Geomembrane thickness (mm)	1.68	1.50	0.50	3	0.45	0.27
Geomembrane asperity height (mm)	0.18	0.00	0.00	1.71	0.45	2.48
Soil relative density (%)	57.83	57.0	20.80	98.0	24.43	0.42
Soil specific unit weight (g/cm³)	1.55	1.54	1.45	2.94	0.29	0.18
Cc	1.03	0.98	0.80	1.68	0.37	0.36
Cu	2.82	1.75	1.11	46.8	5.92	2.09
D_{50} (mm)	0.58	0.52	0.17	3.08	0.60	1.04
ϕ_{soil} (^{o})	38.3	38.0	25.3	49.5	4.99	0.13
$\Phi_{Interface}$ (^{o})	28.2	28.0	7.5	64.5	7.32	0.26

For a better perception of the characteristics of the data, its central tendency is measured by mean and median. The mean is the average of the values of every parameter and it is given by the total of the values divided by the number of samples, as:

$$\bar{X} = \frac{(x_1 + x_2 + x_3 \dots x_n)}{n} \tag{3.1}$$

Where: $x_1, x_2, x_3 \; and \; x_n$ are individual measured values of every parameter, and n is the number of available values for each parameter.

The mean is simply an arithmetic average of the given data, searching for a calculated center. The actual central value of the parameters which is located in equal distance from the minimum and maximum extremes is identified by the median value. The median is the middle value in a sorted, ascending or descending order and can indicate a different central point of the data set than the average. It is the point above and below which half (50%) of the observed data falls, and so represents the midpoint of the data. The median formula of a given set of data having 'n' odd number of observations, can be expressed as:

$$median = \frac{(n+1)_{obs.}}{2} \tag{3.2}$$

Where: n is the total number of registered values.

For the 495 recorded samples the median of each variable is calculated as follows:

$$median = \frac{(495+1)_{obs.}}{2} \tag{3.3}$$

$$median_{495\ obs.} = \frac{496}{2} \tag{3.4}$$

Therefore, the 248^{th} observation is the median value.

As shown in Table 3.2, except for the applied normal stress and the contact area, the mean values are close to the median value of each parameter. This represents the collected data has symmetric behavior that its plot is dissecting at the middle of the graph and the it is not skewed (Swinscow, 1997). Skewed data means one range of measures is more represented in the set than the other measures. The mean and the median values are the indication of the statistical distribution of the gathered information.

3.2. PERFORMANCE MEASURES

In this study, two performance assessment criteria namely Coefficient of Determination (R^2) and Root Mean Square Error (RMSE) are used to evaluate the prediction capability of the RF model. R^2 is used to measure the ability of a RF model performance in the linear regression setting. It is an indication of the proportion of the variance in the dependent variable ($\hat{y}_i$), which is the predicted interface friction angle, based on the observed variable (y_i), which is the friction angle from the laboratory result and computed as:

$$R^2 = 1 - \frac{SS_{Resid}}{SS_{Total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{3.5}$$

Where:

$\bar{y}$ is mean of the observed data given as:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3.6}$$

$SS_{Residuals}$ is the sum of the square of residuals and calculated as:

$$SS_{Resid} = \sum_i (y_i - \bar{y}_i)^2 = \sum_i e_i^2 \tag{3.7}$$

And SS_{Total} is the total sum of the squares and given as:

$$SS_{Total} = \sum_{i} (y_i - \bar{y})^2 \tag{3.8}$$

If the value of R^2 gets close to 1, The values of Y become close to the regression line and similarly if it goes close to 0, the values get away from the regression line. This indicates higher values of R^2 close to 1 indicates a better prediction accuracy and lower values closer to 0 indicates a poor prediction.

The root mean square error (RMSE) is the difference between a predicted value by a model and the actual observed values. RMSE is defined as the square root of differences between predicted values and observed values. The individual differences in this calculation are known as "residuals". The RMSE estimates the magnitude of the errors. It is a measure of accuracy which is used to perform comparison forecasting errors from different estimators for a specific variable. The RMSE of an estimator (y_i) with respect to an estimated parameter ($\hat{y}_i$) is defined as the square root of the mean square error and computed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{3.9}$$

RMSE is a standard way to measure the error of a model in predicting quantitative data and a lower value of RMSE indicates a better performance of the model.

As the residuals which are the variation between observed and predicted interface friction angle are the main elements of the performance measurement, tracing the residuals and indicating their performance distribution is part of the assessment. Usually, residuals are random and unpredictable. However, a good residual distribution can be identified by their plot. A plot for a good residual errors distribution can be characterized by its high density close to zero error and a low density of points away from zero error. The other visual identification is, a good residual errors plot has a symmetric distribution about the origin or to the 0% error axis. Therefore,

residuals are traced and their plot on x-y plane is presented during the performance measure of the developed RF model.

3.3. PEARSON'S CORRELATION MEASURES

Based on the literatures studying geomembrane shear strength, the interface components are influencing the overall shear strength. The influence of each considered interface component on the interface friction angle, may vary depending on the other involved strands. For example; the level of the applied normal stress affects the interface shear strength outcome (Sánchez, 2018; Lashkari, 2021). However, this impact from the normal stress level is affected by the other parameters, such as the sliding velocity of the upper box in the IP tests (Carbone et al. 2015; Sánchez, 2018; Pavanello et al. 2021). This is due to when the ramp proceeds on tilting the upper box tends to slide quickly and the applied stress level tends to decrease. Therefore, there is inter-correlation between the affecting elements and the friction angle and also between the input variables as well.

The linear correlation between the input elements to one another and their influence on the overall shear strength parameter is outlined by using Pearson's correlation coefficient ($\bar{\rho}$). The Pearson coefficient is a type of correlation coefficient that represents the linear relationship between two variables that are measured on the same interval or ratio scale. It is a measure of the strength of the association between two continuous variables and computed as:

$$\bar{\rho} = \frac{Cov(x, y)}{\sigma_x \sigma_y} \tag{3.10}$$

where:

$\bar{\rho}$ = Pearson product-moment correlation coefficient

$Cov(x, y)$ = covariance of variables x and y

σ_x = standard deviation of x

σ_y= standard deviation of y

Based on the data distribution, the R is applied to determine the correlations between each pairwise variable, then a heat map consisting of all correlation coefficients is formed. Pearson's correlation coefficient measures linear relationships between variables. In this study to indicate the presence of a certain level of influence between variables the value $\bar{\rho} \geq |2|$ is considered (Schober et al 2018). For the correlations less than this value, the impact is considered too weak or referred as negligible influence.

3.4. RANDOM FOREST(RF)

In this study, RF which is a supervised Machine Learning algorithm used to estimate the friction angle of interface formed by sand and geomembrane. RF is a bagging technique-based statistical learning theory that uses the bootstrap resampling method and developed by Breiman (2001). In the RF method, the structure where decision trees are formed is called the forest. In the forest, each decision tree is created by selecting samples from the data set by row-sampling with the replacement technique and determining the number of random variables determined from all variables at each node.

RF regression involves the construction of k number of trees $\{Tk(x), m = 1,\ 2, \dots, k\}$. The p-dimensional vector $X = \{x_1, x_2, \dots x_3\}$, where p is the number of features in the dataset, that forms the forest is considered as the input vector of the algorithm. This ensemble generates k outputs corresponding to k number of trees and $\hat{y}_m$, $m = 1,\ 2, \dots, k$ is denoted as the output of each tree (note: $\hat{y}_m = Tk(x)$). Decision trees randomly select features to split the internal nodes and eventually form a random forest and the regression results are expressed as:

$$\bar{H}(x) = \frac{1}{K}\sum_{i=1}^{K}(h_i(x, \theta_k)) \tag{3.11}$$

where $\bar{H}(x)$ represents the prediction result, h_i is defined as a single decision tree; θ_k it is an independent and distributed random variable that determines the growth process of a single decision tree and K is the number of decision trees. Number of trees, maximum depth of trees

and minimum number of samples to split a node are called hyperparameters of a RF. A proper optimization of hyperparameters leads to a more accurate prediction of the algorithm (Zhou et al.2019). The average of the output of each developed tree is considered as the result of the algorithm (Zhang et al. 2021).

In the developed RF model, a new training dataset (bootstrap samples) is selected by replacing the original training dataset for building each regression tree structure. This led to several training data sets being omitted from the sample, which are reused. These omitted data which are known as out-of-bag (OOB) samples constitute one-fifth of new training samples in five-fold cross-validation method. The other four-fifth of the data is used to derive the regression function. Thus, a randomly drawn training sample from the original training set is selected for creating a decision tree each time, and one of the samples outside the bag is used for accuracy testing to develop a generalized RF model. The total learning error is denoted by $\dot{\hat{y}}_e$ and is given in Eq 3.12 and Eq. 3.13, respectively:

$$\hat{y}_i = (x_i)\frac{1}{k}\sum_{m=1}^{k}\hat{y}_m \tag{3.12}$$

$$\hat{y}_e = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{3.13}$$

where $\hat{y}_i$, y_i and n represent the prediction of each tree created by using OOB samples, the true output, and the total number of OOB samples, respectively. This error shows the prediction performance of the RF algorithm. Figure 3.2 illustrates the schematic view of the flowchart of the RF tree building and in order to improve the prediction ability of RF, Differential Evolution algorithm, which is a powerful minimizer of stochastic function is adopted (Pierre et al. 2021) to optimize hyperparameters of the developed model.

