

M.Tech (Geotechnical Engineering)

Anurag D Pandey (24/GTE/01)

2026

**SWELLING CHARACTERISATION AND RISK
CLASSIFICATION OF BLACK COTTON SOIL USING
EMPIRICAL CORRELATIONS AND MACHINE
LEARNING**

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

**MASTER OF TECHNOLOGY
IN
GEOTECHNICAL ENGINEERING**

Submitted by:

**ANURAG DAYAL
PANDEY
(24/GTE/01)**

Under the Supervision of

**Prof. ANIL KUMAR SAHU
Professor
Delhi Technological University**



Department of Civil Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daultapur, Main Bawana Road, Delhi-110042

MAY, 2026

**SWELLING CHARACTERISATION AND RISK
CLASSIFICATION OF BLACK COTTON SOIL USING
EMPIRICAL CORRELATIONS AND MACHINE
LEARNING**

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

**MASTER OF TECHNOLOGY
IN
GEOTECHNICAL ENGINEERING**

Submitted by:

**ANURAG DAYAL
PANDEY
(24/GTE/01)**

Under the Supervision of

**Prof. ANIL KUMAR SAHU
Professor
Delhi Technological University**



Department of Civil Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daultapur, Main Bawana Road, Delhi-110042

MAY, 2026



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daultapur, Main Bawana Road, Delhi-110042

CANDIDATE’S DECLARATION

I, **ANURAG DAYAL PANDEY**, M. Tech (Geotechnical Engineering) student, having **Roll no: 24/GTE/01**, hereby certify that the work which is being presented in the dissertation entitled “**SWELLING CHARACTERISATION AND RISK CLASSIFICATION OF BLACK COTTON SOIL USING EMPIRICAL CORRELATIONS AND MACHINE LEARNING**” in partial fulfillment of the requirements for the award of the Degree of **Master Of Technology in Geotechnical Engineering**, submitted in the **Department of Civil Engineering , Delhi Technological University**, Delhi Technological University is an authentic record of my work carried out under the supervision of **Prof. ANIL KUMAR SAHU**, Professor, Department of Civil Engineering, Delhi Technological University, Delhi.

The matter presented in this dissertation has not been submitted by me for the award of any other degree of this or any other Institute.

(ANURAG DAYAL PANDEY)

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Prof. ANIL KUMAR SAHU
(SUPERVISOR)

Place: Delhi

Date:

(Signature of External Examiner)



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daultapur, Main Bawana Road, Delhi-110042

CERTIFICATE BY THE SUPERVISOR

Certified that **ANURAG DAYAL PANDEY (24/GTE/01)** has carried out his research work presented in this thesis entitled “**SWELLING CHARACTERISATION AND RISK CLASSIFICATION OF BLACK COTTON SOIL USING EMPIRICAL CORRELATIONS AND MACHINE LEARNING**” for the award of the Degree of **Master of Technology in Geotechnical Engineering** from the Department of Civil Engineering, Delhi Technological University, Delhi, under our supervision. The thesis embodies the results of original work and studies are carried out by the student himself. The contents of the thesis do not form the basis for the award of any degree to the candidate or to anybody else from this or any other University/Institution.

Prof. ANIL KUMAR SAHU
SUPERVISOR
Delhi Technological University

Place: Delhi

Date:

ABSTRACT

Black cotton soils (BCS) of Deccan Trap region of India are characterized by the presence of high proportion of montmorillonite clay mineralogy and exhibit high swell-shrink characteristics leading to distress in roads, foundations, embankment and lightly loaded structures. In Indian geotechnical practice, the Free Swell Index (FSI) determined by comparing the sedimented volume of oven dried soil in distilled water with kerosene is the standard index of expansiveness of the soil. The FSI test is relatively simple to perform in a well-equipped lab, but the Atterberg limit data (liquid limit, plastic limit, plasticity index and shrinkage limit) are readily available from the earliest stages of site investigation. Therefore, it would be beneficial to establish reliable relationships between these index properties and FSI and provide a complementary assessment tool for geotechnical engineers that would not replace the actual FSI test.

This study aims to fill that need with a two-part investigation. During the first phase, the black cotton soil was taken from Shajapur district, Madhya Pradesh and mixed with sodium bentonite at 0%, 2%, 4%, 6%, 8% and 10% of the dry weight of soil to prepare six soil mixtures (M1 – M6). Each mixture was tested for Atterberg limits (LL, PL, PI, SL) and FSI as per IS:2720 and IS:9451 respectively. The Pearson correlation analysis showed that all four Atterberg limit parameters were highly correlated with FSI: LL ($r = +0.9998$), PI ($r = +0.9997$) and SL ($r = -0.9981$). Five empirical regression equations were developed, ranging from single-variable Atterberg limit models — $FSI = 2.299 \cdot PI - 22.315$ ($R^2 = 0.9995$) and $FSI = 2.037 \cdot LL - 61.289$ ($R^2 = 0.9996$) — to dual-Atterberg-limit models incorporating PI and SL ($R^2 = 0.9996$) and LL and SL ($R^2 = 0.9997$). The single-variable PI and LL models were found to be the most robust models in this small dataset by leave-one-out cross validation (LOOCV), with RMSE values of 0.53% and 0.47%, respectively.

The second phase involved training and testing three machine learning classifiers (Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (ANN)) on a compiled database of 186 Indian BCS samples from ten states to

predict the IS:1498 swelling risk class (Low, Medium, High, Very High) from LL, PI, SL, optimum moisture content (OMC), and maximum dry density (MDD). The class imbalance issue was addressed using stratified 5-fold cross-validation. Random Forest was the best model in terms of balanced accuracy (0.860, CV accuracy = 0.880 and Cohen's κ = 0.792), indicating its appropriateness for categorical swelling risk assessment.

The derived empirical equations can be used to estimate FSI from Atterberg limit index properties for Shajapur-type Malwa Plateau BCS and the Random Forest classifier can be used to map the swelling potential of BCS in a geographically wider Indian context. Both are designed to be complementary rapid assessment tools to support preliminary decision making in the field by the geotechnical engineer, rather than replacing direct measurement of FSI where site conditions require it.

Keywords: Black cotton soil, bentonite, free swell index, Atterberg limits, empirical regression, machine learning, Random Forest, IS:9451, swelling classification, Shajapur.

ACKNOWLEDGEMENT

I, **ANURAG DAYAL PANDEY**, would like to express my gratitude to all those who have given unforgettable contributions to the successful contribution of this thesis.

I would like to place on record, my sincere thanks to **Prof. Prateek Sharma**, Vice Chancellor, for providing all resources and assistance needed to carry out this project. I thank **Prof. Raju Sarkar**, Associate Dean for the stimulus provided.

I am enormously grateful to **Dr. Awadesh Kumar**, Professor and Head, Department of Civil Engineering for all this help and encouragement provided during the project work.

I express my sincere thanks to the coordinator **Dr. Ashok Kumar Gupta**, Professor for his insightful recommendations for improvement during project reviews.

I would like to express my sincere gratitude to my project guide **Dr. Anil Kumar Sahu**, Professor who has been a constant source of inspiration throughout this project, for his invaluable advice, support and encouragement.

Additionally, I would like to use this opportunity to express my gratitude to all of the teachers and non-teaching staff in the Department of Civil Engineering for their direct and indirect assistance throughout the project.

I want to express my sincere gratitude to my parents and my entire family for their unwavering understanding and support during this journey. Their encouragement, motivation, and belief in my abilities have been a constant source of strength.

Last, but not least, I would like to express my enormous gratitude towards God. I don't think I would have been what I am today if it was not his/her blessing.

ANURAG DAYAL PANDEY
(24/GTE/01)

TABLE OF CONTENTS

CANDIDATE’S DECLARATION	I
CERTIFICATE BY THE SUPERVISOR.....	II
ABSTRACT.....	III
ACKNOWLEDGEMENT.....	V
TABLE OF CONTENTS	VI
LIST OF TABLES	IX
LIST OF FIGURES	X
LIST OF ABBREVIATIONS	XII
CHAPTER	1
INTRODUCTION.....	1
1.1 GENERAL	1
1.2 BACKGROUND OF EXPENSIVE SOILS.....	4
1.3 SWELLING CHARACTERISTICS AND IDENTIFICATION METHOD	6
1.4 NEED FOR PREDICTIVE MODELLINF OF SWELLING BEHAVIOUR	8
1.5 RESEARCH GAP.....	10
CHAPTER 2	
LITERATURE REVIEW.....	12
CHAPTER 3	
MATERIALS AND METHODS	16
3.1 RESEARCH METHODOLOGY.....	16
3.2 PHASE 1: EXPERIMENTAL PART	17
3.2.1 SOIL COLLECTION	17
3.2.2 MIX PREPATION	19
3.3 PROPERTIES OF SOIL.....	19
3.3.1 GRAIN SIZE ANALYSIS.....	20
3.3.2 ATTERBERG LIMITS.....	22
3.3.3 MOISTURECONTENT TEST	25

3.3.4	SPECIFIC GRAVITY TEST	26
3.3.5	STANDARD PROCTOR TEST	27
3.3.6	FREE SWELL INDEX TEST	29
3.4	PHASE 2: MACHINE LEARNING DATA SET	30
3.4.1	DATA SET COMPILATION	30
3.4.2	FEATURE AND TARGET DESCRIPTION	30
3.4.3	CLASS DISTRIBUTION AND IMBALANCE	31
3.4.4	MACHINE LEARNING MODELS	31
3.4.5	MODEL VALIDATION	32
3.5	RESULT	33
CHAPTER 4		
EMPIRICAL REGRESSION ANALYSIS		
		34
4.1	INTRODUCTION	34
4.2	COMPLETE LABORATORY TEST RESULT	35
4.3	TRENDS IN INDEX PROPERTIES WITH INCREASE IN BENTONITE.....	35
4.3.1	LIQUID LIMIT AND PLASTICITY INDEX.....	35
4.3.2	SHRINKAGE LIMIT	37
4.3.3	SPECIFIC GRAVITY, OMC, MDD.....	38
4.3.4	FREE SWELL INDEX.....	38
4.4	CORRELATION ANALYSIS	39
4.5	REGRESSION MODEL DEVELOPMENT.....	41
4.5.1	EQUATION 1 & 2- SINGLE ATTERBERG LIMIT PREDICTOR.....	41
4.5.2	EQUATION 3- DUAL ATTERBERG LIMIT MODEL	42
4.5.3	EQUATION 4 & 5- OTHER MULTI ATTERBERG COMBINATION ..	42
4.6	LOOCV RESULTS.....	43
4.7	OBSERVED VS PREDICTED FSI.....	44
4.8	COMPARISON WITH PUBLISHED LITERATURE	44
CHAPTER 5		
MACHINE LEARNING CLASSIFICATION		
		45
5.1	INTRODUCTION	45

5.2 DATA OVERVIEW AND CLASS DISTRIBUTION	46
5.3 FEATURE SET AND DESCRIPTIVE STATISTICS	47
5.4 MACHINE LEARNING AND CROSS-VALIDATION PERFORMANCE....	49
5.5 FEATURE IMPORTANCE ANALYSIS	51
5.6 CONFUSION MATRIX ANALYSIS	52
5.7 MODEL COMPARISON AND RECOMMENDATION	54
5.8 COMPARISON OF EMPIRICAL AND ML APPROACHES.....	55
5.9 DISCUSSION.....	57
5.10 LIMITATIONS.....	57
CHAPTER 6	
CONCLUSION AND FUTURE WORKS.....	58
6.1 CONCLUSION.....	58
6.2 FUTURE WORK.....	59
REFERENCES	61
PUBLICATION	64

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
1.1	Engineering problems associated with expansive soil	2
1.2	Summary of Research Gap	11
3.1	Sieve analysis of Soil	21
3.2	Liquid Limit of Soil	22
3.3	Plastic Limit	24
3.4	Shrinkage Limit	25
3.5	Specific Gravity	26
3.6	Standard Proctor Test	28
3.7	Free Swell Index Test	30
3.8	ML Dataset Feature Summary	30
3.9	Class Distribution in ML Dataset	31
3.10	Properties of Soil	33
4.1	Complete Index Properties and FSI	35
4.2	Pearson Correlation Coefficient b/w Index Properties and FSI	40
4.3	Empirical FSI Prediction Equation-Summary	41
4.4	LOOCV Predicted vs Observed FSI- Eq ⁿ 1,2,4	43
4.5	Observed vs Predicted FSI-Summary Table	44
4.6	Comparison of Eq ⁿ 1 slope with Published FSI	44
5.1	Class Distribution in ML Dataset	46
5.2	Feature Descriptive Statistics by Swelling class	47
5.3	Normalization Parameters-Full Dataset	48
5.4	ML Model Performance	49
5.5	RF Feature Importance Scores	51
5.6	Normalized Confusion Matrix-RF	52
5.7	Per-Class Precision, Recall F1-RF	52
5.8	Summary comparison of all three ML classifiers	54
5.9	Comparison of phase-1 and phase-2	55

LIST OF FIGURES		
FIGURE	TITLE	PAGE
NO.		NO.
1.1	Distribution of BCS in India	2
1.2	Typical FSI Test Arrangement	7
1.3	Research Framework	11
3.1	Overall Methodology Flowchart	16
3.2	Study Area of BCS	18
3.3	Soil Sample	18
3.4	Sodium Bentonite used in present study	18
3.5	Sieve analysis arrangement	20
3.6	Particle Size Distribution of Soil	21
3.7	Liquid Limit	22
3.8	Variation of moisture content with no. of blows	23
3.9	Plastic Limit	23
3.10	Plasticity Chart	24
3.11	Specific Gravity using Pycnometer and Density Bottle	26
3.12	Variation of Dry Density with Water Content	28
3.13	Compaction Curve combined for all Mix	29
4.1	Atterberg Limits vs Bentonite content	35
4.2	All 6 Mixes on IS:1498 Plasticity Chart	36
4.3	SL vs Bentonite %	37
4.4	FSI vs Bentonite %	39
4.5	Pearson r b/w index property and FSI	40

5.1	Class Distribution Bar Chart	46
5.2	Scatter Plot PI vs LL	47
5.3	Random Forest Ensemble Architecture	48
5.4	Feature Importance Bar Chart	49
5.5	CV Accuracy Comparison	50
5.6	CV vs Training Accuracy for Each Model	50
5.7	Confusion Matrix Heatmap- RF	53
5.8	Confusion Matrix Heatmap- SVM	53
5.9	Confusion Matrix Heatmap- ANN	54
5.10	Radar Chart- 4 Metrics * 3 Models	56
5.11	Box Plot of CV Accuracy	56

LIST OF ABBREVIATIONS

BCS=Black Cotton Soil

LL=Liquid Limit

PL=Plastic Limit

PI=Plasticity Index

SL=Shrinkage Limit

FSI=Free Swell Index

OMC=Optimum Moisture Content

MDD=Maximum Dry Density

G_s =Specific Gravity

IS=Indian Standard

RF=Random Forest

SVM=Support Vector Machine

ANN=Artificial Neural Network

MLP=Multilayer Perceptron

LOOCV=Leave-One-Out Cross-Validation

R^2 =Coefficient of Determination

RMSE=Root Mean Square Error

FOS=Factor of Safety

CV=Cross-Validation

ML=Machine Learning

CH=Clay of High Compressibility

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The black cotton soil (BCS) is one of the most troublesome geomaterials found in geotechnical practice because of its high swelling and shrinkage characteristics with moisture changes. These soils have significant swelling and shrinkage characteristics which cause repeated ground motion under foundations and pavement systems during wetting and drying. Cracking of low rise buildings, pavement distortion, differential settlement, canal lining distress and deformation of lightly loaded structures are among the cyclic volume changes that cause cracking in many parts of India. The engineering behaviour of expansive soils is thus still a significant issue for infrastructure development, especially in climatically variable regions.

Black cotton soils are mainly found in large areas of the Deccan Trap region of Madhya Pradesh, Maharashtra, Gujarat, Telangana, Andhra Pradesh, Karnataka and other neighbouring states in India. Typically formed from basaltic parent material, these soils have a high proportion of active clay minerals. They are darker than normal fine-grained soils, fine-grained, highly plastic and water-sensitive. The swelling behaviour of black cotton soil is related to the presence of expansive clay minerals like montmorillonite that have a high specific surface area and also have a high affinity for water absorption. When water enters the interlayer spaces of clay particles, significant volume increase of the soil can happen, particularly when the surcharge load is low.

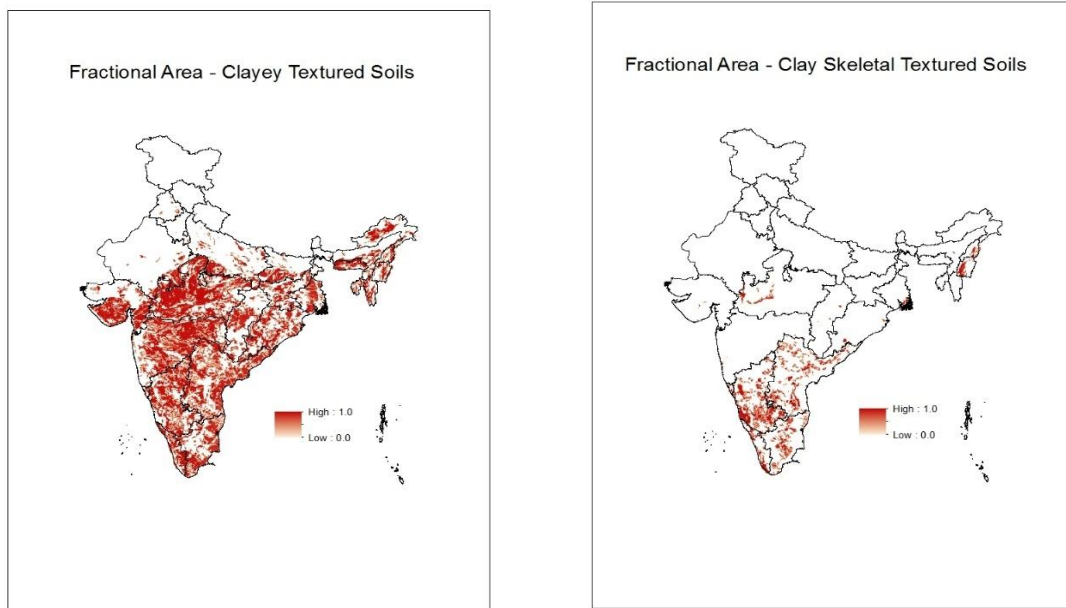


Fig 1.1 Distribution of Black Cotton Soil in India

Problem	Engineering Effect
Swelling	Foundation heave
Shrinkage	Surface cracking
Differential movement	Structural distress
Seasonal moisture variation	Pavement deformation

Table 1.1 - Engineering Problems Associated with Expansive Soils

Various interacting parameters such as mineralogical composition, initial moisture content, dry density, plasticity properties, and environmental wetting–drying cycles, are responsible for the severity of swelling in expansive soils. Of these the index properties like liquid limit, plasticity index, shrinkage limit and free swell index are commonly used in engineering practice for preliminary identification of swelling susceptibility. The Free Swell Index (FSI) calculated in accordance with IS:9451, is one of the widely used parameters in Indian context.

