A Comparative Analysis of Predicting Seismic Liquefaction Susceptibility of Dhaka Subway Project with a Machine Learning Approach

Raiyan Mannan

Omor Seebtaien

Fahmida Mukarroma

Department of Civil and Environmental Engineering

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

2021

A Comparative Analysis of Predicting Seismic Liquefaction Susceptibility of Dhaka Subway Project with a Machine Learning Approach

Raiyan Mannan (170051068)

Omor Seebtaien (170051060)

Fahmida Mukarroma (170051018)

A THESIS SUBMITTED

FOR THE DEGREE OF BACHELOR OF SCIENCE IN CIVIL ENGINEERING

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

ISLAMIC UNIVERSITY OF TECHNOLOGY

2021

Thesis Approval

The thesis titled **A Comparative Analysis of Predicting Seismic Liquefaction Susceptibility of Dhaka Subway Project with a Machine Learning Approach** submitted by Raiyan Mannan (170051068), Omor Seebtaien (170051060), Fahmida Mukarroma (170051018) has been found satisfactory and accepted as partial fulfilment of the requirement for the Degree Bachelor of Science in Civil Engineering.

SUPERVISOR

Dr. Hossain Md. Shahin

Head of Department,
Department of Civil and Environmental Engineering (CEE)
Islamic University of Technology (IUT)
Board Bazar, Gazipur, Bangladesh

Declaration of Candidate

We hereby declare that the undergraduate research work reported in this thesis has been performed by us under the supervision of Professor Dr. Hossain Md. Shahin and this work has not been submitted elsewhere for any purpose (except for publication).

Raiyan Mannan
Student No: 170051068
Academic Year: 2020-2021
Date:

Omor Seebtaien
Student No: 170051060
Academic Year: 2020-2021
Date:

Fahmida Mukarroma
Student No: 170051018
Academic Year: 2020-2021
Date:

Dedication

Our combined thesis work is dedicated towards our respective parents, family and friends. We also express our gratitude the to our respected supervisor Professor Dr. Hossain Md. Shahin. This is a small token of appreciation towards all those who supported us throughout our endeavour and encouraged us to continue our work till the end.

Acknowledgements

"In the name of Allah, Most Gracious, Most Merciful"

All the praises to Allah (SWT) who has blessed us with the opportunity to complete this book. Our earnest gratitude towards our supervisor Professor Dr. Hossain Md. Shahin for giving us rightful instructions and for paying attention to us whenever needed throughout the research work. We are greatly indebted to him for enlightening us with his remarks and guidance in order to complete the thesis. We wish to show our gratefulness to Prosoil Foundation Consultant and GIE for providing us with the necessary data extracted from soil investigation that they performed previously. Our expression of gratitude towards all of the departmental faculty members for their aid and support.

We are sincerely grateful to Samiul Islam, Post Graduate Researcher, Islamic University of Technology (IUT) for his endless patience and care. Without his unyielding supervision, materializing this research into reality would not be possible. Our sincere appreciation to Tahsina Alam, Lecturer, in the Department of Civil & Environmental Engineering, who have continuously guided us in our effort.

We also appreciate all those individuals who have contributed in our work at any scale and those who have showered us with words of encouragement, inspiration and motivation. We are deeply obliged for the collaboration we have received throughout our work.

Table of Content

List of figures

List of tables

Abbreviation

ANN	Artificial Neural Network
AUC	Area Under Curve
BH	Borehole
CPT	Cone Penetration Test
CRR	Cyclic Resistance Ratio
CSR	Cyclic Stress Ratio
FC	Fineness Content
FS	Factor of Safety
GWT	Ground Water Table
LI	Liquefaction Indicator
LPI	Liquefaction Potential Index
MSF	Magnitude Scaling Factor
OA	Overall Accuracy
RL	Reduced Level
ROC	Receiver Operating Characteristic
SPT	Standard Penetration Test

ABSTRACT

Keywords: *Seismic Soil Liquefaction, Standard Penetration Test, Cyclic loading, Liquefaction potential index, Logistic Regression, Support Vector Machine, Artificial Neural Network.*

Seismic soil liquefaction is a dangerous phenomenon that occurs during seismic loading due to earthquakes. In this study, empirical formulas are used to assess liquefaction triggering based on the standard penetration test (SPT) data from the Dhaka Subway Project. After that three machine learning algorithms are applied to predict seismic liquefaction triggering of the obtained dataset. The first machine learning algorithm, Logistic regression, is a linear classification model. It implements the sigmoid function to generate binary outputs. The second machine learning algorithm is the Support Vector Machine (SVM) which represents supervised learning and is widely used as a classification and outliers detection algorithm. The third machine learning algorithm is the Artificial Neural Network (ANN) based on the Multi-layer Perceptron (MLP) theory, which uses a training algorithm called Levenberg-Marquardt backpropagation.

Furthermore, this study also highlights the correlation between different soil parameters in triggering soil liquefaction. The developed models are then evaluated with confusion matrices which are later used to find out Overall Accuracy, Precision, Sensitivity, Recall (Specificity), F1 score, RMSE, and MAE. ROC curves are also used to evaluate these models and establish which model is the most effective.

Chapter 1 INTRODUCTION

1.1 General

The alluvial and deltaic sediments that have been deposited in Bangladesh within the last 6000 to 10000 years dominate the country's surface geology, which belongs to the Holocene epoch. The northern region has coarse-grained mountain front alluvial fan deposits that make up some of the surface sediments. Alluvial sands and silts make up the majority of the sediments in the lowland central region of Bangladesh. On the other hand, deltaic silts and clays predominate in the southern and more coastal regions of Bangladesh.

The city of Dhaka serves as the seat of government for Bangladesh and may be found in the country's geographic centre. Geologically speaking, the city of Dhaka is located on the Dhaka-Gazipur Terrace, which is a southern extension of the Madhupur Tract. The Dhaka City occupies the greater part of this terrace. The Tract is a structural high composed of older sediments, and it is flanked on all sides by extremely young riverine sediments that fill the adjacent alluvial plains. The Tract is home to the most important areas of the city. The land shapes are distinguished by a great number of distinct geomorphic cycles, which in turn are governed by a great number of tectonic waves.

Dendritic drainage patterns may be seen throughout the majority of these terraces. The Ganges-Meghna floodplain can be found to the south of the terraces, the old Brahmaputra floodplain can be found to the east, and the Jamuna floodplain can be found to the west of the terraces.

The process by which a granular material undergoes a transition from a solid state to a liquefied state as a result of an increase in the pore water pressure is referred to as liquefaction (Samui et al., 2011). Because of this, the effective stress placed on the soil decreases, which results in a reduction in bearing capacity. During the process of liquefaction, there are three distinct forms of damage that might occur. Liquefaction could refer to a number of different forms of landslides, the most common of which being ground lateral spreading and collapses of dam embankment (Keefer, 1984). The second point to make is that surface manifestations of liquefaction in soil include sand blows and ground fractures. Liquefaction can have a number of potentially dangerous side effects, the third of which being building settling and/or severe tilting. Damages that can be ascribed to the phenomena of earthquake-induced liquefaction have resulted in hundreds of millions of dollars' worth of costs for society (Seed and Idriss, 1982). Using a variety of regression techniques, including logistic regression and Bayesian

updating, as well as data from earlier studies, we were able to establish probabilistically based triggering associations. It was shown that their Bayesian updating strategy handled the various sources of uncertainty better, however they relied on data processed by prior researchers, resulting in correlations with questionable quality. A considerable gain was made in processing, particularly the use of Bayesian updating algorithms (Juang et al., 2002).

During seismic loading, soil liquefaction resistance can be evaluated using both in situ and soil laboratory tests (e.g., SPT, Vs, CPT data). The laboratory procedures for assessing soil liquefaction resistance necessitate the use of high-quality undisturbed soil samples. However, it might be difficult and expensive to acquire these samples from deteriorated, weakly compacted silty or sandy soils. Since of these shortcomings, geotechnical engineers frequently use in situ testing because they are easy and cost-effective (Seed and Idriss 1971, 1982; Seed et al. 1983, 1984, 1985; Seed and de Alba 1986). Soil liquefaction resistance testing has long relied on the Simplified Procedure, first devised by Seed and Idriss in 1971.

According to the Bangladesh National Building Code (BNBC 2020), the nation has been segmented into four distinct seismic zones, each of which is associated with a unique amount of ground motion. Each zone possesses a seismic zone coefficient (Z), which indicates the maximum considered peak ground acceleration (PGA) on very stiff soil (Site class SA) in units of g. The PGA is a measure of how much the ground moves as a result of an earthquake (acceleration due to gravity). Zone, I have a zone coefficient of 0.12, Zone II has a zone value of 0.20, Zone III has a zone coefficient of 0.28, and Zone IV has a zone coefficient of 0.36. (Zone IV). A map of Bangladesh illustrating the dividing lines between the four distinct zones is provided in Figure 4. Where the city of Dhaka is located, which corresponds to the Zone -II (the moderate seismic intensity zone) and has a Z value of 0.20g.

1.2 Purpose and Objectives

Identifying the factors that have a greater influence on seismic liquefaction triggering can be valuable information for policymakers and geotechnical corporations to implement and execute measures that promote safety restraints during substructure construction. Our study attempts to determine the most efficient machine learning model to predict liquefaction triggering of soil in Bangladesh. The focal objectives of our study are:

- Create reproducible datasets with the empirical formula given by Youd et al., (2001) as well as assess liquefaction layer wise.

- Build the Machine learning Models (Logistic Regression, Support Vector Machine and Artificial Neural Network Model).
- Develop Confusion Matrix and ROC curves for the Machine Learning Models (Logistic Regression, Support Vector Machine and Artificial Neural Network Model).
- Calculate Overall Accuracy, Precision, Sensitivity, Specificity, F1, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

1.3 Scope of the Study

This study will create various opportunities for geotechnical and seismological research for incorporating artificial intelligence. If the validation of the proposed models is achieved then assessing liquefaction for large datasets will become easier. A web-based application can be developed so that minimum effort and resources have to be put to evaluate the ground conditions of a certain area. Since, geotechnical tests for extracting soil parameters can be very expensive, this study might prove to be a very good alternative for acquiring desired results.

1.4 Thesis Outline

The thesis has been organized into six chapters. The chapters are briefly introduced underneath:

Chapter 1: **Introduction**- This chapter explains the background, problem statement, purpose, and objective of the research.

Chapter 2: **Literature Review**- This chapter discusses the relevant pieces of literature that helped in gaining the most suitable work plan for the research.

Chapter 3: **Study Area and Data Collection**- This chapter sheds light on scoping, bounding and data acquiring techniques.

Chapter 4: **Methodology**- This chapter explains the gradual working process of the research and illustrates the method adapted to analyse acquired data.

Chapter 5: **Results and Discussion**- This chapter discusses the analysis of collected data and interprets the obtained results.

Chapter 6: **Conclusion**- This chapter presents the main findings of the research and suggests suitable policy implications.

Chapter 2 LITERATURE REVIEW

2.1 Assessment of Liquefaction based on Empirical formula

A seismically active zone covers most of the part of the location of Dhaka city. Most of the land regions are covered by silty sand, silty clay, Holocene sand, sandy- and clayey-silt, silty sand, silty clay up to a depth of more than 20 m from the surface of the ground. The studies about the potential of liquefaction in Dhaka City are still very underdeveloped, and also the amount of literature available on the liquefaction hazard quantification induced seismically is very limited. For this reason, an attempt on this study was taken to prepare a map of the Dhaka City Corporation area that is prone to seismically induced liquefaction. This research expressed how to compute the liquefaction potential using Simplified Procedure to estimate liquefaction potential index (LPI) of the subsurface geological materials of Dhaka City and by using contour lines of equal LPI values of SPT profiles located at different places of the city and by the cumulative frequency distribution of LPI of different geological materials for preparing a liquefaction hazard map (Rahman et al., 2015). Rahman et al. (2015) and Rahman and Siddiqua (2017) performed potential liquefaction studies in Bangladesh utilizing restricted standard penetration test blow count (SPT-N), cone penetration test (CPT), and shear wave velocity (Vs) data. Studies found that these cities' Holocene alluvium is prone to liquefaction. In these investigations, Youd et al., (2001) empirical equations were utilized to compute the factor of safety factor (FS) of liquefaction, cyclic resistance ratio (CRR), cycle stress ratio (CSR), and the Magnitude Scaling Factor (MSF). Iwasaki et al. (1982) equations were used to determine liquefaction potential index (LPI).

A deterministic connection, proposed by Seed and colleagues (Seed et al. 2001), has been extensively recognized and utilized by the NCEER Workshop (NCEER 1997; Youd et al. 2001), and is one of the most commonly accepted and used SPT-based correlations (1984, 1985). This relationship is depicted in that with very slight modifications for low cycle stress ratios (as recommended by the NCEER Workshop; NCEER 1997 and Youd et al. 2001). In this well-known relationship, the SPT N values are compared to the intensity of cyclic loading, which is expressed as magnitude-weighted equivalent uniform cyclic stress, after being corrected for both effective overburden stress and energy, as well as for equipment and procedural factors affecting SPT testing [to N60 values]. Better handling of r_d in “simplified” assessment of in-situ CSR leads to triggering relationships that are fair when used in conjunction with either (1) direct seismic response studies for evaluation of in-situ CSR, or (2)

better simplified assessment of in-situ CSR, respectively. Furthermore, these findings demonstrate that all previously commonly utilized correlations, when used in conjunction with direct response analyses for the evaluation of CSR, are unconservative skewed due to bias in prior simplified r_d recommendations. This is a significant step forward. The new models give a considerably superior basis for engineering evaluation of the likelihood of liquefaction start when compared to previously existing models, which is a considerable improvement above earlier model. The new models presented and described in this paper explicitly address the issues of: (1) FC, (2) magnitude-correlated, and (3) effective overburden stress (Ks effects), and they provide both (1) an unbiased basis for evaluating liquefaction initiation hazard and (2) a significant reduction in overall model uncertainty, as well as (1) an unbiased basis for evaluating liquefaction initiation hazard and (2) a significant reduction in overall model uncertainty. (Cetin et al., 2018).

2.2 Application of Machine Learning in Geotechnical Engineering

Past research suggests that AI is the future for data analysis in civil engineering. In the studies related to soil liquefaction, a lot of algorithms representing AI were used and shown to be very effective. In a study SVM and ANN was used to predict soil liquefaction susceptibility based on SPT data (Samui et al., 2011). Many studies have found a way to accurately measure soil liquefaction capability by utilizing machine learning models in general and neural network models in specific for the moment, further research is needed to address the issues of model overfitting and the stability of models under a variety of sampling approaches (Tuan Anh Pham, 2021). This research uses an updated, robust and promising approach to measure soil liquefaction resistance for Dhaka City. While training an artificial neural network (ANN), it is necessary to collect SPT-N and Vs. data from sites with historical liquefaction and non-liquefaction data to determine the liquefaction indicator (LI) function and the points of the limit state function (LSF). The errors associated with the LPI computation may be decreased by using more SPT-N, Vs. data, fluctuation in groundwater level, the correct definition of surface geological units, and suitable ground motion. Finally, this Dhaka City liquefaction hazard map can be used to guide future urban development and planning efforts aimed at mitigating liquefaction-related damage and loss. (Fahim et al., 2021). However, the ANN models were integrated with this investigation based on Juang et al. (2003, 2002, 2000) to predict CRR. Inherently, ANN models are going to generate more realistic results, plus SPT-N and Vs. Profiles were utilized to describe subsurface heterogeneity better accurately.

2.3 Neural Networks

An artificial neural network (ANN) is a type of computing system that is used to model how the human brain evaluates and processes data (Chadha et al., 2022). Each artificial neuron, or node, receives and processes a signal that is then delivered to the neurons to which it is connected. A neural network comprises an input layer, an output layer, and one or more hidden layers. The input layer receives information from the training features, which are carried out via hidden layers, and the output layer delivers the prediction. The input layer's number of neurons is equal to the number of features. In regression, a single output neuron is employed, whereas in classification, the number of output neurons equals the total number of class labels. An artificial neural network first undergoes a training phase during which it learns patterns from data. Following that, the loss is determined using the model's forecast and the actual data. Backpropagation is used to change the weights of the attributes. ANNs are a type of information processing system that, by simulating the processes and connectivity of biological neurons, approximates the behaviour of the human brain. ANNs represent complex, non-linear functions with a large number of parameters that are changed (calibrated or trained) until the ANN's output resembles that of a known data set. ANNs require large quantities of data to train; after training successfully, an ANN should be able to provide output for a new set of inputs. The primary distinctions among the many types of ANNs are the network design and the process used to determine the weights and functions for inputs and neurodes (training) (Caudill et al., 1992). When solving a variety of issues, researchers have utilized a training set that consists of a different percentage of the total accessible data. The ANN model's performance can be improved by maintaining a statistically consistent training and testing dataset, which also makes it easier to evaluate the models in the long run (Shahin et al., 2000). For the purpose of training, for example, Kurup and Dudani (2002) used 63 percent of the data, Tang et al. (2005) used 75 percent, and Padmini et al. (2008) used 80 percent. In this investigation, we have put seventy percent of the data to use for instructional purposes.