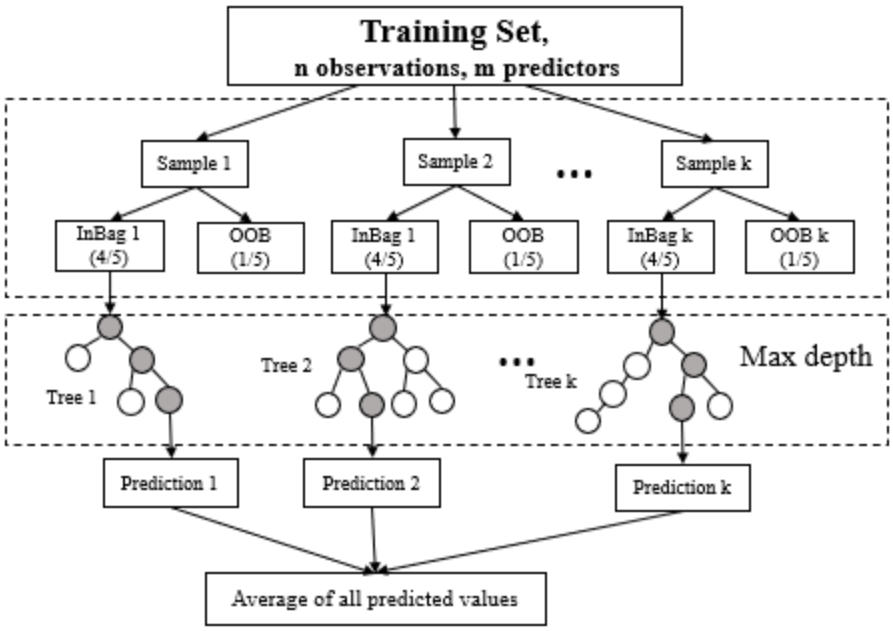


Figure 3.2. RF tree building (Pant and Ramana, 2021).

3.4.1. FEATURE IMPORTANCE ANALYSIS

The developed random forest model was tasked to estimate the relative importance of input features. Calculating the variable importance implies which of the features of the data are the most influential on the target variable. There is some randomness assigned to this process as the features that enter the contest for being selected on a node are chosen randomly and the variables chosen at each node are the ones that maximize the decrease of a certain error. For an individual decision tree, the most important variables are ranked by how much they reduce the error when they appear, having this error reduction weighted by the number of observations on the node, this is done for every tree in the forest, and then averaged to find the importance of an individual feature.

In RF for each tree in the forest, there is a misclassification rate for the out-of-bag observations. To assess the importance of a specific predictor variable, the values of the variable are randomly permuted for the out-of-bag observations, and then the modified out-of-bag data are passed down the tree to get new predictions. The difference between the misclassification rate for the modified and original out-of-bag data, divided by the standard error, is a measure of the importance of the

variable. Random forests compute how much each variable decreases the node impurity. The most important variable is the one that decreases the impurity the most (Nyongesa 2020). The final importance of the variable is the average of the impurity decrease for each variable across all the trees. Random forests use the out-of-bag samples to construct a different variable importance measure, apparently to measure the prediction strength of each variable. When the b^{th} tree is grown, the OOB samples are passed down the tree, and the prediction accuracy is recorded. Then the values for the j^{th} variable are randomly permuted in the out-of-bag samples, and the accuracy is again computed. The decrease in accuracy as a result of this permuting is averaged over all trees, and is used as a measure of the importance of variable j in the random forest (Hastie et al. 2017). The permutation feature importance directly measures feature importance by observing how random re-shuffling of each predictor influences model performance (Liaw & Wiener, 2019).

The approach can be described in the following steps:

1. Train the baseline model and record the MSE score by passing the OOB set. This can also be done on the training set, at the cost of sacrificing information about generalization.

2. Re-shuffle values from one feature in the selected dataset, pass the dataset to the model again to obtain predictions and calculate the metric for this modified dataset. The feature importance is the difference between the benchmark score and the one from the modified (permuted) dataset.

3. Repeat 2. for all features in the dataset.

The MSE computed as:

$$MSE_{OOB} = n^{-1} \sum_{1}^{n} \left\{ y_i - \widehat{y_i}^{OOB} \right\}^2 \tag{3.14}$$

Where $\widehat{y_i}^{OOB}$ is the average of the OOB predictions for the i^{th} observation.

And the percent variance for each variable computed as:

$$Percent\ Variance = 1 - \frac{MSE_{OOB}}{\partial_y^2} \quad (3.15)$$

3.5. DIFFERENTIAL EVOLUTION (DE)

Differential Evolution (DE) is a heuristic approach for the global optimization of nonlinear and non-differentiable continuous space functions (Storn and Price 1997). It starts with an initial population of candidate solutions. These candidate solutions are iteratively improved by introducing mutations into the population, and retaining the fittest candidate solutions that yield a lower objective function value. DE can handle nonlinear and non-differentiable multi-dimensional objective functions, while requiring very few control parameters to steer the minimization. These characteristics make the algorithm easier and more practical to use. The algorithm works in two phases, i.e., (i) Initialization: population is generated randomly and, (ii) Evolution: the generated population goes through mutation, crossover and selection processes which are repeated until a termination criterion is met.

(i) Initialization: During initialization, a set of uniformly distributed population is generated. Let $S^G = \{X_j^G : j = 1,2, \dots, NP\}$ be the population at any generation G, NP denotes the size of population. Here, X_j^G denotes a D-dimensional vector as $X_j^G = \{X_{1,J}^G, X_{2,J}^G, \dots, X_{D,J}^G\}$. X_j^G is generated using uniformly distributed random number $rand\ (0,1)$

$$X_j^G = X_{low} + (X_{upp} - X_{low}) * rand\ (0,1) \quad (3.16)$$

Where X_{low} and X_{upp} are lower and upper bounds of search space S^G.

Once the initial population is generated, the next phase of evolution is activated.

(ii) Evolution: This is the second phase where mutation, crossover and selection operations are performed.

(a) Mutation: In mutation a mutant vector V_j^G is generated for each target vector X_j^G at generation G as:

$$V_j^G = X_{r1}^G + F * (X_{r2}^G - X_{r3}^G) \tag{3.17}$$

where F is the scaling factor and value of is F is vary from 1 to 0 and $r_1, r_2, r_3 \in \{1, 2, \ldots, NP\}$ are mutually different, randomly chosen vectors as shown in Figure 3.3.

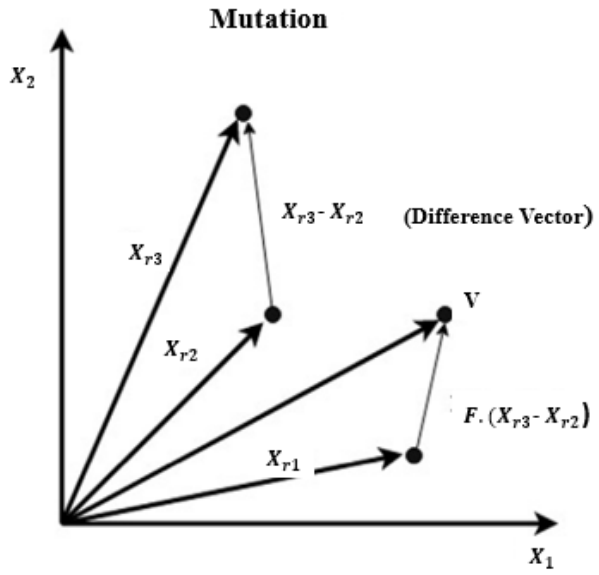


Figure 3.3. Mutation scheme of DE (Bilal et al. 2020).

(b) Crossover: After mutation, crossover is done to generate a new vector called trial vector denoted as $V_j^G = \{u_{1,J}^G, u_{2,J}^G, \ldots, u_{D,J}^G\}$. As shown in Figure 3.4, Crossover is performed between target vector $X_j^G = \{X_{1,J}^G, X_{2,J}^G, \ldots, X_{D,J}^G\}$ and mutant vector $V_j^G = \{v_{1,J}^G, v_{2,J}^G, \ldots, v_{D,J}^G\}$ using a crossover probability C_r whose value is between 01. U_j^G is generated as:

(c)

$$u_{i,j}^G = \begin{cases} u_{i,j}^G \; if \; rand_j \leq C_r \\ X_{i,j}^G \; otherwise \end{cases} \tag{3.18}$$

Where i ∈ {1, 2, …, D} and C_r ∈ [0,1].

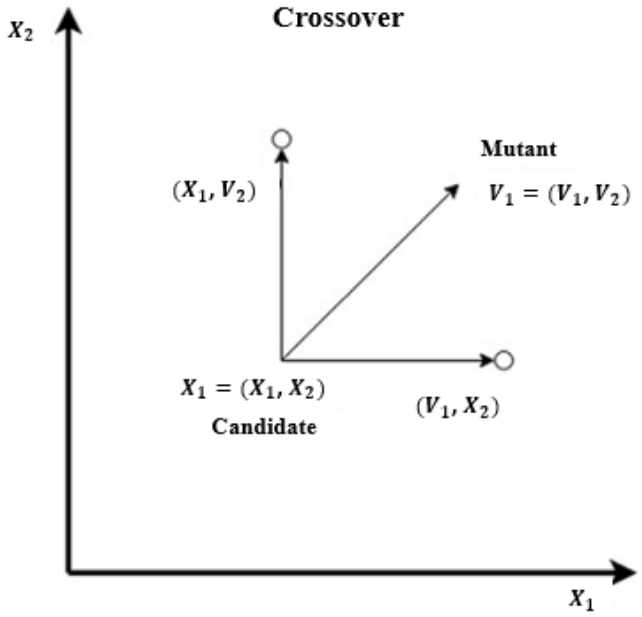


Figure 3.4. Crossover scheme of DE (Bilal et al. 2020).