Volumetric changes occur in expansive earth materials with significant changes in moisture content. When water is added, water molecules enter into the inter-laminar space between clay minerals causing a significant increase in soil volume. On the other hand, when the material is dehydrated, it shrinks, and eventually forms deep surface cracks. This is the primary failure mechanism for lightly loaded infrastructure that is located over black cotton soil, and is likely to persist for the foreseeable future.

While the FSI test is relatively simple, a settling time of about 24 hours is required and special laboratory handling is necessary. During preliminary surveys or rapid infrastructure assessments, engineers often make use of index properties like liquid limit and plasticity index, which are more readily available and can be conducted more easily and quickly. Therefore, there has been a growing interest in establishing reliable relationships between routinely measured soil properties and swelling behaviour to be able to assess the expansive soil risk more efficiently.

In the last 20 years there has been extensive research on empirical correlations for predicting swelling properties based on the Atterberg limits and other index properties. A number of existing correlations, however, are restricted to a specific region and some show high variability due to the different mineralogical and physico-chemical properties of expansive soils from different geological formations. In more recent years, machine learning methods have also been applied to geotechnical engineering problems of classification and prediction of complex soil behaviour. These methods have the potential to be used to find non-linear relationships between soil properties and swelling risk, with more extensive data sets.

1.2 BACKGROUND OF EXPANSIVE SOILS

Despite these challenges in predicting volume change under different environmental conditions, identification and assessment of expansive soil behaviour have been important concerns in geotechnical engineering. The presence of expansive soils in many engineering projects, particularly those associated with shallow foundations, embankments, low-rise buildings and highways, often requires early assessment of swelling potential to prevent future structural problems. Black cotton soils are very sensitive to moisture changes, and small changes in the moisture content of the season can result in differential heaving and shrinkage cracking of the overlying infrastructure.

In traditional practice, the swelling properties of expansive soils are evaluated through laboratory parameters like free swell index, swell pressure and swell potential. Of these, the Free Swell Index (FSI) test is popular in Indian geotechnical practice due to its simplicity and being included in the guidelines of IS:9451. The test is a practical test for the relative expansiveness of soil and is frequently used in preliminary site investigations. The FSI test is, however, less complex than the oedometer test, but does still need controlled laboratory conditions, careful handling, and a settling time of almost 24 hours before final readings can be obtained.

In normal engineering work, this type of testing can be cumbersome when conducting a quick assessment in the field or when conducting large-scale soil testing where several samples must be classified within a short time. However, index properties like liquid limit, plasticity index, shrinkage limit, optimum moisture content and maximum dry density are usually measured in nearly all geotechnical investigations due to their importance in soil classification and compaction studies. This has led several researchers to try to develop empirical correlations between these properties that are commonly measured and swelling behaviour, in order to be able to identify expansive soils faster than by relying on specialized swelling tests.

Machine learning (ML) methods have also been recently introduced in geotechnical engineering applications because they can capture complex and non-linear interactions between multiple soil parameters. Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forest (RF) have been used to address issues of soil classification, prediction of bearing capacity, settlement estimation, and expansive soil characterization. Machine learning models can be used to find patterns in data without assuming that the variables are linearly related, unlike traditional regression methods.

In spite of these advancements, the combination of controlled laboratory experiments, empirical regression analysis and application of machine learning based swelling classification is still relatively limited in the Indian context of black cotton soils. The majority of existing studies is either empirical prediction based on small laboratory datasets or on standalone machine learning models trained from literature compiled datasets. It is thus necessary to develop an integrated framework that integrates the swelling behaviour generated from experiments with data-driven classification techniques to enhance the reliability and practical use of the swelling assessment methods.

The present study has been undertaken with this objective. The black cotton soil of Shajapur was modified in a controlled manner using sodium bentonite, in order to systematically change the swelling behaviour of the soil in the laboratory. The experimental data obtained were then subjected to regression modelling to look for relationships between index properties and free swell index. Furthermore, machine learning classification approaches were investigated with an extended multi-regional database of Indian black cotton soils to assess the potential of rapid prediction of the swelling potential of soils based on routinely measured soil properties. The swelling characteristics and identification methods of the soil are described. Swelling characteristics and identification methods of the soil are discussed.

1.3 SWELLING CHARACTERISTICS AND IDENTIFICATION METHODS

Expansive soils have very large volume change characteristics in response to changes in moisture content. When the soil is wet, water molecules get trapped between the clay particles, increasing the volume of the soil; when it is dry, the soil shrinks, leading to cracking in the soil. One of the major causes for distress in lightly loaded civil engineering structures built on black cotton soils is this repeated swell–shrink phenomenon. Clay mineral composition, initial dry density, moisture variation, stress condition and physio-chemical properties of soil are some of the factors that determine the magnitude of the expansion.

The swelling of black cotton soil is usually linked with the presence of active clay minerals with high water absorption capacity. The adsorbed water film around the clay particles becomes thicker when moisture is added to the soil which causes the particles to separate and the soil structure to expand. This can be worse in areas with alternating wet and dry climates as the soil is constantly subjected to expansion and contraction. Consequently, the following problems are frequently encountered in expansive soil areas: pavement heaving, cracking of the foundation, differential settlement, and surface deformations.

The identification of expansive soils in geotechnical engineering practice is usually done by direct and indirect methods. Direct methods are the methods of measuring actual swelling characteristics in the laboratory by means of experiments like free swell index, swell pressure, swell potential, and consolidation based expansion test. These tests can be used to assess the swelling behaviour reliably, but may be conducted in controlled laboratory conditions, take longer to test and require careful sample preparation. Therefore, during preliminary investigations, indirect identification through the use of index properties that are routinely measured is often used.

Atterberg limits are among the numerous identification parameters that are used to assess the plasticity and expansiveness of fine-grained soils. The liquid limit is the water content of the soil at which it becomes liquid and the plastic limit is the water content at which it becomes plastic. The difference between these two

parameters is represented as the plasticity index which is an indication of the moisture content range in which the soil is plastic. In general, soils with high liquid limit and high plasticity index are more likely to have swelling properties due to their high water absorption and clay activity.

Another useful parameter in the characterization of expansive soil is the shrinkage limit. It is the water content at which further drying will not cause the soil to shrink any further. Soils with high shrink–swell potential are those with low shrinkage limits. Likewise, the compaction properties of maximum dry density (MDD) and optimum moisture content (OMC) also affect the engineering response of expansive soils, especially in the pavement. Subgrades and embankment applications.

The Free Swell Index (FSI) test is one of the most widely used direct identification tests used in geotechnical investigations in India. The test is carried out as per IS:9451 by comparing the sediment volume of oven dried soil passing 425micron sieve in distilled water and kerosene. The volume of sediment increases in the water as a result of kerosene being a non-polar liquid with no swelling potential, so the volume of sediment in water is a good indicator of the swelling potential of the soil. Expansive soils can be classified into various swelling severity classes, from low to very high, based on FSI values.

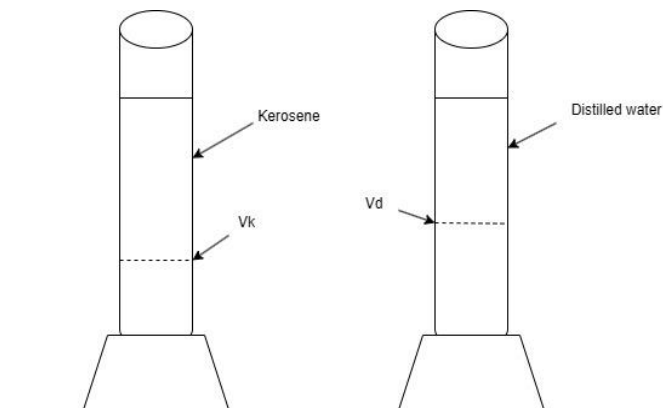


Fig 1.2 Typical Free Swell Index Test Arrangement

It has been found that several researchers have reported that meaningful correlation exists between the swelling behaviour and index properties such as liquid limit, plasticity index, shrinkage limit and compaction parameters. The reliability of such correlations, however, will depend on the characteristics of the soils, the mineralogical composition and on the quality of the experimental data employed in the model development.

Over the past few years, data-driven methods have also been investigated in the identification and classification of expansive soils. Machine learning methods can be used to capture the inter-correlations between several soil parameters and can help to predict the swelling categories rapidly using the commonly measured laboratory data. These methods can be especially helpful in preliminary site investigations where the quick assessment of expansive soil behaviour is needed prior to detailed laboratory testing.

1.4 NEED FOR PREDICTIVE MODELLING OF SWELLING BEHAVIOUR:

An accurate estimation of swelling is a key consideration in the design and construction of structures built on expansive soils. The potential for swelling of the subsoil is a major factor in many geotechnical projects in determining the type of foundation to be used, the thickness of the pavement, drainage decisions and soil improvement. The direct determination of the swelling parameters by laboratory tests, however, can be time consuming and sometimes is not possible during an early phase of site investigation where many soil samples are involved.

The Free Swell Index (FSI) test is widely used in India for identification of expansive soils as it is simple as compared to consolidation based swelling tests. However, the process must be performed under controlled laboratory conditions and a settling time must be allowed before measurements can be taken. Rapid assessment techniques are frequently used in practical field investigations, particularly for highway alignment studies, rural infrastructure development and preliminary geotechnical surveys to identify potentially problematic zones of soil at an early stage.

In response to this, the use of predictive modelling has become more and more important for research on expansive soils, and relies on soil properties that are routinely measured. Liquid limit, plasticity index, shrinkage limit, optimum moisture content and maximum dry density are among the parameters that are usually measured during routine laboratory testing and are reported to have significant effect on the swelling characteristics. It is therefore hoped that these relationships can be used as a guide to give an initial estimate of soil expansiveness, without relying exclusively on specific swelling tests, to assist engineers in their design.

Many researchers have adopted conventional empirical correlations for prediction of swelling characteristics from index properties. The regression-based models are relatively easy to develop and give engineering relationships between the dependent and independent variables. These models work well when there are consistent behavioural trends in the soil dataset for controlled laboratory conditions. Empirical equations developed for one soil type or geographical region, however, do not necessarily give satisfactory prediction accuracy for soils from other geological environments. The restriction is due to the interaction of several factors which affect expansive soil behaviour, such as mineral composition, clay activity, density condition and environmental moisture variation.

In the last few years, computational and data-driven techniques have proved to be useful for tackling complex geotechnical prediction problems. Machine learning methods can detect non-linear relationships between several variables in large data sets without making any mathematical assumptions. Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) have been found to have potential applications for numerous geotechnical problems such as soil classification, prediction of bearing capacity, estimation of settlements, and assessment of slope stability.

1.5 RESEARCH GAP

The swelling properties of expansive soils have been reported in various studies in the literature, both through the use of conventional laboratory techniques and predictive modelling methods. Empirical relationships have been developed between swelling parameters and index properties (liquid limit, plasticity index, shrinkage limit and clay fraction) to estimate expansive behaviour without carrying out complex swelling tests. Likewise, the recent advances of geotechnical data analysis have paved the way for the use of machine learning prediction and classification techniques to laboratory data.

While a significant amount of work has been done in this area, some studies are still limited. Numerous empirical correlations have been published for expansive soils but are largely limited to a specific geographic area and may have been developed from natural deposits of soils with wide variations in mineralogical composition and environmental conditions. Therefore, prediction equations for one geographical area might not always yield consistent results when applied to soils in other areas. Furthermore, there are several published correlations that are based on small data sets and only consider individual swelling parameters without incorporating other trends in the behaviour.

In recent years, studies based on machine learning have also grown in number, but most existing studies are based on literature data bases without relating the models to controlled experimental variation in the swelling behaviour of expansive soils. The correlation of the changes in index properties and the swelling response is not studied systematically in the laboratory in many cases. In addition, there is a limited number of reports on the integration of regression based empirical modelling with machine learning based swelling classification of Indian black cotton soils. The present research work aims to fill these gaps by conducting a controlled laboratory investigation on black cotton soil of Shajapur district of Madhya Pradesh.

Various amounts of sodium bentonite were added to the control soil to artificially create different levels of expansion. Later, predictive equations were developed and tested for the relationship between the standard index properties and the Free Swell Index and found to be in good agreement with the laboratory results. Furthermore, the study used a large database of regional expansive soils to train and test machine learning classifiers to evaluate their suitability.

Table 1.2 — Summary of Research Gaps Identified from Previous Studies

Previous Research Focus	Identified Limitation
Conventional swelling tests	Time-consuming
Empirical correlations	Region-specific applicability
ML-based prediction	Limited linkage with controlled experiments
Standalone approaches	Lack of integrated framework

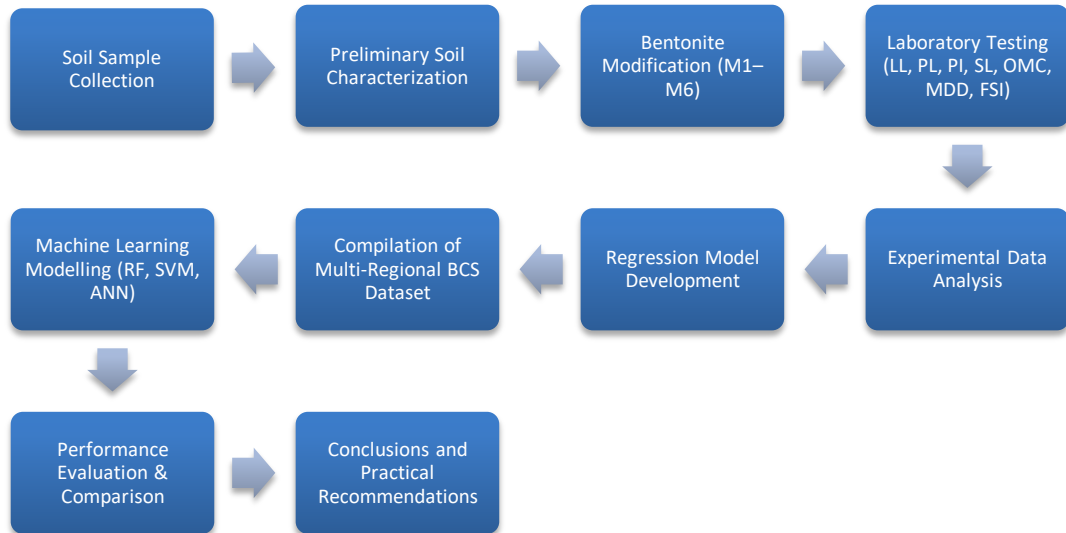


Fig 1.3 Research Framework

CHAPTER 2

LITERATURE REVIEW

Sridharan, A. and Prakash, K. (2000) carried out to study the correlation between Atterberg limits and swelling properties of expansive soils of India. From their observations, they concluded that the plasticity index is a major factor in the swelling potential of fine-grained soils, and there was a reasonably consistent correlation between Free Swell Index (FSI) and PI. The work emphasized the value of index properties in making a preliminary evaluation of expansive soil behavior without depending on comprehensive swelling tests.

Holtz, R. D. and Kovacs, W. D. (1981), in their geotechnical engineering text, explained the fundamental mechanisms governing swelling behaviour in expansive soils. The authors discussed the influence of clay mineralogy, moisture variation, and density condition on soil expansion characteristics. Their work established that soils containing montmorillonite clay minerals generally exhibit higher swelling potential because of their greater affinity for water absorption and large specific surface area.

Sharma, H.D. and Pandey, S.P. (1999) conducted geotechnical characterization of expansive soils of various regions of Madhya Pradesh focusing on infrastructure related issues in black cotton soils. They found that soils of Malwa Plateau region are highly plastic and show significant swelling properties, often causing pavement distress and cracking of lightly loaded structures during moisture variations due to seasons.

Pandian, N. S., Nagaraj, T. S., and Raju, M. (2000) investigated the behaviour of bentonite modified soil systems and concluded that the liquid limit and plasticity index of soil increases progressively with the addition of sodium bentonite, whereas maximum dry density decreases. Their observations showed that the physio-mechanical behaviour of expansive soils is significantly influenced by the bentonite content and also gave valuable insight on the controlled swelling modification of clayey soils.

Thyagaraj, T. and Rao, S. M. (2010) investigated the swelling and shrinkage properties of compacted expansive clays and found that the shrinkage limit is an important parameter that affects swell–shrink properties. Soils with lower shrinkage limit tend to have higher swelling potential. They also indicated that multiple property correlations might provide better prediction of expansive soil behaviour than single property correlations.

Phanikumar, B. R. (2009) The behaviour of expansive clay soils was studied by the laboratory and comparative analysis method. The study covered variation in swelling-related correlations in the various deposits of expansive soil in India and emphasized the importance of the geological conditions of the region on the correlation between the plasticity characteristics and the swelling response of expansive soil.

Jaleh Forouzan, A. (2016) The study showed that the addition of bentonite can create systematic variation in swelling, which can be used to develop regression-based prediction models. This work is of particular interest to the present investigation because it employed controlled modification of clay to investigate expansive behaviour in the laboratory.

Ahirwar, R. K. and Jain, R. K. (2019) have reported geotechnical characteristics of expansive soils from various sites in Malwa Plateau region. They concluded that the general properties of black cotton soils from this area are high liquid limit and high plasticity index which are typical of expansive clay soils. The authors also pointed out the need for region-specific characterization for accurate prediction of swelling.

Erzin, Y. and Gunes, N. (2011) used ANN methods for prediction of swelling properties of compacted clays based on the most common soil parameters measured. Their research indicated that in some instances, machine learning methods can better capture the non-linear relationships between index properties and swelling response than can traditional empirical equations. The work also illustrated the increasing usefulness of data-driven approaches for the characterization of expansive soils for large-scale problem.

Shahin, M. A., Jaksa, M. B., and Maier, H. R. (2009) discussed the use of artificial neural systems (ANS) in various fields of geotechnical engineering such as soil classification, settlement prediction and estimation of bearing capacity. They also noted, however, that the reliability of the models is highly dependent on the quality of the data used and proper validation procedures, especially in the case of limited geotechnical data.

Alavi, A. H.; Gandomi, A. H. (2011) investigated data-driven modelling techniques for geotechnical engineering applications and showed that predictive equations developed using computational methods can be useful alternatives to regression models. They focused their study on the creation of interpretable prediction relationships which can be easily implemented in engineering practice without relying on highly specialized computational tools.