Chapter 3 STUDY AREA AND DATA COLLECTION

3.1 Study Area

Dhaka, the capital city of Bangladesh, is consistently ranked as one of the fastest expanding megacities in the world. The city's population is growing at a rate of about one percent each year, which is causing severe challenges including traffic congestion throughout the entire city. The primary reasons for prolonged instances of traffic congestion include drivers' propensity to break the law, antiquated approaches to navigating traffic, and congested road conditions. The economic losses that are experienced as a result of these congestions are not something that can be neglected. Additionally, it is responsible for significant amounts of air pollution and noise pollution, both of which contribute to the deterioration of the ecosystem as a whole.

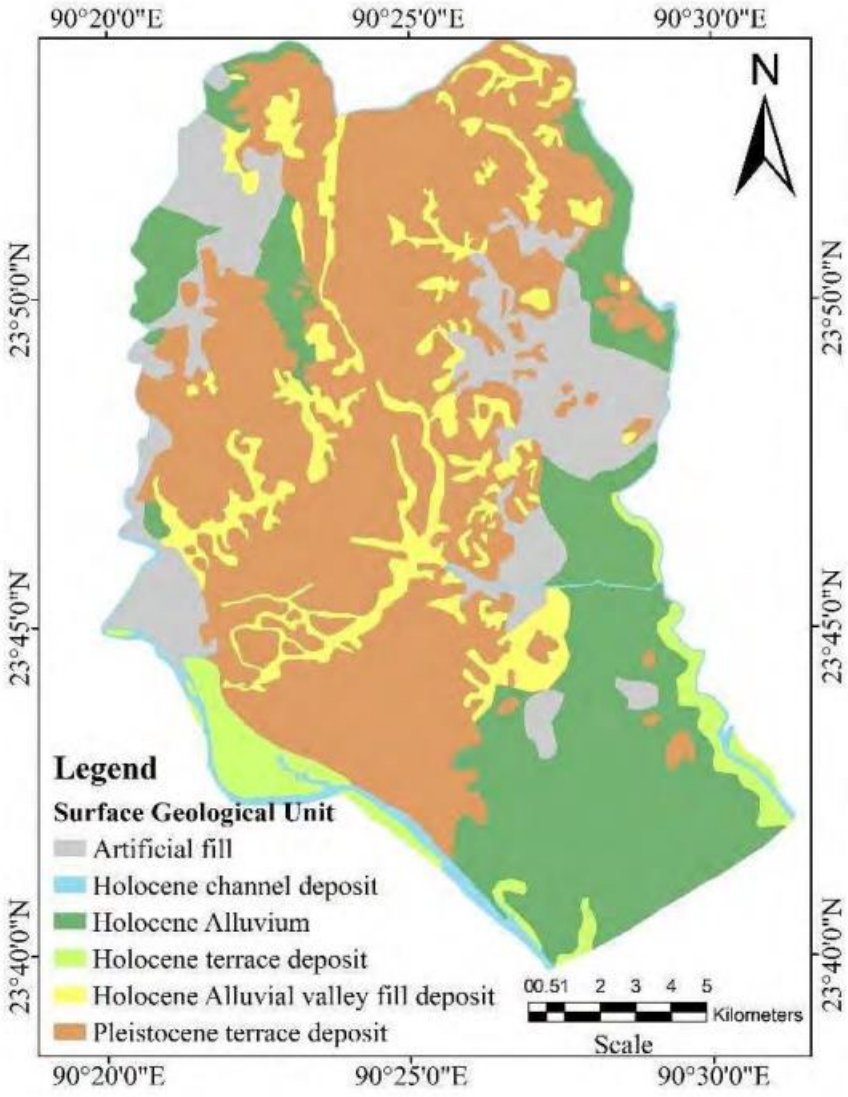

Figure 1 Surface geology map of Dhaka city (modified from Rahman et al., 2015)

The plan for the Dhaka city subway is not limited in any manner by the blueprint itself or the paper it is drawn on. The initiatives are being carried out in accordance with the plan that was developed by the Government, and this will be accomplished throughout the term of current government. According to the government, construction on the subway system will begin while this administration is still in office. The surface geology of Dhaka city is shown in **figure 1**.

In July of 2018, the Ministry of Bangladesh Road Transport and Bridges and the Spanish consulting organization TYPSA inked a deal for the building of a subway system in Dhaka. The initial stage of the subway's construction is anticipated to incur a cost of approximately $5.62 billion. TYPSA, a premier consulting engineering organization specializing in transportation, urban development, and renewable energy, will investigate four distinct avenues that could be taken by the subway. When it is finished, around 4 million of the approximately 8 million working population of Dhaka city will be able to use the underground subway on four routes, and the traffic situation in Dhaka city will significantly improve as a result.

3.2 Data Collection

A total of 180 boreholes were provided to us as secondary data. These boreholes were later to be transformed in tabulated format as a means to produce the data sets required for the machine learning models. The location of these boreholes is scattered and confined in greater Dhaka city. Various types of geotechnical tests were performed on these boreholes to find out different soil parameters and soil classification of different locations in Dhaka.

In **table 1** below represents the location of the boreholes and their corresponding reduced level and ground water table:

Table 1 Borehole information

Borehole	Latitude/Easting	Longitude/Northing	RL	GWT
SW BH 01	229362.6	2632641	5.11	7
SW BH 02	235070.1	2631467	6.19	7
SW BH 03	233400	2640510	3.66	3
SW BH 04	235971.4879	2624792.634	5.57	2
SW BH 05	235971.4879	2623990.669	5.8	3.5
SW BH 06	2625054.684	237683.3035	7.43	3.7
SW BH 07	233664.195	2636787.19	6.94	12
SW BH 08	237554.486	2627164	2.26	14
SW BH 09	234433.491	2643502.9	7.75	4
SW BH 10	239870.88	2636901.09	5.09	5
SW BH 11	236782.969	2632265.27	6.84	6
SW BH 12	234472.8125	2627007.679	7.75	1.5
SW BH 13	238639.705	2627045.55	4.34	2
SW BH 14	229680.205	2630655.615	7.98	9
SW BH 15	236703.807	2629408.595	6.46	3
SW BH 16	237288.144	2635846.355	7.863	6.36
SW BH 17	236113.1863	2636139.104	7.24	5.1
SW BH 18	230401.378	2631452.847	5.455	5
SW BH 19	233019.7747	2637163.765	3.828	8
SW BH 20	231669.71	2636851.964	8.662	8
SW BH 21	230303.0657	2636246.384	5.8235	8.9
SW BH 22	229629.353	2635026.293	10.58	9
SW BH 23	229406.3956	2633767.049	6.88119	3.6
SW BH 24	229211.352	2632693.601	5.87	10
SW BH 25	230044.4108	2631918.916	4.39099	2.6
SW BH 26	233713.8574	2641438.065	6.33	13
SW BH 27	230575.0084	2629690.264	7.88336	7
SW BH 28	230785.829	2628694.448	7.94	5.6
SW BH 29	231061.5619	2627452.143	8.09399	6.8
SW BH 30	231656.159	2626474.506	7.98	18.7
SW BH 31	232509.2344	2625834.313	8.14681	8.9
SW BH 32	233633.533	2625320.644	5.594	2.6
SW BH 33	234778.1143	2624692.308	6.30753	10.4

SW BH 34	236151.108	2645648.66	7.517	8.5
SW BH 35	235452.6775	2644782.841	6.4987	1
SW BH 36	234713.503	2644480.53	4.397	9.3
SW BH 37	234285.1522	2642360.76	7.01	3.1
SW BH 38	230776.292	2630800.702	5.882	3
SW BH 39	237061.1597	2628754.054	6.002	9
SW BH 40	236137.839	2627518.482	5.942	7.5
SW BH 41	236969.7828	2624304.583	5.83	3
SW BH 42	235706.298	2627604.939	6.798	1.5
SW BH 43	237036.2118	2627742.978	5.807	5.5
SW BH 44	238127.358	2627827.81	5.791	4.5
SW BH 45	2628206.096	239110.927	5.822	1.5
SW BH 46	240301.379	2628308.075	5.654	4.3
SW BH 47	241407.1933	2628330.813	5.967	3.1
SW BH 48	242447.933	2628417.609	5.161	4
SW BH 49	243504.713	2628543.805	3.676	3
SW BH 50	233107.1477	2639361.801	3.098	3.8
SW BH 51	232794.2567	2638374.367	3.94355	5.3
SW BH 52	233230.502	2636365.733	5.874	2.7
SW BH 53	233513.2265	2635445.635	4.2	12.5
SW BH 54	233868.8252	2633223.708	7.368	14.82
SW BH 55	234650.965	2632015.583	6.026	30
SW BH 56	235957.697	2630632.131	6.049	2
SW BH 57	233796.139	2634310.969	6.35	4.6
SW BH 58	229846.209	2628592.203	6.608	11
SW BH 59	229344.9468	2629609.17	5.70016	8.45
SW BH 60	231978.5132	2631061.281	7.459	1.5
SW BH 105	245787.317	2633933.292	6.35248	8.25
SW BH 106	245849.313	2634585.427	7.092	6
SW BH 108	246038.384	2637613.416	6.18	3.5
SW BH 109	245844.022	2640411.331	7.507	13
SW BH 110	245827.9557	2641678.397	7.212	4.15
SW-BH-111	246117.101	2642559.181	4.169	6
SW BH 121	243731.6228	2640480.756	6.707	15.3
SW BH 122	245898.716	2639696.681	6.99	2
SW BH 123	246899.5254	2640390.258	8.922	3.3
SW BH 124	247894.931	2640358.794	8.501	3
SW BH 125	248956.696	2640345.02	5.955	10
SW BH 126	250219.44	2640312.684	6.951	2
SW BH 127	251328.982	2640283.724	3.876	1
SW BH 128	236925.9366	2623137.833	6.662	1
SW BH 129	237516.183	2622393.234	6.48	4
SW BH 130	234747.0752	2620128.794	7.536	18.02
SW BH 131	235799.707	2621774.337	5.0295	17
SW BH 133	236028.24	2626217.999	7.57	4.1
SW BH 136	237766.267	2620733.467	5.77048	9
SW BH 137	245475.59	2613982.419	5.41301	6.3
SW BH 138	229447.4241	2626784.806	5.4	18.45
SW BH 140	227081.9803	2627104.284	6.006	16.02
SW BH 142	225086.5554	2627024.849	5.97	10.02

SW BH 143	242859.385	2619075.436	3.392	30
SW BH 144	242043.231	2620326.886	3.07	3
SW BH 145	241417.482	2621108.128	3.671	2
SW BH 146	240944.2496	2621965.456	4.489	2.7
SW BH 147	240843.884	2622997.156	4.3064	23.6
SW BH 148	240876.878	2624032.285	4.45	2.4
SW BH 149	240880.457	2652002.754	6.392	6.5
SW BH 150	239811.835	2617779.349	4.47	10
SW BH 151	240330.546	2626843.506	6.307	10
SW BH 152	240369.926	2630565.27	5.589	9
SW BH 153	240158.3231	2633792.049	5.753	6.7
SW BH 154	240044.704	2634465.318	7.19631	7
SW BH 155	233102.135	2628533.757	7.96699	4.5
SW BH 156	239719.512	2638894.085	5.33	21
SW BH 157	238839.0109	2639777.524	3.877	1.9
SW BH 158	238316.964	2640457.713	5.921	3.1
SW BH 159	238189.4754	2641664.005	6.481	4.5
SW BH 160	235961.8297	2642016.468	8.007	4.2
SW BH 161	235061.1761	2642015.352	7.79275	18
SW BH 162	232876.8071	2642200.731	5.278	3
SW BH 163	231742.2979	2642130.755	5.349	0.5
SW BH 164	230729.004	2642272.937	4.594	10
SW BH 165	229573.5891	2642593.795	5.412	7.4
SW BH 166	228260.759	2642926.328	6.946	2
SW BH 167	227330.6072	2642963.66	12.463	4.95
SW BH 168	226051.076	2643011.952	13.094	14
SW-BH-169	224973.4831	2643081.608	12.079	5.4
SW BH 170	223958.889	2643107.362	13.532	3
SW-BH-171	222966.4943	2642867.821	9.957	7.2
SW BH 172	222291.2727	2643508.443	13.261	2
SW BH 173	235567.6895	2620787.205	7.109	20.1
SW BH 174	236613.7696	2633723.168	7.6684	26.5
SW BH 175	236158.855	2634709.376	6.26561	15
SW BH 176	234384.3759	2635785.73	6.53352	4.1
SW BH 177	234146.723	2636827.19	6.141	7.2
SW BH 178	233701.6847	2638284.544	3.572	1.8
SW BH 179	237120.8656	2642074.054	8.799	12
SW BH 180	240006.7745	2637800.735	6.971	4.4

The given boreholes can be used in data tabulation and liquefaction hazard mapping.

Chapter 4 METHODOLOGY

4.1 Work Flow

The work flow of the study is outlined below in figure 2:

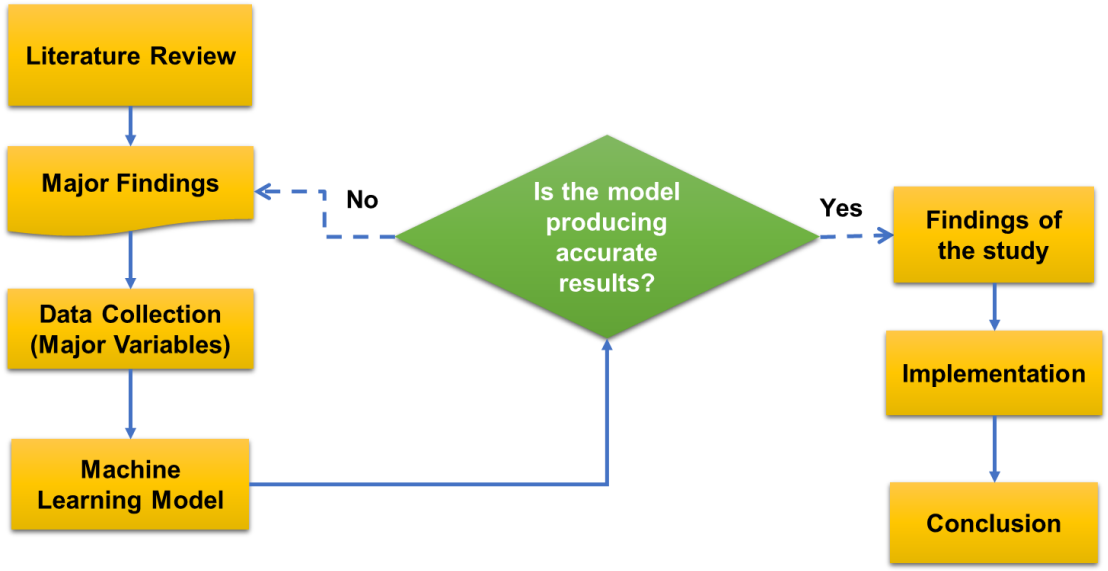

Figure 2 Workflow diagram of study

4.2 Data Tabulation

In the NCEER workshops, Youd et al. (2001) proposed a simplified procedure for evaluating the liquefaction resistance of granular soils. This method has become a standard practice for everything related to seismic liquefaction. In this study we used this simplified procedure on criteria based on (1) Standard Penetration Tests (SPT), (2) Magnitude scaling factors, (3) Correction factors for overburden stress and sloping ground, (4) Earthquake magnitude and Peak ground acceleration. To evaluate the seismic triggering of liquefaction, we calculate a factor of safety based on CSR and CRR.