(d) Selection: In this operation, a comparison is done between the target vector and trial vector according to their fitness value. The one having better fitness survives to the next generation. This operation is performed as:

$$X_j^{G+1} = \begin{cases} u_j^G \; if \; f(u_j^G) \leq f(X_j^G) \\ X_j^G \; otherwise \end{cases} \tag{3.19}$$

Mutation, Crossover and selection of evolution phase are repeated till a satisfying result is found and their flowchart of is given in Figure 3.5.

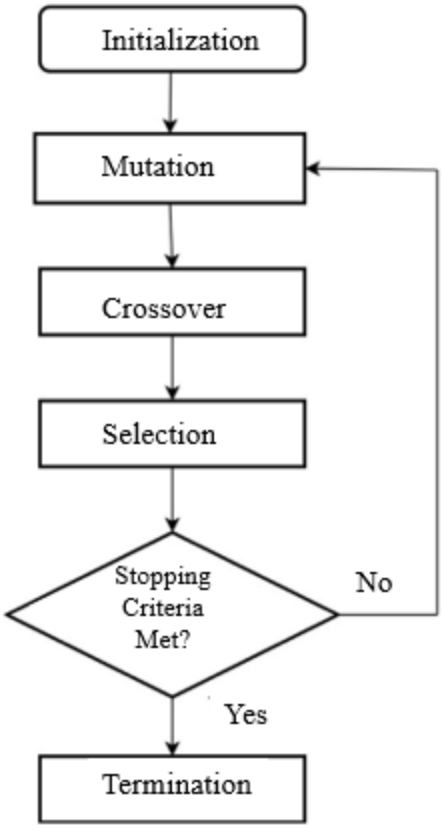


Figure 3.5. Initialization and evolution of DE (Bilal et al. 2020).

In this work the population analyzed by DE are not the samples from the data. Instead, the RF hyperparameters; number of trees, maximum depth and minimum number of samples to split a node are analyzed by mutation, crossover and selection steps of DE.

After the data behavior is observed, RF is model initialized. From the total gathered data, 80% is utilized for training the suggested model and the remaining 20% is processed for testing the ability (Pham et al.2020; Pant & Ramana2021). The number of trees, the maximum depth of a tree, and the minimum number of features to split a node are all optimized by DE. The main framework of the implemented methodology has been presented in Figure 3.6. and the optimized amount of each RF hyperparameter is summarized in Table 3.3.

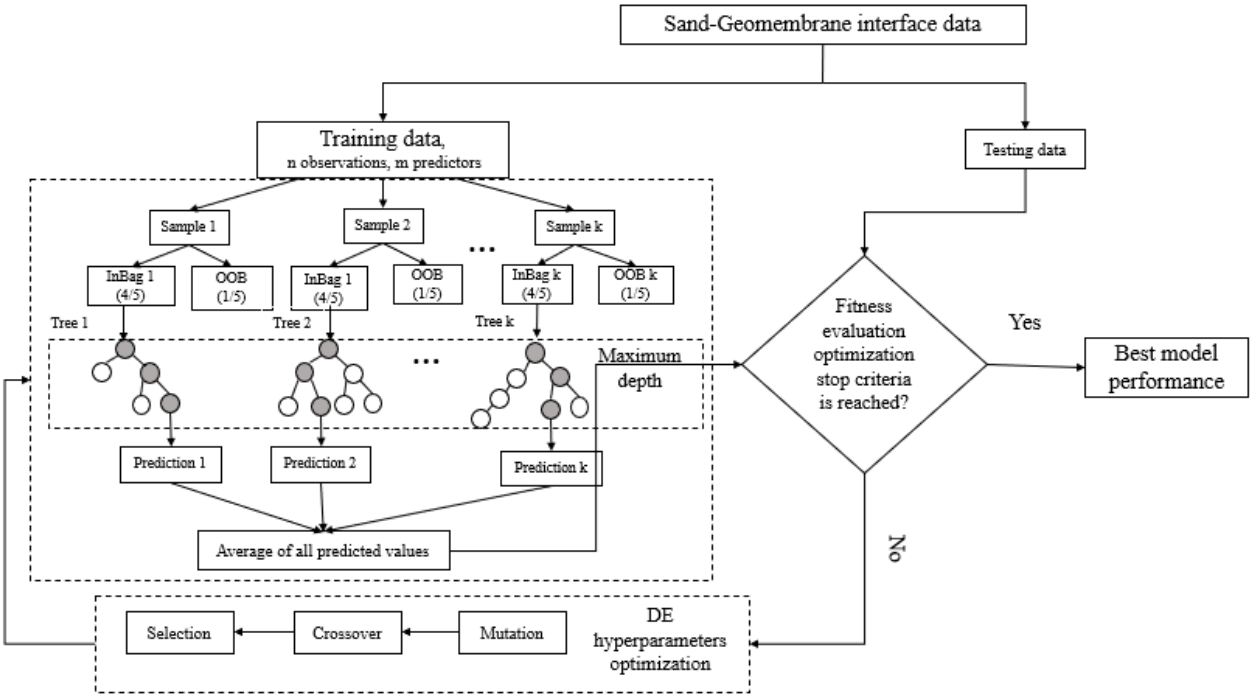


Figure 3.6. Flow chart of model training and hyperparameter optimization (modified from Zheng et al. 2019).

Table 3.3. RF hyperparameters tuned by DE.

RF hyperparameters	Optimum amount optimized by DE
Number of trees	103
Maximum depth	30
Minimum no. of samples to split a node	4

4. RESULTS

In this chapter the results from the applied methodologies are mentioned, interpreted and compared with the findings of the laboratory studies. The analysis is based on the statistical data accumulated from the actual laboratory investigations and the end results are compared and contrasted with the indications of the studies on interface shear strength. Before developing a RF, model and initialize the prediction process, the available data is scrutinized. The survey was first focused on measuring the Pearson's correlation coefficient between every contemplated parameter and interpreting the influence level in accordance with the laboratory results. Afterwards, the RF forest is initialized and linear and non-linear approaches are followed, and then the gathered data is divided in two sets of training and testing. Preceding the prediction, the main developed RF technique has been trained and the necessary hyperparameters were optimized by a DE algorithm. Following the training phase, the algorithm is tested and it is detected that the developed model predicted the interface friction angle adequately. The estimated values are in a very close range with the friction angles of the laboratory investigations. The performance of the applied RF technique is measured by two regression metrics, R^2 and RMSE and the model accuracy as well surveyed by tracing the residuals. After the prediction, feature importance analysis which is ranking the input variables based on their influence level is performed. Sorting out the parameter by their importance is further reviewed and contrasted with the literatures. Therefore, the basic feature of the analysis is examining the statistical and ML results side by side with the outcomes of the literatures.

4.1. RESULTS FROM LINEAR REGRESSION ANALYSIS AND RF LINEAR APPROACH

The influence extent of each considered interface parameter on its adjacent pair is herein firstly analyzed by simple linear regression. Pearson's correlation values have measured statistical interdependence and corresponding measurements are illustrated by Pearson's correlation coefficient matrix heat map as shown in Figure 4.1. There are various ranges for classifying Pearson's coefficient results based on their correlation strength. From studies utilizing Machine Learning algorithms for geotechnical parameters prediction, Zhang et al. (2021) considered

values $|\bar{\rho}| > 0.6$ as attestation of strong correlations. In Pant & Ramana (2021), a Machine Learning study for pullout coefficient of geogrids values, $|\bar{\rho}| > 0.5$ is considered as an indication of strong correlations. In this study, the classifications from Pant & Ramana (2021) are adopted and listed in Table 4.1. As listed in Table 4.1, there are 44 pairs of variables which are showing the presence of influence among the 105 pairs of variables analyzed herein. Of the 44 pairs, 37 pairs show moderate, 6 of them are showing strong and one of them is showing very strong linear correlations. This designates that 58% of the pairs have negligible linear relationships. With particular scrutiny on pairs involving interface friction angle and influencing elements, Pearson's correlation results are indicating negligible correlations for only 6 pairs and moderate to strong relationships for the remaining 8 matches. The majority of pairs between interface friction angle and affecting elements are stipulating the presence of influence by the considered strands on shear strength outcome. If the ruling observation focuses only on the Pearson's coefficient results it can mislead to a stipulation of a lack of influence between studied variables. However, the analysis can be interpreted as an implication of non-linearity between the pairs (Spanos 2003).

	Strain rate	Normal stress	Area	Thickness	Asperity	Relative Density	Soil unit weight	Cc	Cu	D_{50}	φ soil	CDS	MDS	IP	Φ Interface
Strain rate	1.00														
Normal stress	-0.46	1.00													
Area	0.50	-0.32	1.00												
Thickness	0.02	0.13	-0.24	1.00											
Asperity	-0.04	0.08	0.00	-0.02	1.00										
Relative Density	0.01	0.06	0.01	-0.15	0.24	1.00									
Soil unit weight	-0.26	0.23	0.02	0.15	0.34	0.24	1.00								
Cc	-0.09	0.04	0.26	-0.31	-0.33	0.16	0.14	1.00							
Cu	-0.04	0.21	0.03	0.10	0.00	-0.22	0.41	0.50	1.00						
D_{50}	-0.02	0.00	-0.04	-0.04	-0.06	0.17	-0.05	-0.18	-0.22	1.00					
φ soil	0.18	-0.02	-0.33	0.13	0.06	0.28	-0.24	-0.39	-0.07	0.12	1.00				
CDS	-0.63	0.46	-0.28	0.14	-0.01	0.02	0.19	-0.08	-0.09	-0.17	0.02	1.00			
MDS	-0.31	0.11	-0.12	-0.01	0.35	0.02	0.29	0.00	0.33	0.32	-0.16	-0.33	1.00		
IP	0.84	-0.49	0.34	-0.11	-0.30	-0.04	-0.41	0.07	-0.18	-0.07	0.10	-0.58	-0.58	1.00	
Φ Interface	-0.25	0.10	-0.07	-0.04	0.64	0.04	0.24	-0.02	0.16	-0.24	0.31	0.22	0.23	-0.39	1.00

Figure 4.1. Correlation coefficient matrix heatmap of the feature variables and their labels.

