

COMPARATIVE EVALUATION OF MACHINE LEARNING MODELS FOR RAINFALL-INDUCED SLOPE STABILITY PREDICTION IN BIHAR EMBANKMENTS USING FACTOR OF SAFETY

A Dissertation

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF TECHNOLOGY IN GEOTECHNICAL ENGINEERING BY

**AKASH PUSHKAR
24/GTE/16**

Under the Supervision of

Dr. Ashok Kumar Gupta

Professor, Department of Civil Engineering
Delhi Technological University



Department of Civil Engineering

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daultpur, Main Bawana Road, Delhi-110042, India

May, 2026



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I owe an enormous debt of gratitude to Dr. Ashok Kumar Gupta, Professor in the Department of Civil Engineering at Delhi Technological University, whose guidance shaped this dissertation at every stage. Whether it was a half-formed idea about which model to try next or a structural question about the validation approach, his door was always open. His deep familiarity with geotechnical engineering and his practical sense of what machine learning can and cannot do in the field kept this work grounded throughout.

I am equally grateful to Ms. Shivangi Bhardwaj, Ph.D. Research Scholar, who offered both technical input and moral support during the more frustrating stretches of this work. Her ability to spot logical gaps in an argument — and to do so kindly — made the thesis considerably better than it would otherwise have been.

The Department of Civil Engineering at Delhi Technological University provided computing resources and access to research literature without which none of the computational work in Chapters 4 and 5 would have been possible. I also want to thank my fellow students in the Geotechnical Engineering group, whose questions during our informal lab discussions forced me to explain my assumptions more carefully than I might have done on my own.

Finally, I am grateful to the India Meteorological Department for making the RF25 gridded rainfall dataset publicly accessible, and to the open-source communities behind scikit-learn, XGBoost, LightGBM and SHAP. This work sits on foundations others built and made freely available; that is not something I take for granted.

Place: Delhi

Date: May 2026

AKASH PUSHKAR

24/GTE/16

M.Tech(GTE)

Delhi Technological University

Delhi-110042



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **“COMPARATIVE EVALUATION OF MACHINE LEARNING MODELS FOR RAINFALL INDUCED SLOPE STABILITY PREDICTION IN BIHAR EMBANKMENTS USING FACTOR OF SAFETY”** in partial fulfillment of the requirements for the award of the Degree of Master of Technology, submitted in the Department of Civil Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from August 2024 to May 2026 under the supervision of Dr. Ashok Kumar Gupta.

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

Candidate's Signature

This is to certify that the candidate 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.

Signature of Supervisor(s)

Signature of External Examiner



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

CERTIFICATE BY SUPERVISOR(s)

This is to certify that the Project Dissertation titled “Comparative Evaluation of Machine Learning Models for Rainfall-Induced Slope Stability Prediction in Bihar Embankments Using Factor of Safety” submitted by AKASH PUSHKAR, Roll No. 24/GTE/16, Department of Civil Engineering, Delhi Technological University, Delhi, in partial fulfillment of the requirement for the award of the degree of Master of Technology (Geotechnical Engineering), is a record of the project work carried out by the student under my supervision.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Dr. Ashok Kumar Gupta

Date:

Professor

Department of Civil Engineering
Delhi Technological University
Delhi-110042

ABSTRACT

Bihar sits at the heart of the Indo-Gangetic Plain and bears the full force of Himalayan river flooding every year. The Bihar Water Resources Department has built over 3,500 earthen embankments along the Kosi, Gandak, Bagmati, Kamla-Balan, Burhi Gandak and Ganga river systems to protect more than 40 million people, yet roughly 28% of these structures suffer some form of failure — breaching, slumping or seepage-induced piping — each monsoon season. The economic toll reaches up to Rs. 2,400 crores annually, and between 2 and 5 million people are displaced from their homes as a direct result. The core problem is straightforward but difficult to solve at scale: site-by-site geotechnical assessment costs Rs. 50,000–2,00,000 per embankment and takes 1–2 weeks, making comprehensive pre-monsoon screening of the full network financially and logistically impossible.