4.2.1 Evaluation of CSR

For the calculation of Cyclic Stress Ration (CSR), Seed and Idriss (1971) proposed the following equation:

$$CSR = \left(\frac{\tau_{av}}{\sigma'_{v_0}}\right) = 0 \cdot 65 \left(\frac{a_{max}}{g}\right)\left(\frac{\sigma_{v_0}}{\sigma'_{v_0}}\right) \cdot r_d$$

Here, a_{max} = peak horizontal acceleration at the ground surface produced by a seismic event; g = acceleration due to gravity; σ_{v_0} and σ'_{v_0} are total and effective overburden stress; r_d = stress reduction coefficient.

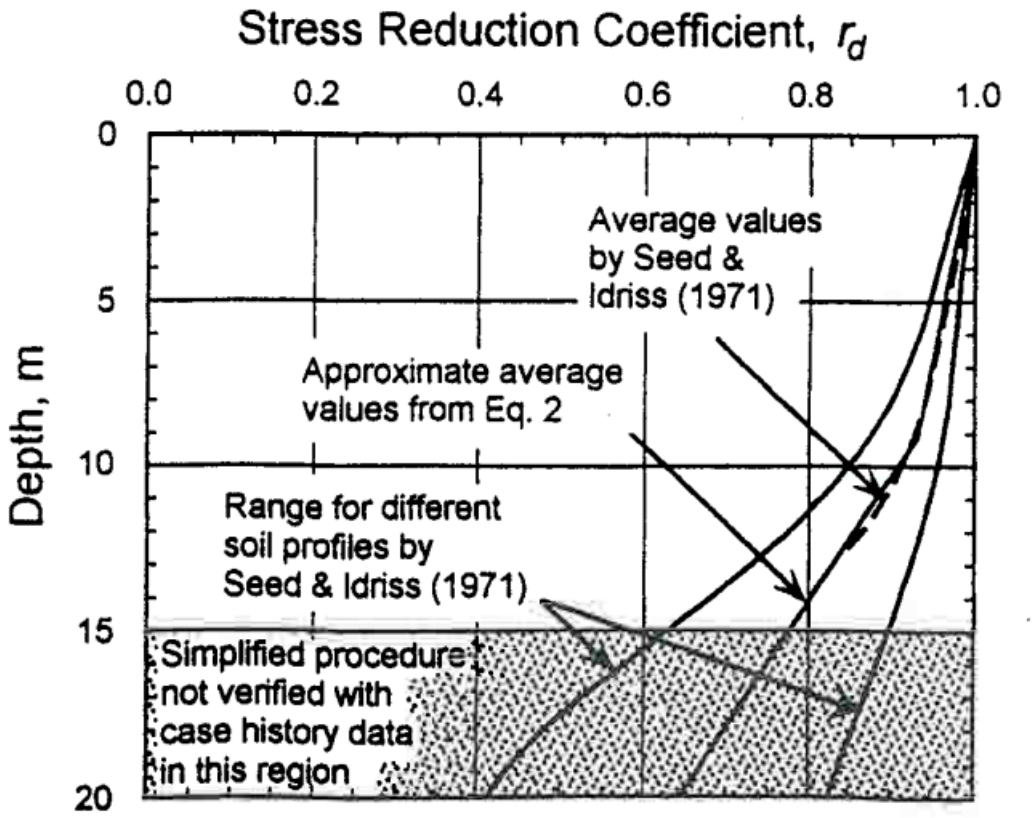

Figure 3 rd vs Depth curves developed by Seed and Idriss (1971)

The stress reduction coefficient was estimated using empirical curves in figure 3 produced against depth developed by Seed and Idriss (1971). But T. F. Blake (1996) approximated the mean of the curve so that r_d can be computed easily by the following equation:

$$r_d = \frac{1-0.4113z^{0.5}+0.04052z+0.001753z^{1.5}}{1-0.4177z^{0.5}+0.05729z-0.006205z^{1.5}+0.001210z^2}$$

Here z = depth beneath ground surface in meters.

4.2.2 Evaluation of CRR

SPT Correction

For the correction of the field SPT-N values, we use various factors, and finally calculate the normalized SPT value. The following equation was used to find the normalized SPT value $(N_1)_{60}$:

$$(N_1)_{60} = N_M C_N C_E C_R C_B C_S$$

Here, N = field SPT value; C_N = correction for effective overburden stress; C_E = correction for hammer energy ratio (ER); C_B = correction for borehole diameter; C_R = correction for rod length; C_S = correction for samplers. For the calculating correction for effective overburden stress the following equation proposed by Liao and Whitman (1986) was used:

$$C_N = \left(\frac{p_a}{\sigma'_{v_o}}\right)^{0.5} \qquad \rho_\alpha = 100kPa$$

$$0.4 \leq C_N \leq 1.7$$

Here C_N normalizes N for an effective overburden stress of 100 kPa.

To calculate C_E, we used the following equation:

$$C_E = \frac{ER(\%)}{60}$$

Here ER is the hammer efficiency ratio used in the apparatus of extracting SPT values.

Finally, the we used a correction for Fines to correct the normalized value again using the following equation:

$$C_{FINES} = (1+0.004 \cdot \text{FC}) + 0.05\left(\frac{FC}{N_{1,60}}\right)$$

$$(N_1)_{60,Cs} = (N_1)_{60} \cdot C_{FINES}$$

Here, FC is fineness content in decimals.

An approximation of the clean- sand based curve was developed by A.F. Raunch (1998) and the following equation was proposed to evaluate CRR:

$$CRR_{7.5} = \left(\frac{1}{34 - (N_1)_{60}}\right) + \left(\frac{(N_1)_{60}}{135}\right) + \frac{50}{(10 \cdot (N_1)_{60} + 45)^2} - \frac{1}{200}$$

Here, this equation will only be valid for $(N_1)_{60} < 30$. If it is greater than 30, clean granular soils will be too dense to liquefy and are classed as non-liquefiable.

4.2.3 Magnitude Scaling Factor

Previously, MSF were estimated using various curves against the magnitude of the Earthquake. In the NCEER workshop an approximation was done using those curves and the following equation was proposed for calculating MSF:

$$MSF = \frac{10^{2.24}}{M_W{}^{2.56}}$$

Here, MW is the magnitude of the earthquake.

4.2.4 Factor of Safety

For assessing the influence of magnitude scaling factors in an artificial hazard, the equation of factor of safety against liquefaction is as follows:

$$F_L = \frac{CRR_{7.5}}{CSR} \cdot MSF$$

where, the if the factor of safety is less than 1 then liquefaction will occur and if it is greater than 1 then it will be termed as non-liquefiable.

4.2.5 Data Used

The following table was formulated using the given equations and then sorted out to include only the feature parameters that represents the dataset to be used in the machine learning algorithms.

Table 2 Dataset for Machine Learning Application

SL no.	r_d	CSR	$(N_1)_{60}$	FC	Liquefaction
1	0.983	0.128	10.457	95.400	No
2	0.976	0.127	12.699	95.400	No
3	0.998	0.130	3.600	16.760	Yes
4	0.990	0.129	7.986	16.760	Yes
5	0.983	0.144	13.151	16.760	No
6	0.976	0.168	18.449	16.450	No
7	0.969	0.185	6.240	16.450	Yes
8	0.998	0.130	9.000	20.700	Yes
9	0.990	0.129	11.911	20.700	No
10	0.983	0.128	9.226	20.700	Yes
11	0.976	0.151	14.283	20.700	No

12	0.969	0.168	11.833	95.650	No
13	0.962	0.181	13.907	95.650	No
14	0.953	0.190	3.949	96.050	Yes
15	0.953	0.124	9.445	66.000	No
16	0.943	0.123	4.110	66.000	Yes
17	0.953	0.124	10.544	41.000	No
18	0.943	0.123	10.332	41.000	No
19	0.998	0.130	1.800	30.000	Yes
20	0.990	0.192	14.280	30.000	No
21	0.998	0.130	10.800	7.000	No
22	0.990	0.129	5.889	7.000	Yes
23	0.983	0.128	6.082	5.000	Yes
24	0.976	0.127	5.745	5.000	Yes
25	0.969	0.126	6.333	5.000	Yes
26	0.943	0.150	2.294	73.000	Yes
27	0.923	0.159	2.180	73.000	Yes
28	0.883	0.165	0.000	73.000	Yes
29	0.827	0.164	0.972	81.000	Yes
30	0.761	0.158	7.407	81.000	Yes
31	0.998	0.130	3.600	83.000	Yes
32	0.987	0.128	6.833	83.000	No
33	0.973	0.126	10.789	81.000	No
34	0.958	0.125	12.109	81.000	No
35	0.943	0.127	4.225	81.000	Yes
36	0.923	0.137	14.136	39.000	No
37	0.827	0.145	16.391	23.000	No
38	0.696	0.134	10.032	25.000	Yes
39	0.641	0.128	16.866	25.000	No
40	0.998	0.130	7.200	30.000	Yes
41	0.983	0.145	3.309	30.000	Yes
42	0.998	0.130	7.200	12.000	Yes
43	0.990	0.129	8.032	12.000	Yes
44	0.983	0.128	6.222	12.000	Yes
45	0.976	0.127	5.877	30.000	Yes
46	0.969	0.126	1.296	87.000	Yes
47	0.962	0.125	1.172	87.000	Yes
48	0.953	0.124	1.078	64.000	Yes
49	0.943	0.123	1.057	64.000	Yes
50	0.923	0.133	8.106	88.000	No
51	0.883	0.141	6.742	88.000	No
52	0.998	0.130	1.800	35.000	Yes
53	0.990	0.129	1.997	35.000	Yes
54	0.983	0.128	1.547	64.000	Yes
55	0.976	0.138	3.050	64.000	Yes
56	0.998	0.130	7.200	40.990	No
57	0.990	0.129	8.128	40.990	No
58	0.983	0.128	1.574	40.990	Yes
59	0.998	0.130	1.800	35.000	Yes
60	0.990	0.129	2.032	35.000	Yes
61	0.983	0.128	1.574	35.000	Yes

62	0.998	0.130	7.200	35.000	No
63	0.990	0.129	8.025	76.100	No
64	0.983	0.128	12.408	75.100	No
65	0.976	0.149	4.707	94.570	Yes
66	0.969	0.165	13.180	94.570	No
67	0.953	0.187	15.589	89.820	No
68	0.998	0.130	7.200	85.300	No
69	0.990	0.129	5.939	85.300	No
70	0.983	0.128	4.600	85.300	Yes
71	0.976	0.127	4.345	85.300	Yes
72	0.969	0.134	7.917	61.100	No
73	0.962	0.148	12.562	61.100	No
74	0.983	0.128	7.598	95.100	No
75	0.976	0.127	8.621	95.100	No
76	0.969	0.126	7.607	93.700	No
77	0.998	0.130	14.400	27.800	No
78	0.990	0.129	7.986	26.800	Yes
79	0.983	0.128	6.186	85.000	No
80	0.976	0.127	4.372	84.000	Yes
81	0.969	0.126	6.437	88.300	No
82	0.962	0.125	5.811	88.900	No
83	0.953	0.135	2.225	87.900	Yes
84	0.990	0.129	5.905	95.550	No
85	0.983	0.128	7.624	95.550	No
86	0.976	0.127	10.082	95.550	No
87	0.969	0.126	5.081	95.870	Yes
88	0.962	0.125	5.745	95.870	No
89	0.953	0.124	3.171	94.130	Yes
90	0.998	0.130	7.200	45.000	No
91	0.990	0.129	8.103	45.000	No
92	0.962	0.125	4.631	97.200	Yes
93	0.953	0.124	11.701	96.200	No
94	0.943	0.123	12.550	96.000	No
95	0.761	0.099	6.687	14.400	Yes
96	0.969	0.126	11.213	91.930	No
97	0.962	0.125	10.138	91.930	No
98	0.953	0.124	11.394	94.510	No
99	0.943	0.123	12.178	94.510	No
100	0.969	0.126	8.857	35.400	No
101	0.962	0.125	6.826	35.400	No
102	0.953	0.124	13.549	45.100	No
103	0.990	0.297	8.160	20.660	Yes
104	0.976	0.308	11.400	74.660	Yes
105	0.962	0.294	15.010	50.800	Yes
106	0.953	0.291	6.830	50.800	Yes
107	0.943	0.290	10.055	93.080	Yes
108	0.923	0.282	10.603	74.000	Yes
109	0.953	0.249	17.754	18.050	Yes
110	0.923	0.266	14.301	42.360	Yes
111	0.998	0.130	7.200	4.200	Yes

112	0.990	0.129	6.114	4.200	Yes
113	0.976	0.127	10.438	16.600	No
114	0.990	0.129	13.611	11.800	No
115	0.983	0.128	9.037	11.800	Yes
116	0.976	0.135	4.408	10.700	Yes
117	0.969	0.152	8.300	4.300	Yes
118	0.998	0.130	3.600	90.800	Yes
119	0.990	0.129	3.997	90.800	Yes
120	0.998	0.130	12.600	19.640	No
121	0.990	0.129	15.449	19.640	No
122	0.998	0.130	12.600	24.770	No
123	0.990	0.129	11.713	24.770	No
124	0.983	0.128	1.512	24.770	Yes
125	0.923	0.120	6.883	60.340	No
126	0.962	0.125	9.848	41.560	No
127	0.943	0.123	8.007	48.780	No
128	0.923	0.161	9.529	90.320	No
129	0.883	0.165	15.953	45.080	No
130	0.990	0.129	7.375	14.120	Yes
131	0.953	0.124	8.277	96.420	No
132	0.943	0.123	10.077	96.420	No
133	0.998	0.130	10.800	28.860	No
134	0.990	0.129	15.881	28.860	No
135	0.983	0.128	3.075	19.630	Yes
136	0.976	0.127	4.357	19.630	Yes
137	0.969	0.126	14.091	42.730	No
138	0.962	0.125	9.270	42.730	No
139	0.923	0.120	14.523	30.720	No
140	0.883	0.115	13.901	30.720	No
141	0.953	0.141	14.119	45.760	No
142	0.943	0.148	10.973	29.280	Yes
143	0.923	0.156	16.603	29.280	No
144	0.883	0.160	17.486	24.660	No
145	0.962	0.139	8.368	93.420	No
146	0.953	0.149	8.038	93.420	No
147	0.943	0.157	9.316	81.700	No
148	0.923	0.166	16.579	89.220	No
149	0.883	0.171	13.504	89.220	No
150	0.983	0.128	12.183	95.340	No
151	0.998	0.130	3.600	12.440	Yes
152	0.990	0.129	3.865	12.440	Yes
153	0.983	0.128	2.994	12.440	Yes
154	0.998	0.130	3.600	14.600	Yes
155	0.990	0.129	3.862	14.600	Yes
156	0.998	0.130	9.000	7.640	Yes
157	0.990	0.129	13.521	7.640	No
158	0.983	0.128	8.977	6.140	Yes
159	0.976	0.127	9.893	6.140	Yes
160	0.969	0.126	6.232	7.470	Yes
161	0.962	0.125	5.637	7.470	Yes

162	0.998	0.130	16.200	14.980	No
163	0.990	0.129	15.465	14.980	No
164	0.983	0.128	10.482	14.980	No
165	0.976	0.127	11.315	5.120	No
166	0.969	0.126	4.990	5.120	Yes
167	0.962	0.125	3.385	12.120	Yes
168	0.883	0.151	2.870	97.620	Yes
169	0.827	0.149	2.679	97.620	Yes
170	0.998	0.130	16.200	24.280	No
171	0.953	0.166	16.037	98.600	No
172	0.943	0.174	8.710	100.000	Yes
173	0.923	0.182	5.872	96.300	Yes
174	0.998	0.130	7.200	2.680	Yes
175	0.990	0.129	5.796	2.680	Yes
176	0.983	0.128	16.463	2.680	No
177	0.969	0.166	11.435	20.640	Yes
178	0.923	0.200	6.066	95.940	Yes
179	0.883	0.201	10.157	97.160	Yes
180	0.998	0.130	3.600	82.220	Yes
181	0.990	0.129	5.952	82.220	No
182	0.983	0.128	4.610	82.220	Yes
183	0.976	0.127	13.064	73.840	No
184	0.969	0.126	6.401	73.840	No
185	0.962	0.125	9.234	64.060	No
186	0.953	0.124	13.774	64.060	No
187	0.962	0.145	12.480	89.810	No
188	0.962	0.125	13.010	26.440	No
189	0.953	0.129	15.349	26.440	No
190	0.943	0.136	17.711	21.900	No
191	0.998	0.130	5.400	90.020	Yes
192	0.990	0.129	5.953	90.020	No
193	0.983	0.128	13.683	14.000	No
194	0.976	0.127	2.859	14.000	Yes
195	0.969	0.134	5.190	13.280	Yes
196	0.962	0.147	7.384	13.280	Yes
197	0.953	0.157	1.177	91.320	Yes
198	0.827	0.174	1.990	93.100	Yes
199	0.761	0.166	3.779	93.100	Yes
200	0.983	0.128	17.965	13.520	No
201	0.976	0.127	18.383	13.520	No
202	0.969	0.126	11.224	9.240	No
203	0.962	0.131	10.411	9.240	Yes
204	0.953	0.142	8.871	26.240	Yes
205	0.943	0.149	7.858	26.240	Yes
206	0.923	0.158	10.739	80.880	No
207	0.883	0.164	17.302	80.880	No
208	0.998	0.130	10.800	17.520	No
209	0.990	0.129	5.796	17.520	Yes
210	0.983	0.128	2.993	17.520	Yes
211	0.976	0.127	14.135	15.780	No