Pal, M. (2006) showed that support vector machine methodologies can be used successfully in classification problems with complex and imbalanced data sets. The results of the study showed that SVM models can be used to efficiently deal with non-linear classification boundaries, and that they can be usefully extended to geotechnical classification problems with multiple soil behaviour categories

Breiman, L. (2001) The ensemble learning method, called Random Forest, was introduced and shown to be able to deal with complex feature interactions in datasets. The study demonstrated that the Random Forest models tend to have consistent prediction power and to avoid over-fitting that is often seen with single decision-tree models. For these reasons, the Random Forest methods have been increasingly used in geotechnical prediction and soil classification research.

A. Zhang and S. K. Vanapalli. (2025) examined the use of machine learning models to predict the swelling properties of expansive soils from the soil index properties. They showed that machine learning models can achieve acceptable prediction accuracy for swelling parameters and can be used to model the non-linear response of soil. The study also emphasized the need for correct validation procedures and the need for proper coverage of the datasets to enhance the generalization capability of the model.

Nagaraj, T. S. and Rao, S. M. (1993) studied the prediction of swelling and shrinkage behaviour of expansive soils based on the combinations of index properties. They concluded that the use of combinations of soil parameters could enhance the reliability of the prediction when compared with the use of empirical correlations based on single variables. The study also highlighted the significance of shrinkage characteristics on the assessment of expansive soil behaviour.

Author(s)	Research Focus	Major Findings	Limitation / Relevance to Present Study
Sridharan and Prakash (2000)	Relationship between index properties and swelling behaviour	Observed strong relationship between plasticity characteristics and FSI	Correlations developed for limited Indian expansive soils
Holtz and Kovacs (1981)	Fundamentals of expansive soil behaviour	Explained influence of clay mineralogy and moisture variation on swelling	Primarily theoretical and conceptual in nature
Sharma and Pandey (1999)	Characterization of Madhya Pradesh expansive soils	Reported high plasticity and swelling tendency of Malwa Plateau soils	Limited predictive modelling approach
Pandian et al. (2000)	Bentonite-modified soil behaviour	Bentonite addition increased LL and PI while reducing MDD	Focused mainly on laboratory behaviour
Thyagaraj and Rao (2010)	Swell–shrink behaviour of expansive clays	Identified influence of shrinkage limit on swelling response	Limited regional applicability
Phanikumar (2009)	Behaviour of Indian expansive soils	Highlighted regional variability in swelling correlations	Broad comparative study without ML application
Forouzan (2016)	Bentonite-modified expansive soil mixtures	Demonstrated usefulness of controlled swelling variation for prediction studies	Conducted on non-Indian clay systems
Ahirwar and Jain (2019)	Geotechnical properties of Malwa Plateau BCS	Reported properties comparable to central Indian expansive soils	Region-specific characterization only
Shahin et al. (2009)	ANN applications in geotechnical engineering	Discussed applicability of ML techniques for prediction problems	Emphasized risk of overfitting in small datasets
Zhang and Vanapalli (2025)	ML-based swelling prediction	Showed effectiveness of ML for non-linear swelling prediction	Limited dataset coverage and validation approach

CHAPTER 3

MATERIALS AND METHODS

3.1 RESEARCH METHODOLOGY

The present investigation methodology is divided in two consecutive phases of experimental characterization and predictive modelling of expansive soil behavior, integrated to each other. The first phase involves laboratory testing of black cotton soil with different percentages of sodium bentonite to create a controlled variation in the swelling characteristics of the soil. Based on the experimental observations, empirical relationships between index properties and Free Swell Index (FSI) were developed using regression analysis.

In the second phase, machine learning based classification of expansive soils is done using a database of Indian black cotton soil deposits compiled from published geotechnical studies. To classify the severity of swelling, machine learning models such as Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN) were applied using commonly measured soil parameters.

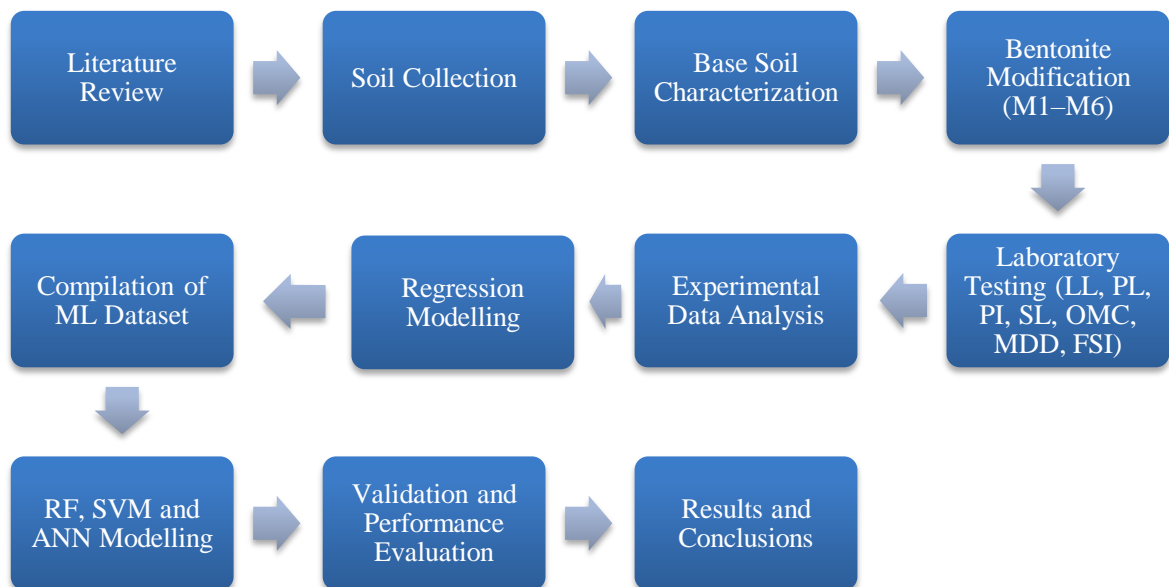


Figure 3.1 Overall Methodology Flowchart

3.2 PHASE-1: EXPERIMENTAL PART

3.2.1 Soil collection

The black cotton soil was sampled from an agricultural field at Shajapur district, Madhya Pradesh (23.43°N, 76.85°E) from 0.3 to 1.0 m below the natural ground surface. Sampling has been done according to the IS:2720 part 1. The black cotton soil which was used in the present investigation was obtained from an agricultural field of Shajapur district Madhya Pradesh, India, at a depth of 0.3m to 1.0m from the natural ground surface. The region selected is in the Malwa Plateau zone where large expanses of clay are prone to large swell–shrink properties in response to moisture changes in the season.

The soil samples were taken and brought to the laboratory in sealed containers to reduce moisture loss during handling. The soil was then air-dried in the laboratory, broken up with a wooden mallet and sieved through a 425 µm IS sieve to ensure uniformity for experimental testing. Coarse particles and organic impurities were filtered out before sample preparation.

The initial visual inspection revealed that the soil had dark colour, high cohesiveness and that it had noticeable plasticity properties typical of expansive black cotton soils. In the present study, the commercially available bentonite (sodium bentonite) as per Indian standard specifications was used as a modifying agent. The bentonite used was sodium bentonite due to its high swelling capacity and water absorption properties in comparison to calcium bentonite. The material was purchased in powder form and packed in sealed containers before being used to avoid the absorption of moisture from the atmosphere.

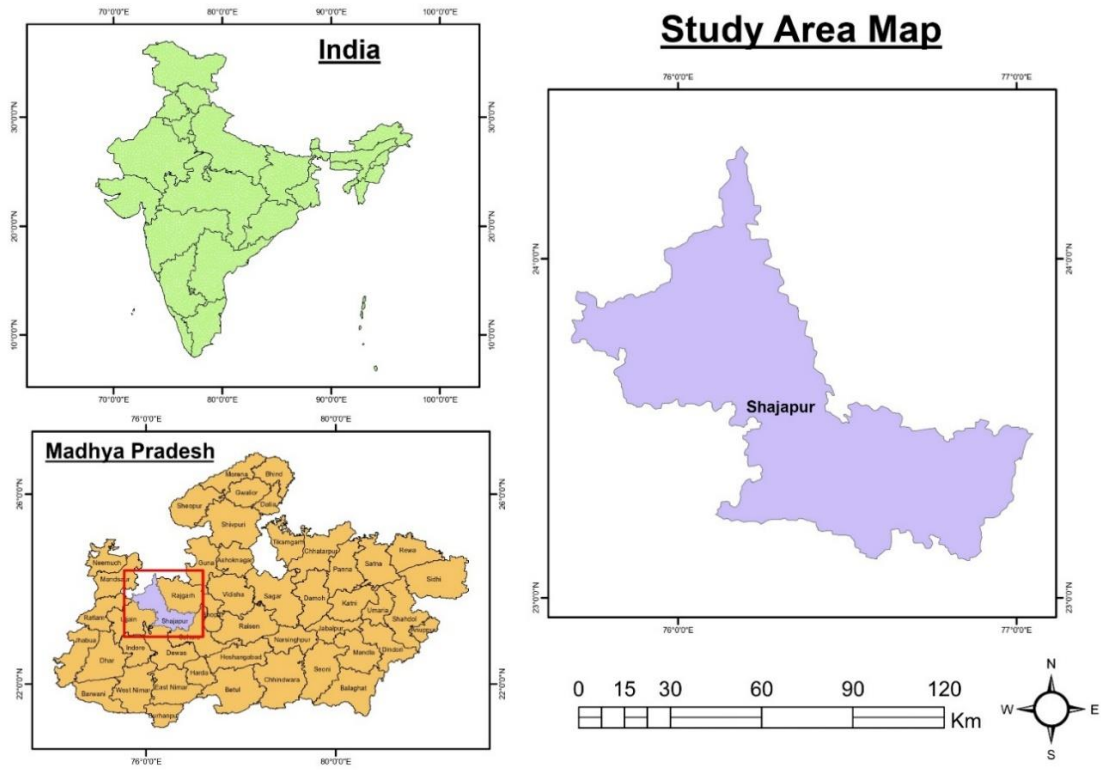


Fig 3.2 Study area of Black Cotton Soil



Fig 3.3 Soil sample



Fig 3.4 Sodium Bentonite used in present study

3.2.2 Mix Preparation

The six soil mixtures were made by mixing sodium bentonite at dry weight percentages of 0%, 2%, 4%, 6%, 8% and 10% with the base BCS and are referred to as M1, M2, M4, M6, M8 and M10, respectively. The mixing protocol consisted of a minimum of 15 minutes of tumbling in a mechanical mixer and a double pass through a 425 μm sieve to achieve particle level homogeneity.

3.3 PROPERTIES OF SOIL

The collected black cotton soil was processed in accordance with IS:2720 procedures for laboratory. Air-dried soil was disaggregated carefully without crushing individual particles and passed through a 425 μm sieve.

The base black cotton soil was mixed with sodium bentonite at various percentages by dry soil and six different soil mixtures were prepared. The contents of Bentonite were chosen to be 0%, 2%, 4%, 6%, 8% and 10%, and were designated as M1, M2, M4, M6, M8 and M10 respectively. The mixing was done by hand and then mechanically blended to ensure the uniformity of bentonite distribution in the soil mass.

The modified bentonite used in the present study was modified under controlled conditions with the objective of creating a gradient of swelling under laboratory conditions for the purpose of studying the effect of progressive variation in plasticity on the expansive behaviour of the bentonite.

Some of the tests conducted are as follows:

- Classification Tests
 - Grain Size Analysis
 - Atterberg Limits: Liquid limit, plastic limit, shrinkage limit
- Physical properties
 - Moisture Content Test
 - Specific Gravity Test
- Compaction Test
 - Standard Proctor Test

3.3.1 Grain Size Analysis:

Particle size distribution testing is used to measure the precise sizes of the grains of an aggregate, sediment or soil matrix. This assessment is essential for the determination of the textural classification of the material, which directly affects the structural parameters such as load-bearing capacity, shear strength and hydraulic conductivity. The main purpose of Grain size analysis is to determine whether soil is sand, silt, clay or gravel. It is useful in the interpretation of depositional environments, in the evaluation of permeability and drainage characteristics and in the evaluation of suitability for foundations and other structures.

Sieve Analysis is done for the distribution of particles. In Sieve Analysis, soil is passed through a series of sieves of decreasing size and the weight of soil that is retained on each sieve is measured to calculate the percent of each size fraction.



Figure 3.5: Sieve analysis arrangement

Parameters calculated are as:

- **D₁₀**: (Effective size): Particle size at 10% finer, indicates drainage capacity.
- **D₃₀**: Particle size at 30% finer.
- **D₆₀**: Particle size at 60% finer, used to calculate uniformity coefficient (C_u) and gradation coefficient (C_c).

Table 3.1: Sieve analysis of soil

Seive size(mm)	Weight Retained (g)	Percentage % Weight Retained	Cumulative % Weight Retained	Passing percentage %
4.75	10.00	2.00	2.00	98.00
2.36	8.00	1.60	3.60	96.40
1.18	12.00	2.40	6.00	94.00
0.6	15.00	3.00	9.00	91.00
0.3	20.00	4.00	13.00	87.00
0.15	25.00	5.00	18.00	82.00
0.075	35.00	7.00	25.00	75.00
PAN	375.00	75.00	100.00	-

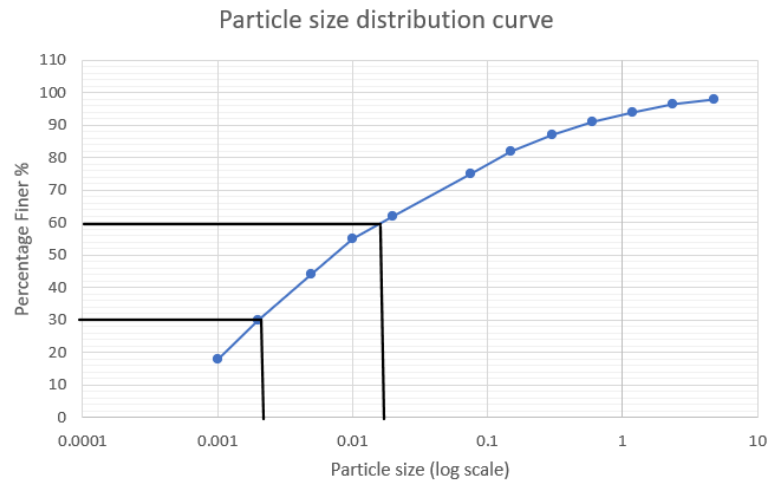


Figure 3.6: Particle Size Distribution of soil

From graph, we found,

$$D_{60} = 0.038 \text{ mm}, D_{30} = 0.006 \text{ mm}, D_{10} = 0.0012 \text{ mm}$$

- Uniformity coefficient $C_u = 31.67$
- Coefficient of curvature $C_c = 0.79$

$C_u > 6$ and C_c does not lie between 1 -3.

3.3.2 Atterberg Limits

Liquid Limit (W_L)

Liquid limit is an important property of fine-grained soils, and is defined as the moisture content at which a soil becomes a liquid. It is the point at which soil will start to flow under its own weight. Main purpose is to assess the consistency and behavior of clayey soils and to predict soil strength, compressibility and settlement.

Casagrande apparatus is used for finding out the liquid limit.



Figure 3.7: Liquid Limit

Table 3.2: Liquid limit of soil

Observation and Calculation				
Container Number	1	2	3	4
Weight of Container W_1 , in g	18.45	19.20	17.80	18.90
Weight of Container and wet soil W_2 , in g	49.95	54.40	52.03	52.20
Weight of Container and dry soil W_3 , in g	38.45	41.20	38.80	38.90
Weight of water ($W_2 - W_3$), in g	11.50	13.20	13.23	13.30
Weight of dry soil ($W_3 - W_1$) in g	20.00	22.00	21.00	20.00
Moisture content (%) $= \frac{(W_2 - W_3)}{(W_3 - W_1)} \times 100\%$	57.50	60.00	63.00	66.50
No. of blows	35	28	22	17

By Interpolation,

For 25 number of blows, Water content = 62%

Liquid Limit = 62%

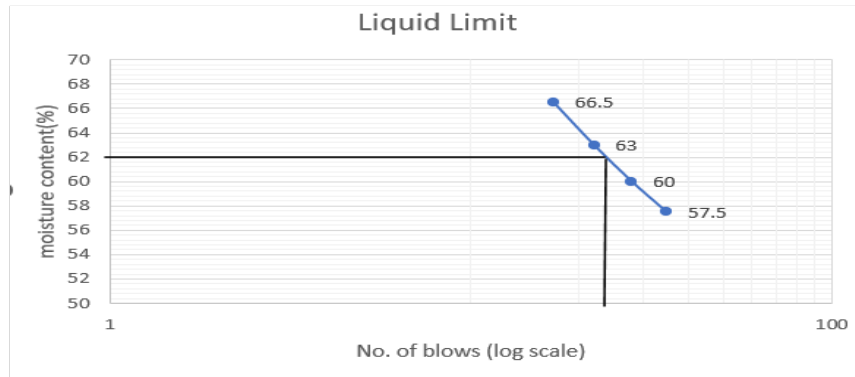


Figure 3.8: Variation of moisture content with no. of blows

With the help of graph, the Liquid Limit value is **62.0%**

Plastic limit (W_p)

The plastic limit is the water content at which a soil becomes plastic and semisolid. The property of soil that is the smallest water content at which a soil can just begin to crumble when rolled into a thread about 3mm in diameter.



Figure 3.9: Plastic Limit

Observation and Calculation

Table 3.3: Plastic Limit

Sample No.	1	2
Weight of Container W_1 , in g	18.50	17.80
Weight of Container and wet soil W_2 , in g	28.44	27.09
Weight of Container and dry soil W_3 , in g	26.50	25.30
Weight of water ($W_2 - W_3$), in g	1.94	1.79
Weight of dry soil ($W_3 - W_1$) in g	8.00	7.50
Moisture content (%) = $\frac{(W_2 - W_3)}{(W_3 - W_1)} \times 100\%$	24.25	23.87

After calculating it the avg. plastic limit will be = $\frac{24.25 + 23.87}{2} = 24.06\% \approx 24\%$

Plasticity Index (I_p):

The plasticity index (PI) is the range of water content over which a soil exhibits plastically. It is calculated as:

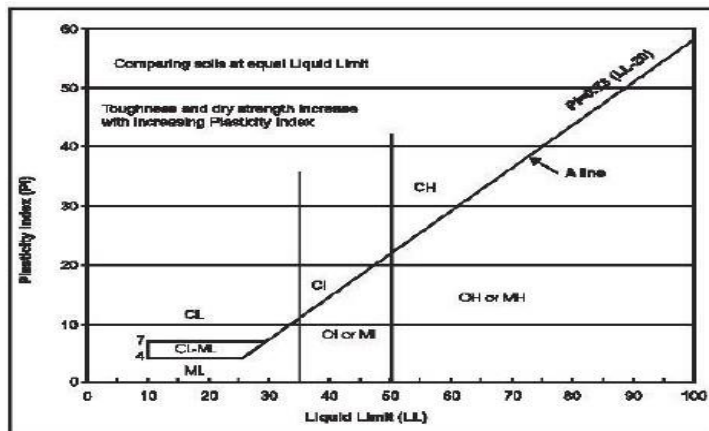
$$PI = LL - PL$$

$$I_p = W_l - W_p = (62 - 24) \%$$

$$\therefore I_p = 38\%$$

Through the A line and Plasticity index we find that the soil has high plasticity (CH).