This thesis builds a machine learning framework aimed squarely at that gap. A synthetic dataset of 10,000 slope stability scenarios was generated using the Infinite Slope Model, drawing on 14 years (2011–2025) of daily gridded rainfall data from IMD’s RF25 product ($0.25^\circ \times 0.25^\circ$ resolution). Eight input features were used: effective cohesion (c'), friction angle (ϕ'), unit weight (γ), slope angle (β), slope height (z), saturation ratio (m derived from IMD rainfall), mean daily rainfall and maximum daily rainfall. Each Factor of Safety was classified into three stability classes per IS:7894-1975: *Stable* ($\text{FoS} \geq 1.50$), *Marginal* ($1.25 \leq \text{FoS} < 1.50$) and *Unstable* ($\text{FoS} < 1.25$).

Five classifiers were implemented and compared under a single unified framework: Logistic Regression, Random Forest, SVM with RBF kernel, XGBoost and LightGBM. The dataset was split 70/15/15 (train/validation/test) with stratification preserved across all partitions. Class imbalance — the Marginal class accounts for only 13.2% of scenarios — was handled by applying SMOTE exclusively to the training set. All models were tuned via GridSearchCV or RandomizedSearchCV and validated using Repeated Stratified K-Fold Cross-Validation (5 folds $\times$ 5 repeats = 25 splits).

The SVM with RBF kernel proved the strongest classifier, reaching **97.1% accuracy**, a **macro F1-score of 0.953** and an **AUC-ROC of 0.997** — ahead of all baselines by 2.3–11.7 percentage points in accuracy. Cross-validation confirmed the result was not a lucky split: $99.40\% \pm 0.002$ across 25 folds.

The dataset labels were independently checked against Bishop’s Simplified Method (86% class agreement, $r = 0.998$) and the Fellenius Method (94%, $r = 0.996$). The best model was also validated against the **Durgawati Earthen Dam** (Kaimur, Bihar), where it produced a mean FoS = 1.864 — consistent with published Bishop’s values of 1.771–2.017.

SHAP analysis confirmed the model had learned genuine geotechnical relationships rather than statistical noise: slope angle ($|\overline{\text{SHAP}}| = 0.438$) and slope height (0.184) dominate predictions, with cohesion (0.137) and friction angle (0.112) capturing soil strength. Rainfall features contribute 8.5% in total, confirming that integrating real IMD data adds measurable predictive value beyond geotechnical parameters alone.

Contents

Acknowledgement	i
Candidate's Declaration	ii
Certificate	iii
Abstract	v
Contents	viii
List of Tables	ix
List of Figures	xi
List of Symbols and Abbreviations	xii
1 INTRODUCTION	1
1.1 Background and Motivation	1
1.2 Problem Statement	2
1.3 Aim of the Thesis	2
1.4 Scope and Organization of the Thesis	3
2 RELATED WORK	4
2.1 Overview	4
2.2 Classical Geotechnical Methods	4
2.3 Machine Learning Approaches to Slope Stability	5
2.3.1 Early Neural Network Approaches	5
2.3.2 Support Vector Machines	5
2.3.3 Ensemble Methods — Random Forest and Gradient Boosting .	5
2.3.4 Class Imbalance and SHAP Interpretability	5
2.4 Rainfall-Slope Stability Coupling	6
2.5 Summary and Research Gaps	6
3 STUDY AREA AND DATASET GENERATION	7
3.1 Study Area: Bihar, India	7
3.1.1 Geographical and Topographical Setting	7
3.1.2 Soil Characteristics of Bihar Alluvial Embankments	7
3.2 IMD RF25 Rainfall Dataset	8