Table 4.1. Pearson's correlation results summary (Pant & Ramana 2021)

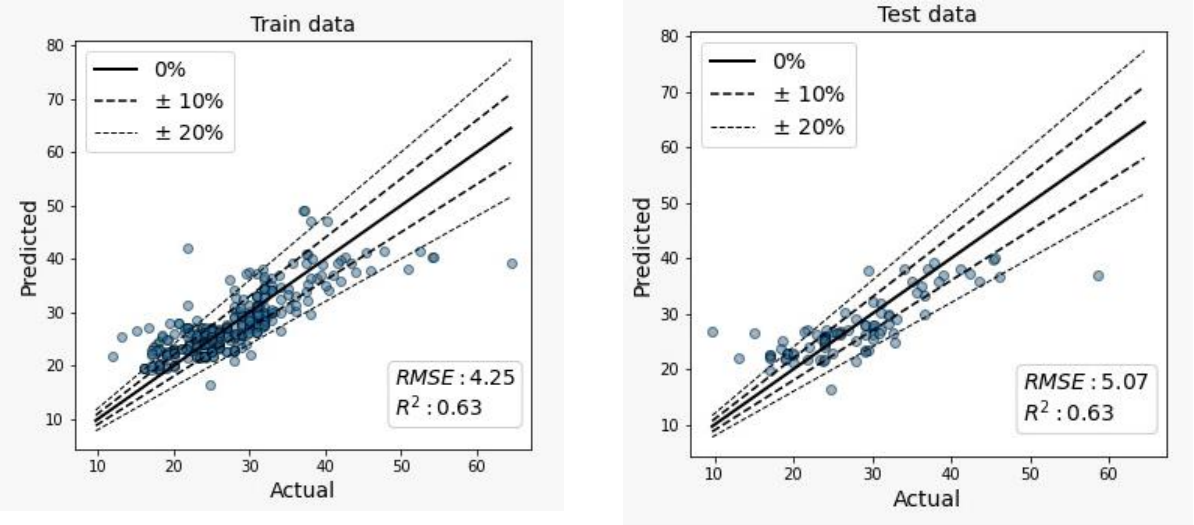
	Pearson's correlation value	Number of pairs
Very weak (negligible)	$0 < \|\bar{\rho}\| < 0.09$	38
Weak (negligible) correlations	$0.1 < \|\bar{\rho}\| < 0.19$	22
Moderate correlations	$0.2 < \|\bar{\rho}\| < 0.49$	37
Strong correlations	$0.5 < \|\bar{\rho}\| < 0.69$	6
Very strong	$0.7 < \|\bar{\rho}\| < 1$	1

According to Figure 4.1, IP tests presented a negative value ($\bar{\rho}$ = -0.39) correlated to interface shear strength. On the other hand, CDS and MDS presented positive correlations ($\bar{\rho}$ = 0.22 and $\bar{\rho}$ = 0.23; respectively), outlining the instrument's tendency of determining interface shear strength higher than IP tests (Izgin & Wasti, 1998; Fereira et al.2016). An increase in strain rate showed a tendency of reducing the interface shear strength ($\bar{\rho}$ = -0.25) and this inverse relationship agrees with Carbone et al. (2015), Sánchez (2018), and Pavanello et al. (2021). The correlation between soil unit weight and the interface friction angle has a positive result ($\bar{\rho}$ = 0.24) which agree with the phenomena that an increase in relative density improves the interface shear. This positive relationship between relative density and interface shear goes along with the findings of O'Rurke et al. (1990) and Afonso (2009). The increase in D_{50} is causing a decrease in the interface friction angle by having a negative value of $\bar{\rho}$ = -0.24, and this relationship agrees with Lopes et al. (2015). The Pearson's correlation coefficients are inferring particle size and relative density of soil have negligible influence on the interface friction angle by resulting in a value of $|\bar{\rho}| < 2$. The actual laboratory studies do not comply with this conclusion by identifying particle size and relative density as factors influencing interface shear strength (Frost et al. 2002, Choudhary et al. 2016). Asperity height highly increases the interface shear strength, as expected ($\bar{\rho}$ = 0.64). Bacas et al. (2015) and Araujo et al. (2022) concluded a similar relationship between asperity and interface shear. An increase in ϕ_{soil} improves the interface shear strength ($\bar{\rho}$ = 0.31) and it is already presented in Vangla et al. (2015).

Increasing the normal stress show influences on the results of IP and CDS tests, resulting $|\bar{\rho}|$ = 0.49 and 0.46; respectively and this positive and $|\bar{\rho}| > 0.2$ values agree with Sánchez, (2018) and Lashkari & Gamali (2021). However, the increase in applied normal stress impacted MDS in a

lower magnitude resulting, $|\bar{\rho}| = 0.11$, which is in the margin of no influence measure of $|\bar{\rho}| < 0.2$. Similarly, strain rate highly influences the IP ($\bar{\rho} = 0.84$) and CDS ($\bar{\rho} = -0.63$) tests and has lower impact on the MDS test ($\bar{\rho} = -0.31$). These Pearson's values are taken into account a linear correlation only, which may not always be correct to affirm.

As aforementioned above, the majority of matches pairing interface friction angle and input parameters are showing the presence of influence between the variables by applying simple linear correlation measurements. Considering these correlations RF algorithm is utilized to make prediction following a linear approach, aiming to compare the RF linear performance versus its inclusive nonlinear approach. The result of this prediction is targeted to evaluate the impact of available data behavior, which is linearity, and the effect of missing elements on the interface friction angle prediction. As shown in Figure 4.2 regression metric measures show $R^2 = 0.63$ and RMSE= 4.25 for the training set. For the testing set the prediction is measured as $R^2 = 0.63$ and RMSE=5.07.

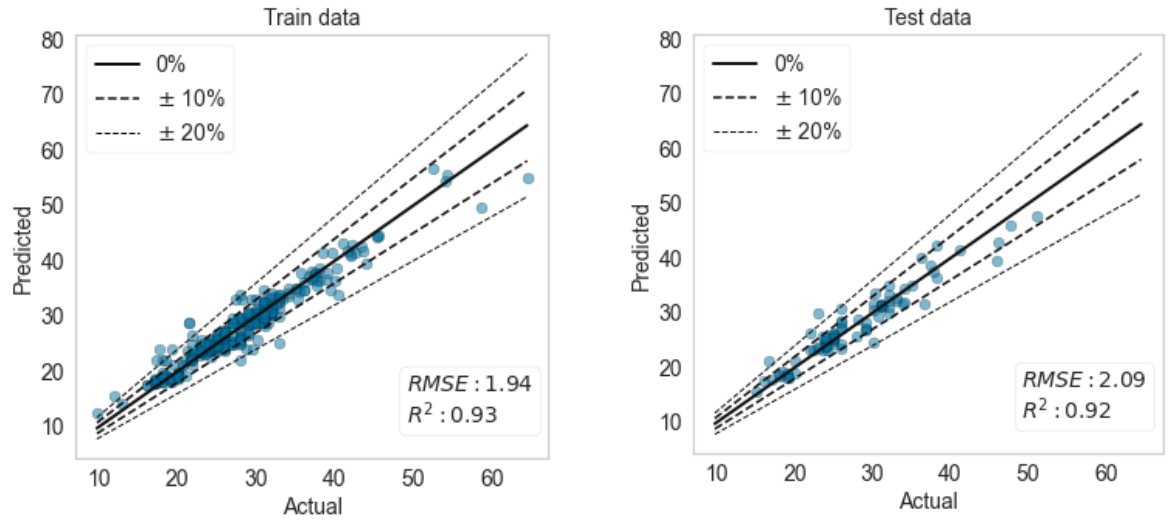


(a) (b)

Figure 4.2. RF linear prediction of interface friction angle, (a) for training set, (b) for testing set.

4.2. RANDOM FOREST INCLUSIVE APPROACH RESULTS

The prediction performance of RF is assessed by measuring the Coefficient of Determination (R^2) and Root Mean Square Error (RMSE). As shown in Figure 4.3, the interface friction angle estimation for the training data resulted in $R^2 = 0.93$ and RMSE = 1.94. And for prediction of the testing set, it is measured $R^2 = 0.92$ and RMSE = 2.09. This indicates a good agreement and less error between interface friction angle from RF estimation and laboratory test results. In civil and geotechnical engineering parameters prediction utilizing ML algorithms, $R^2 > 0.90$ is considered as an indication of a valid fit (Dutta et al. 2019; Farooq et al. 2020; Liu et al. 2020; Zhang et al. 2021). Even though there are variations in the type of data and applied algorithms through the reference, this value is acceptable for the studied interface parameter.