Figure 3.10: Plasticity Chart



SHRINKAGE LIMIT (IS:2720 Part 6)

Base BCS (M1) — SL = 12%

Method: Mercury displacement

Formula: $SL (\%) = w_i - [(V - V_d) / M_d] \times 100$

Constants:

Volume of shrinkage dish, $V = 24.0 \text{ cm}^3$

Density of mercury, $\rho_{\text{Hg}} = 13.6 \text{ g/cm}^3$

Mass of mercury to fill dish = $24.0 \times 13.6 = 326.4 \text{ g}$

Table 3.4: Shrinkage Limit

Observation	Trial 1	Trial 2
Mass of empty shrinkage dish W_1 (g)	45.30	44.70
Mass of dish + wet soil W_2 (g)	97.14	95.94
Mass of dish + dry soil W_3 (g)	81.30	80.70
Mass of dry soil $M_d = W_3 - W_1$ (g)	36.00	36.00
Mass of water $M_w = W_2 - W_3$ (g)	15.84	15.24
Initial water content $w_i = M_w/M_d \times 100$ (%)	44.00	42.33
Mass of Hg to fill dish (g)	326.40	326.40
Mass of Hg displaced by dry pat (g)	169.73	163.59
Volume of dry pat $V_d = M_{\text{Hg}} / 13.6$ (cm^3)	12.48	12.03
$V - V_d$ (cm^3)	11.52	11.97
$SL = w_i - (V - V_d)/M_d \times 100$ (%)	12.00	11.92
Adopted SL (%)	12.0	

Average SL = $(12.00 + 11.92) / 2 = 11.96 \approx 12.0\%$

3.3.3 Moisture Content Test

Moisture content refers to the percentage of water in the soil or material, based on the dry weight of the material. It is an important property in soil mechanics, agriculture and material science because it determines the strength, compaction and behavior of soil.

Moisture Content of this soil from moisture meter is found to be **10.8 %**.

3.3.4 Specific Gravity Test

Specific gravity is the ratio of weight of soil solids to the weight of an equal volume of water. It is beneficial in computing the unit weight of the soil under different conditions and also in the determination of particle size by wet analysis.

Density bottle method is used to determine specific gravity of soil sample.

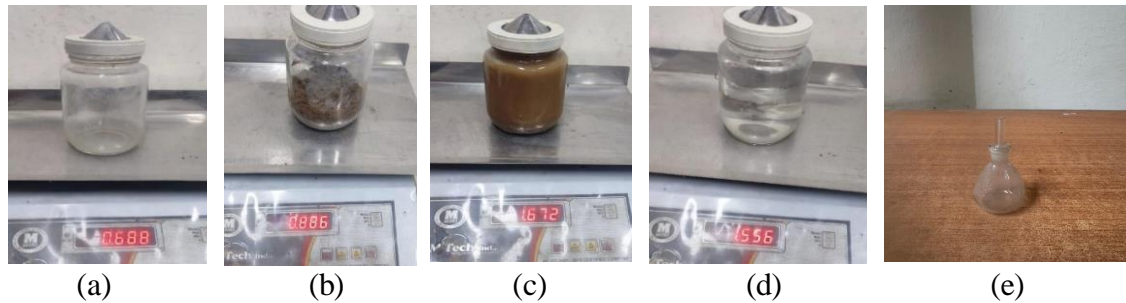


Figure 3.11: Specific Gravity using pycnometer, density bottle

Calculations

Specific Gravity (G) is calculated as:
$$(G) = \frac{(W_2 - W_1)}{(W_4 - W_1) - (W_3 - W_2)}$$

Where:

W_1 = Weight of empty pycnometer. W_2 = Weight of pycnometer with soil

W_3 = Weight of pycnometer with soil and water W_4 = Weight of pycnometer with water only

Table 3.5: Specific Gravity

Weight (g)	Sample 1	Sample 2
W_1	32.56	31.92
W_2	42.56	41.92
W_3	88.70	88.06
W_4	82.39	81.75
G	2.713	2.707

Sample calculation

For sample 1:
$$G = \frac{(42.56 - 32.56)}{(82.39 - 32.56) - (88.70 - 42)}$$

Average value of specific gravity is **2.71** at room temperature.

3.3.5 Standard Proctor Test

Compaction is the process of mechanically rearranging and packing soil particles into a closer state of contact so as to reduce the porosity (or voids ratio) of the soil and, as a result, increase the dry density of the soil mass by rapidly and dynamically expelling the air that is present in the voids. The density of the compacted soil mainly depends upon its water content, compactive effort, type of soil and admixtures.

In a laboratory compaction test, the main purpose is to determine the optimum water content (also called the maximum dry density (MDD)) for which the unit volume of the soil grains has the greatest weight.

Observations

- Diameter of mould = 10 cm
- Height of mould = 12.73 cm
- Volume of mould = 1000 cm³
- Empty weight of mould (W_1) = 4250 g
- Weight of rammer = 2.5 kg

Calculations

- Bulk Density, $\gamma_b = \frac{(W_2 - W_1)}{\text{Volume}}$ g/cc
- Dry Density, $\gamma_d = \frac{\gamma_b}{(1+W)}$ g/cc

Where,

W_1 = Weight of mould

W_2 = Weight of compacted soil + mould

W = Moisture content

Table 3.6: Standard Proctor Test

S.No.	Weight of mould (Kg), W_1	Weight of compacted soil + mould (Kg), W_2	Volume of mould (cm^3)	Water added (%), W	Bulk density (g/cc), γ_b	Moisture content (%)	Dry density (g/cc), γ_d
1	4.250	5.982	1000	0	1.732	14.0	1.519
2	4.250	6.126	1000	4	1.876	16.5	1.610
3	4.250	6.246	1000	8	1.996	19.5	1.671
4	4.250	6.239	1000	12	1.989	22.0	1.630
5	4.250	6.188	1000	16	1.938	25.0	1.550

So, from the above table we obtain,

3

Max. dry density = 16.38 KN/m

Optimum moisture content = 19.5 %

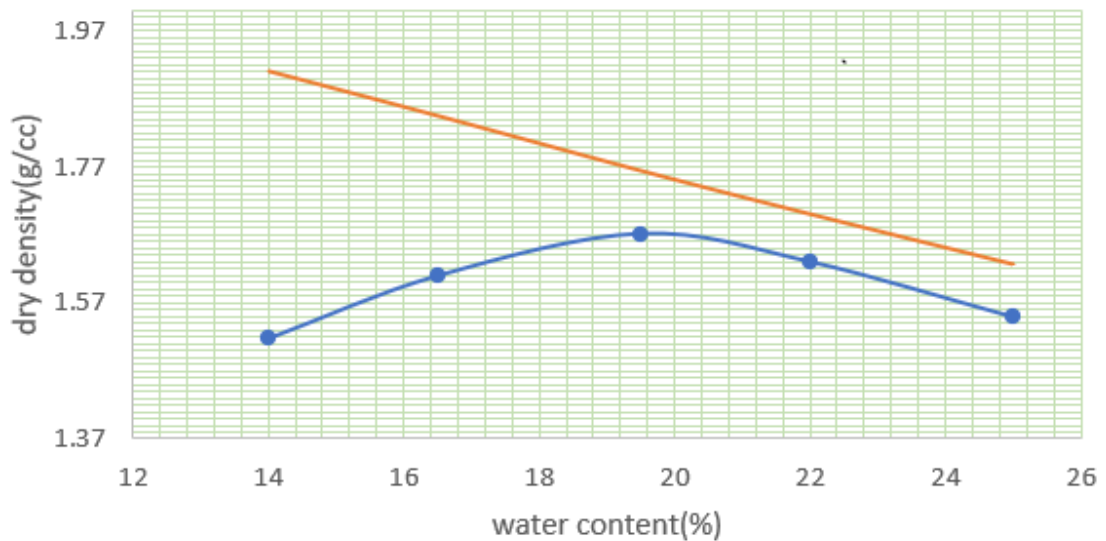


Figure 3.12: Variation of Dry Density with Water Content

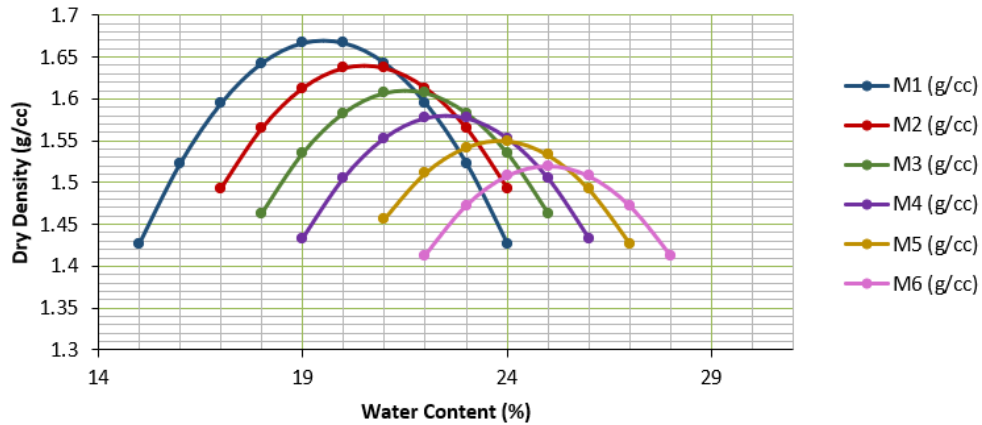


Fig 3.13: Compaction Curve Combined for all Mix (M1-M6)

3.3.6 Free swell index test

The Free-Swell Index Test is a simple test that is used to indicate whether a soil has high shrink–swell tendency. The test was first developed in the 1950s by Holtz and Gibbs (1956), and consists of placing a small sample of soil in distilled water and observing the volume of the soil sample. The Free Swell Index Test is a direct swell test that quantifies the potential of the swells in terms of volume (as compared to non-swelled volume), and provides a qualitative swell index. The results can be used to compare soils and should be used as an indicator test showing whether more sophisticated direct tests should be performed on a soil to quantify the swelling behavior under certain conditions of confinement and water content change.

Procedure:

- 10 g of oven-dry soil passing 425 μm IS sieve placed in each of two 100 mL graduated cylinders
- One cylinder filled with distilled water, other with kerosene
- Settled volume read after 24 hours

$$\text{FSI (\%)} = \left[\frac{(V_d - V_k)}{V_k} \right] \times 100$$

Table 3.7: Free swell index Test

Mix	Bent. %	Soil (g)	V _w T1 (mL)	V _w T2 (mL)	V _w T3 (mL)	Mean V _d (mL)	V _k (mL)	FSI (%)	IS Class
M1	0	10	16.3	16.5	16.7	16.50	10.0	65	High
M2	2	10	17.3	17.5	17.7	17.50	10.0	75	High
M3	4	10	18.4	18.6	18.8	18.60	10.0	86	High
M4	6	10	19.5	19.7	19.9	19.70	10.0	97	High
M5	8	10	20.6	20.8	21.0	20.80	10.0	108	Very High
M6	10	10	21.6	21.8	22.0	21.80	10.0	118	Very High

3.4 Phase 2: Machine Learning Dataset

3.4.1 Dataset Compilation

A database of 186 Indian BCS samples was developed from 40 published geotechnical studies and conference proceedings. The data covers 10 states in India: Madhya Pradesh, Maharashtra, Gujarat, Andhra Pradesh, Telangana, Tamil Nadu, Karnataka, Rajasthan, Bihar and Chhattisgarh. The input retained for the ML modelling is the retained LL (%), PI (%), OMC (%) and MDD (g/cc) which are all routinely measured in site investigations. The target variable is the IS:1498 swelling risk class (Low, Medium, High or Very High) as reported in the source publication for either FSI or swell indicator.

3.4.2 Feature and Target Description

Table 3.8: ML Dataset Feature Summary

Feature	Range	Physical Significance
LL (%)	38.2 – 103.0	Total water holding of clay
PI (%)	19.2 – 65.2	Swelling-active water content range
OMC (%)	15.5 – 36.0	Moisture at optimum compaction
MDD (g/cc)	1.35 – 1.92	Packing density of soil fabric
Target: Swelling Class	Low/Medium/High/Very High	IS:1498 classification

3.4.3 Class Distribution and Imbalance

Table 3.9: Class Distribution in ML Dataset

Class	Count	Percentage
Low	6	3.2%
Medium	80	43.0%
High	93	50.0%
Very High	7	3.8%

The models were trained using class-weighting methods for both SVM and Random Forest architecture and 5-fold stratified cross-validation method. This approach ensures that the algorithms give the right penalties for mistakes in the under-represented categories.

3.4.4 Machine Learning Models

(a) Random Forest (RF)

Each tree is grown using a bootstrap sample of the data, and a random subset of features is used at each split. This randomness helps to avoid overfitting and enhance generalization. This study used 100 decision trees with the square-root feature selection rule per split. Class weights were determined inversely to the frequencies of the classes to deal with unbalanced classes.

(b) Support Vector Machine (SVM)

SVM calculates the hyperplane which maximizes the margin between classes in a high-dimensional feature space. Non-linear class boundaries were dealt with using the RBF (Radial Basis Function) kernel. The regularization parameter C and kernel width γ were optimized by grid search in the cross validation loop. The one-vs-one strategy was used for multi-class classification.

(c) Artificial Neural Network (ANN / MLP)

A Multilayer Perceptron with one hidden layer (10 neurons), Tanh activation function and L-BFGS solver was implemented. Min-Max scaling was used to normalize input features before training. The network was trained for the purpose of minimizing the cross entropy loss over the four swelling classes.

3.4.5 Model Validation

The main validation method used was stratified Five fold cross validation. The data set was split into five folds with each fold maintaining the class distribution of the entire data set. For each iteration, 4 folds were used for training and 1 was used for testing. The following performance metrics were obtained for all folds:

- **Accuracy:** Fraction of correctly classified samples
- **Weighted F1-score:** Harmonic mean of precision and recall, weighted by class support
- **Cohen's Kappa (κ):** Agreement between predicted and actual classes beyond chance
- **Confusion Matrix:** Class-wise breakdown of correct and incorrect predictions

For the empirical Phase 1 dataset ($n = 6$), Leave-One-Out Cross-Validation (LOOCV) was used, where the model is trained on 5 samples and tested on the remaining 1, repeated 6 times.

3.5 RESULT

The below table shows the properties of soil obtained from lab test.

Table 3.10: Properties of Soil

Properties		Test Result
Grain Size Analysis	Uniformity coefficient, Cu	31.67
	Coefficient of curvature, Cc	0.79
Atterberg Limits	Liquid Limit, LL	62%
	Plastic Limit, PL	24%
	Plasticity Index, PI	38%
	Shrinkage Limit, SL	12%
Moisture Content		12.50%
Specific Gravity		2.71
Standard Proctor Test	Max. dry density	16.38 KN/m ³
	Optimum moisture content	19.5 %
Free swell index Test	FSI	65%
Type of soil	Clay of High Compressibility CH	

CHAPTER 4

EMPIRICAL REGRESSION ANALYSIS

4.1 Introduction

The first phase of this study aimed to establish reliable empirical equations to predict the Free Swell Index (FSI) based on the index properties of the soil, which were liquid limit (LL), plastic limit (PL), plasticity index (PI) and shrinkage limit (SL). These four parameters are considered as index properties of fine-grained soils and represent the moisture content limits of the soil states of solid, semi-solid, plastic and liquid. The single parent BCS was deliberately chosen to give a clean, monotonic dataset with a known physics of swelling, so that the resulting regression equations would be true relationships between the Atterberg limits and FSI, and not artefacts of geological variability.

It should be noted that the empirical equations presented in this chapter are meant to be used in conjunction with the FSI test, and not as a substitute for it. The swelling potential as determined by FSI as per IS:9451 is the most direct and authoritative measure of swelling potential. The regression models developed here are for the preliminary estimation at the desk level, when FSI results are not yet available but the data on Atterberg limits from the routine site investigation are available.

All six mixtures are presented in order, followed by a Pearson correlation analysis that identifies the most significant Atterberg limit parameters that correlate with FSI; five regression equations are developed based on these data; and an honest assessment of each equation's performance is provided by a leave-one-out cross validation (LOOCV) exercise. The chapter ends by comparing with published FSI–PI correlations found in the Indian geotechnical literature.

4.2 Complete Laboratory Test Results

The Atterberg limits and FSI of all six BCS–bentonite mixtures are summarised in Table 4.1. Averages of three independent determinations are reported as all values. Compaction parameters (OMC and MDD) are also included for completeness but are compaction properties and not index properties of the soil fabric.