3.2.1	Data Source and Spatial Resolution	8
3.3	Dataset Generation Using the Infinite Slope Model	8
3.3.1	Model Selection Justification	8
3.3.2	Factor of Safety Computation	9
4	MACHINE LEARNING METHODOLOGY	11
4.1	Overall Framework	11
4.2	Data Preprocessing	12
4.2.1	Train-Validation-Test Split	12
4.2.2	Feature Scaling	12
4.2.3	SMOTE Oversampling	12
4.3	Model Architectures	12
4.3.1	Parameter Settings	13
4.4	Evaluation Metrics	13
4.5	Cross-Validation Strategy	14
5	RESULTS AND DISCUSSION	15
5.1	Overview	15
5.2	Classification Performance	15
5.3	Detailed SVM Performance Analysis	16
5.3.1	Per-Class AUC-ROC	16
5.3.2	Confusion Matrix Analysis	17
5.4	Cross-Validation Results	18
5.5	Model-Wise Discussion	19
6	GEOTECHNICAL VALIDATION	20
6.1	Overview	20
6.2	Cross-Method Validation of Dataset Labels	20
6.2.1	Bishop's Simplified Method Validation	20
6.2.2	Fellenius Method Validation	20
6.3	Durgawati Earthen Dam Case Study	22
6.3.1	Site Description and Input Parameters	22
6.3.2	Results and Cross-Method Comparison	22
6.4	District-Level Real-Time Early Warning System (EWS)	22
6.4.1	EWS Architecture and Data Pipeline	23
6.4.2	District-Specific Critical Rainfall Thresholds	23
6.4.3	Four-Tier Alert Classification	25
6.4.4	Seven-Day Forward-Looking FoS Window	26
6.4.5	EWS Dashboard and Operational Demonstration	26
7	SHAP ANALYSIS AND ENGINEERING INTERPRETATION	28
7.1	Overview	28
7.2	Global Feature Importance	28
7.3	Physical Consistency Verification	28
7.4	Directional Feature Effects and Engineering Recommendations	30

8	DISCUSSION	31
8.1	Comparison with Prior Literature	31
8.2	Physical Consistency of Results	32
8.3	Practical Implications for Deployment	32
8.4	Limitations	32
9	CONCLUSION AND FUTURE SCOPE	33
9.1	Conclusion	33
9.2	Future Work	34
A	Research Contributions and Dissemination	36
A.1	List of Publications	36
A.2	Paper Acceptance Proof	36
A.3	Infinite Slope Model: Complete Derivation	37
	A.3.1 Model Assumptions	37
	A.3.2 Force Equilibrium Derivation	37
A.4	Python Code: Complete ML Training Pipeline	37

List of Tables

3.1	Geotechnical Parameter Ranges for Bihar Alluvial Embankment Soils	7
3.2	Dataset Class Distribution Across 10,000 Generated Scenarios	9
4.1	Hyperparameter Configurations for All Five Models	13
5.1	Classification Performance of Five ML Models on the Holdout Test Set	15
5.2	SVM Confusion Matrix on the Holdout Test Set	18
5.3	Repeated Stratified K-Fold Cross-Validation Results (5 Folds $\times$ 5 Repeats = 25 Splits)	18
6.1	Geotechnical Cross-Method Validation of Synthetic Dataset Labels ($n = 50$)	21
6.2	Durgawati Earthen Dam Geotechnical Input Parameters for Case Study Validation	22
6.3	Cross-Method Validation on Durgawati Earthen Dam, Kaimur, Bihar .	22
6.4	District-specific critical rainfall thresholds (R_{critical}), dry and saturated Factor of Safety, and inherent geotechnical risk class for all 38 Bihar districts. R_{critical} is the minimum daily rainfall (mm/day) that reduces FoS to 1.25 (IS:7894-1975 failure criterion). <i>Stable</i> (>50) denotes districts where full saturation ($m = 1.0$) does not reduce FoS below 1.25.	24
7.1	SHAP Global Feature Importance Rankings for the SVM Model . . .	29
8.1	Comparison of This Study's Results with Prior ML-Based Slope Stability Studies	31