212	0.969	0.126	4.986	15.780	Yes
213	0.998	0.130	7.200	8.900	Yes
214	0.990	0.129	3.862	8.900	Yes
215	0.983	0.128	8.976	3.080	Yes
216	0.976	0.127	7.065	3.080	Yes
217	0.969	0.126	3.739	16.420	Yes
218	0.962	0.125	2.254	16.420	Yes
219	0.998	0.130	14.400	8.760	No
220	0.990	0.129	9.656	8.760	Yes
221	0.983	0.128	8.976	11.380	Yes
222	0.976	0.127	12.718	11.380	No
223	0.969	0.126	3.739	3.300	Yes
224	0.962	0.125	3.382	3.300	Yes
225	0.998	0.130	3.600	19.900	Yes
226	0.990	0.129	7.724	19.900	Yes
227	0.983	0.143	4.750	19.900	Yes
228	0.976	0.139	7.515	46.820	No
229	0.998	0.130	14.400	5.240	No
230	0.990	0.129	7.724	5.240	Yes
231	0.983	0.128	8.975	5.240	Yes
232	0.976	0.127	9.890	43.360	No
233	0.969	0.134	11.565	43.360	No
234	0.983	0.128	6.874	82.760	No
235	0.953	0.124	6.637	90.230	No
236	0.943	0.123	5.392	90.230	No
237	0.923	0.120	8.810	85.780	No
238	0.883	0.115	12.334	85.780	No
239	0.998	0.130	9.000	15.220	Yes
240	0.990	0.129	9.682	15.220	Yes
241	0.983	0.128	10.499	15.220	No
242	0.976	0.127	9.985	96.760	No
243	0.962	0.125	13.740	61.960	No
244	0.998	0.130	9.000	12.320	Yes
245	0.990	0.129	11.586	12.320	No
246	0.983	0.128	7.479	12.140	Yes
247	0.976	0.127	2.826	12.140	Yes
248	0.923	0.166	8.455	30.140	Yes
249	0.998	0.130	16.200	12.280	No
250	0.990	0.207	14.280	12.280	Yes
251	0.990	0.129	15.461	21.500	No
252	0.983	0.128	10.479	13.080	No
253	0.976	0.127	4.242	13.080	Yes
254	0.969	0.126	6.235	16.440	Yes
255	0.923	0.132	10.733	87.920	No
256	0.883	0.138	10.133	87.920	No
257	0.998	0.130	5.400	18.840	Yes
258	0.990	0.129	9.661	18.840	Yes
259	0.983	0.128	8.980	23.740	Yes
260	0.976	0.127	12.724	23.740	No
261	0.998	0.130	9.000	92.370	No

262	0.990	0.129	13.891	92.370	No
263	0.983	0.128	13.835	86.660	No
264	0.976	0.127	11.616	86.660	No
265	0.969	0.126	14.086	92.020	No
266	0.953	0.124	12.733	23.580	No
267	0.943	0.123	11.403	23.580	No
268	0.923	0.120	9.433	19.230	No
269	0.990	0.129	7.725	19.620	Yes
270	0.983	0.128	15.591	21.320	No
271	0.883	0.115	12.275	22.510	No
272	0.827	0.107	7.662	22.510	Yes
273	0.998	0.130	3.600	7.480	Yes
274	0.990	0.129	5.793	7.480	Yes
275	0.962	0.125	2.315	15.590	Yes
276	0.883	0.206	18.083	37.980	No
277	0.998	0.130	1.800	25.440	Yes
278	0.990	0.129	3.864	25.440	Yes
279	0.983	0.128	1.497	26.680	Yes
280	0.976	0.127	8.482	26.680	Yes
281	0.953	0.124	15.770	29.840	No
282	0.883	0.121	16.751	8.800	No
283	0.998	0.130	5.400	8.120	Yes
284	0.990	0.129	17.380	8.120	No
285	0.983	0.128	7.479	8.260	Yes
286	0.976	0.127	5.652	8.260	Yes
287	0.969	0.126	3.738	10.580	Yes
288	0.962	0.125	2.254	10.580	Yes
289	0.998	0.130	9.000	12.580	Yes
290	0.990	0.129	9.492	12.580	Yes
291	0.983	0.128	8.823	12.400	Yes
292	0.976	0.127	9.723	12.400	Yes
293	0.969	0.126	6.125	14.420	Yes
294	0.962	0.125	2.216	14.420	Yes
295	0.998	0.130	3.600	99.893	Yes
296	0.990	0.129	11.599	99.798	No
297	0.990	0.129	2.040	17.600	Yes
298	0.998	0.130	3.600	17.740	Yes
299	0.990	0.129	5.795	17.740	Yes
300	0.983	0.147	6.413	17.740	Yes
301	0.976	0.168	4.882	16.660	Yes
302	0.969	0.183	7.507	16.660	Yes
303	0.998	0.130	1.800	32.360	Yes
304	0.990	0.129	3.867	32.360	Yes
305	0.998	0.130	18.000	16.660	No
306	0.990	0.129	15.453	16.660	No
307	0.983	0.128	4.489	12.060	Yes
308	0.976	0.137	16.180	12.060	No
309	0.969	0.154	12.383	20.200	No
310	0.962	0.165	1.297	20.200	Yes
311	0.998	0.281	7.200	14.200	Yes

312	0.990	0.279	16.320	14.200	Yes
313	0.983	0.277	14.280	12.760	Yes
314	0.976	0.275	16.633	12.760	Yes
315	0.969	0.273	9.168	15.140	Yes
316	0.998	0.130	12.600	11.060	No
317	0.990	0.129	3.863	11.060	Yes
318	0.983	0.128	2.992	20.180	Yes
319	0.976	0.127	4.240	20.180	Yes
320	0.998	0.130	10.800	6.240	No
321	0.990	0.129	9.658	6.240	Yes
322	0.983	0.128	2.992	6.240	Yes
323	0.976	0.127	7.066	5.950	Yes
324	0.969	0.126	8.725	5.950	Yes
325	0.883	0.140	16.863	30.370	No
326	0.883	0.179	17.257	27.360	No
327	0.827	0.176	12.832	27.360	Yes
328	0.976	0.166	16.427	93.120	No
329	0.998	0.130	10.800	17.660	No
330	0.990	0.129	5.795	17.660	Yes
331	0.983	0.128	5.985	17.660	Yes
332	0.976	0.127	15.546	21.010	No
333	0.969	0.126	8.725	21.010	Yes
334	0.962	0.125	7.892	35.450	No
335	0.953	0.124	9.334	35.450	No
336	0.943	0.126	7.223	84.980	No
337	0.923	0.135	13.803	13.470	No
338	0.883	0.141	7.466	13.470	Yes
339	0.827	0.142	12.434	13.470	No
340	0.761	0.138	8.491	20.280	Yes
341	0.943	0.123	11.225	26.480	No
342	0.923	0.120	11.172	26.480	No
343	0.883	0.115	12.626	26.480	No
344	0.953	0.124	9.181	37.500	No
345	0.943	0.123	8.015	37.500	No
346	0.923	0.120	14.667	26.500	No
347	0.962	0.125	12.402	9.920	No
348	0.998	0.130	18.000	16.350	No
349	0.983	0.150	12.983	17.740	No
350	0.976	0.172	9.862	17.740	Yes
351	0.998	0.130	7.200	20.120	Yes
352	0.990	0.129	1.932	20.120	Yes
353	0.983	0.128	13.466	17.230	No
354	0.976	0.127	16.959	17.230	No
355	0.969	0.128	12.539	19.360	No

From table 2 we can observe that a total of 355 datasets were extracted from 180 bore holes, where 182 cases were found to be liquefied and non-liquefied cases were 173.

4.3 Logistic Regression

Logistic regression expresses the multiple linear regression equation in logarithmic terms (the logit), thereby overcoming the problem of violating the linearity assumption. Instead of predicting the value of the outcome variable from one or more predictor variables, we predict the probability of the outcome occurring, given known values of the predictors. We used a training set of 70 percent randomly selected and 30 percent of data as the testing set. The available hyperparameters to be tuned for this algorithm were C value, Solver and a penalty. Using various cross-validation techniques we found the best accuracy that can be achieved with this algorithm in correspondence to the best parameters used for achieving it. The logistic regression is basically a statistical model where the equation of linear regression is put through a sigmoid function. This makes the model give binary outputs i.e., predicts a value that can represent two different classes.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots\dots\dots \beta_n x_n$$

This equation above expresses a linear regression according to statistics, where β_0 is the intercept and x is the input feature. But in case of machine learning the intercept is considered as overall bias of the model and the coefficient of x expresses the weights assigned to the corresponding feature. Therefore, the transformed equations are:

$$y = \sum_{i=1}^{m} w_i \, x_i$$

$$\emptyset(y) = \frac{1}{1 + e^{-y}}$$

Here, $\emptyset$ function represents the sigmoid function that was used for prediction in this machine learning model.

4.4 Support Vector Machine

The support vector machine (SVM) classifier uses statistical learning theory as its foundation Vapnik (1995) and searches for the best hyperplane to use as a decision function in high-dimensional space (Boser et al., 1992) (Cristianini and Shawe-Taylor 2000). Instead of employing empirical risk minimization, the SVMs use structural risk minimization to get their results. Empirical risk reduces the probability of making a misclassification on the training set, whereas structural risk reduces the probability of making a misclassification on a data point that has never been seen before and is drawn at random from a probability distribution that is either fixed or unknown Vapnik (1998).

As a result of the fact that the samples from each of the three sets of data are already separated into liquefied cases and non-liquefied cases, the classification of the samples from the three

sets of data is a problem of binary classification. The support vector machine (SVM) is a technique for automatic learning that was developed by Vapnik (1995) and is predicated on the statistical learning theory. When it comes to dealing with binary classification issues, the SVM has had a great degree of success for a significant amount of time in the past. Any problem involving binary classification can be written as (x_i, y_i), where i is an integer between 1 and l and represents the number of cases in the specified test set. The value of y_i can take one of two possible values: either +1 or -1. When utilizing SVM to solve a classification problem, an ideal classification hyperplane must be determined from the outset in order to guarantee accurate classification. Once this is done, the blank area on both sides of the hyperplane must be maximized in order to meet this requirement (Zhou et al., 2018). So we set two boundary conditions to establish a hyperplane by which classification will be don on two kinds of samples.

$$w^T \times x_i + b \geq +1 \text{ for } y_i = +1 \rightarrow \text{liquefaction}$$

$$w^T \times x_i + b \leq -1 \text{ for } y_i = -1 \rightarrow \text{no liquefaction}$$

Around this defined hyperplane there will be many sample points. These points are known as support vectors. The sum of the distance from the hyperplane to the support vectors is called Margin which is expressed by the following equation:

$$Margin = \frac{2}{\|w\|}$$

4.5 Artificial Neural Network

One or more dependent variables can be predicted from a set of independent variables using an Artificial Neural Network, which combines the architecture of the human brain with statistical learning methods. It is referred to as a multilayer perceptron, or MLP, and it is comprised of a straightforward network of nodes or neurons (Ahmed & Pradhan, 2019, p. 7). A weight, denoted by (w_i), and a bias, denoted by b, were each associated with their respective neurons. Using the neural network libraries Keras and Tensorflow in Python, in this study the model used a Multi-layer Perceptron (MLP) network architecture and a training approach called Levenberg-Marquardt Backpropagation. The training algorithm is the approach that is utilized for determining and optimizing the weight that should be assigned to each node in the network. The execution of this technique normally takes more time, but it can lead to a good generalization of the liquefaction dataset. The architecture of the neural network used in study is outlined below in figure 4:

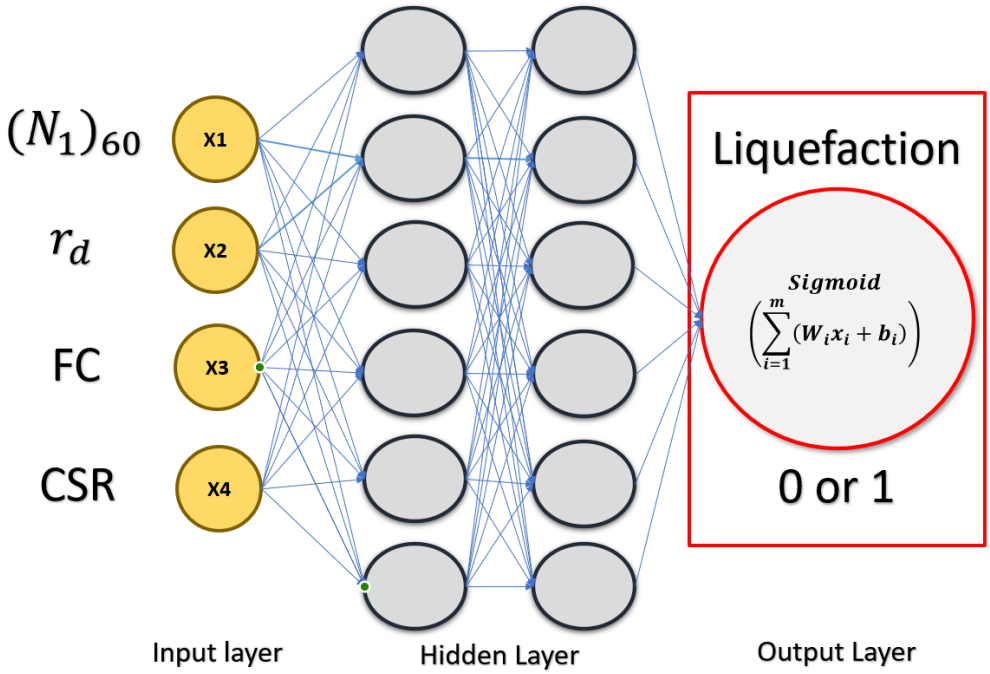

Figure 4 ANN model Architecture

The prediction accuracy of the model is placed as the primary focus of this strategy for selecting variables. It begins with a model that has one independent variable and yields the maximum predicted accuracy. This model is used throughout the process. After that, the outcomes of each possible combination of the remaining variables with the first variable are analysed in order to determine which of the two-variable models is the most accurate. The process is continued

until the inclusion of new variables no longer contributes to an improvement in the model's predictive ability (Anderson & Bro, 2010).

The hyperparameters available for tuning in ANN models are an optimizer, batch size, epochs, learning rate, neuron number and activation functions. For our input layer we used a regular keras dense layer with 4 neurones. Then 2 hidden layers were provided where each layer comprised of 6 neurones and the activation function used was ReLU (Rectified Linear Unit). The output of the network used a sigmoid function and comprised of 1 neuron. When the model was compiled a loss function, an optimizer, batch size and epochs were assigned to it. For this particular model we used a loss function called binary cross entropy, the optimizer was Adam which uses an approach known as stochastic gradient descent, that is predicated on the adaptive estimate of first-order and second-order moments. Finally, a batch size of 32 and 1000 epochs was used to finish compiling the model.