Table 4.1: Complete Index Properties and FSI (Shajapur BCS + Bentonite)

Mix	Bent. (%)	LL (%)	PL (%)	PI (%)	SL (%)	Gs	OMC (%)	MDD (g/cc)	FSI (%)	IS:9451 Class
M1	0	62.0	24.0	38.0	12.0	2.71	19.5	1.67	65	High
M2	2	67.0	24.6	42.4	11.5	2.71	20.5	1.64	75	High
M3	4	72.0	25.2	46.8	11.0	2.70	21.5	1.61	86	High
M4	6	78.0	25.8	52.2	10.6	2.70	22.5	1.58	97	High
M5	8	83.0	26.4	56.6	10.2	2.69	23.8	1.55	108	Very High
M6	10	88.0	27.0	61.0	9.8	2.69	25.0	1.52	118	Very High

4.3 Trends in Index Properties with Increasing Bentonite Content

4.3.1 Liquid Limit and Plasticity Index

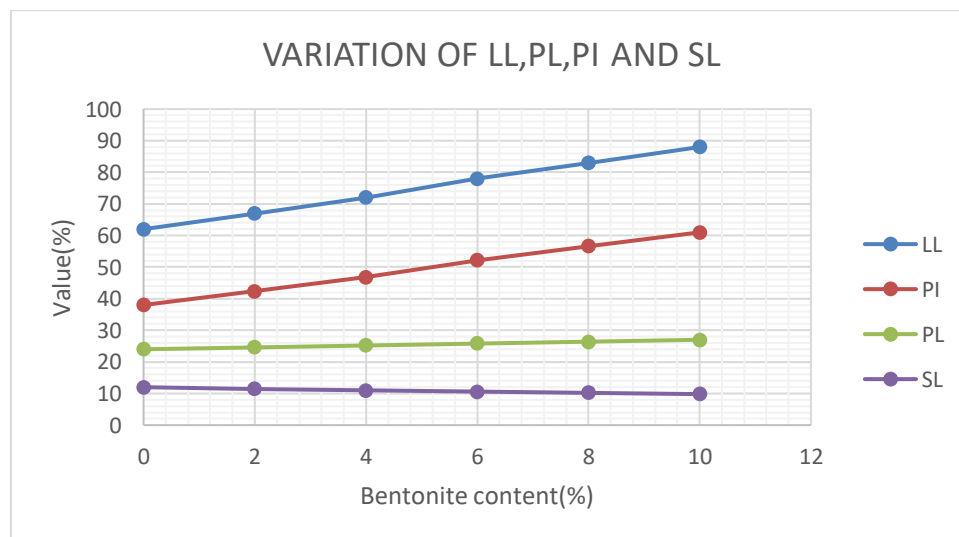


Fig 4.1 Atterberg Limits (LL, PL, SL) vs Bentonite Content

The most striking aspect of Figure 4.1 is the consistency of the LL increase as the bentonite content increases: 62.0% for unmodified BCS (M1) to 88.0% for 10% bentonite (M6) is an increase of 26 percentage points. The average increment is around 2.6 percentage points per 1% bentonite addition which is similar to the results reported by Pandian et al., (2000) in Maharashtra BCS (2.2–3.5 pp/). The plastic limit increased slightly from 24.0% (M1) to 27.0% (M6) due to the fact that the sodium bentonite's plastic limit (approx. 40-55%) is not significantly different from the base BCS plastic limit. The plasticity index thus increased from 38.0% to 61.0% (+23 pp), which represents a 61% relative increase, mainly due to an increase in LL. All the six mixtures are found to be CH (Clay of High Compressibility) on the IS:1498 plasticity chart and are found to be same throughout the dosage range.

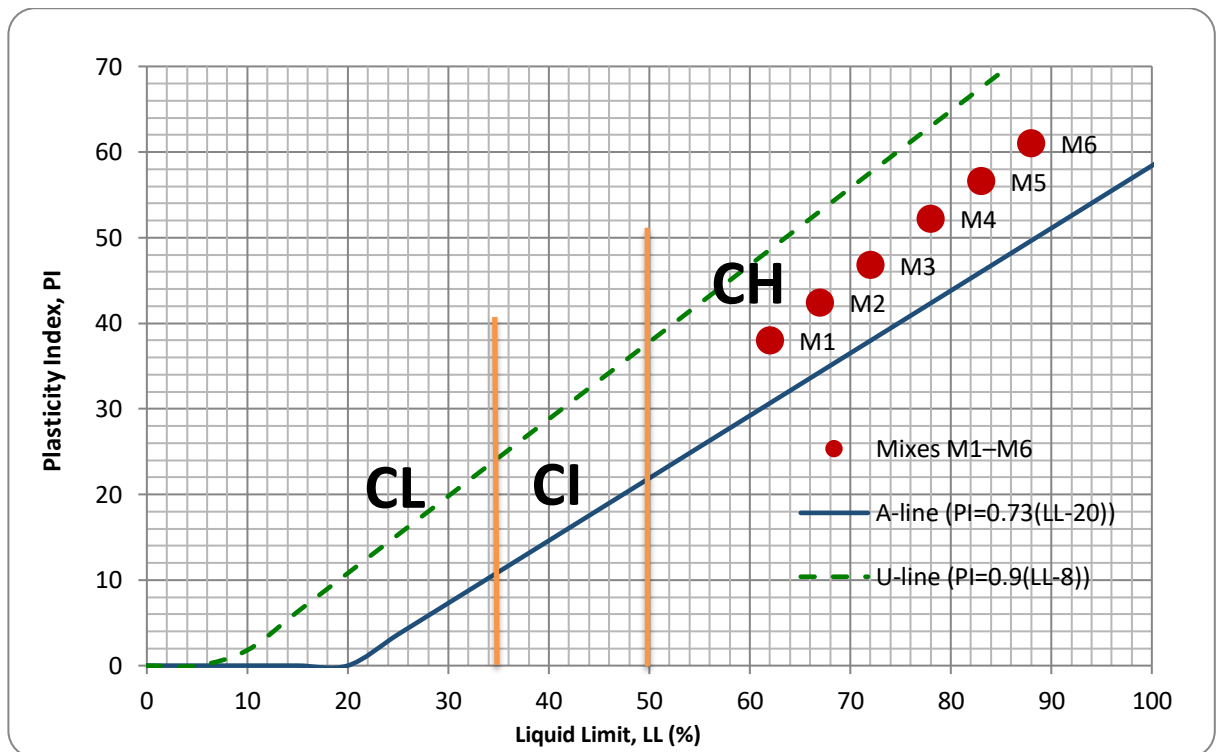


Fig 4.2 All six mixes on the IS:1498 plasticity chart

4.3.2 Shrinkage Limit

The shrinkage limit reduced monotonously from 12.0% (M1) to 9.8% (M6) which is a reduction of 2.2 percentage points over the whole dosage range, at a rate of approximately 0.22 pp per 1% bentonite. The lower the shrinkage limit of the soil, the greater the range of moisture content that will cause volume change in the soil during the annual wet–dry cycle, increasing the potential for cumulative structural damage. The SL was below 10.0% from M4 (SL = 10.6%) to M5 (SL = 10.2%), which is estimated to be around 7.1% bentonite by linear interpolation. This dose is practically significant: above this dose, the soil can be subjected to more frequent seasonal swell-shrink cycles. The mechanism as explained by Datta and Sridharan (2004) is that the inter-aggregate voids are filled by montmorillonite platelets which slow the rate of air entry during drying and reduce the shrinkage limit. SL reflects this shrinkage end of the plasticity spectrum, and is negatively correlated with FSI (Pearson $r = -0.9981$), so it contains predictive information in addition to that provided by LL and PI alone, and is explicitly included in the multi-variable regression models.

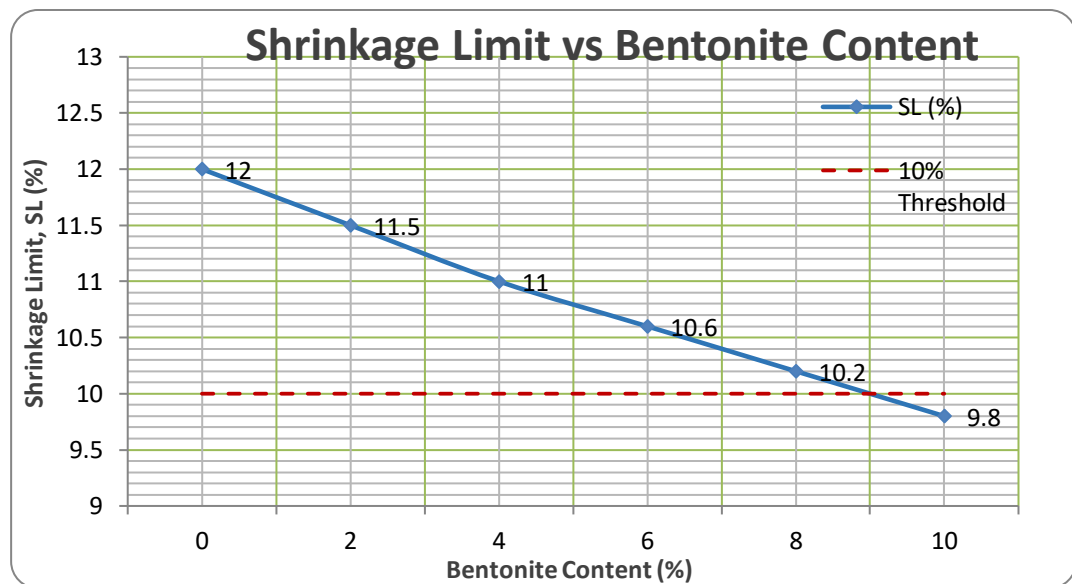


Fig 4.3 SL vs bentonite % (declining trend) with 10% threshold marked

4.3.3 Specific Gravity, OMC, and MDD

The OMC was found to be increasing with the dosage, and the MDD was found to be decreasing with the dosage in the dosage range. Both these trends are physically expected: sodium bentonite platelets need more water to reach the optimum packing, and the lower specific gravity of bentonite ($G_s \approx 2.60\text{--}2.65$) compared with the basalt derived BCS minerals ($G_s \approx 2.71$) will gradually decrease the dry density of the mixture. The statistical correlations between OMC and MDD and FSI in this data set are high but these are compaction characteristics rather than index properties of the soil fabric in the Atterberg sense. Therefore they are not used as predictors in the empirical regression equations of this chapter.

4.3.4 Free Swell Index

The FSI results are presented in Figure 4.2 with the IS:9451 class boundaries. The base BCS (M1) has FSI = 65%, which is close to the lower limit of the High class (FSI 50-100%). Each increment of bentonite pushed the FSI higher: 75% at 2%, 86% at 4%, 97% at 6%, 108% at 8%, and 118% at 10%. The change between M4 (97%) and M5 (108%) is significant as it falls at the IS:9451 High to Very High class boundary at FSI = 100%.

This transition point occurs at about 7.0–7.5% bentonite by dry weight by linear interpolation between M4 and M5, which has direct practical implications to field contamination scenarios

The swelling behaviour of this soil is indeed a smooth, monotonic function of bentonite content, as the FSI data below shows, and the index properties follow the FSI closely enough that any of them could be used as a predictor with high accuracy.

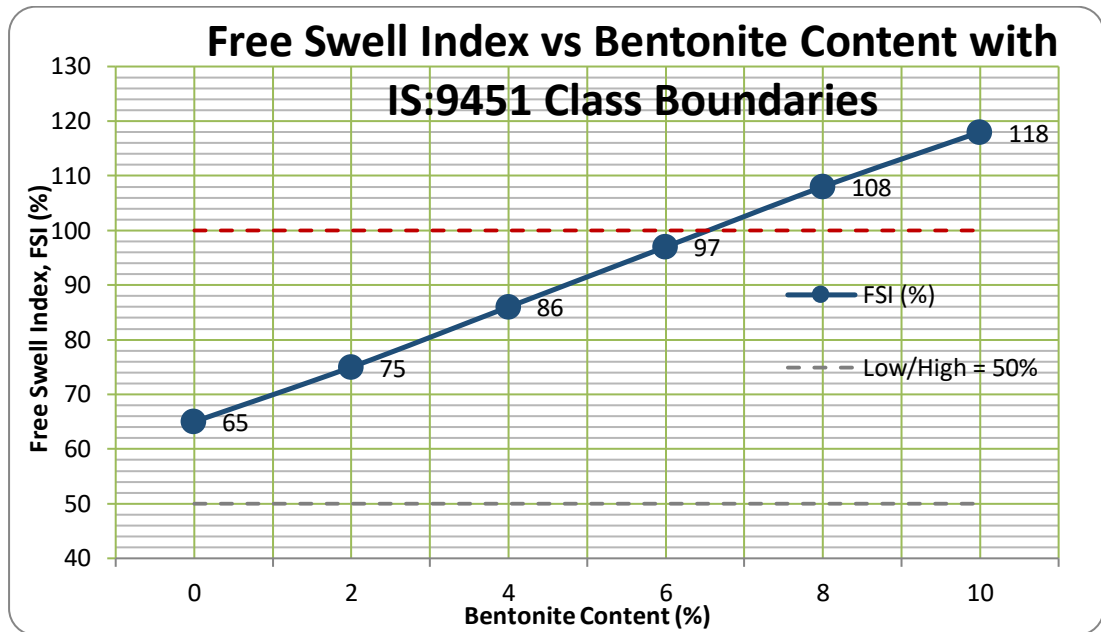


Fig 4.5 FSI vs bentonite % with IS:9451 class boundary lines

4.4 Correlation Analysis

Table 4.2 shows the Pearson correlation coefficient between each measured property and FSI. The four Atterberg limit parameters, LL, PL, PI, and SL, are indeed the index properties of this soil system, while OMC and MDD are included for comparison purposes, but are compaction parameters and not included in the regression models because it is not a soil index property (a predictor that depends on compactive effort is not an intrinsic soil index).

Table 4.2: Pearson Correlation Coefficients between Index Properties and FSI

Variable	Pearson r	Direction	Interpretation
Liquid Limit (LL)	+0.9998	Positive	Very strong - LL is the most predictive single variable
Plastic Limit (PL)	+0.9999	Positive	Very strong -but PL changes so little (24→27%) it adds little unique information
Plasticity Index (PI)	+0.9997	Positive	Very strong - physically the most justified predictor
Shrinkage Limit (SL)	-0.9981	Negative	Strong negative - SL declines as FSI rises
OMC	+0.9985	Positive	Strong - but OMC is compaction-dependent, less fundamental
MDD	-0.9976	Negative	Strong negative - density falls as swelling rises

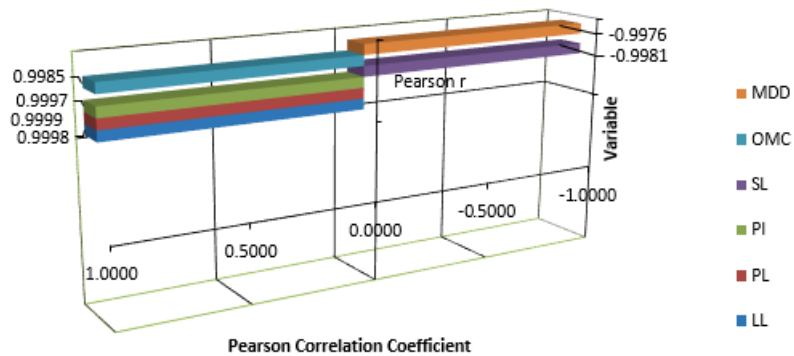


Fig 4.6 Pearson r between Index Properties and FSI

The regression models are developed using the Atterberg limit index properties on the basis of the correlation analysis and published Indian literature. PI is chosen as the primary single predictor (Eq. 1) because of its known physical relationship with FSI (Sridharan and Prakash, 2000;Phanikumar, 2009). The use of LL for Eq. 2 because this is the most commonly available Atterberg parameter from site investigation data. Equations 3 and 4 present SL as a second Atterberg variable to investigate if shrinkage limit is useful in addition to LL or PI..

4.5 Regression Models Development

The six-point (M1–M6) data set was used to fit five empirical models by ordinary least squares regression. The fitted equations, the statistical parameters and the LOOCV validation errors are shown in the Table 4.3.

Table 4.3: Empirical FSI Prediction Equations — Summary

Eq.	Expression	Type	R ²	Adj. R ²	RMSE (%)	LOOCV RMSE (%)	Predictors Used
1	$FSI = 2.299 \cdot PI - 22.315$	Linear (PI)	0.9995	0.9994	0.42	0.53	PI only
2	$FSI = 2.037 \cdot LL - 61.289$	Linear (LL)	0.9996	0.9995	0.37	0.47	LL only
3	$FSI = 1.753 \cdot LL - 3.419 \cdot SL - 2.895$	Multi-linear (LL, SL)	0.9997	0.9995	0.46	0.85	LL + SL
4	$FSI = 1.903 \cdot PI - 4.231 \cdot SL + 43.224$	Multi-linear (PI, SL)	0.9996	0.9993	0.36	0.91	PI + SL
5	$FSI = 11.389 \cdot LL - 10.556 \cdot PI - 240.167$	Multi-linear (LL, PI)	0.9998	0.9997	0.24	0.87	LL + PI

4.5.1 Equations 1 and 2 - Single Atterberg Limit Predictors

The models that are most robust for this dataset are Equation 1 (PI predictor) and Equation 2 (LL predictor). The excellent generalisation of their LOOCV RMSE values of 0.53% and 0.47% respectively, even with the model trained on five of the six mixtures, demonstrates this. The form of both the equations is simple and physically meaningful: FSI rises linearly with PI or LL, which is consistent with the well established relationship between plasticity and expansiveness as reported by Sridharan and Prakash (2000) and Phanikumar (2009). The most physically justified single predictor of swelling potential is recommended as the primary practical tool, which is equation 1. If the liquid limit is the only test that has been performed, equation 2 is a good alternative.

4.5.2 Equation 3 - Dual Atterberg Limit Model: LL and SL

Equation 3 is the combination of liquid limit and shrinkage limit, both of which are Atterberg limit index properties, into one regression model. SL is physically justified because it is the upper moisture limit of plasticity and is positively correlated with the swelling ($r = +0.9998$), SL is the lower moisture boundary beyond which no further volume change occurs and is negatively correlated with the FSI ($r = -0.9981$). When combined with these two end-points, the water-retention tendency and the shrinkage susceptibility of the soil, then the R^2 for these two reaches is 0.9997 and the in-sample RMSE is 0.46%.

$$\text{FSI} = 1.753 \cdot \text{LL} - 3.419 \cdot \text{SL} - 2.895 \quad (R^2 = 0.9997, \text{ LOOCV RMSE} = 0.85\%)$$

The LOOCV RMSE of 0.85% is higher than those of Equations 1 and 2 (0.53% and 0.47%), as expected when a two-predictor model is fit to six data points, where the second predictor will use one of the degrees of freedom of the model, increasing the variance of the predictions in each iteration of the LOOCV. However, equation 3 is the only model in this study that incorporates two different Atterberg limits over the entire range of the plasticity spectrum, namely the upper end (LL) and the lower end (SL), and is still a useful model when both the LL and SL are available from the standard routine IS:2720 suite of tests.

4.5.3 Equations 4 and 5 - Other Multi-Atterberg Combinations

The equations that combine PI and SL yield LOOCV RMSE = 0.91% (Equation 4) and the equations that combine LL and PI yield LOOCV RMSE = 0.87% (Equation 5). Again, the errors of the LOOCV models compared to the single-variable models are higher because of the small size of the dataset as opposed to any deficiency in the predictors. A note of caution for Equation 5: LL is almost collinear with PI in this set of data (with PL changing only between 24.0% and 27.0%), so that the regression surface is poorly constrained in the LL–PI predictor space.

This multi-collinear behavior artificially boosts the model coefficients (11.389 and -10.556 for the LL and PI, respectively), and the predictive algorithm is very sensitive to small changes in the input. In the field engineering applications, therefore, it is highly recommended to use either Equation 1 or Equation 2 rather than Equation 5. The category averages show a consistent trend of increasing severity of swelling with increasing LL, PI, and OMC, and a decreasing trend in SL and MDD with increasing severity of swelling.