List of Figures

3.1	Bihar District-Wise Embankment Breach Risk Map Showing Rainfall-Induced Slope Failure Hotspots (1987–2025) Based on WRD Bihar Data	8
3.2	Bihar IMD Rainfall Analysis (2011–2025): Yearly Average, Monsoon vs Non-Monsoon Distribution, and Monthly Pattern	9
3.3	Dataset Class Distribution Across Three Stability Classes (10,000 Scenarios)	10
4.1	Overall Workflow of the Proposed Machine Learning Framework for Bihar Embankment Slope Stability Prediction	11
5.1	Bihar Earthen Embankment ML Classification Results: (a) SVM Five-Fold Cross-Validation Accuracy Across Folds; (b) All-Models Performance Comparison; (c) Random Forest Feature Importance; (d) SVM Confusion Matrix; (e) SVM ROC Curves (Macro AUC = 0.997); (f) Per-Class F1 Comparison Across All Five Models	16
5.2	ROC Curves (One-vs-Rest) for All ML Models Across Three Stability Classes	17
5.3	SVM Confusion Matrix: (a) Raw Counts; (b) Class-Normalised Percentages	18
5.4	Actual vs Predicted Stability Class for All Five ML Models on the Bihar Slope Stability Holdout Test Set	19
6.1	Geotechnical Verification of the Infinite Slope Model Compared to Bishop’s Simplified Method and Fellenius Method on 50 Selected Scenarios: (a) Correlation Between Infinite Slope and Bishop’s FoS; (b) Correlation Between Infinite Slope and Fellenius FoS; (c) FoS Values for All 50 Cases; (d) Confusion Matrix for Class Agreement; (e) Distribution of FoS Differences	21
6.2	(a) Daily IMD RF25 Rainfall for Kaimur District (2011–2025); (b) Daily FoS for Durgawati Dam Embankment Core with Published Bishop’s FoS Range from Himanshu and Burman [9]	23
6.3	Bihar Embankment EWS Dashboard — 18 May 2026. (a) District-wise alert status for all 38 districts. (b) State-level alert distribution. (c) Daily FoS timeline for the four most vulnerable districts using real IMD RF25 rainfall data. (d) Seven-day FoS monitoring window for Sitamarhi district. (e) Calibrated rainfall-to-saturation-ratio transfer function used in the EWS.	27

7.1	SHAP Global Feature Importance Scores for All Eight Input Features of the SVM Model	30
A.1	Conference Acceptance Proof — ICSCCC-2026	36

List of Symbols and Abbreviations

ANN	Artificial Neural Network
API	Antecedent Precipitation Index
AUC	Area Under the Curve
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BIS	Bureau of Indian Standards
CV	Cross-Validation
DEM	Digital Elevation Model
DTU	Delhi Technological University
EFB	Exclusive Feature Bundling
FEM	Finite Element Method
FoS	Factor of Safety
GIS	Geographic Information System
GOSS	Gradient-based One-Side Sampling
IMD	India Meteorological Department
IS	Indian Standard
LEM	Limit Equilibrium Method
LightGBM	Light Gradient Boosting Machine
LR	Logistic Regression
ML	Machine Learning
NDMA	National Disaster Management Authority
RBF	Radial Basis Function
RF	Random Forest
RF25	IMD Gridded Rainfall Data at 0.25° Resolution
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
WRD	Water Resources Department (Bihar)
XGBoost	eXtreme Gradient Boosting
c'	Effective Cohesion (kPa)
ϕ'	Effective Internal Friction Angle (°)
γ	Bulk Unit Weight of Soil (kN/m ³)
γ_w	Unit Weight of Water, 9.81 (kN/m ³)
β	Slope Angle (°)
z	Slope Height / Depth to Failure Surface (m)

m Saturation Ratio (0–1)

Chapter 1

INTRODUCTION

1.1 Background and Motivation

Bihar's geography makes flooding close to inevitable. The state stretches across the eastern Indo-Gangetic Plain between latitudes 24.3° – 27.5° N and longitudes 83.3° – 88.3° E, drained by the Kosi, Gandak, Bagmati, Kamla-Balan, Burhi Gandak and the Ganga — rivers that descend steeply from the Himalayas, carry enormous sediment loads and have historically produced some of the most dynamic and hazardous flood-plain conditions anywhere in the world. In response, the Bihar Water Resources Department has constructed over 3,500 earthen embankments along these river banks, collectively spanning thousands of kilometres and representing one of the largest flood-control infrastructure systems in India.

The problem is that these embankments fail at a rate that should be alarming. According to the Bihar Water Resources Department's Annual Report [2], about 28% of embankments experience some degree of failure — breaching, slumping, overtopping or seepage-induced piping — every monsoon season. The consequences are severe: the National Disaster Management Authority [21] estimates annual economic damages of up to Rs. 2,400 crores, with 2–5 million people displaced from their homes each year as a direct result.

The failure mechanism is well understood. Prolonged or intense rainfall infiltrates the embankment slope and the underlying foundation soils, progressively raising the degree of saturation. As saturation increases, pore water pressure rises within the soil skeleton, reducing effective normal stresses. The Mohr-Coulomb failure criterion tells us that this reduction in effective stress directly diminishes the shear strength available on potential failure surfaces. When mobilised shear stress along any failure plane exceeds the available shear strength, the Factor of Safety drops below unity and slope failure follows.