4.5.1 ANN network diagram

The network diagram of the ANN models is outlined below in figure 5:

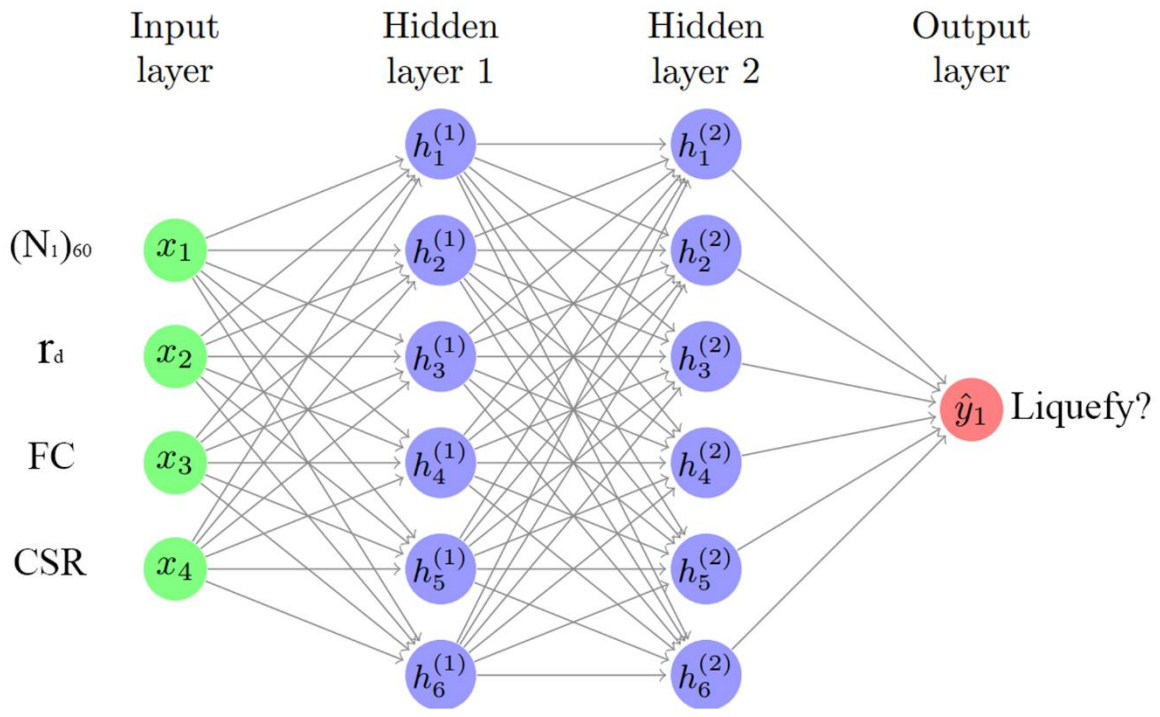

Figure 5 ANN Model Network Diagram

4.6 Model Work Flow

The work flow of the machine learning models is outlined below in figure 6:

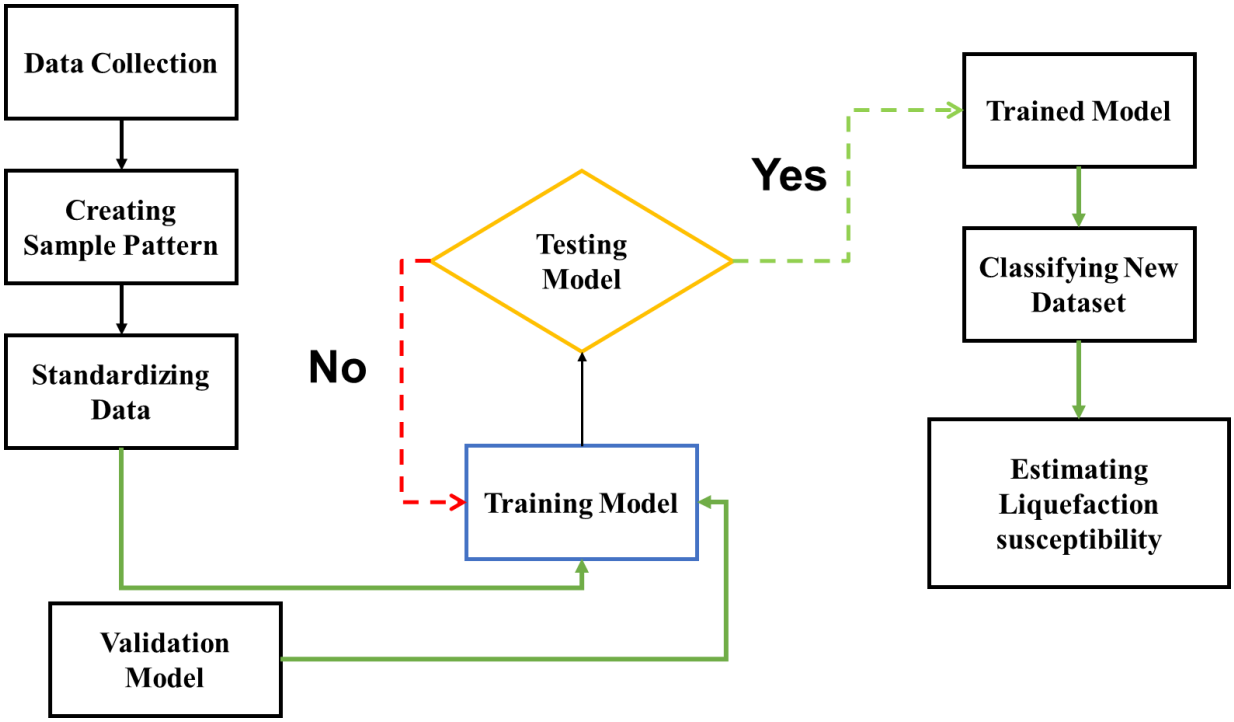

Figure 6 Work Flow of the machine learning Model

4.7 Model Optimization

4.7.1 Feature Scaling

To scale our features, we used a method called standardization. The process of converting the structures of numerous datasets into a single, standardized data format is known as data standardization. It is concerned with the modification of datasets, which takes place after the data have been gathered from a variety of sources but before the data are fed into target systems.

$$z = \frac{x - \mu}{\sigma}$$

Standardizing the dataset would give the machine learning algorithm the opportunity to work more efficiently and give accurate results. In the given equation μ is the mean of the data and σ is the standard deviation.

4.7.2 L2 Regularization

In machine learning, overfitting is a problem where a model is too complex and fits the training set so badly that when it is applied to the testing set unsatisfactory results are produced. Different methods can be used to help these machine learning models to avoid the overfitting problem. L2 regularization is one of the techniques which is used to minimize this overfitting and make the prediction of the model more enhanced.

4.7.3 K-fold Cross Validation

Utilizing the K-fold CV technique, this investigation attempts to solve the issue of the model being overfit to the data that was supplied in order to conduct its analysis. The K-fold CV technique is frequently used in the areas of machine learning, even when there is only a small amount of data available to work with. This method is used to train and modify the model before it is put through its paces against the definitive testing set. In order to cross validate our models, we used 10-fold resampling. This indicates that 9 different data sets were used to train the model, and the 10th different set of data was used to validate the model's performance. This process is repeated ten times with a variety of validation folds, and it was carried out in the absence of a testing set.

4.7.4 Grid Search

Grid Search employs a unique combination of each of the hyperparameters that have been specified and the values for those hyperparameters. It then assesses the performance of each combination and chooses the one that results in the best performance for the hyperparameters. In addition to the Grid Search operation, the cross-validation procedure is carried out using Grid Search CV. During the training of the model, cross-validation is performed. Because of this, the processing will take a significant amount of time and will be prohibitively expensive due to the high number of hyperparameters involved. Before we train the model with the data, the data are first split into two parts: the train data and the test data. This is standard procedure. The train data is further divided into two pieces throughout the cross-validation process, which are referred to as the train data and the validation data respectively. This technique shows best performance for Ensemble Learning. In this study, we used grid search for logistic regression and SVM

4.7.5 Random Search

Random Search uses a variety of different random combinations of hyperparameters in an effort to locate the optimal answer for the constructed model. Although it is quite comparable to grid search, it has been demonstrated to produce superior results in comparison. Because random values are chosen at each occurrence, it is quite possible that the entirety of the action space has been traversed as a result of the randomness. Because of this, it requires an enormous amount of time to cover every facet of the combination when performing grid search. The premise that not all hyperparameters are equally essential yields the best results for this application. Random search has the disadvantage of producing a significant variance when it is being used in computing. Luck plays a role given that the selection of the parameters is entirely arbitrary and there is no use of intelligence in the sampling of the various combinations of parameters. During each iteration of this search procedure, a variety of different parameter combinations are shuffled at random. Because random search follows a pattern in which the model may end up being trained on the optimized parameters without any aliasing, the chances of discovering the optimal parameter are relatively better in random search. This is due to the fact that random search patterns.

If we compare random search with grid search, we can see that grid search combines instances of parameters in accordance with fixed meshes, whereas random search combines those parameters in an erratic manner. This can be demonstrated if we compare random search with grid search. Therefore, it would appear that random search has the capability to locate the best combination of parameters more effectively provided that the number of search options is sufficiently high. According to the findings of a large number of studies, random search produces superior outcomes to grid search in a variety of specific contexts. A randomized search will show best performance for neural networks. In this study, we used Random Search for our ANN algorithm.

4.8 Model Evaluation

For evaluating the performance of a model for classification, different types of techniques are used. These include Overall Accuracy (OA), Precision, Sensitivity, Specificity, F1, RMSE and MAE. These indicators are intended to measure categorical accuracy between the actual and the predicted results. Usually, anything above 0.8 represents a good model. For calculating these indicators, the following formulas are used:

$$OA = \frac{(True\ Negative + True\ Positive)}{(True\ Negative + False\ Positive + False\ Negative + True\ \ Positive)}$$

$$Precision = \frac{True\ Positive}{False\ Positive + True\ Positive}$$

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

$$F1 = \frac{2\ True\ Positive}{True\ Positive + False\ Positive + False\ Negative}$$

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

$$MAE = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n}$$

For understanding these the parameters for calculating the indicators, the following table 1 expresses a typical confusion matrix:

Table 3 Confusion matrix between cluster labels

Actual	Predicted	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Chapter 5 RESULT AND DISCUSSION

5.1 Exploratory Data analysis

The term "Exploratory Data Analysis" refers to the critical process of performing initial investigations on data in order to discover patterns, to spot anomalies, to test hypothesis, and to check assumptions with the assistance of summary statistics and graphical representations. This is a very important step in the analysis process. Understanding the data first and making an effort to gain as much insight as possible from it is a recommended course of action. EDA is used largely to examine what the data can disclose beyond the formal modelling or hypothesis testing work, and it also provides a better knowledge of the variables contained inside a data collection as well as the relationships between those variables. In addition to this, it can assist in determining whether or not the statistical methods that are being considered for the data analysis are suitable. This means before engaging ourselves into actual modelling first we try to visualize the data and then try to interpret the visualization.

5.1.1 Scatterplots

The scatterplots of our 4 input features are shown below. Here we can see the cases for which liquefaction has occurred and for which it has not. The range of values can be interpreted from this type of graphs for assessing the occurrence of an event.