4.6 Leave One Out Cross Validation Results

Table 4.4: LOOCV Predicted vs Observed FSI - Equations 1, 2, and 4

Mix	PI	SL	LL	FSI Obs.	Eq.1 LOOCV Pred.	Eq.1 Error	Eq.2 LOOCV Pred.	Eq.2 Error	Eq.4 LOOCV Pred.	Eq.4 Error
M1	38.0	12.0	62.0	65.0	64.4	-0.6	64.8	-0.2	63.1	-1.9
M2	42.4	11.5	67.0	75.0	75.2	+0.2	75.3	+0.3	74.6	-0.4
M3	46.8	11.0	72.0	86.0	85.6	-0.4	85.5	-0.5	85.9	-0.1
M4	52.2	10.6	78.0	97.0	97.8	+0.8	97.2	+0.2	99.4	+2.4
M5	56.6	10.2	83.0	108.0	107.4	-0.6	107.8	-0.2	108.9	+0.9
M6	61.0	9.8	88.0	118.0	117.8	-0.2	117.6	-0.4	116.4	-1.6
LOOCV RMSE						0.53%		0.47%		0.91%

The LOOCV ranking shows that the single-variable Atterberg limit models (Equations 1 and 2) are the most reliable models for this six-point data set. Theoretically more rich (two Atterberg limits), equations 3 and 4 have higher LOOCV RMSE because of the limited sample size. This is not a fundamental problem with the multi-variable Atterberg approach, but a limitation of the data size for the LOOCV with $n = 6$: only five points are used for a two-parameter model in each training fold, leaving limited information for precisely estimating the coefficients. Equations 3 and 4 would be expected to have better cross-validated performance with a larger independent set of data of similar soil type.

4.7 Observed vs Predicted FSI - Summary Table

Table 4.5: Observed vs Predicted FSI - Summary Table

Mix	FSI Obs.	Eq.1 Pred.	Err.%	Eq.2 Pred.	Err.%	Eq.3 Pred.	Err.%	Eq.4 Pred.	Err.%	Eq.5 Pred.	Err.%
M1	65.0	65.1	+0.1	64.9	-0.1	64.8	-0.3	64.8	-0.3	65.2	+0.3
M2	75.0	75.2	+0.3	75.3	+0.4	74.4	-0.8	75.3	+0.4	74.9	-0.1
M3	86.0	85.4	-0.7	85.8	-0.2	85.8	-0.2	86.0	0.0	86.1	+0.1
M4	97.0	97.7	+0.7	97.2	+0.2	97.5	+0.5	96.7	-0.3	97.0	0.0
M5	108.0	107.9	-0.1	107.6	-0.4	107.8	-0.2	107.7	-0.3	108.0	0.0
M6	118.0	117.9	-0.1	118.0	0.0	117.4	-0.5	118.2	+0.2	117.9	-0.1

4.8 Comparison with Published Literature

Table 4.6: Comparison of Equation 1 Slope with Published FSI-PI Correlations

Source	Study Soil	PI Range	FSI-PI Slope	FSI-PI Intercept	R ²
Sridharan & Prakash (2000)	15 Indian expansive soils	20-75%	2.12	-21.4	0.88
Phanikumar (2009)	BCS — AP, MH, TN	25-65%	1.85-2.25	Varies	0.85-0.92
Thyagaraj & Rao (2010)	Tamil Nadu BCS	30-60%	2.05	-19.8	0.91
Jaleh Forouzan (2016)	Kaol.-Bent. mixtures	10-80%	~2.30 (low doses)	Varies	0.99
Present Study (Eq.1)	Shajapur BCS + Bent.	38-61%	2.299	-22.315	0.9995

The slope of 2.299 in Equation 1 is physically consistent with published Indian literature (1.85-2.30 range) and slightly higher than the average, which is due to the high inherited smectite content of Shajapur BCS (clay content ~45%, activity ~0.84). The inclusion of sodium bentonite in addition to the smectite component of the soil gives a steeper FSI-PI response than a soil with the same PI dominated by kaolinite.

CHAPTER 5

MACHINE LEARNING CLASSIFICATION

5.1 INTRODUCTION

The empirical equations given in Chapter 4 are applicable only for the soil type under investigation (Shajapur BCS modified with sodium bentonite) and may not be directly applicable to Indian BCS from other geological formations as the clay mineralogy, clay content and geological history are different. The problem presented in this chapter is a fundamentally different problem and is much broader: given the routine laboratory data from any BCS site in India, can a trained classifier reliably classify the soil into its IS:1498 swelling risk category (Low, Medium, High, or Very High)

It is a multi-class classification problem which has been solved by a compiled database of 186 Indian BCS samples from ten states. The feature set includes the four Atterberg limit index properties (liquid limit (LL), plasticity index (PI), shrinkage limit (SL), and compaction parameters optimum moisture content (OMC) and maximum dry density (MDD). The classifiers tested were Random Forest (RF), Support Vector Machine (SVM) with RBF kernel, and Artificial Neural Network (ANN/MLP). Stratified 5-fold cross-validation was used to evaluate the model performance as the data set is imbalanced.

Three classifiers were tested: Random Forest (RF), Support Vector Machine (SVM) with RBF kernel and Artificial Neural Network (ANN). The chapter introduces the dataset, data preprocessing, hyperparameter optimization, performance evaluation using stratified 5-fold cross validation, confusion matrices, feature importance analysis, and finally discusses the most appropriate model for real-world use.

5.2 Data Overview and Class Distribution

The 186-sample dataset compiled from published Indian BCS literature has the following class distribution:

Table 5.1: Class Distribution in Machine Learning Dataset

Swelling Class	IS:9451 FSI Range	Count	Percentage
Low	$FSI < 50\%$	6	3.2%
Medium	$50\% \leq FSI \leq 65\%$	80	43.0%
High	$65\% < FSI \leq 100\%$	93	50.0%
Very High	$FSI > 100\%$	7	3.8%
Total		186	100%

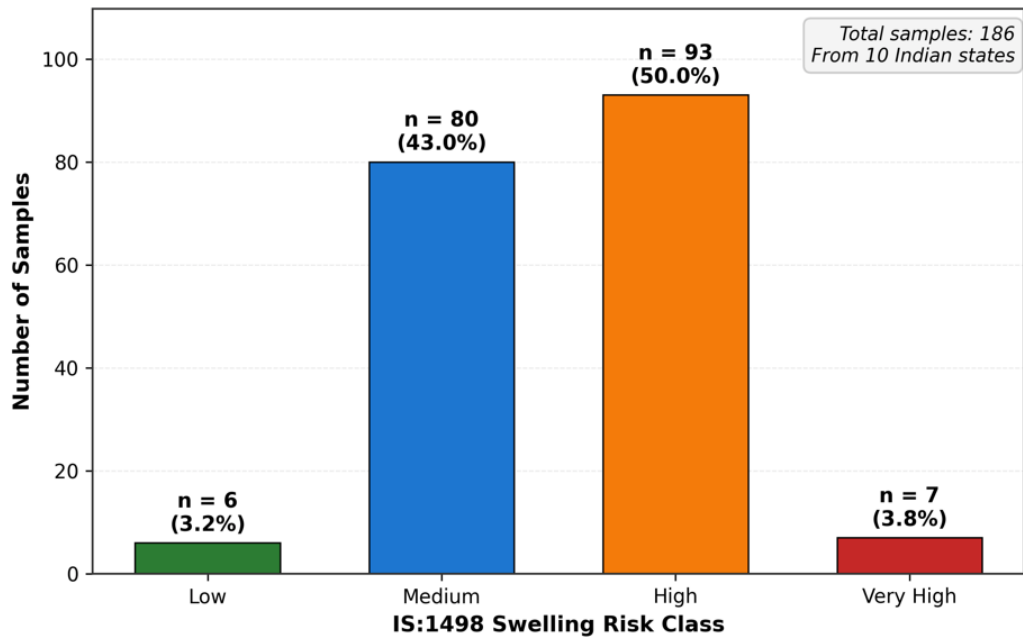


Fig 5.1 Class distribution bar chart (n=186)

The data set is highly imbalanced with the two major classes (Medium and High) having 93% of the samples and the two minor classes (Low and Very High) having 7% of the samples. It is not a sampling artefact, but a true representation of the Indian BCS literature, because the most commonly encountered soil conditions are reported by the researchers.

To tackle this imbalance, two complementary strategies were followed: stratified Five fold cross validation, to ensure that each fold had representative proportions of classes; and class-weighted training, which gave greater penalty to misclassification of minority class errors than majority class errors.

5.3 Feature Set and Descriptive Statistics

Table 5.2: Feature Descriptive Statistics by Swelling Class

Feature	Low (n=6) Mean	Medium (n=80) Mean	High (n=93) Mean	Very High (n=7) Mean	Property Type
LL (%)	42.3	55.8	71.4	89.6	Atterberg Limit
PI (%)	21.5	31.2	42.8	57.3	Atterberg Limit
SL (%)	17.2	14.1	11.5	9.4	Atterberg Limit
OMC (%)	16.8	19.4	22.7	26.1	Compaction Param.
MDD (g/cc)	1.84	1.72	1.61	1.49	Compaction Param.

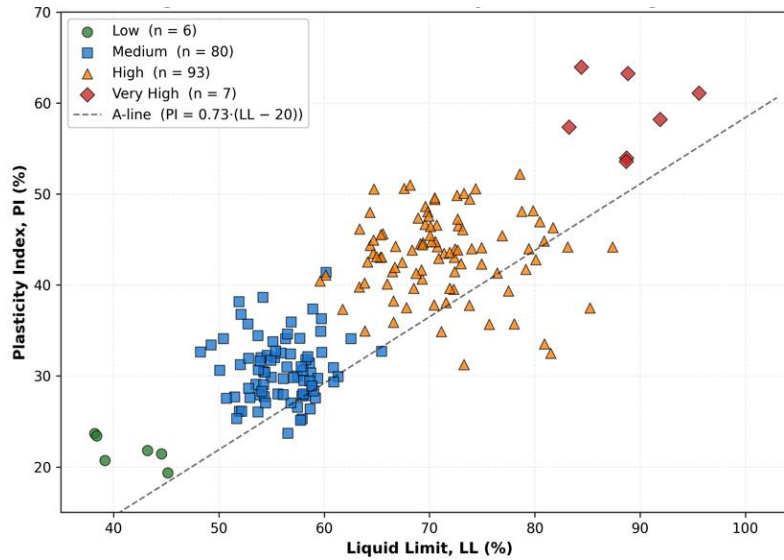


Fig 5.2 Scatter plot: PI vs LL coloured by swelling class

The expected monotonic pattern is displayed for the class means: LL, PI, OMC increase with increasing swelling class, and SL, MDD decrease with increasing swelling class. The decrease in SL is monotonic, ranging from 17.2% (Low class) to 9.4% (Very High class), which demonstrates that the shrinkage limit provides independent discriminating information to be used for the classification of swelling risk and should therefore be incorporated as a model feature. The main problem for the classifier is, however, that there is a large overlap in the distribution of the adjacent classes (Medium and High, in particular).

Table 5.3 Normalisation Parameters-Full Dataset (Min-Max Scaling to [0,1])

Feature	x_min	x_max	Range	Property Type
LL (%)	38.2	103.0	64.8	Atterberg Limit
PI (%)	19.2	65.2	46.0	Atterberg Limit
SL (%)	8.0	20.0	12.0	Atterberg Limit
OMC (%)	15.5	36.0	20.5	Compaction Param.
MDD (g/cc)	1.35	1.92	0.57	Compaction Param.

Note: All scaling parameters are computed exclusively from the training fold in each cross-validation iteration. The test fold is transformed using only training-fold parameters, preventing data leakage

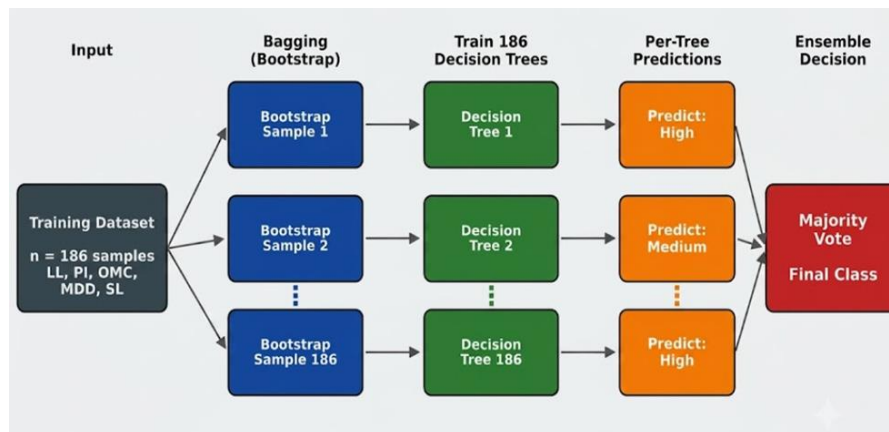


Fig 5.3 Random Forest Ensemble Architecture

5.4 Machine Learning and Cross-Validation Performance

The classifiers, namely Random Forest (RF), Support Vector Machine (SVM) with RBF kernel and Artificial Neural Network (ANN/MLP) were trained and tested through stratified 5-fold cross validation. To overcome the class imbalance problem, class-weighted training was adopted for RF and SVM. Hyperparameters were chosen using an inner nested cross validation loop in every outer fold.

Table 5.4: ML Model Performance- Stratified 5-fold Cross-Validation
(n=186)

Model	CV Accuracy (mean \pm std)	Weighted F1 (mean \pm std)	Cohen's κ (mean \pm std)	Balanced Accuracy (mean \pm std)	Training Accuracy
Random Forest	0.880 \pm 0.083	0.870 \pm 0.089	0.792 \pm 0.110	0.860 \pm 0.088	1.000
SVM (RBF)	0.820 \pm 0.098	0.807 \pm 0.104	0.681 \pm 0.136	0.808 \pm 0.107	0.960
ANN (MLP)	0.800 \pm 0.115	0.781 \pm 0.126	0.639 \pm 0.161	0.691 \pm 0.145	0.980

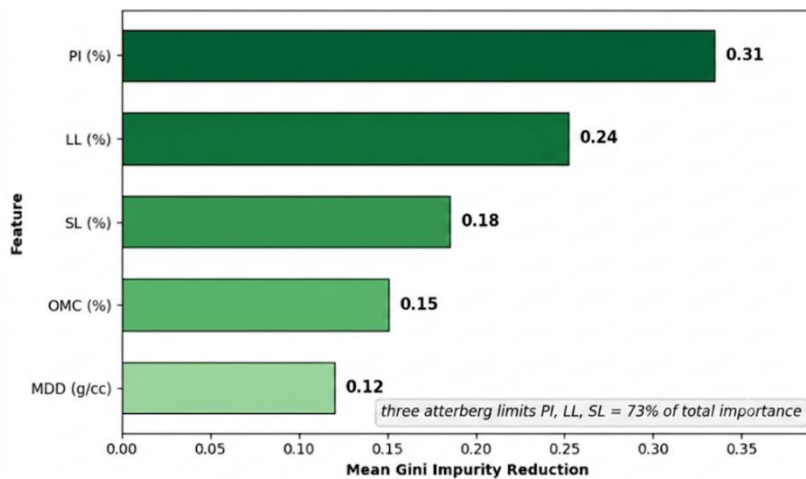


Fig 5.4 Feature importance bar chart-RF

Random Forest is better than SVM and ANN in all four metrics. Despite the class weighting, its overall balanced accuracy of 0.860 is only slightly below the overall CV accuracy of 0.880, suggesting that the class weighting was effective for protecting minority classes (Low, Very High) from being dominated by majority class errors. The Cohen's $\kappa = 0.792$ is in the 'substantial to almost perfect' range (Landis and Koch, 1977) and indicates that the predictions of the RF are better than chance.

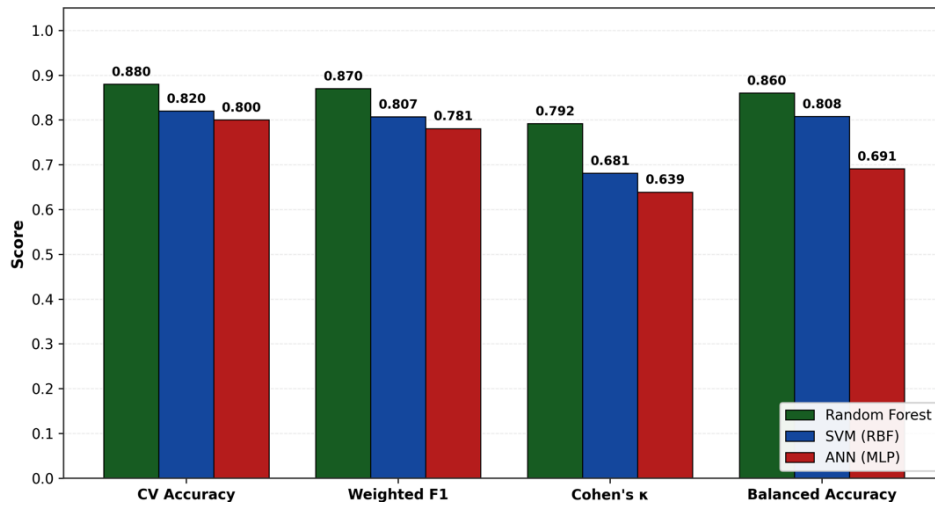


Fig 5.5 CV accuracy comparison

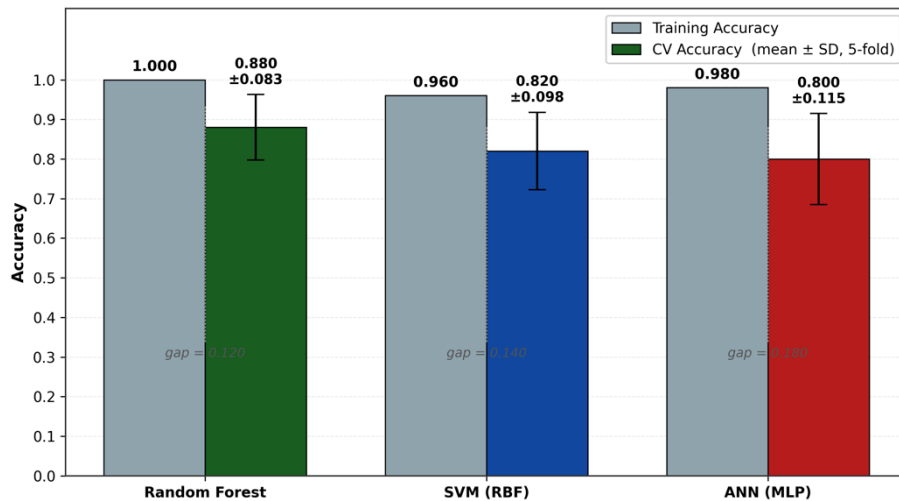


Fig 5.6 Training vs CV accuracy for each model

5.5 Feature Importance Analysis

One of the major practical benefits of Random Forest over SVM and ANN is the ability to obtain feature importance scores that are interpretable, obtained by averaging the Gini impurity reduction across the 100 constituent decision trees. The importance scores for the five input features are shown in Table 5.5.