The Factor of Safety is the standard benchmark in geotechnical practice. Per IS:7894-1975 [5], $\text{FoS} \geq 1.50$ indicates a stable slope; $1.25 \leq \text{FoS} < 1.50$ indicates marginal stability; $\text{FoS} < 1.25$ indicates an unstable slope requiring immediate intervention. Computing the FoS by traditional methods — the Infinite Slope Model, Bishop's Simplified Method, the Fellenius Method or the Janbu Generalised Method — requires site-specific geotechnical investigations costing Rs. 50,000–2,00,000 per site and 1–2 weeks per embankment [5]. With over 3,500 embankments to screen before each monsoon, comprehensive traditional assessment is simply not feasible.

This gap is what motivates the use of machine learning here. Ensemble methods, support vector machines and interpretability tools such as SHAP have shown real promise for geotechnical prediction [10, 23, 28], but prior studies on slope stability have consistently worked with small datasets of 249–600 samples, used no real rainfall data, focused on non-alluvial soils and reported inadequate multi-class metrics. This thesis addresses all of those gaps within a single framework designed specifically for Bihar’s embankments.

1.2 Problem Statement

Bihar’s earthen embankments protect more than 40 million people but fail at roughly a 28% annual rate due to rainfall-induced slope instability. Traditional geotechnical methods are too slow and too expensive to screen all 3,500-plus structures before each monsoon season. What is needed is a machine learning framework capable of predicting slope stability rapidly, accurately and at scale — one that uses real meteorological data and geotechnical parameters calibrated for India’s most flood-prone state.

A second problem is worth naming directly: the geotechnical machine learning literature lacks rigorous side-by-side comparisons. Most studies evaluate a single model and then cite numbers from other papers that used different datasets, different metrics and different preprocessing pipelines. Drawing reliable conclusions about which algorithm genuinely works best for slope stability classification is therefore impossible from the existing literature. Running a controlled five-algorithm comparison under identical experimental conditions is the second major objective of this work.

1.3 Aim of the Thesis

This thesis has six aims. The first is to build a large, physically realistic dataset of 10,000 slope stability scenarios using the Infinite Slope Model, incorporating Bihar-specific geotechnical parameter ranges and 14 years of real IMD RF25 daily rainfall data. The second is to implement and systematically compare five machine learning classifiers — Logistic Regression, Random Forest, SVM, XGBoost and LightGBM — under a unified evaluation framework. The third is to optimise model hyperparameters and validate performance using Repeated Stratified K-Fold Cross-Validation. The fourth is to independently validate dataset labels against Bishop’s Simplified Method and the Fellenius Method, and to validate the best model on the real-world Durgawati Earthen Dam. The fifth is to apply SHAP analysis to quantify feature importance and confirm that the best-performing model is physically interpretable. The sixth is to assess practical scalability for district-wise embankment risk prioritisation and to formulate specific recommendations for the Bihar Water Resources Department.

1.4 Scope and Organization of the Thesis

The thesis is organised into nine chapters. Chapter 2 reviews the existing literature on slope stability analysis and machine learning-based geotechnical prediction. Chapter 3 describes the study area, the IMD RF25 rainfall data and the dataset generation procedure using the Infinite Slope Model. Chapter 4 covers the machine learning methodology: preprocessing, SMOTE balancing, model architectures, hyperparameter tuning and the cross-validation strategy. Chapter 5 presents classification results on the holdout test set and across cross-validation folds, with a detailed breakdown of the SVM's performance. Chapter 6 covers geotechnical validation of the dataset labels against Bishop's and Fellenius methods, and presents the Durgawati Earthen Dam case study. Chapter 7 presents the SHAP analysis and engineering interpretation of feature importance. Chapter 8 discusses the results against the existing literature, examines practical implications and acknowledges limitations. Chapter 9 summarises the main conclusions and outlines directions for future research.