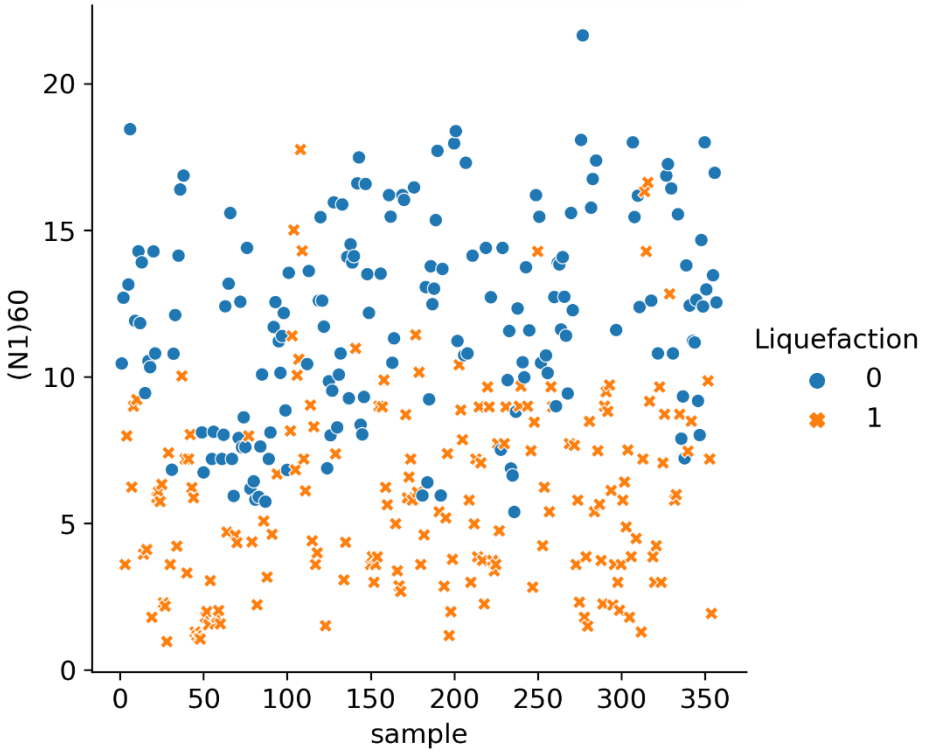

Figure 7 Scatterplot of $(N_1)_{60}$

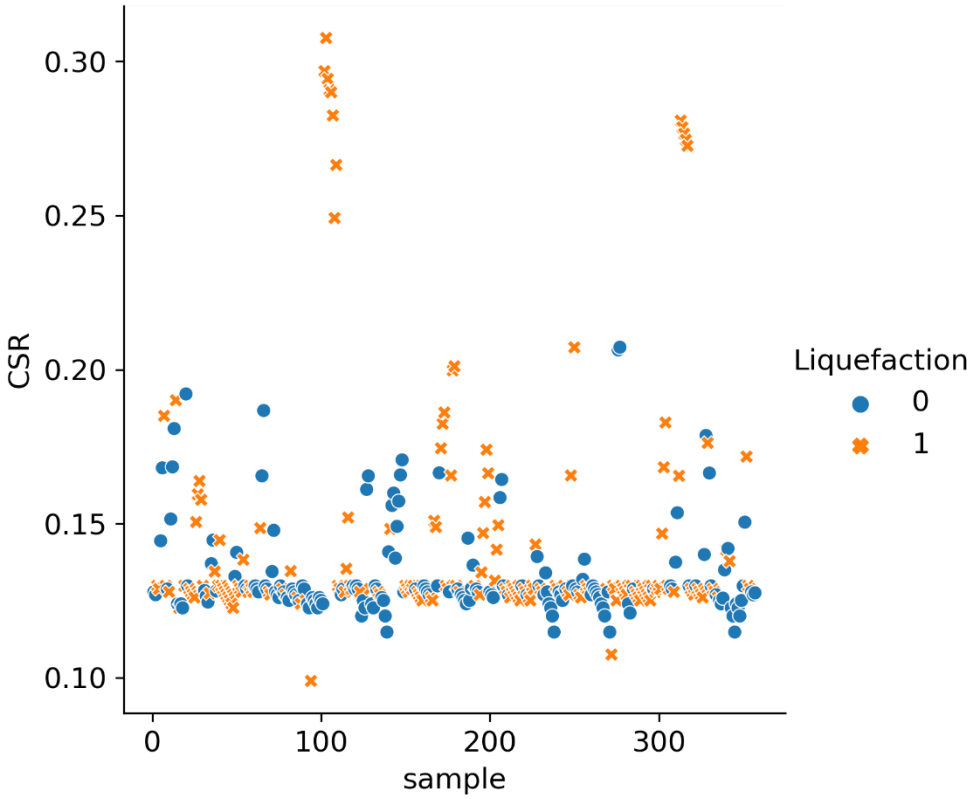

Figure 9 Scatterplot of CSR

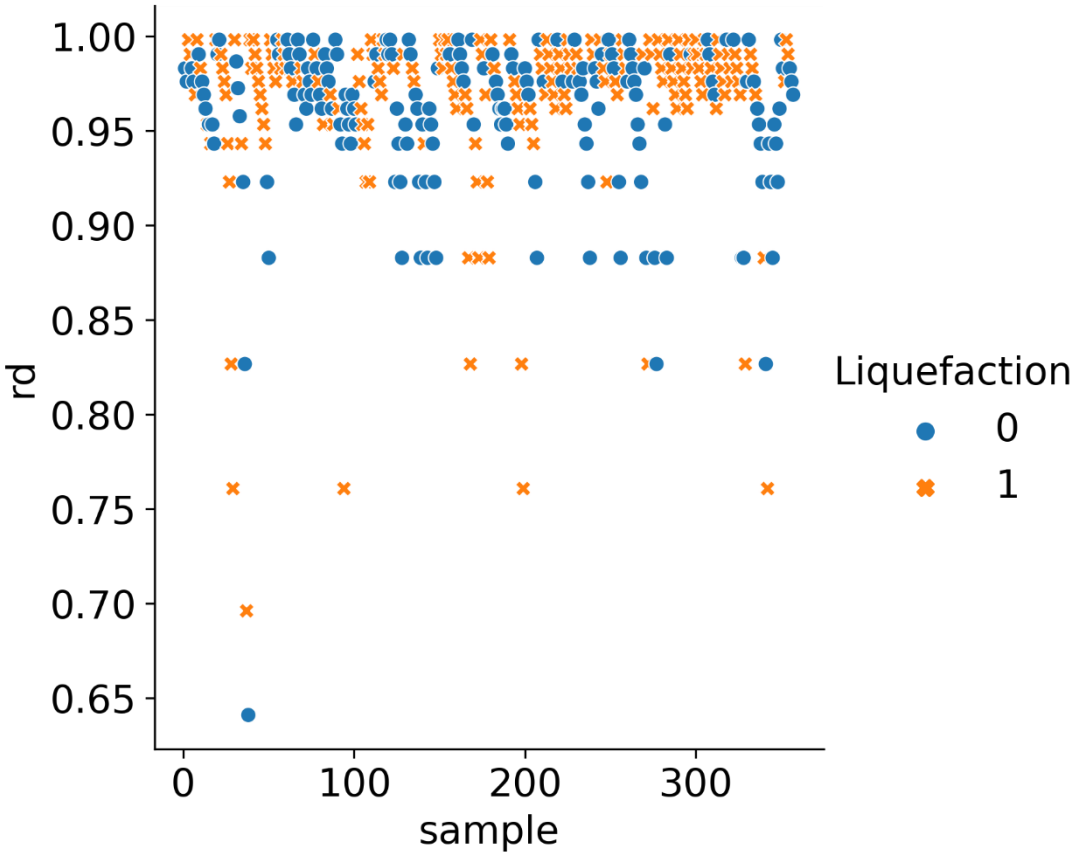

Figure 8 Scatterplot of rd

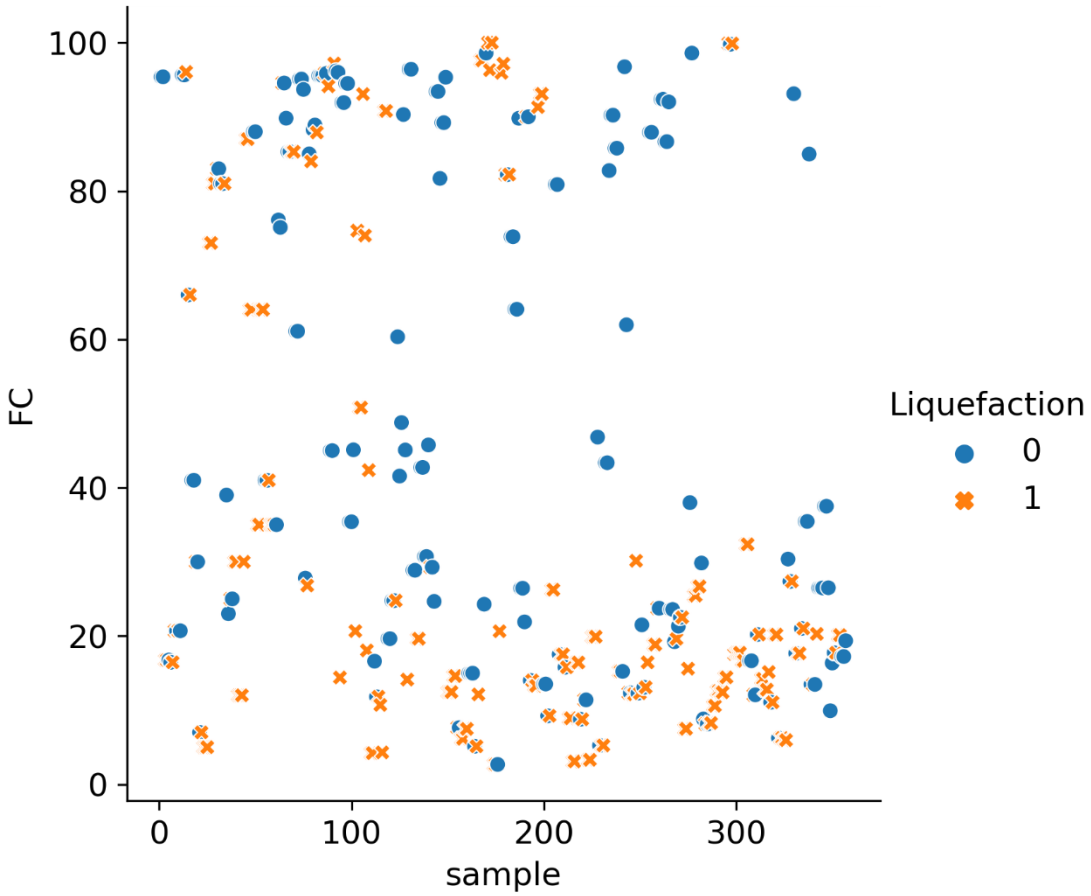

Figure 10 Scatterplot of FC

5.1.2 Histograms with KDE

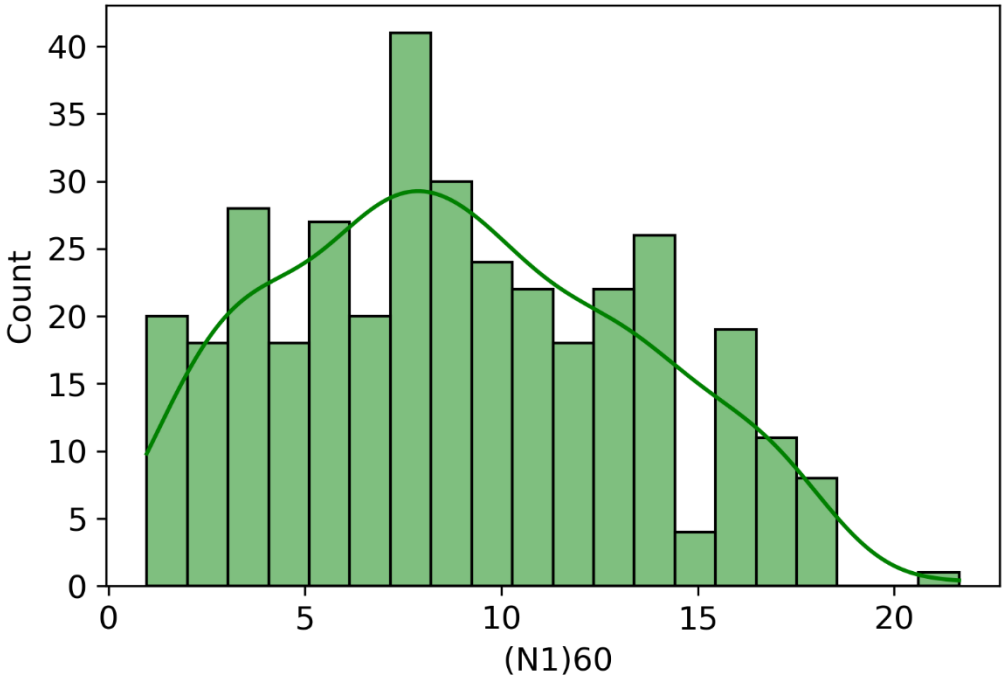

Figure 11 Histogram of $(N_1)_{60}$

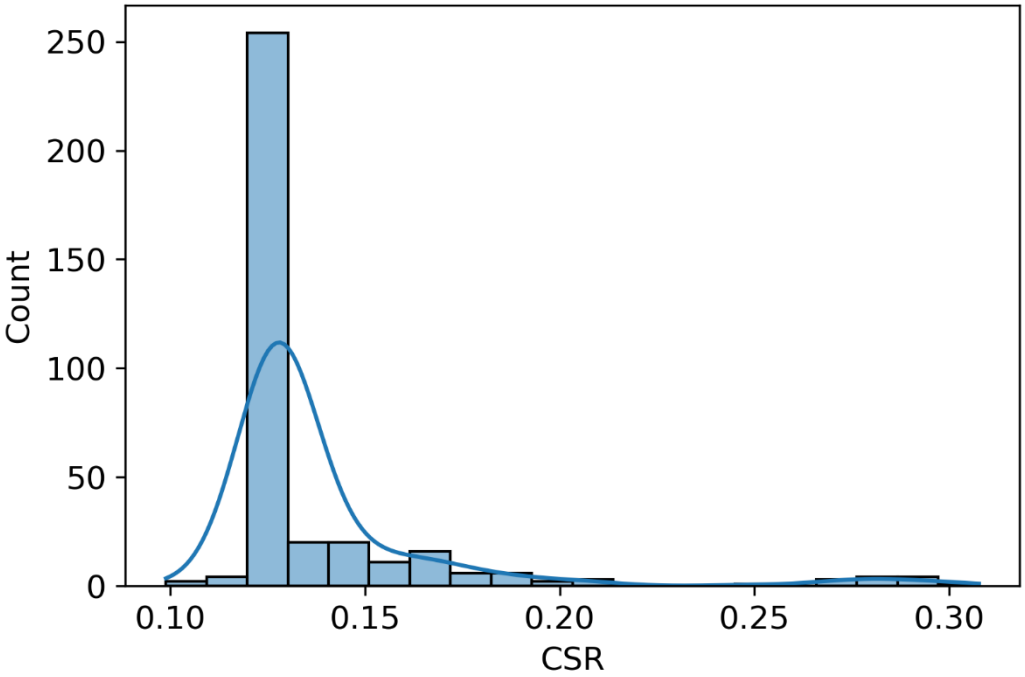

Figure 12 Histogram of CSR

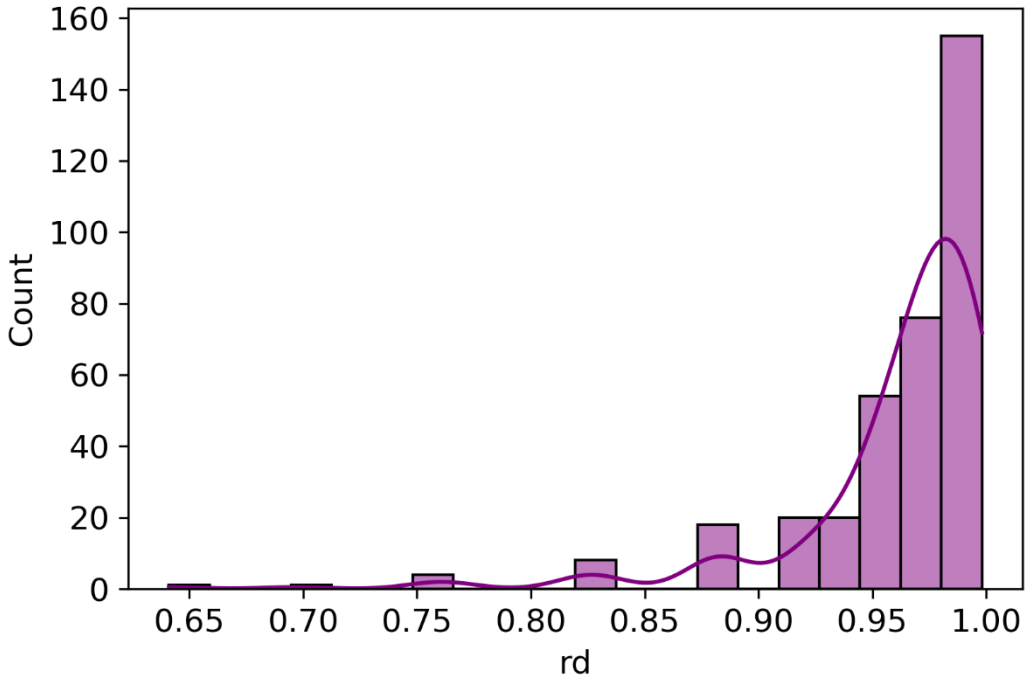

Figure 14 Histogram of rd

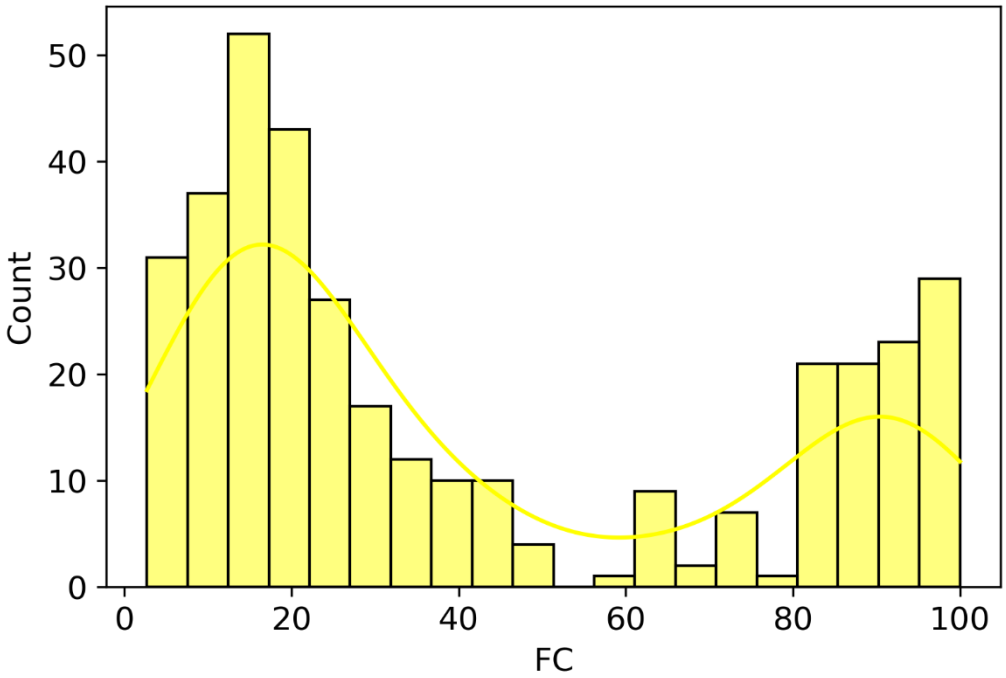

Figure 13 Histogram of FC

5.2 Statistical Information

Descriptive statistics can also be used to describe a population in its entirety. In a nutshell, descriptive statistics are used to help describe and comprehend the characteristics of a particular data set by providing concise summaries of the data set's samples as well as its measurements.

Measures of centre, such as the mean, median, and mode, are the most well-known types of descriptive statistics, and they are utilized almost universally across all educational levels of mathematics and statistics. To determine the mean, also known as the average, first add up all of the figures that are contained within the data set. Next, divide this total by the total number of figures that are contained within the data set. The statistical category known as descriptive statistics is subdivided into measures of central tendency and measures of variability (spread). The mean, the median, and the mode are all examples of measures of central tendency. On the other hand, measures of variability include the standard deviation, variance, as well as minimum and maximum variables.