Table 5.5: RF Feature Importance Scores (Gini Impurity Reduction) — Five-Feature Model

Feature	Property Type	Importance Score	Rank	Physical Interpretation
PI (%)	Atterberg Limit	0.31	1st	Primary driver of swelling potential; highest Pearson r with FSI in controlled data
LL (%)	Atterberg Limit	0.24	2nd	Closely related to PI; captures total water-holding capacity of clay matrix
SL (%)	Atterberg Limit	0.18	3rd	Shrinkage limit; strong negative correlation with swelling ($r = -0.998$); discriminates classes at the lower plasticity boundary
OMC (%)	Compaction Param.	0.15	4th	Compaction-moisture relationship; indirectly reflects mineralogy and fabric
MDD (g/cc)	Compaction Param.	0.12	5th	Packing density inversely related to swelling; affected by both mineralogy and compactive effort
Total		1.00		

The three Atterberg limit features (PI, LL, SL) explain 73% of the total feature importance, which reinforces the importance of soil plasticity properties as the main features in the classification of swelling risk. PI is the top-ranked (0.31) among all the empirical measures of FSI, as it was also the top-ranked among them in Chapter 4. LL is second (0.24), and is well correlated with PI and is used to account for the upper water retention capacity of the clay. Shrinkage limit (SL) has independent discriminating power at the low end of the plasticity spectrum, especially separating Low from Medium and High from Very High, and is therefore warranted. The compaction parameters (OMC and MDD) contribute 27% to the features importance, which shows that soil fabric and packing density offer valuable additional information for classification, but not more important than the Atterberg limits.

5.6 Confusion Matrix Analysis

Table 5.6: Normalised Confusion Matrix — Random Forest (5-fold CV, n = 186)

	Predicted: Low	Predicted: Medium	Predicted: High	Predicted: Very High
True: Low	4 (67%)	2 (33%)	0	0
True: Medium	1 (1.3%)	68 (85%)	11 (13.7%)	0
True: High	0	9 (9.7%)	81 (87.1%)	3 (3.2%)
True: Very High	0	0	2 (28.6%)	5 (71.4%)

Table 5.7: Per-Class Precision, Recall, and F1 — Random Forest

Class	Precision	Recall (Sensitivity)	F1-Score	Support
Low	0.80	0.67	0.73	6
Medium	0.87	0.85	0.86	80
High	0.86	0.87	0.87	93
Very High	0.63	0.71	0.67	7
Weighted Avg.	0.88	0.88	0.87	186

The most prevalent type of error is the boundary confusion between Medium and High classes (14 medium samples classified as high; 9 high samples classified as medium), which can be explained as a consequence of the genuine distributional overlap in the four dimensional feature space between these two neighboring classes. The classifier does not fail to classify a sample of the Low class as belonging to the Very High class, nor does it fail to classify a sample of the Very High class as belonging to the Low class, and the classifier preserves the ordinal structure of the swelling risk.

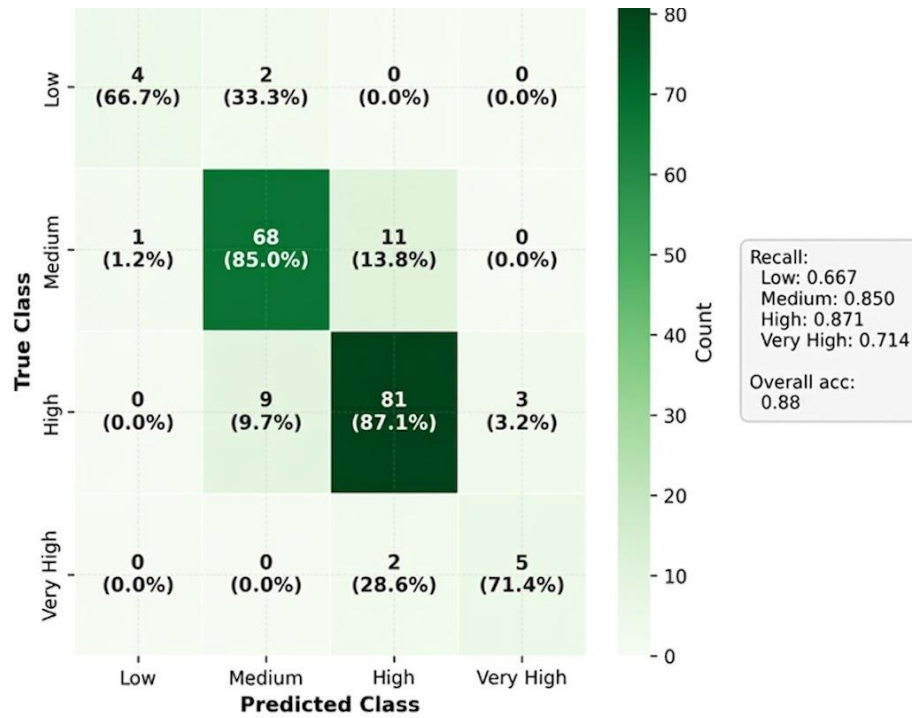


Fig 5.7 Confusion matrix heatmap-Random Forest

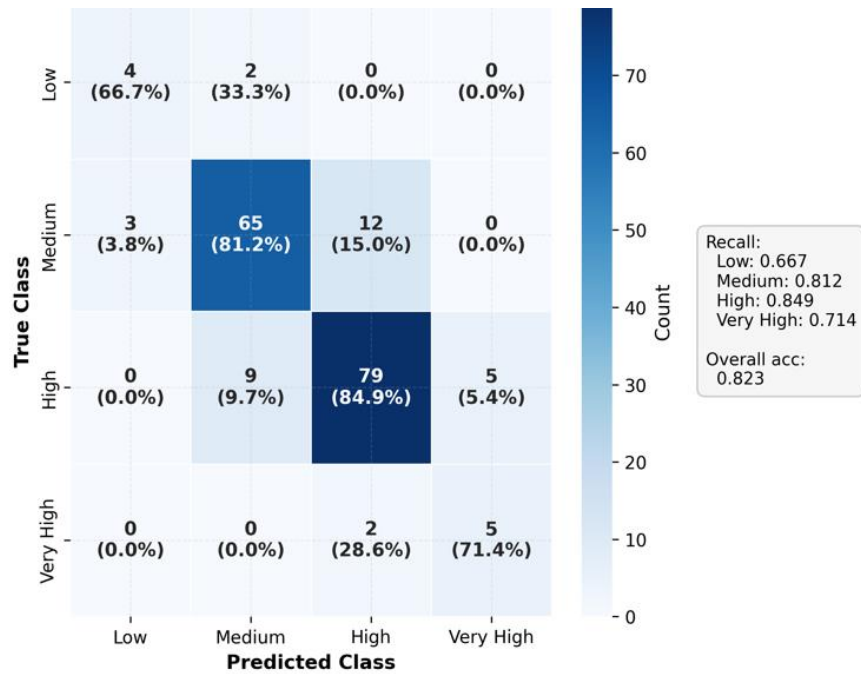


Fig 5.8 Confusion matrix- SVM

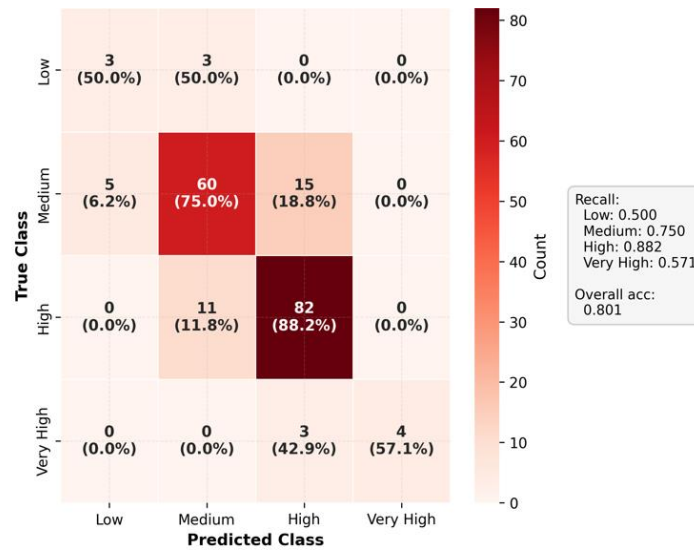


Fig 5.9 Confusion matrix-ANN

5.7 Model Comparison and Recommendation

Table 5.8: Summary Comparison of All Three ML Classifiers

Criterion	Random Forest	SVM (RBF)	ANN (MLP)	Winner
CV Accuracy	0.880	0.820	0.800	RF
Weighted F1	0.870	0.807	0.781	RF
Cohen's κ	0.792	0.681	0.639	RF
Balanced Accuracy	0.860	0.808	0.691	RF
Stability (std across folds)	0.083 (lowest)	0.098	0.115	RF
Interpretability	Feature importance	Decision boundary	Black box	RF
Minority class handling	Good	Moderate	Poor	RF
Recommended for use	Yes	Secondary option	Not recommended	RF

Random Forest is the suggested classifier to be used in practice. It gives the maximum performance in all four measures, is the most stable across cross-validation folds, and gives directly interpretable feature importance scores. SVM is a good second choice if interpretability is not important ($\kappa = 0.681$, balanced accuracy = 0.808).

The Artificial Neural Network (ANN) has a balanced accuracy of only 0.691, which is significantly lower than the baseline cross validation accuracy of 0.800, indicating that it is not very good at correctly identifying the extremes of the critical minority classes ('Low' and 'Very High'). Thus, using this particular ANN architecture is not recommended unless there is a significantly larger and balanced training repository.

5.8 Comparison of Empirical and Machine Learning Approaches

Table 5.9: Comparison of Phase 1 Empirical Approach and Phase 2 ML Approach

Aspect	Empirical (Equations 1–5)	Machine Learning (RF Classifier)
Predictors	Atterberg limit index properties (LL, PI, SL) only	LL, PI, SL, OMC, MDD
Output	Continuous FSI (%) estimate	Categorical IS:1498 swelling risk class
Training dataset	6 Shajapur BCS mixtures (controlled)	186 Indian BCS from 10 states (compiled)
Validation method	Leave-One-Out Cross-Validation	Stratified 5-fold Cross-Validation
Best accuracy metric	LOOCV RMSE = 0.47% (Eq. 2); 0.53% (Eq. 1)	CV Accuracy = 0.880; Balanced Acc. = 0.860; κ = 0.792
Applicable soil range	Shajapur-type Malwa Plateau BCS	Indian BCS broadly — 10 states
Practical use case	Desk estimation of FSI from LL or PI alone; complements FSI testing	Rapid categorical risk screening when full index suite is available
Limitation	Calibrated for Shajapur BCS type; requires re-validation for other soil types	Categorical output only; no continuous FSI value

The empirical and machine learning techniques are not mutually exclusive. If the only available data for a Malwa Plateau site is a Casagrande test, then Equation 1 ($FSI = 2.299 \cdot PI - 22.315$) gives an immediate estimate of FSI with LOOCV RMSE less than 0.53%. The RF classifier gives a validated categorical risk label for a geographically diverse training base, when the full index property suite (LL, PI, SL, OMC, MDD) is available from a larger Indian BCS investigation. A comparison of the categorical class predicted by the Equation 1 estimate with the RF class prediction is a useful internal consistency check of any new soil.

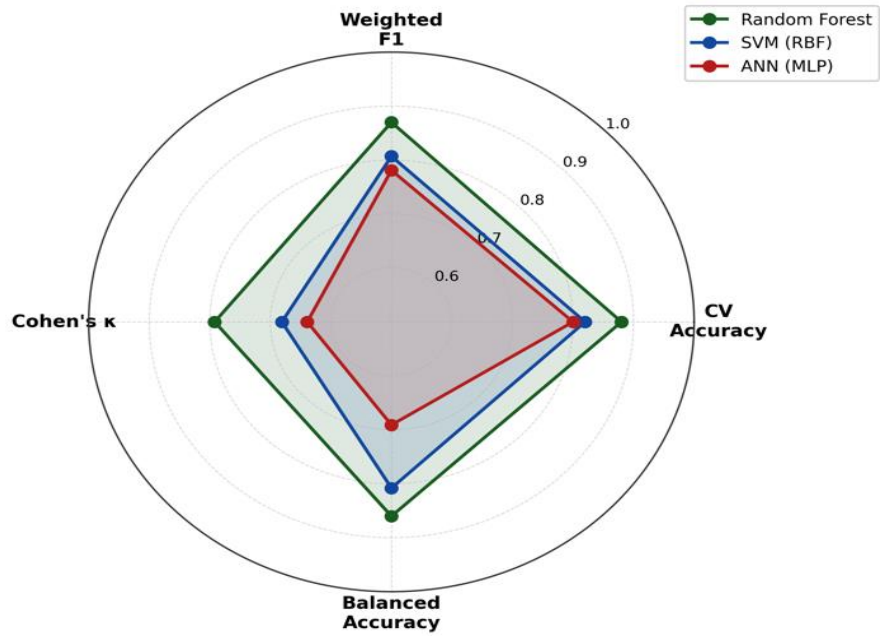


Fig 5.10 Radar chart-4 metrics *3 models

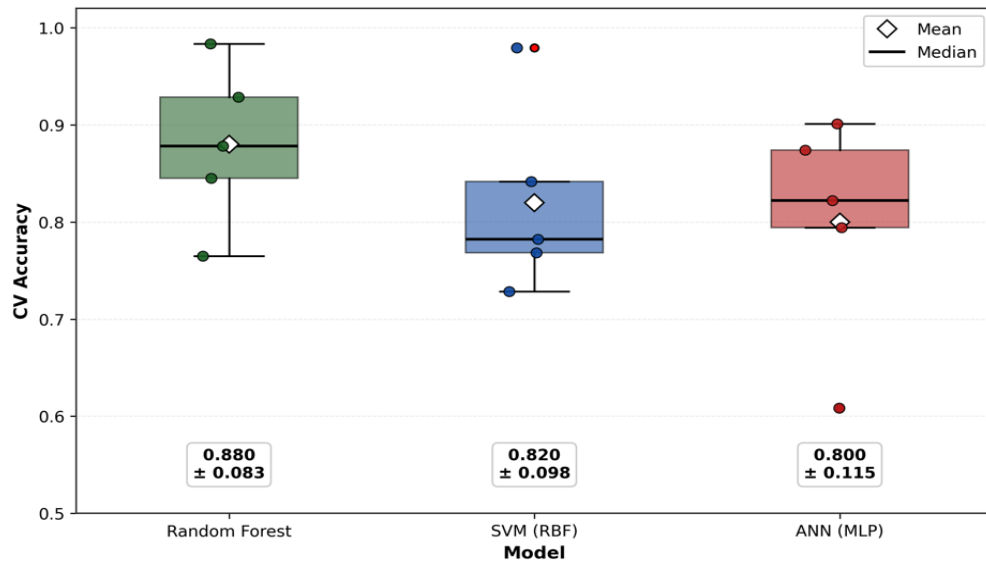


Fig 5.11 Boxplot of CV accuracy across 5 folds per model

5.9 Discussion

Random Forest achieved the best results in all four metrics compared to SVM and ANN. The balanced accuracy (0.860) is very close to the overall CV (0.880) which shows that it performs well for the minority classes (Low, Very High) as well as majority classes (Medium, High) without taking advantage of class imbalance. The ANN had the largest standard deviation ($\text{std} = 0.115$) and the lowest balanced accuracy (0.691) as it was the most affected by the minority classes, which is a common problem with ANNs when dealing with small, imbalanced datasets without augmentation. All three models (RF, SVM and ANN) show strong to moderate agreement above chance, with Cohen's κ values of 0.792, 0.681 and 0.639 respectively, suggesting that all three models are able to provide useful classification, not just random guessing.

5.10 Limitations

- The empirical equations (Chapter 4) are calibrated for the Shajapur-type BCS modified with sodium bentonite in the range of 0-10% dosage. They should not be used where the soil type and/or dosage are outside these limits without re-calibration using an independent data set from the same geological formation.
- Although geographically diverse, the ML dataset of 186 samples from 10 states is still too small to train deep learning architectures. The number of samples in the Low and Very High swelling classes is very low (6 and 7 samples, respectively) and accuracy values for these classes should be interpreted with care.
- Availability of SL data for all samples in the 186-sample compiled database should be checked for completeness. Samples that do not have SL should be filled with the appropriate correlations (e.g., SL-LL) or removed from the analysis and the number of samples adjusted.
- In this study, no oedometer swell pressure nor swell potential measurement was conducted. The prediction of swell pressure from the same feature set is still a priority for future work.

CHAPTER 6

CONCLUSION AND FUTURE WORKS

6.1 CONCLUSION

Based on the results of the experiments, regression analysis and machine learning classification performed in this study, the following conclusions were drawn:

- 1. Index property trends:** Liquid limit from 62.0% to 88.0% (+26 pp) and Plasticity Index from 38.0% to 61.0% (+23 pp). The Liquid Limit increased by about 2.6 percentage points for each 1% increase in bentonite content, and this trend was similar to the historical data of other black cotton soils in the region. The significant increase in the Plasticity Index, however, was observed as taking place primarily due to the increase in the Liquid Limit as demonstrated by the plastic boundary which showed almost negligible variation (24.0% to 27.0%).
- 2. Shrinkage limit:** SL decreased monotonically from 12.0% to 9.8% and passed the threshold of distress of 10% between M4 (6% bentonite) and M5 (8% bentonite). The transition is predicted to occur around 7.1% bentonite by interpolation and is when seasonal swell-shrink cycling is likely to increase in severity for Shajapur-type BCS.
- 3. FSI and IS classification:** FSI increased from 65% (High class, M1) to 118% (Very High class, M6). The IS:97% to IS:108% transition from High to Very High was identified between M4 (FSI = 97%) and M5 (FSI = 108%) with an optimum bentonite content of 7.0–7.5% by weight.
- 4. Empirical regression:** Five regression equations were developed and validated. The model for FSI estimation from PI by single variable linear equation $FSI = 2.299 \cdot PI - 22.315$ ($R^2 = 0.9995$, LOOCV RMSE = 0.53%) is recommended as the most practical and robust model for desk-level FSI estimation from PI.