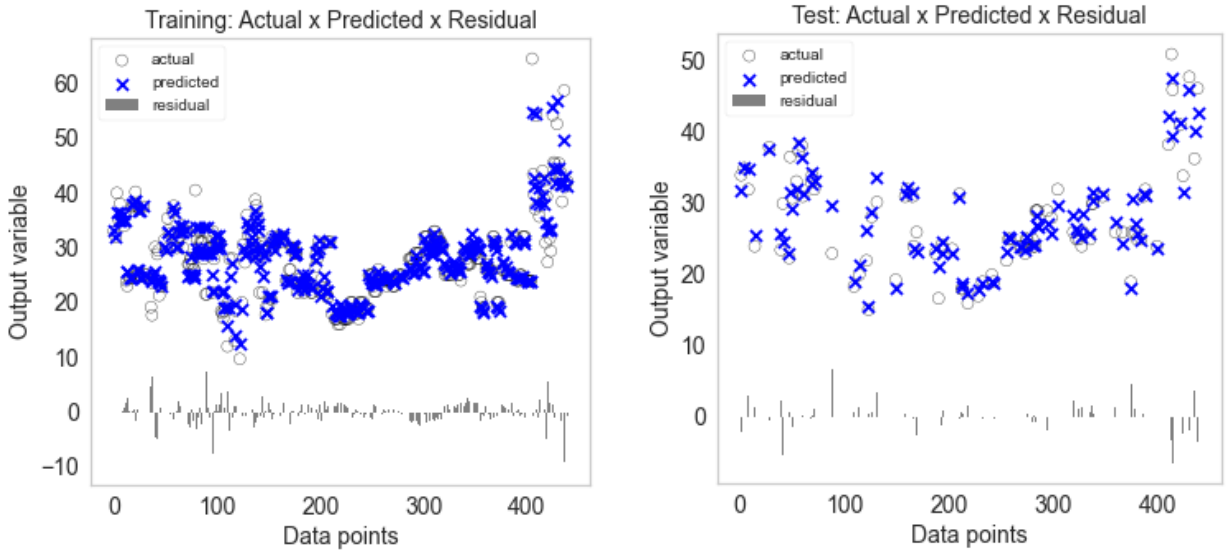


(a) (b)

Figure 4.3. RF nonlinear performance measurement: (a) for training set, and (b) for testing set.

A strong R^2 measurement can be confirmed by observing the residuals and with the analysis of prediction intervals (Frost 2019). The decrease in R^2 and the increase in RMSE from training to testing set shows the model is not over-fitted and this is also an indication that the model has functioned adequately.

The validation process of the RF algorithm was performed by tracing residual errors and is shown in Figure 4.4. It can be observed that the laboratory results (O marks) and the predicted interface friction angles (X marks) obtained by the RF model are in strong coherence. From the total collected interfaces, 80% of the data is utilized in the training set. As shown in Figure 4.4a, from 396 predicted interface shear strength, only for 12 of them the friction angle has a ±5° variation from laboratory test results. For the remaining 384 interfaces the difference between interface friction angles of RF estimation and laboratory test is lower than 5°. As aforementioned, 20% of the data is utilized in the testing set and Figure 4.4b shows that from 99 predicted interfaces, only for 8 of them the friction angle has outlined a ±5° difference from the literatures. For the remaining 91 interfaces, the difference between predicted and laboratory test friction angles remained lower than 5°.

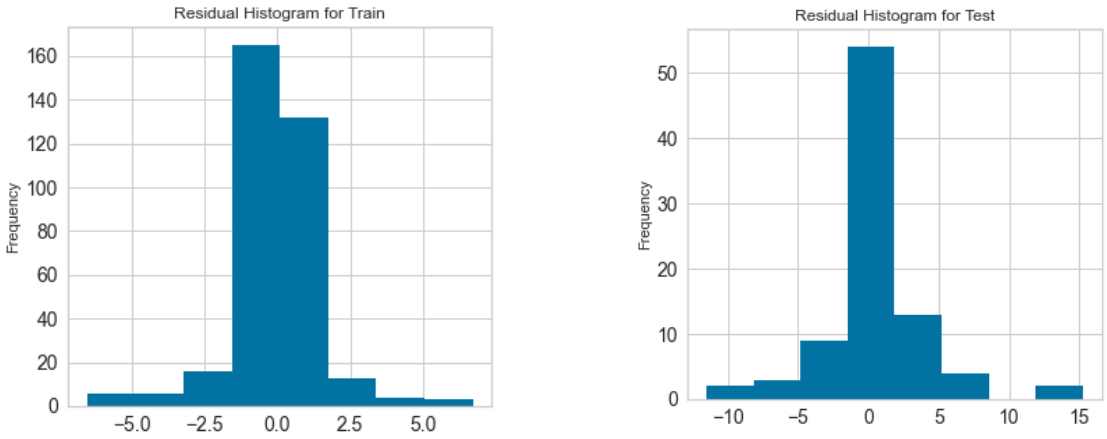


(a) (b)

Figure 4.4. Validation of RF for actual vs predicted interface friction angles comparison: (a) training set, (b) testing set.

Figure 4.4 also depicts the trace for the residuals which are the vertical distance between the predicted and laboratory result is shown for both training and testing sets. The results show that a larger residual error is indicated closer to the left and right end of the plot. This is due to the developed model distributing the errors symmetrically farther from the center which is keeping an approximately normal distribution. Smaller deviations which are located around the center of

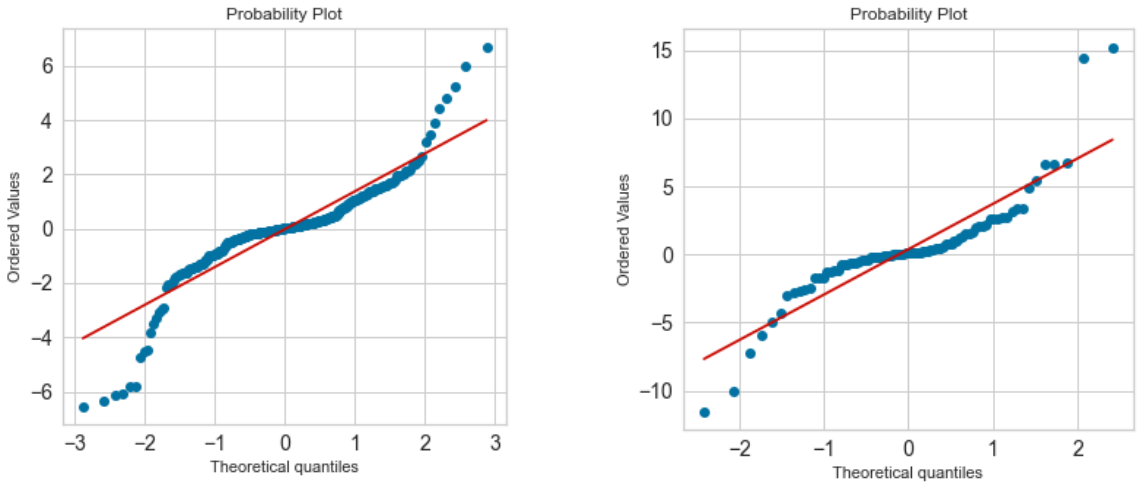
the plot can be considered as indications of a normal distribution as well. In the Histogram of the residuals, as shown in Figure 4.5 it is visible that the friction angle difference between predicted and laboratory results weighted around 0^o angle variation and that the friction angle showed ±5^o difference for fewer predicted interfaces. For the training, approximately 35 residual values were between ±5 and ± 2.5 degrees, and nearly 12 samples were higher than ±5 degrees. For the testing, approximately 20 residual values were between ±5 and 0 degrees and nearly 8 samples exhibit a difference higher than ±5 degrees. However, the prediction has shown differences mostly of about ±5 degrees. Most differences can be associated with various factors related to the applied tests, such as; - the size of the sample, type of the equipment, presence of asperities on the geomembranes, and size of sand particles. (Izgin et al. 1998; Frost et al. 2002; Gourc et al. 2003; Viana 2007; Pitanga et al.2009 & 2011; Carbone et al. 2012; Bacas et al. 2015; Carbone et al. 2015; Ferreira et al. 2016; Vangla et al. 2016; Punetha et al. 2017; Adeleke 2021; Araujo et al. 2022).



(a) (b)

Figure 4.5. Histogram of the residuals, training set, (b) testing set

Finally, to analyze the RF prediction reliability, the normality of the residuals is also measured. and is shown in Figure 4.6. The probability plot of the residual shows a thin-tailed distribution, which has a negligible deviation at the ends (Thode 2002). Thin-tailed probability plots have weighed the concentration of samples at the middle but not at the ends. Therefore, the probability plot can be considered as normally distributed and this indicates that the model is fit and stable.



(a) (b)

Figure 4.6. Probability plot of residuals, (a) training set and (b) testing set.

`

4.3. FEATURE IMPORTANCE ANALYSIS

The random forest model also can estimate the relative importance of input features which is an important reference for model selection and interpretation. As shown in Figure 4.7 for the collected data and the geomembrane and cohesionless soil interface, asperity height has the highest influencing factor. The remaining influencing parameters are listed through their rank accordingly and this indication is about which factor has a major influence on the developed model. Although most laboratory test results did not intend to make a comparison between factors on impact level, it is important to identify which input variable influenced most in the developed model. RF algorithm clearly shows the impact rank and quantified the influence level as well.