Table 4 Statistical Description of data

Features	Count	Mean	Standard Deviation	Minimum	Q_1	Median	Q_2	Maximum
r_d	355.0	.9623	4.623597e-02	.640990	.953344	.976006	.990420	.998156
CSR	355.0	.1396	3.209568e-02	.098898	.126881	.128755	.134452	.307640
$(N_1)_{60}$	355.0	8.8293	4.468369e+00	.972050	5.577806	8.555852	12.387329	18.449158
FC	355.0	40.987	3.308514e+01	2.68	14.35	25.44	81	97.8

5.3 Model Validation

5.4 ROC Curve

The evaluation yielded some findings, which were then presented in the form of a Receiver Operating Characteristic (ROC) curve. The ROC curve is a plot of sensitivity versus false positive rate, where the line along the diagonal represents a pure 50 percent chance of accurate prediction of a model, and the Area Under the ROC Curve (AUC) represents a value between 0 and 1, where a value closer to 1 suggests the better performance of the model. In summary, the ROC curve shows a plot of sensitivity versus false positive rate (Park, Goo, & Jo, 2004). For the purposes of model validation, an AUC value that is greater than 0.7 is generally

considered to be an acceptable value. In our research, the AUC value for the Logistic Regression model was 0.969, the AUC value for the SVM model was 0.992, and the AUC value for the ANN model was 0.97, both of which are indicative of a well-functioning model. The ROC curves of the individual models and an ROC curve of the compilation of all models are as follows:

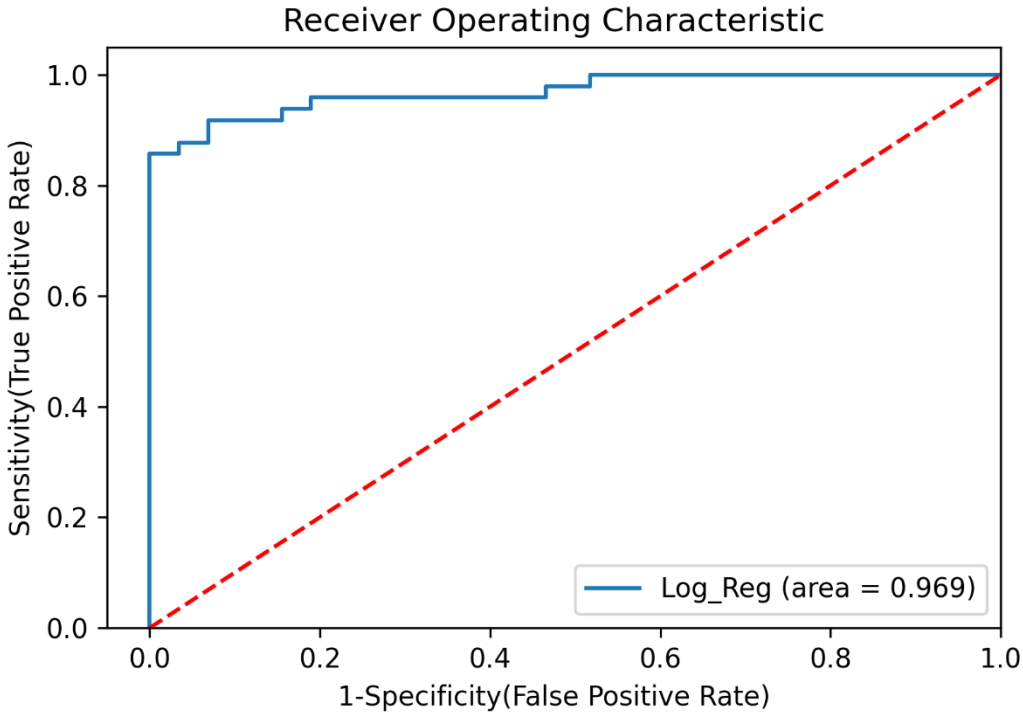

Figure 15 ROC curve Logistic Regression

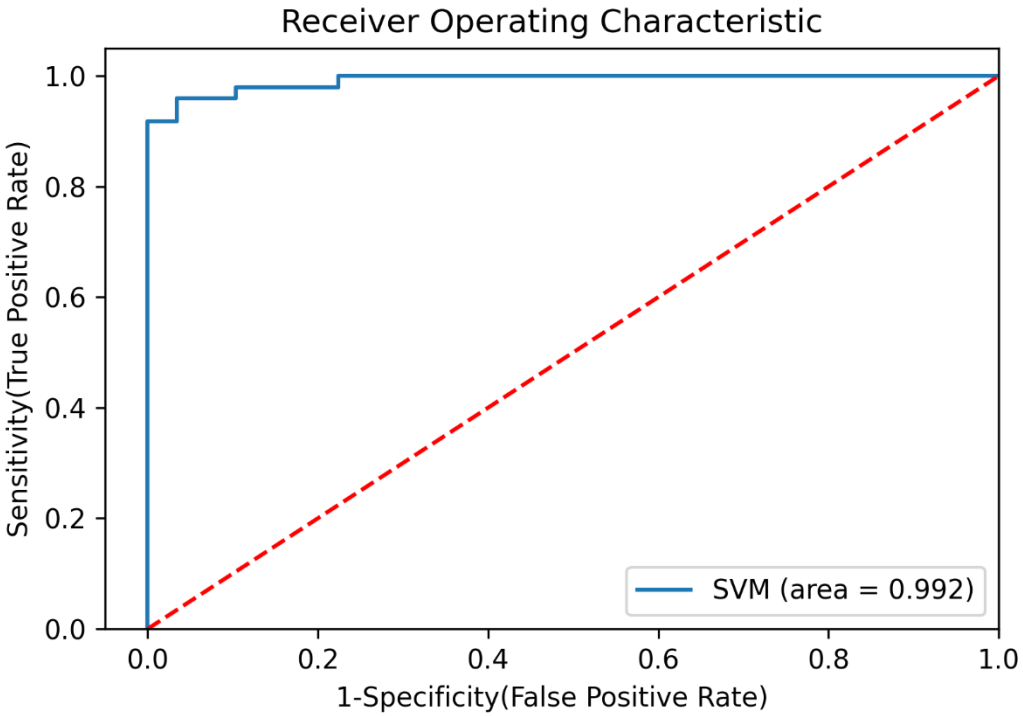

Figure 16 ROC curve SVM

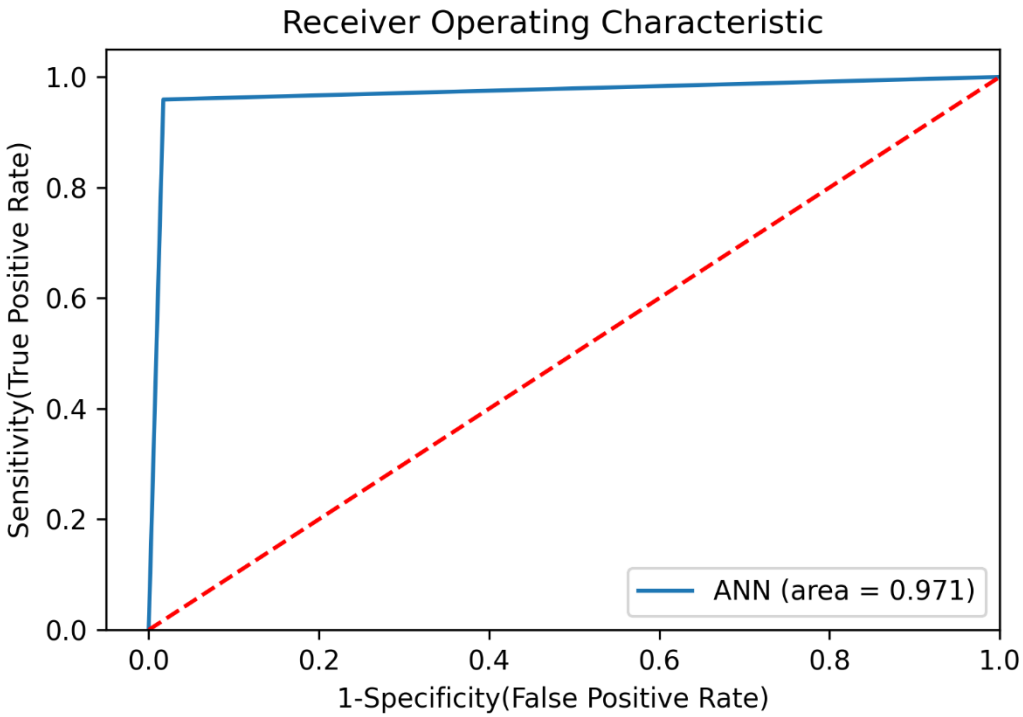

Figure 17 ROC curve ANN

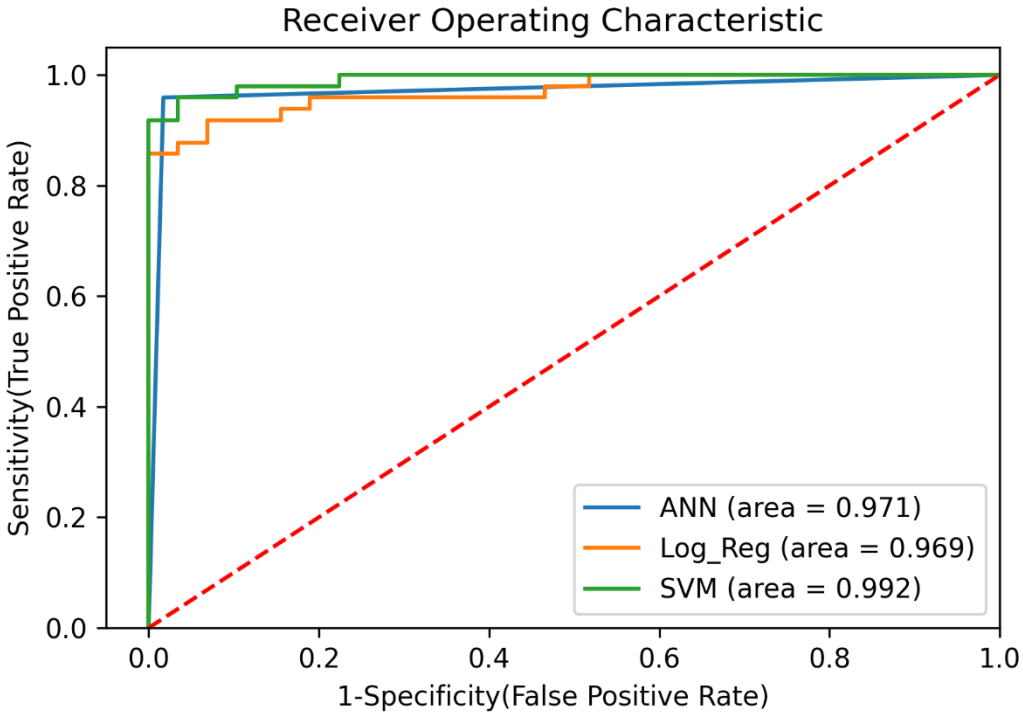

Figure 18 ROC curve all Models Comparison

5.5 Confusion Matrix

A method for analysing and summarizing the performance of a classification techniques is known as a confusion matrix. If you have an unbalanced number of observations in each class or if you have more than two classes in your dataset, relying solely on classification accuracy could lead to inaccurate conclusions. The calculation of a confusion matrix can provide you with a clearer picture of the aspects of your classification model that are functioning correctly as well as the mistakes that it is producing. A summary of the results of predictions made on a classification task is known as a confusion matrix. Count values are used to compile a summary of both the number of accurate and incorrect predictions, which is then broken down by each class. The confusion matrix is unsolvable without this piece of information. The confusion matrix illustrates the various ways in which your classification model can get its predictions wrong when it is used to create forecasts. It provides you with an understanding not only of the faults that your classifier is making, but also, and perhaps more crucially, the categories under which those errors fall. The drawback of relying solely on categorization accuracy is circumvented thanks to the aforementioned breakdown. The confusion matrices developed for the models implemented in this study are as follows:

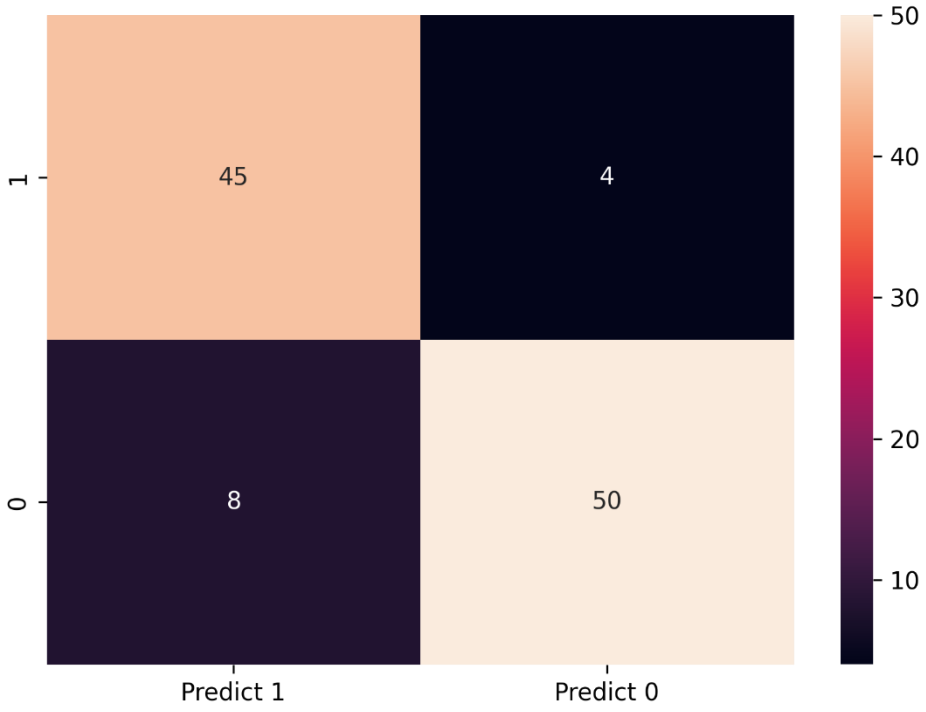

Figure 19 Confusion Matrix Logistic Regression

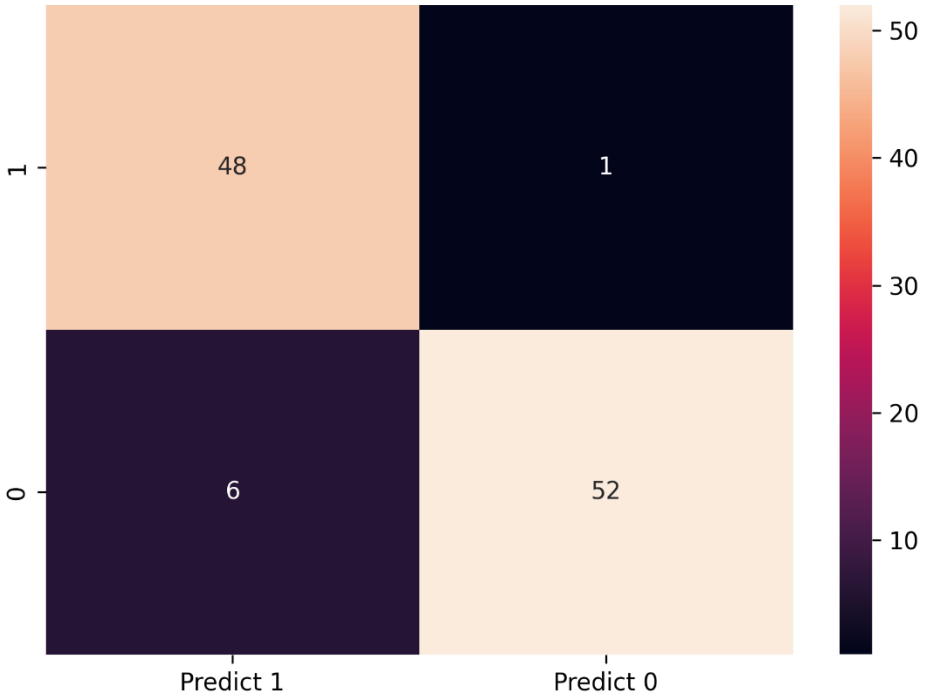

Figure 21 Confusion Matrix SVM

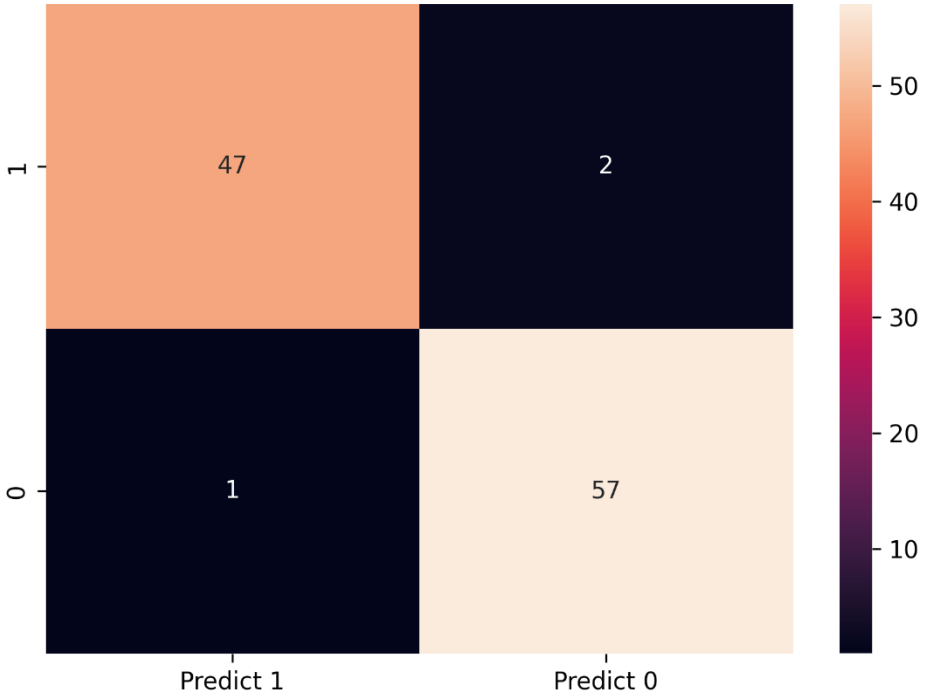

Figure 20 Confusion Matrix ANN

5.5.1 Correlation

The relationship between the feature parameters against liquefaction can be established with a correlation matrix.