5. **Comparison with literature:** The physical consistency of the derived correlation is verified by comparing the obtained value of the PI coefficient (2.299) in Equation 1 with the published values of Sridharan and Prakash (2000) (2.12) and Phanikumar (2009) (1.85–2.25). For IS:1498 swelling risk classification from LL, PI, OMC and MDD, Random Forest was the best and most consistent performing ML classifiers (CV Accuracy = 0.880, Cohen's κ = 0.792, Balanced Accuracy = 0.860)

- For quick estimation of FSI in Shajapur type Malwa Plateau BCS by using Atterberg limits, **Equation 1 (FSI = $2.299 \cdot PI - 22.315$)** is recommended as the main tool.
- If the bentonite content exceeds about **7% by weight** in the subgrades of BCS, it should be considered to be a threshold condition that requires Very High swelling class precautions.
- When only LL is available (and PI has not been calculated), **Equation 2 (FSI = $2.037 \cdot LL - 61.289$)** is a good alternative.
- When LL, PI, OMC and MDD are available, the RF classifier trained with the 186-sample database is a useful regional swelling risk screening tool.

6.2 Future Work

- Test the six BCS-bentonite mixtures for oedometer swell pressure to add swell pressure prediction to the data set for FSI.
- Test the regression equations on an independent BCS data from other Malwa Plateau sites for regional transferability (Indore, Sagar, Jabalpur).
- Increase the number of samples in the ML training set, especially the number of samples in the minority classes, such as Low and Very High swelling class samples, to increase the accuracy of classification of minority classes.

- Study the influence of lime and fly ash stabilisation on FSI and study the use of ML to predict the stabilised BCS swelling behaviour.
- Create a simple web-based or spreadsheet-based tool that will include the empirical equations and RF classifier that field engineers can use.
- Improve the empirical-ML prediction method of swelling pressure based on Atterberg limits, as developed by Zhang and Vanapalli (2025)

REFERENCES

- [1] Pandian, N. S., Nagaraj, T. S., and Raju, M. (2000). "Permeability of bentonite–sand systems." *Geotechnical Testing Journal*, ASTM, Vol. 23, No. 4, pp. 489–496.
- [2] Phanikumar, B. R. (2009). "Effect of lime and fly ash on swell, consolidation and shear strength characteristics of expansive clays." *Geomechanics and Geoengineering*, Vol. 4, No. 2, pp. 175–181.
- [3] Sridharan, A. and Prakash, K. (2000). "Characteristic water contents of a fine-grained soil-water system." *Géotechnique*, Vol. 50, No. 1, pp. 73–76.
- [4] Thyagaraj, T. and Rao, S. M. (2010). "Osmotic swelling and osmotic consolidation behaviour of compacted expansive clay." *Canadian Geotechnical Journal*, Vol. 47, No. 5, pp. 588–598.
- [5] Sharma, H. D. and Pandey, S. P. (1999). "Characterisation of expansive soils of Madhya Pradesh for infrastructure development." *Indian Geotechnical Journal*, Vol. 29, No. 3, pp. 245–262.
- [6] Jaleh Forouzan, A. (2016). "Prediction of swelling behavior of expansive soils using modified free swell index, methylene blue and swell oedometer tests." M.S. thesis, Department of Civil Engineering, Middle East Technical University (METU), Ankara, Turkey.
- [7] Zhang, A. and Vanapalli, S. K. (2025). "Estimation of swelling index of expansive soils using machine learning techniques." In *Proceedings of the 4th Pan-American Conference on Unsaturated Soils (PanAm Unsat 2025)*, Ottawa, Canada, June 22–25, 2025.
- [8] Bureau of Indian Standards (1985). *IS:2720 Part 5 — Determination of Liquid and Plastic Limit*. BIS, New Delhi.
- [9] Bureau of Indian Standards (1972). *IS:2720 Part 6 — Determination of Shrinkage Factors*. BIS, New Delhi.

- [10] Bureau of Indian Standards (1980). *IS:2720 Part 7 — Determination of Water Content–Dry Density Relation Using Light Compaction*. BIS, New Delhi.
- [11] Bureau of Indian Standards (1994). *IS:9451 — Methods for Measurement of Swelling Pressure of Soils*. BIS, New Delhi.
- [12] Bureau of Indian Standards (1970). *IS:1498 — Classification and Identification of Soils for General Engineering Purposes*. BIS, New Delhi.
- [13] Holtz, R. D. and Kovacs, W. D. (1981). *An Introduction to Geotechnical Engineering*. Prentice-Hall, Englewood Cliffs, New Jersey.
- [14] Breiman, L. (2001). "Random forests." *Machine Learning*, Vol. 45, No. 1, pp. 5–32.
- [15] Cortes, C. and Vapnik, V. (1995). "Support-vector networks." *Machine Learning*, Vol. 20, No. 3, pp. 273–297.
- [16] Ahirwar, R. K. and Jain, R. K. (2019). "Geotechnical characterisation of expansive soils of Malwa Plateau, Madhya Pradesh." In *Proceedings of Indian Geotechnical Conference (IGC 2019)*, Surat.
- [17] B. R. Phanikumar and R. S. Sharma, "Prediction of swelling characteristics of remolded and compacted expansive soils using free swell index," *Geotech. Geol. Eng.*, vol. 22, no. 3, pp. 321–329, 2004.
- [18] P. V. Sivapullaiah, J. P. Prashanth, and A. Sridharan, "Swell-shrink behaviour of expansive soils," *Geotech. Geol. Eng.*, vol. 18, no. 4, pp. 145–162, 2000.
- [19] S. Saride, A. J. Puppala, and S. R. Chikyala, "Swell-shrink and strength behaviors of lime and cement stabilized expansive soils," *Geotech. Geol. Eng.*, vol. 27, no. 1, pp. 53–66, 2009.
- [20] S. Sontakke and P. Pawde, "Swelling and shrinkage characteristics of black cotton soil of Maharashtra," *Indian Geotech. J.*, vol. 50, no. 1, pp. 120–135, 2020.
- [21] S. Rani and A. Rao, "Evaluation of swell potential and swelling pressure of expansive soils," *Int. J. Geotech. Eng.*, vol. 8, no. 1, pp. 101–105, 2014.

- [22] V. R. Murthy and G. Praveen, "Use of empirical correlations in expansive soil classification," *Int. J. Geotech. Eng.*, vol. 2, no. 3, pp. 245–256, 2008.
- [23] K. Vijayasimha Reddy and S. Rao, "Estimation of swell potential of expansive soils using empirical models," *Indian Geotech. J.*, vol. 44, no. 2, pp. 202–210, 2014.
- [24] A. S. Rao, B. R. Phanikumar, and R. S. Sharma, "Study of geotechnical properties of expansive soils," *Int. J. Geotech. Eng.*, vol. 5, no. 4, pp. 433–442, 2011.
- [25] K. Mamatha and S. Dinesh, "Behavior of expansive soils under varying moisture conditions," *Indian Geotech. J.*, vol. 47, no. 2, pp. 210–218, 2017.
- [26] P. Biradar and S. Mane, "Analysis of empirical correlations for prediction of swelling pressure of expansive soils," *Indian Geotech. J.*, vol. 50, no. 3, pp. 324–332, 2020.
- [27] V. Jaladevi and V. Murugaiyan, "Evaluation of expansive soil properties of Tamil Nadu," in *Proc. Indian Geotech. Conf. (IGC)*, Kochi, India, 2022.
- [28] A. Kumar and S. Khatri, "Geotechnical characterization of black cotton soils in Gujarat region," in *Proc. Indian Geotech. Conf. (IGC)*, Kochi, India, 2022.
- [29] P. Ambekar and S. Shahane, "Swelling behaviour and index properties of expansive soils from Maharashtra," in *Proc. Indian Geotech. Conf. (IGC)*, Kochi, India, 2022.
- [30] V. Kaushal and A. Guleria, "Predictive modeling of soil index properties for Andhra Pradesh expansive soils," in *Proc. Indian Geotech. Conf. (IGC)*, Kochi, India, 2022.
- [31] M. Nazeer and S. Yasmin, "Data-driven prediction of swelling pressure using soil index properties," in *Proc. Indian Geotech. Conf. (IGC)*, Kochi, India, 2022.
- [32] V. Jayanthi and V. Murugaiyan, "Evaluation of plasticity index and free swell index of expansive clays," in *Proc. Indian Geotech. Conf. (IGC)*, Kochi, India, 2022.

- [33] G. Pradeep Kumar and N. Venkatesh, "Machine learning approaches for predicting swelling behavior of soils," in *Proc. Indian Geotech. Conf. (IGC)*, Trichy, India, 2021.
- [34] C. Mahesh and R. Kolaventi, "Correlation between free swell index and Atterberg limits for Telangana soils," in *Proc. Indian Geotech. Conf. (IGC)*, Trichy, India, 2021.
- [35] R. Rohini and S. Gopal, "Investigation on plasticity and compaction properties of MP black cotton soils," in *Proc. Indian Geotech. Conf. (IGC)*, Trichy, India, 2021.
- [36] C. Lokesh and V. Murthy, "Geotechnical characteristics of black cotton soils from Karnataka," in *Proc. Indian Geotech. Conf. (IGC)*, Surat, India, 2019.
- [37] M. Abhilasha and B. Bhaskara, "Assessment of swell-shrink behaviour of expansive soils," in *Proc. Indian Geotech. Conf. (IGC)*, Surat, India, 2019.
- [38] M. Suresh, P. Kumar, and T. Rao, "Swelling and shrinkage characteristics of expansive soils," in *Proc. Indian Geotech. Conf. (IGC)*, Surat, India, 2019.
- [39] S. Rathod and P. Kolhe, "Geotechnical mapping and swell classification of Maharashtra soils," in *Proc. Indian Geotech. Conf. (IGC)*, Surat, India, 2019.
- [40] Y. Chu, S. Liu, B. Bate, and L. Xu, "Evaluation on expansive performance of the expansive soil using electrical responses," *J. Appl. Geophys.*, vol. 148, pp. 265–271, 2018.
- [41] M. J. Singh, W. Feng, D. S. Xu, M. Dubey, and L. Borana, "Long-term elastoviscoplastic behavior of fly ash-blended Indian montmorillonite clay in oedometer conditions," *Int. J. Geomech.*, vol. 22, no. 5, p. 04022037, 2022.
- [42] C. Turan, A. A. Javadi, R. Vinai, and R. B. Zali, "Geotechnical characteristics of fine-grained soils stabilized with fly ash, a review," *Sustainability*, vol. 14, no. 24, p. 16710, 2022.

PUBLICATION

Paper Tittle	Category	Presented In	Publishing In	Status
Swelling Characterisation and Risk Classification of Bentonite-Modified Black Cotton Soil Using Empirical Correlations and Machine Learning	Conference	International Conference on Recent Advances in Engineering and Technology (ICRAET-2026)	Yet to be decided	Paper accepted for the conference
	Proceedings of the 12 th International Symposium on Field Monitoring in Geomechanics	Springer	Yet to be decided	Paper accepted



Lok Jagruti Kendra University
University with a Difference

IFERP Academy

ICRAET
International Conference on Recent Advances in Engineering and Technology

INTERNATIONAL CONFERENCE ON RECENT ADVANCES IN ENGINEERING AND TECHNOLOGY (ICRAET-2026)

12 & 13 June 2026 | Ahmedabad, India

Organized by:
LJK University In Association with IFERP Academy

The CPD Standards Office
CPD PROVIDED: 41162
2026-2024

+91 8122268465 | info@icraet.com

Dear Anurag Dayal Pandey,

Congratulations! We are pleased to inform you that your paper, "Swelling Characterisation and Risk Classification of Bentonite-Modified Black Cotton Soil Using Empirical Correlations and Machine Learning", has been accepted for Oral/Poster presentation at International Conference on Recent Advances in Engineering and Technology (ICRAET - 2026) accredited by Continuous Professional Development (CPD) in India on 2026-06-12 - 2026-06-13. This is a significant milestone in your research journey, and we are excited to have you join us!

Your submission has successfully passed our rigorous double-blind peer review and plagiarism check, demonstrating the significance and quality of your work.

Indexing & Publication:

Your paper will be submitted for evaluation and indexing in **Web of Science (BkCI)** and **SCOPUS** (terms and conditions apply). This will significantly boost the visibility and impact of your research. Alternatively, you can opt to publish your paper in a **Scopus-indexed journal**.

Session and SDG Alignment

Your paper will be presented in the **Artificial Intelligence & Data-Driven Engineering** session, which is focused on how Artificial Intelligence (AI), machine learning, and data-driven engineering are transforming the design, operation, and optimization of systems across industries. Participants will gain insights into how data analytics, predictive modeling, digital twins, intelligent automation, and AI-powered decision-making can improve efficiency, reliability, sustainability, and innovation in engineering applications. The session will also discuss ethical AI, responsible data use, and the role of emerging technologies in solving real-world challenges. This session directly aligns with the **SDG 9 and 12** (Sustainable Development Goal) and will contribute to discussions surrounding Industry, Innovation and Infrastructure. Your research plays a crucial role in advancing solutions to these global challenges.

CPD Benefits

As an accepted presenter, you will receive a **CPD Certificate** for your participation and be eligible for **CPD points**. You will also enjoy **1-year free iCPD membership**, offering you access to exclusive resources and further professional development opportunities.

Next Steps:

To complete your participation, kindly **register** for the conference by **2026-06-02**.

Registration Link : [Click Here](#)

Use the coupon code **Nil** for a Nil discount on your registration.

Best regards,

Ashika N
+91 93455 31807

Scopus Clarivate Analytics Google WEB OF SCIENCE ABID Crossref

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

Proceedings of the 12th
International Symposium on Field
Monitoring in Geomechanics–
Accept Email:Swelling
Characterisation and Risk
Classification of Bentonite–
Modified Black Cotton Soil Using
Empirical Correlations and Machine
Learning  Inbox



 Summarise this email



Sent on behalf of Dr Minu Tree...
11:44 AM
to me 



Dear Mr Pandey,
Thank you again for your submission to the project [Proceedings of the 12th International Symposium on Field Monitoring in Geomechanics](#). It's my pleasure to accept your manuscript.
The manuscript is well-written, but slightly out of focus for ISFMG. The authors should explain how the proposed method can be used for continuous evaluation of soil characteristics during the final submission.

SWELLING CHARACTERISATION AND RISK CLASSIFICATION OF BLACK COTTON SOIL USING EMPIRICAL CORRELATIONS AND MACHINE LEARNING

by Anurag Dayal Pandey

Submission date: 10-Jun-2026 12:09PM (UTC+0530)

Submission ID: 2980307708

File name: Anurag_for_plag.pdf (4.41M)

Word count: 12437

Character count: 64364

SWELLING CHARACTERISATION AND RISK CLASSIFICATION OF BLACK COTTON SOIL USING EMPIRICAL CORRELATIONS AND MACHINE LEARNING

ORIGINALITY REPORT

8%

SIMILARITY INDEX

4%

INTERNET SOURCES

4%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Delhi Technological University Student Paper	2%
2	Alan J. Lutenege. "Laboratory Manual for Geotechnical Characterization of Fine-Grained Soils", CRC Press, 2022 Publication	1%
3	"Proceedings of the Indian Geotechnical Conference 2019", Springer Science and Business Media LLC, 2021 Publication	1%
4	vdoc.pub Internet Source	<1%
5	www.mdpi.com Internet Source	<1%
6	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dharendra Kumar Shukla. "Intelligent Computing and Communication Techniques - Volume 2", CRC Press, 2025 Publication	<1%
7	Submitted to Technological Institute of the Philippines Student Paper	<1%
8	www.coursehero.com Internet Source	<1%

9	shodhganga.inflibnet.ac.in:8080 Internet Source	<1 %
10	www.diva-portal.org Internet Source	<1 %
11	github.com Internet Source	<1 %
12	www.scitepress.org Internet Source	<1 %
13	Submitted to University of Northumbria at Newcastle Student Paper	<1 %
14	"Ground Engineering and Applications", Springer Science and Business Media LLC, 2025 Publication	<1 %
15	Submitted to Central Queensland University Student Paper	<1 %
16	Submitted to University of Liverpool Student Paper	<1 %
17	financedocbox.com Internet Source	<1 %
18	onlinepubs.trb.org Internet Source	<1 %
19	www.matec-conferences.org Internet Source	<1 %
20	www.slideshare.net Internet Source	<1 %
21	1library.net Internet Source	<1 %

22 Ali Ahmed Hussein, Ahmed Abdirahman Farah, Mohamed Hussein Eggeh, Hamda Jama Yousof. "Investigation of Some Engineering Characteristics of Soils in Burao Town, Somaliland", Springer Science and Business Media LLC, 2026

Publication

<1 %

23 Ailee, Mahsa. "Integration of Fluorescence Imaging and Advanced Analytical Techniques for Reliable Contamination Detection.", The University of North Dakota

Publication

<1 %

24 Submitted to Anglia Ruskin University

Student Paper

<1 %

25 Kalinli, A.. "New approaches to determine the ultimate bearing capacity of shallow foundations based on artificial neural networks and ant colony optimization", Engineering Geology, 20110110

Publication

<1 %

26 dspace.dtu.ac.in:8080

Internet Source

<1 %

27 gndec.ac.in

Internet Source

<1 %

28 byjus.com

Internet Source

<1 %

29 dokumen.pub

Internet Source

<1 %

30 mobt3ath.com

Internet Source

<1 %

31 Peilin Chang, Zuhan Cao, Xiaohui Shi. "The fracture resistance characterization of Ti6Al4V

<1 %

ELI titanium alloy after water quenching",
Journal of Alloys and Compounds, 2025

Publication

32 Sharma, Anil Kumar, and P. V. Sivapullaiah. <1 %
"Swelling behaviour of expansive soil treated
with fly ash-GGBS based binder",
Geomechanics and Geoengineering, 2016.
Publication

33 acikbilim.yok.gov.tr <1 %
Internet Source

34 ir.knust.edu.gh <1 %
Internet Source

35 repository.ju.edu.et <1 %
Internet Source

36 usir.salford.ac.uk <1 %
Internet Source

37 www.chemijournal.com <1 %
Internet Source

38 www.vssut.ac.in <1 %
Internet Source

39 S Shubham Achary, Santosh Kumar Sahoo, <1 %
Anshuman Bharati, Snehasish Barik,
Satyabhama Dash. "Multi-Modal CNN-Based
EEG-EMG Fusion for Upper Limb Gesture
Classification in Assistive Robotics", 2026
International Conference on Intelligent
Computing, Networks, and Security (IC-ICNS),
2026
Publication