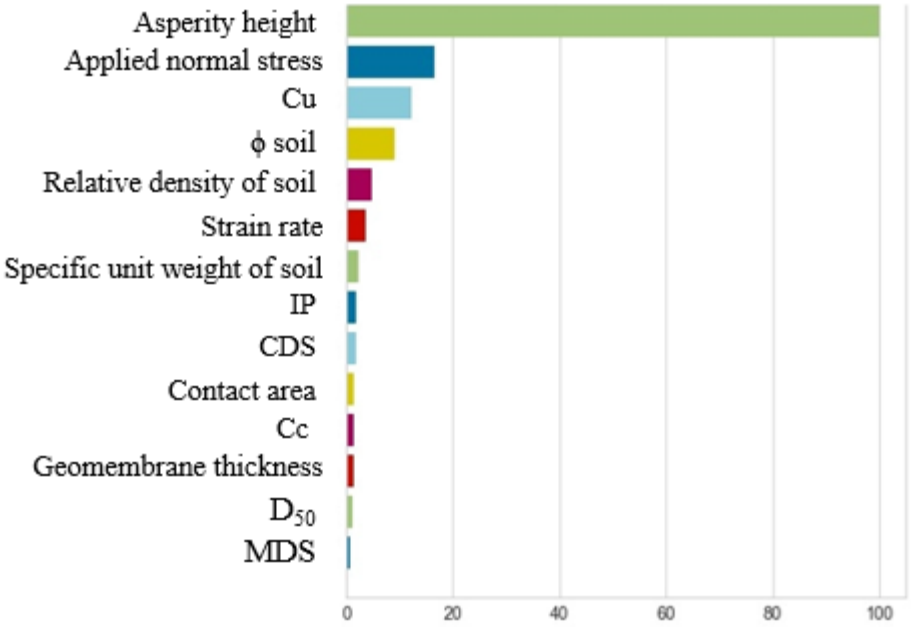


Figure 4.7. Relative importance measure for each influencing element from RF.

Asperity height of geomembrane has a definitive impact on interface shear strength results. As Bacas et al. (2015) showed smooth geomembranes with small asperity height exhibited lesser interface friction strength than textured geomembranes with higher asperity and roughness have resulted in higher interface shear strength. The increase in shear strength is resulted from a higher asperity structure and allows a greater interlocking mechanism with soil particles. The increase in shear strength by increasing asperity height is also indicated in Cen et al. (2018), in which the authors concluded that there was a 12-18% interface friction angle increase comparing smooth to a textured geomembrane. However, increasing asperity height is not always a factor in increasing interface friction strength. There is an optimum asperity height for each applied geosynthetics material (Adeleke et al. 2021, Araujo et al. 2022 for instance) and optimum asperity density will also contribute to higher shear strength besides height (Bacas et al.2015). Cen et al. 2018 studied that the influence of asperity height is also dependent on applied normal stress, gradation, and size of soil in the interface which is per the results exhibited by RF. Therefore, even though asperity height has a direct impact on the overall interface shear strength, it doesn't work solely (Bacas et al. 2015; Cen et al. 2018; Araujo et al. 2022). RF also presented that there are parameters with a high influence on the interface shear strength. Laboratory test results have indicated the impact of the remaining properties from RF relative importance rank.

Punetha et al. (2017) concluded that an increase in the relative density of sand improves interface strength. It is indicated by Lashkari & Jamali (2021) decreasing in relative density of soil particles exhibited decreasing in an interface shear strength.

Applied normal stress is the main feature of most geosynthetic interface studies. The applied stress level can differ based on the utilized equipment. Higher normal stresses can be applied for CDS and MDS apparatus and lower normal stresses should be applied for studies employing IP (Palmeira et al. 2004; Sánchez 2018; Pavanello et al. 2022). The applied normal stress affects a studied interface. Sánchez (2018) has shown that an increase in normal stress decreases the interface friction angle slightly. Lashkari & Jamali (2021) also showed that for sand and geomembrane interfaces peak interface friction angle increases when applied normal stress decreases. This normal stress-interface shear strength relation is dependent on the other parameters as well. Carbone et al. (2015), Sánchez (2018) and Pavanello et al. (2021) indicated that the effect of applied normal stress is dependent on the sliding velocity of the IP upper box. In this study, RF depicted variation of the interface shear strength as a function of the test velocity among other parameters.

The influence of particle size and relative density of soil on the shear strength interface shown in Fig, 4.7 is indicated by Frost et al. (2002), Vangla et al. (2015), and Choudhary et al. (2016).

Based on the findings of the kinds of literature the influence of interface parameters on the shear strength outcome is confirmed. RF analysis also outlined that every parameter investigated herein has an impact depending on the other interface elements, with some of them presenting a higher influence than others.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. CONCLUSIONS

This paper outlines the application of Random Forest for geomembrane interface shear strength analysis based on different types of tests and interfaces. A total number of 495 interface friction angles and 14 influencing factors for each interface are used to create the data set. Overall, the RFR algorithm worked adequately in predicting interface friction angle which consists of geomembrane and cohesionless soil.

Based on the analysis from RF regression results the following conclusions are driven;

Pearson's correlation coefficient measurements showed limitations for the investigation of geomembrane-sand interface shear strength. Even though 8 out of 14 pairs between interface friction angle and affecting elements showed the presence of influence, it is not possible to affirm all the indications according to the actual laboratory results.

The optimized Random Forest regression algorithm predicted geomembrane-sand interface friction angle adequately.

The Random Forest regression model functioned better with the nonlinearity of the data and delivered a prediction in strong coherence with laboratory test results

The Random Forest regression algorithm has identified the relative importance of input elements for the developed model. Asperity height of geomembrane is by far the most important feature in the selected model, followed mainly by Applied normal stress, C_u, ϕ_{soil}, Soil relative density, and the strain rate with their relative importance of 18%, 13%, 8%, 5%, and 4% respectively. The remaining parameters are labeled for having small relative importance and having a negligible influence on the developed model.

5.2. RECOMMENDATIONS FOR FURTHER STUDIES

Based on the results obtained in this study, some recommendations can be proposed for further research on the study of the sand-geomembrane interface:

- To perform a RF prediction based on the data from a single test type, with a maximum possible registration of interface components. Referring to that, there were some missing elements in the gathered study data and the results from three different equipment compiled together.

- To perform a sand-geomembrane interface shear strength prediction by utilizing Gradient boosting RF and comparing with OOB (out-of-bag) technique.

- To employ other heuristic hyperparameter optimization algorithms and comparing with the results of RF model optimized by DE.

- To execute geomembrane interface laboratory studies with a better data registration aiming to contribute more for analyses and predictions utilized by Machine Learning algorithms.

REFERENCES

Abu-Farsakh, M., Colonel, J., & Tao, M., (2007). Effect of soil moisture content and dry density on cohesive soil-geosynthetic interactions using large direct shear tests. Journal of Materials in Civil Engineering, 19 (7), 540–549.

Afonso, M.R.F.L. (2009). Ensaios de corte direto na caracterização da interface solo-geossintético: efeito da variação da tensão normal. Dissertação de Mestrado. Universidade de Porto, Portugal. 82p.

Afzali-Nejad, A., Lashkari, A., & Shourijeh, P.T. (2017). Influence of particle shape on the shear strength and dilation of sand-woven geotextile interfaces. Geotextile and Geomembrane 45 (1), 54-66.

Akan, R., Keskin, S.N., & Uzundurukan, S. (2015). Multiple Regression model for the prediction of unconfined compressive strength of jet grout columns. Procedia Earth and Planetary science 15, 299-303.

Akbar, S., Massoud, T., & Raziyeh, F., (2014). A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks. KSCE J. Civil Eng. 18 (4), 941–948.

Araujo, G.L.S., Sánchez, N.P., Palmeira, E.M., & Maria, G.G.A., (2022). Influence of micro and macroroughness of geomembrane surfaces on soil-geomembrane and geotextile interface strength. Geotextiles and Geomembrane, 03/015.

ASTM D7466. (2015). Standard Test Method for Measuring Asperity Height of Textured Geomembranes. ASTM International. (Reapproved):4.

Bacas, B.M., Cañizal, J., & Konietzky, H. (2015). Frictional behavior of three critical geosynthetic interfaces. Geosynthetics International, 22 (5), 355–365.

Blond, E., & Elie, G. (2006). Interface shear-strength properties of textured polyethylene geomembranes, Proceedings of the Sea to Sky Geotechnique, (pp, 898–904). Quebec, Canada.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J., (1984). Classification and Regression Trees.

Breiman, L., (1996). Bagging predictors. Machine Learning. 24, 123–140.

Breiman, L., (1999). Random Forests--Random Features. Technical Report, Statistics Department, University of California, Berkeley. Cite seer. 567p.

Breiman, L., (2001). Random forests. Machine Learning 45 (1), 5–32.

Carbone, L., Briançon, L., Gourc, J.P., Moraci, N. & Carrubba, P. (2012). Geosynthetic interface friction using force procedure at the tilting plane, 93-98.

Brownlee, J. 2020. Linear regression for Machine Learning. © 2022 Machine Learning Mastery.

Carbone, L., (2014). Interface behavior of geosynthetics in landfill cover systems under static and seismic loading conditions. PhD Dissertation. Department of Civil Engineering, Energy, Environment and Materials Engineering, University Mediterranea of Reggio Calabria. 245p.

Carbone, L., Gourc, J.P., Carrubba, P., Pavanello, P., & Moraci, N., (2015). Dry friction behavior of a geosynthetic interface using inclined plane and shaking table tests. Geotextile and Geomembranes 43 (4), 293–306.

Cen, W., Wang, H., Sun, Y., (2018). Laboratory Investigation of Shear Behavior of High-Density Polyethylene Geomembrane Interfaces. Polymers, 10, 734.