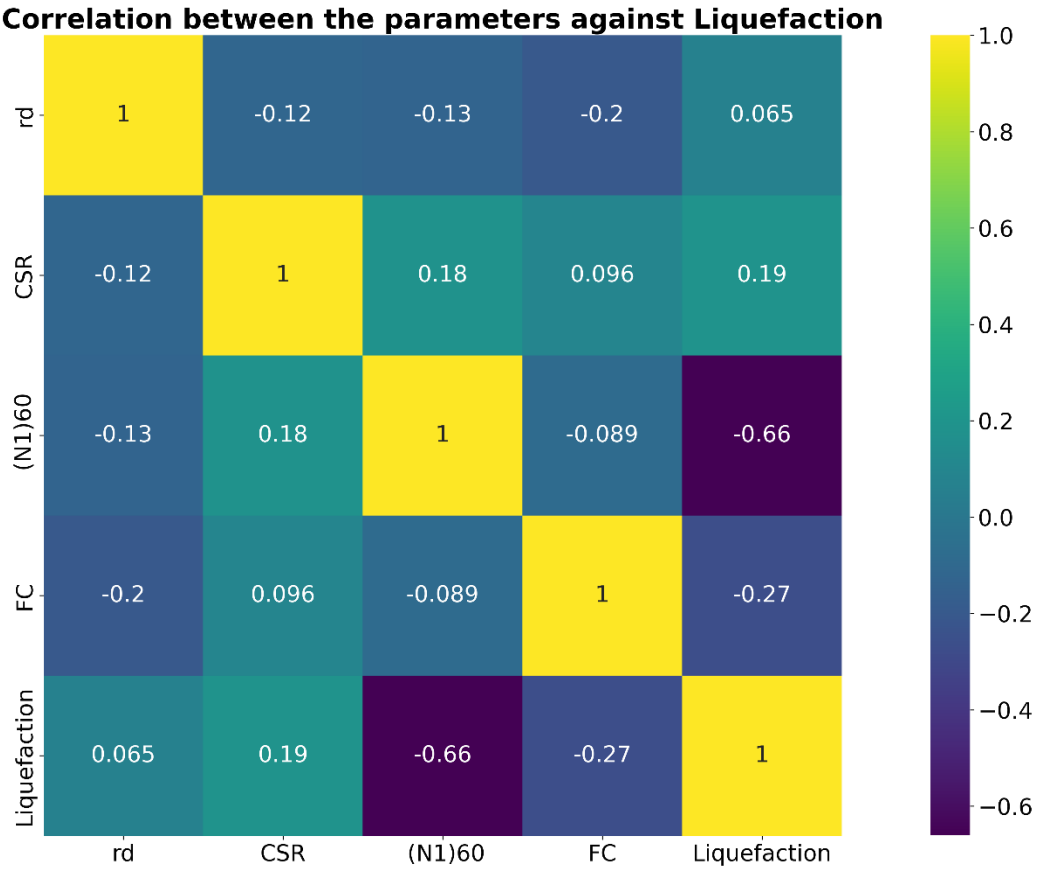

Figure 22 Correlation Between input features against Liquefaction

From this matrix we can observe that $(N_1)_{60}$ (r = -0.66) and FC (r = -0.27) has the most influence in triggering liquefaction. While the least influential parameter to trigger liquefaction is r_d (r = 0.065). This confirms that the selection of the input parameters was satisfactory and that most of the parameters except rd has quite a big amount of influence in liquefaction triggering.

5.6 Model Comparison

The comparison of these models is traditionally done with different types of scores. The scores express the model performance in classifying different datasets from one or more input features. In this study we used seven types scoring techniques to justify the reason of choosing one specific model for predicting the triggering of liquefaction accurately. There are Overall Accuracy, Precision, Sensitivity, Specificity, F1, RMSE and MAE. With these indicators we can easily compare the performance of the models and choose one to make accurate predictions for the developed dataset.

Table 5 Performance Evaluation Indicators

Model	`Indicators						
	OA	Precision	Sensitivity	Specificity	F1	RMSE	MAE
LogReg	0.8878	0.85	0.92	0.926	0.8823	0.33488	0.1121
SVM	0.93457	0.89	0.9796	0.8965	0.932	0.2557	0.0654
ANN	0.972	0.9791	0.9592	0.9827	0.9691	0.16744	0.028

From table 5, we can observe that the highest model performance is exhibited by the ANN mode. The SVM model is not that far behind and the least performance is from the logistic regression model. This proves that neural network has an advantage over ensemble learning algorithms, which is why they are preferred as one the best classification algorithms.

Chapter 6 CONCLUSION

6.1 Key Findings

In this study, we found that the classification performance evaluation cannot be done using only performance indicators. Techniques like ROC curves and confusion matrices are also better ways to assess the classification capability and model validation. So, during this study, we came across an event that even though the ANN algorithm shows high performance indicators the SVM model performs better in classifying the testing dataset we provided. Further findings are listed below:

- This study shows that even though the ANN model has the highest accuracy, SVM has better chances of predicting the right outcome. From ROC curve, we can see that SVM has less chances of predicting false negative values.
- 3 out of 5 of the input parameters show that they're skewed when represented graphically. This more or less makes the model biased and these models are likely to predict one-sided outcomes, so the models can't be validated even if though they exhibit a high accuracy. For reducing the biasness, we can introduce some more feature parameters, such as water table depth and Unit weight of soil.
- Another reason, for these models to produce such high accuracies is because these models were regularized and their input features were standardized.
- For smaller dataset i.e., data less than 1000, we can confirm that ensemble learning can be prioritized over neural networks.

6.2 Limitations

In this study, we materialized the datasets using empirical equations. Although this type of simplified procedure has acquired approval from a lot of institutions, they are still in consistent. For producing more validated models we need observed data from the field, where liquefaction has actually occurred. This type of instances will prove the model to be more applicable in real life situations. Without actual proof of the event occurring, assessing liquefaction based on empirical methods cannot be accepted for application in real life scenarios.

Although this research implements cutting edge technology for data analysis, we still cannot implement it due to lack of practical data. Other than that, the models can be applied anywhere else for deciding ground improvement degree.

6.3 Future Study

For further study in this topic the following methods can be implement for better and more developed implementation of machine learning in geotechnical and seismological fields:

- This study can further be improved by adding different classes to asses a deeper degree of Liquefaction instead of just predicting a binary output.
- A future study can be conducted where a damage assessment can be done for future situations so that engineers can decide the degree of ground mitigation required for a specific site.
- Resources will be far less consumed if this study can be implemented in practical field.
- Samples from places other than Dhaka city can be collected to make the models more diverse and generalized.

REFERENCE

Juang, C. H., Jiang, T., & Andrus, R. D. (2002). Assessing probability-based methods for liquefaction potential evaluation. *Journal of Geotechnical and Geoenvironmental Engineering*, *128*(7), 580–589.

Samui, P., & Sitharam, T. G. (2011). Machine learning modelling for predicting soil liquefaction susceptibility. *Natural Hazards and Earth System Science*, *11*(1), 1–9. https://doi.org/10.5194/nhess-11-1-2011

Robertson, P. K., & Wride, C. E. (1998). Evaluating cyclic liquefaction potential using the cone penetration test. *Canadian Geotechnical Journal*, *35*(3), 442–459.

Ahmed, A. A., & Pradhan, B. (2019). Vehicular traffic noise prediction and propagation modelling using neural networks and geospatial information system. *Environmental Monitoring and Assessment*, *191*(3), 1–17.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Vapnik, V. (1998). *Statistical learning theory wiley new york google scholar*.

Iwasaki, T. (1978). A practical method for assessing soil liquefaction potential based on case studies at various sites in Japan. *Proc. Second Int. Conf. Microzonation Safer Construction Research Application, 1978*, *2*, 885–896.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152.

Idriss, I. M., & Boulanger, R. W. (2006). Semi-empirical procedures for evaluating liquefaction potential during earthquakes. *Soil Dynamics and Earthquake Engineering*, *26*(2–4), 115–130.

Seed, H. B., & De Alba, P. (1986). Use of SPT and CPT tests for evaluating the liquefaction resistance of sands. *Use of in Situ Tests in Geotechnical Engineering*, 281–302.

Samui, P., & Sitharam, T. G. (2011). Machine learning modelling for predicting soil liquefaction susceptibility. *Natural Hazards and Earth System Science*, *11*(1), 1–9. https://doi.org/10.5194/nhess-11-1-2011

Robertson, P. K., Woeller, D. J., & Finn, W. D. L. (1992). https://doi.org/10.1061/(ASCE)GT.1943-5606.0000631. *Canadian Geotechnical Journal*, *29*(4), 686–695.

Keefer, D. K. (1984). Landslides caused by earthquakes. *Geological Society of America Bulletin*, *95*(4), 406–421.

Shahin, M. A., Jaksa, M. B., & Maier, H. R. (2000). *Predicting the settlement of shallow foundations on cohesionless soils using back-propagation neural networks*. Department of Civil and Environmental Engineering, University of Adelaide ….

Cetin, K. O., Seed, R. B., Kayen, R. E., Moss, R. E. S., Bilge, H. T., Ilgac, M., & Chowdhury, K. (2018). Examination of differences between three SPT-based seismic soil liquefaction triggering relationships. *Soil Dynamics and Earthquake Engineering*, *113*(July 2017), 75–86. https://doi.org/10.1016/j.soildyn.2018.03.013

Andersen, C. M., & Bro, R. (2010). Variable selection in regression—a tutorial. *Journal of Chemometrics*, *24*(11-12), 728–737.

Seed, H. B. (1982). Ground motions and soil liquefaction during earthquakes. *Earthquake Engineering Research Insititue*.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Zhou, J., Huang, S., Wang, M., & Qiu, Y. (2021). Performance evaluation of hybrid GA–SVM and GWO–SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation. *Engineering with Computers*, *0123456789*. https://doi.org/10.1007/s00366-021-01418-3

Idriss, I. M., & Boulanger, R. W. (2010). SPT-based liquefaction triggering procedures. *Rep. UCD/CGM-10*, *2*, 4–13.

Haque, D. M. E., Khan, N. W., Selim, M., Kamal, A. S. M. M., & Chowdhury, S. H. (2020). Towards Improved Probabilistic Seismic Hazard Assessment for Bangladesh. *Pure and Applied Geophysics*, *177*(7), 3089–3118. https://doi.org/10.1007/s00024-019-02393-z

Seed, H. B., & Idriss, I. M. (1967). Analysis of soil liquefaction: Niigata earthquake. *Journal of the Soil Mechanics and Foundations Division*, *93*(3), 83–108.

Lenz, J. A., & Baise, L. G. (2007). Spatial variability of liquefaction potential in regional mapping using CPT and SPT data. *Soil Dynamics and Earthquake Engineering*, *27*(7), 690–702. https://doi.org/10.1016/j.soildyn.2006.11.005

Pham, T. A. (2021). Application of Feedforward Neural Network and SPT Results in the Estimation of Seismic Soil Liquefaction Triggering. *Computational Intelligence and Neuroscience*, *2021*. https://doi.org/10.1155/2021/1058825

Zhou, Z., Zhang, R., Wang, Y., Zhu, Z., & Zhang, J. (2018). Color difference classification based on optimization support vector machine of improved grey wolf algorithm. *Optik*, *170*, 17–29.

Robertson, P. K., & Campanella, R. G. (1985). Liquefaction potential of sands using the CPT. *Journal of Geotechnical Engineering*, *111*(3), 384–403.

Seed, H. B., & Idriss, I. M. (1971). Simplified procedure for evaluating soil liquefaction potential. *Journal of the Soil Mechanics and Foundations Division*, *97*(9), 1249–1273.

Hsein Juang, C., Chen, C. J., Jiang, T., & Andrus, R. D. (2000). Risk-based liquefaction potential evaluation using standard penetration tests. *Can Sci Publ J*, *37*.

Park, S. H., Goo, J. M., & Jo, C. (n.d.). Receiver Operating Characteristic (ROC) Curve: Practical Guide for Radiologists. *Korean J Radiol*, *5*(1).

Chadha, J., Jain, A., & Kumar, Y. (2022). Artificial intelligence techniques in wireless sensor networks for accurate localization of user in floor, building and indoor area. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-12979-w

Tang, Y., Zhang, Y.-Q., Huang, Z., & Hu, X. (2005). Granular SVM-RFE gene selection algorithm for reliable prostate cancer classification on microarray expression data. *Fifth IEEE Symposium on Bioinformatics and Bioengineering (BIBE'05)*, 290–293.

Fear, C. E., & McRoberts, E. C. (1993). *Report on liquefaction potential and catalogue of case records*. Geotechnical Engineering Library, Department of Civil Engineering ….

Kurup, P. U., & Dudani, N. K. (2002). Neural networks for profiling stress history of clays from PCPT data. *Journal of Geotechnical and Geoenvironmental Engineering*, *128*(7), 569–579.

J., C. C., & Hsein, J. C. (2022). Calibration of SPT- and CPT-Based Liquefaction Evaluation Methods. In *Innovations and Applications in Geotechnical Site Characterization* (pp. 49–64). https://doi.org/doi:10.1061/40505(285)4

Youd, T. L., & Idriss, I. M. (2001). Liquefaction Resistance of Soils: Summary Report from the 1996 NCEER and 1998 NCEER/NSF Workshops on Evaluation of Liquefaction Resistance of Soils. *Journal of Geotechnical and Geoenvironmental Engineering*, *127*(4), 297–313. https://doi.org/10.1061/(asce)1090-0241(2001)127:4(297)

Fahim, A. K. F., Rahman, M., Hossain, M., & Kamal, A. S. M. (2022). Liquefaction resistance evaluation of soils using artificial neural network for Dhaka City, Bangladesh. *Natural Hazards*, 1–31.

Caudill, M., & Butler, C. (1992). *Understanding neural networks; computer explorations*. MIT press.

Pal, M. (2006). Support vector machines-based modelling of seismic liquefaction potential. *International Journal for Numerical and Analytical Methods in Geomechanics*, *30*(10), 983–996. https://doi.org/10.1002/nag.509

Rahman, M. Z., Siddiqua, S., & Kamal, A. S. M. M. (2015). Liquefaction hazard mapping by liquefaction potential index for Dhaka City, Bangladesh. *Engineering Geology*, *188*, 137–147. https://doi.org/10.1016/j.enggeo.2015.01.012

Seed, H. B., Idriss, I. M., & Arango, I. (1983). Evaluation of liquefaction potential using field performance data. *Journal of Geotechnical Engineering*, *109*(3), 458–482.

Padmini, D., Ilamparuthi, K., & Sudheer, K. P. (2008). Ultimate bearing capacity prediction of shallow foundations on cohesionless soils using neurofuzzy models. *Computers and Geotechnics*, *35*(1), 33–46.

Cetin, K. O., Seed, R. B., Der Kiureghian, A., Tokimatsu, K., Harder, L. F., Kayen, R. E., & Moss, R. E. S. (2004). Standard Penetration Test-Based Probabilistic and Deterministic Assessment of Seismic Soil Liquefaction Potential. *Journal of Geotechnical and Geoenvironmental Engineering*, *130*(12), 1314–1340. https://doi.org/10.1061/(asce)1090-0241(2004)130:12(1314)

Juang, C. H., Yuan, H., Lee, D.-H., & Lin, P.-S. (2003). Simplified cone penetration test-based method for evaluating liquefaction resistance of soils. *Journal of Geotechnical and Geoenvironmental Engineering*, *129*(1), 66–80.