Choudhary, A., K., & Krishna, A., M., (2016). Experimental Investigation of Interface Behavior of Different Types of Granular Soil/Geosynthetics. International Journal of Geosynthetic and Ground Engineering 2(4), 1-11.

Cortes, C. & Vapnik, V. (1995). Support Vector Networks. Machine Leaming, © Kluwer Academic Publishers, Boston. Manufactured in The Netherlands, (20), 273-297.

DeJong, J.T., & Westgate, Z.J., (2005). Role of over consolidation on sand-geomembrane interface response and material damage evolution. Geotextile and Geomembranes, 23 (6), 486–512.

Dinh, T. V., Nguyen, H., Tran, X., & Hoang, N. (2021). Predicting Rainfall-Induced Soil Erosion Based on a Hybridization of Adaptive Differential Evolution and Support Vector Machine Classification. Hindawi Mathematical Problems in Engineering, 6647829. 20p.

Dutta, R.K., Gnananandarao, T., & Sharma, A. (2019). Application of Random Forest regression in the prediction of ultimate bearing capacity of strip footing resting on dense sand overlying loose sand deposit. Journal of Soft Computing in Civil Engineering (3-4), 28-40.

Eberhart, R., & Kennedy, J. (1995). A new optimizer using particle swarm theory. In Proceedings of the MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 39–43.

Farooq, F., Amin, M.N., Khan, K., Sadiq, M.R., Javed, M.F., Aslam, F., & Alyousef, R. (2020). A Comparative study of Random Forest and Genetic Engineering Programming for the

prediction of compressive strength of High Strength Concrete (HSC). Applied Science 10, (7330).

Feng, S.J., Liu, X., Chen, H.X., & Zhao, T., (2018). Micro-mechanical analysis of geomembrane sand interactions using dem. Computer Geotech. (94), 58–71.

Ferreira, F. B., Vieira, C. S., & Lopes, M. L., (2012). Experimental investigations on shear strength of soil–geogrid interfaces. Proceedings of the 5th European Geosynthetics Congress, Valencia, Spain, (5), 211–217.

Ferreira, F.B., Vieira, C. S., Lopes, M. L. (2015). Direct shear behavior of residual soil–Geosynthetic interfaces - influence of soil moisture content, soil density and geosynthetic type. Geosynthetics International, 22 (3), 257–272.

Ferreira, F.B., Vieira, C.S., & Lopes, M.L. (2016). Soil-geosynthetic interface strength properties from inclined plane and direct shear tests-A comparative analysis. In Proceedings of the 6th Asian Regional Conference on Geosynthetics: Geosynthetics for Infrastructure Development, 925–937.

Fuggle, A.R., & Frost, J.D., (2010). Particle size effects in interface shear behavior and geomembrane wear. Proceedings of International Symposium on Characterization and Behavior of Interfaces, Atlanta, 51-57.

Frost, J.D., DeJong, J.T., & Recalde, M. (2002). Shear failure behavior of granular and continuum interfaces. Engineering Fracture Mechanics, 69 (17), 2029-2048.

Frost, D.J., Kim, D., & Lee, S.W., (2012). Microscale geomembrane-granular material interactions. KSCE Journal of Civil Engineering, 16(1), 79–92.

Frost, J. (2019). Regression analysis, an intuitive guide for using and interpreting linear models. © by Jim Frost. 358p.

Girard, H., Fischer, S., & Alonso, E., (1990). Problems of friction posed by the use of geomembranes on dam slopes - examples and measurements. Geotextiles and Geomembranes, 9 (2), 129-143.

Ghahramani, Z., (2015). Probabilistic machine learning and artificial intelligence. Nature 521 (7553), 452.

Gourc, J.P., Reyes & Ramirez, R., (2003). Use of the inclined plane test in measuring geosynthetic interface friction relationship. Geosynthetics International, 10(5), 165–175.

Han, Q., Gui, C., Xu, J., & Lacidogna, G. (2019). A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. Construction and Building Materials, © Elsevier Ltd. All rights reserved (226), 734–742.

Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning Data Mining, Inference, and Prediction. Springer series in statistics. Second edition. 764p.

Ho, T.K., (1995). Random decision forests. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition IEEE, (1), 278–282.

Hoang, N.D., Pham, A.D., & Quốc, L.N. (2016). Estimating Compressive Strength of High-performance Concrete with Gaussian Process Regression Model. Advances in Civil Engineering.

Huang, J., Duan, T., Zhang, Y., Liu, J., & Zhang, J. (2020). Predicting the permeability of pervious concrete based on the Beetle Antennae Search algorithm and Random Forest model. Advances in Civil Engineering, Article ID 8863181.

Izgin, M. & Wasti, Y. (1998). Geomembrane–sand interface frictional properties as determined by inclined board and shear box tests. Geotextiles and Geomembranes, 16(4): 207–219.

Jewell, R., (1996). Soil Reinforcement with Geotextiles. Special Publication. Construction Industry Research & Information Association (CIRIA).

Jeremiah, J. J., Abbey, S.J., Booth, C. A., & Kashyap, A. (2021). Results of Application of Artificial Neural Networks in Predicting Geo-Mechanical Properties of Stabilized Clays—A Review. Geotechnics, (1) 147–171.

Jiang, A., Guo, X., Zheng, S., & Xu, M. (2021). Parameters Identification of Tunnel Jointed Surrounding Rock Based on Gaussian Process Regression Optimized by Difference Evolution Algorithm. Tech Science Press, 127 (3),1177-1199.

Kang, K., Song, K., An, J., & Kim, B. (2020). Optimization of soil slope stabilization with evolution algorithm Applied Numerical Modeling in Geomechanics, Billaux, Hazzard, Nelson & Schöpfer (eds.),15-03.

Koutsourais, M. M., Sprague, C. J., & Pucetas, R.C. (1991). Interfacial Friction Study of Cap and Liner Components for Landfill Design. Geotextiles and Geomembranes 10 (1), 531-548.

Lashkari, A. & Jamali, V. (2021). Global and local sand–geosynthetic interface behavior. Géotechnique 71, (4) 346–367.

Li, S., He P., Li, L., Shi, S., Zhang, Q., Zhang, J., & Hu, J. (2017). Gaussian process model of water inflow prediction in tunnel construction and its engineering applications. Tunnelling and Underground Space Technology, (69),155-161.

Liaw, A. & Wiener, M. (2002). Classification and Regression by Random Forest. Vol 2/3, ISSN 1609-3631.

Lima Júnior, N.R., (2000). Estudo da interação solo-geossintético em obras de proteção ambiental com o uso do equipamento de plano inclinado. Dissertação de Mestrado, Universidade de Brasília, Brasil. 148p.

Linares-Unamunzaga, A., Pérez-Acebo, H., Rojo, M., Gonzalo-Orden, H. (2019). Flexural Strength Prediction Models for Soil–Cement from Unconfined Compressive Strength at Seven Days. Materials, 12,387.

Liu, C.N., Ho, Y.H., & Huang, J.W., (2009). Large scale direct shear tests of soil / PET-yarn geogrid interfaces. Geotextiles and Geomembranes, 27 (1), 19–30.

Liu, H., & Martinez, J., (2014). Creep behavior of sand–geomembrane interfaces. Geosynth. Int. 21 (1), 83–88.

Liu J., & Qiao, S. (2015). The Image segmentation algorithm based on Differential Evolution Particle Swarm Optimization Fuzzy C-Means Clustering. Computer Science and Information Systems 12(2),873–893.

Liu, P., Xianguo, W., Hongyu, C., & Tiemei, Z. (2020). Prediction of compressive strength of High-Performance Concrete by Random Forest algorithm. Conference Series: Earth and Environmental Science 552 (01/2020).

Lopes, C.P.F.C. (2001). Study of soil and geosynthetic interaction by shear tests and inclined board tests. MSc Dissertation, Civil Engineering University, Porto City, Portugal, (p, 186).

Lu, P., Chen, S., Sheng, X., & Gao, Y. (2022). Application of the Differential Evolutionary Algorithm to the Estimation of Pipe Embedding Parameters. Sensors 22(10),3942.

Ly, H., & Pham, B. T. (2020). Soil unconfined compressive strength prediction using Random Forest (RF) Machine Learning model. The Open Construction & Building Technology Journal, (14), 278-285.

Ly, H-B., Nguyen T-A., & Pham, B.T. (2022). Investigation on factors affecting early strength of high-performance concrete by Gaussian Process Regression. PLoS ONE 17(1): e0262930.

Nair, S.J., KishoShahina., K.S., & Brightson P.C. (2018). Geotechnical investigation of different soil samples using Regression Analysis. International Research Journal of Engineering and Technology (IRJET), 05, (04).

Nyongesa, D. (2020). Variable selection using Random Forests in SAS. SAS Global Forum: 4826.

Manhour, E.A. (2020). Prediction of compression index of the soil of Al-Nasiriya city using Simple Linear Regression Model. Geotechnical and geological engineering © Springer Nature Switzerland AG.

Mello, L.G.R., (2001). Estudo da interação solo-geossintético em taludes de obras de provisão de resíduos. Dissertação de Mestrado, Universidade de Brasília, Brasil. 137p.

Mello, L.G.R., Lima Júnior, N.R., & Palmeira, E.M., (2003). Study of the interaction between soil and geosynthetic interfaces in waste provision areas. Soils and Rocks, 1(26), 19-35.

Murphy, K.P., (2012). Machine learning: a probabilistic perspective. Adaptive computation and machine learning series. ISBN 978-0-262-01802-9.

Ocak, I., & Seker, S. E., (2013). Calculation of surface settlements caused by EPBM tunneling using artificial neural network, SVM, and Gaussian processes. Environ. Earth Sci. 70, 1263–1276.

O'Rourke, T.D., Druschel, S.J., & Netravali, A.N., (1990). Shear strength characteristics of sand-polymer interfaces. Journal of Geotechnical Engineering, 116 (3), 451–469.

Ou, Z. (2015). Differential Evolution's Application to Estimation of Soil Water Retention Parameters. Agronomy, 5(3), 464-475.

Pal, M., Singh, N.K., Tiwari, N.K., (2013). Pier scour modeling using random forest regression. ISH J Hydraulic Engineering 19,69–75.

Palmeira, E.M. (1987). The study of soil-reinforcement interaction by means of large-scale laboratory tests. Thesis. University of Oxford. 237p.

Palmeira, E.M., Lima Junior, N.R., Viana, H.N.L., Mello, L.G.R., (2004). Large scale ramp test for the study of soil-geosynthetic interaction in slope of waste disposal areas. EuroGeo (3). 6p.

Palmeira, E.M., (2009). Soil–geosynthetic interaction: Modeling and analysis. Geotextiles and Geomembranes, 27 (5), 368–390.

Pan, Q., Qu, X., Liu, L., Dias, D., (2020). A sequential sparse polynomial chaos expansion using Bayesian regression for geotechnical reliability estimations. International Journal for Numerical and Analytical Methods in Geomechanics, 44, 874–889.

Pant, A. & Ramana, G.V. (2021). Novel application of machine learning for estimation of pullout coefficient of geogrid. Thomas Telford Ltd. Geosynthetics International, 1072-6349.

Pavanello, P.C., Moraci, N., (2021). The characterization of geosynthetic interface friction by means of the inclined plane test. Geotextiles and Geomembranes, 49, 257–275.

Pavanello, P., Carruba, P., Moraci, N. (2022). Geosynthetic interface friction at low normal stress: Two approaches with increasing shear loading. Applied Science, 12, 1065.

Pham, B. T., Qi, C., Ho, L. S., Nguyen-Thoi, T., Al-Ansari, N., Nguyen, M. D., Nguyen, H.D., Ly, H., Le, H.V., & Prakash, I. (2020). A Novel Hybrid Soft Computing Model using Random Forest and Particle Swarm Optimization for Estimation of Undrained Shear Strength of Soil. Sustainability, 12, 2218.

Pham, B. T., Nguyen M. D., Ly H., Pham T. A., Hoang V., Le H. V., Le T., Nguyen H. Q., & Bui G. L. (2020b). Development of Artificial Neural Networks for prediction of compression coefficient of soft soil. CIGOS, Innovation for Sustainable Infrastructure, Lecture Notes in Civil Engineering 54.

Pierre, G. A. N., Shui-Long, S., Annan, Z., Giuseppe, M. (2021). Artificial neural network optimized by differential evolution for predicting diameters of jet grouted columns. Journal of Rock Mechanics and Geotechnical Engineering (13),1500-1512.

Pintelas, P., & Livieris, I.E. (2020). Special Issue on Ensemble Learning and Applications. Algorithms, 13(6), 140. 4p.

Pitanga, H.N., Gourc, J.P., & Vilar, O.M., (2009). Interface shear strength of geosynthetics: Evaluation and analysis of inclined plane tests. Geotextile and Geomembranes, 27 (6), 435–446.

Pitanga, H.N., Gourc, J.P., & Vilar, O.M. (2011). Enhanced Measurement of Geosynthetic Interface Shear Strength Using a Modified Inclined Plane Device. Geotechnical Testing Journal, 34(6), 643–652.

Puri N, Prasad HD, & Jain A. (2018). Prediction of Geotechnical Parameters Using Machine Learning Techniques. Proceeded by Computational Science (125), 509–517.

Punetha, P., Mohanty, P.P., & Samanta, M. (2017). Microstructural investigation on mechanical behavior of soil-geosynthetic interface in direct shear test. Geotextiles and Geomembranes (45), 197-210.

Punetha, P., Samanta, M., Mohanty, P. (2019). Evaluation of the dynamic response of geosynthetic interfaces. International Journal of Physical Modeling in Geotechnics 19 (3), 141-153.

Ramya, D. & Vinodhkumar, S. (2017). Development of Support Vector Machine Model to Predict Stability of Slopes Based on Bound Theorems. International Journal of Engineering and Technology, Vol 9 (2).

Rasmussen, C.E., & Williams, C.K.I. (2006). Gaussian Processes for Machine Learning. The MIT Press, Cambridge.

Rohmer, J., & Foerster, E. (2011). Global sensitivity analysis of large-scale numerical landslide models based on Gaussian-Process metamodeling. Comput. Geosci. 37 (7), 917–927.

Roy, S. (2016). Assessment of Soaked California Bearing Ratio value using geotechnical properties of soils. Resources and Environment, 6(4), 80-87.

Sánchez, N.P., (2018). Estudo de alguns aspectos que influenciam a adesão entre geossintéticos e diferentes materiais. Tese de Doutorado, Programa de Pós-Graduação em Geotecnia, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF. 168p.

Schober, P., Boer, C., Schwarte, L. A., (2018). Correlation coefficients: Appropriate use and interpretation. Anesthesia and Analgesia 126 (5), 1763-1768.

Schonlau, M., & Zou, R.Y. (2020). The random forest algorithm for statistical learning, The Stata Journal, 20(1), 3–29.

Shin, T. (2021). Understanding Feature Importance and How to Implement it in Python. Towards Data Science 26/2/2021.

Spanos, A. (2003) Probability theory and statistical inference: Econometric modeling with observational data. © Cambridge University Press (Virtual Publishing). 844p.

Storn, R., & Price, K. (1997). Differential Evolution – A simple and heuristic for global optimization over continuous spaces. Journal of global optimization (11), 341-359.

Sun, Y., Li, G., Zhang, J., Qian, D. (2019). Prediction of the strength of rubberized concrete by an Evolved Random Forest model. Advances in Civil Engineering, volume, Article ID 5198583. 7p.

Suwansawat, S., Einstein, H.H. (2006). Artificial neural networks for predicting the maximum surface settlement caused by EPB shield tunneling. Tunneling and Underground Space Technology incorporating Trenchless Technology, 21, (2)133-150.

Swinscow, T. D.V. (1997). Statistics at square one. Revised by M J Campbell, University of Southampton, Copyright BMJ Publishing Group.

Thode Jr., H. C. (2002). Testing for normality. Copyright © by Marcel Dekker, Inc. All Rights Reserved. 368p.

Truong, V.H., & Pham, H.A. (2021). Support Vector Machine for Regression of Ultimate Strength of Trusses: A Comparative Study. Engineering Journal. (25), 7.

Vangla, P. & Madhavi, L.G. (2014). Characterization of soil-geosynthetic interaction based on surface roughness studies. Geo-Congress Technical Papers, GSP 234.

Vangla, P., Latha, G.M. (2015). Influence of particle size on the friction and interfacial shear strength of sands of similar morphology. International Journal of Geosynthetic and Ground Engineering, 1(1), 1-12.

Vangla, P., & Gali, M.L., (2016). Shear behavior of sand-smooth geomembrane interfaces through micro-topographical analysis. Geotext. Geomembranes 44 (4), 592–603.

Viana, H.N.L. (2007). Estudos da estabilidade e condutividade hidráulica de sistemas de revestimento convencional e alternativos para obras de disposição de resíduos. Tese de Doutorado. Universidade de Brasília. Brasil. 259p.

Vieira, C.S., Lopes, M.L & Caldeira L. (2013). Soil-Geosynthetic Interface Shear Strength by Simple and Direct Shear Test. International Conference on Soil Mechanics and Geotechnical Engineering, Paris, (18),5.

Vu, V. T. (2022). Assessment of Slope Stability under Bounded Uncertainties Using Differential Evolution and Particle Swarm Optimization. 18, (1) 59 - 65.

Williams, C.G, & Ojuri, O.O. (2021). Predictive modelling of soils' hydraulic conductivity using artificial neural network and multiple linear regression. SN Applied Sciences (3),152.

Yesiller, N. 2005. Core thickness and asperity height of textured geomembranes: a critical review Experimental case study. GFR Engineering Solutions. 23(4), 4–7.

Zhang, W., Wu, C., Li, Y., Wang, L. & Samui, P. (2019). Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, (15): 27-40.

Zhang, W., Wu, C., Zhong, H., Li, Y., & Wang, L. (2021). Prediction undrained shear strength using extreme gradient boosting and Random Forest on Bayesian optimization. Geoscience Frontiers (12), 469–477.

Zhang, W., Zhang, R., Wu, C., Goh, A.T.C., & Wang, L. (2022). Assessment of basal heave stability for braced excavations in anisotropic clay using extreme gradient boosting and random forest regression. Underground Space 7, 233–241.

Zhou, J., Shi, X., Du, K., Qiu, X., Li, X., & Mitri, H.S. (2016). Feasibility of random-forest approach to prediction of ground settlements induced by the construction of a shield-driven tunnel. International Journal GeoMech. 17 (6), 04016129.

Zhou, J., Li, E. Wei, H., Li, C., Qiao, Q., & Armaghani, D. J. (2019). Random Forest and Cubist algorithms for predicting shear strengths of rockfill materials. Soft computing techniques in structural engineering and materials, Appl. Sci. 9(8), 1621.