Chapter 2

RELATED WORK

2.1 Overview

Slope stability analysis has been studied for over a century, moving from hand-calculated limit equilibrium methods through finite element simulation to, more recently, data-driven machine learning. This chapter reviews key contributions in each of these areas thematically, which allows the present work to be positioned clearly against what has come before. The review covers classical geotechnical methods, early neural network approaches, ensemble methods, support vector machines and interpretability-focused frameworks.

2.2 Classical Geotechnical Methods

The standard tools for slope stability assessment are the limit equilibrium methods, which treat failure as occurring along a predetermined surface and compute the Factor of Safety as the ratio of total resisting to driving forces. Bishop's Simplified Method [3], introduced in 1955, extended this framework to circular failure surfaces and remains the most widely used method in practice. The Fellenius Method [29] simplifies the analysis further by neglecting interslice forces, producing a conservative estimate well suited to preliminary assessment. The Infinite Slope Model, developed for shallow translational failures in long, uniform slopes, provides a closed-form FoS expression that is particularly appropriate for Bihar's alluvial embankments.

These methods are rigorous, but they have real limitations at scale. Fredlund and Krahn [7] showed that the choice of method influences the computed FoS by 5–15% depending on geometry and soil parameters. Finite element packages such as PLAXIS, ABAQUS and SLOPE/W are more rigorous but require detailed mesh generation and are far too slow for screening thousands of embankments. Probabilistic approaches, reviewed by Zhang et al. [26], treat geotechnical parameters as random variables and yield a probability of failure, but demand parameter distribution data that is almost never available across a network as large as Bihar's.

2.3 Machine Learning Approaches to Slope Stability

2.3.1 Early Neural Network Approaches

Machine learning for slope stability started with neural networks in the mid-1990s. Goh [8] applied a back-propagation network to slope stability data from 45 samples and achieved 85–90% accuracy — a proof of concept that a trained parametric model could approximate the nonlinear relationship between soil parameters and stability without any explicit analytical formulation. The contribution was genuinely foundational, even if the tiny dataset made the results difficult to generalise.

2.3.2 Support Vector Machines

Samui [24] applied support vector machines to slope stability prediction on 80 samples, achieving $R^2 = 0.88$. The study argued that SVM’s maximum-margin principle, combined with a nonlinear kernel, could capture the complex decision boundaries arising from interactions among geotechnical parameters. Notably, SVMs generalised well even on small training sets, a property the study attributed to structural risk minimisation — which is particularly useful in geotechnical work where collecting labelled data is expensive.

2.3.3 Ensemble Methods — Random Forest and Gradient Boosting

Hoang and Pham [10] achieved 90.3% accuracy using Random Forest on 279 samples, establishing ensemble methods as a credible alternative to neural networks for slope stability classification. Qi and Tang [23] demonstrated XGBoost at 88.6% on 325 samples, showing that regularised gradient boosting handles tabular geotechnical data well. Zhou et al. (2022) reported 92.4% accuracy with LightGBM on 600 samples, noting that the leaf-wise growth strategy produced faster training without sacrificing accuracy.

Lin et al. [16] carried out one of the more systematic comparative studies in this area, evaluating four methods — ANN, SVM, Decision Tree and Random Forest — on the same dataset under identical conditions. Both SVM and Random Forest outperformed the simpler models, but the study found no consistent winner between them across different datasets.

2.3.4 Class Imbalance and SHAP Interpretability

Huang et al. [11] combined Random Forest with SMOTE oversampling and reported an 8.3% improvement in F1-score on an imbalanced slope stability dataset. This was among the first studies to explicitly acknowledge and address class imbalance in geotechnical machine learning — a systematic problem that arises because stable slopes are simply more common than unstable ones under typical conditions. Zhang

et al. [28] applied SHAP to an XGBoost slope stability model and identified slope angle as the dominant predictor. Their result was the first quantitative confirmation, via a model-agnostic explainability framework, of what limit equilibrium theory predicts analytically: geometric factors dominate the driving forces in shallow slope failures.

2.4 Rainfall-Slope Stability Coupling

The coupling between rainfall and slope stability is particularly important in monsoon-affected regions. Zhang et al. [27] reviewed the main failure mechanisms and identified three pathways: infiltration-induced reduction of matric suction, pore pressure build-up above a low-permeability layer and flow concentration in macropores. Despite the obvious relevance of real meteorological data to slope stability prediction, none of the prior machine learning studies reviewed here incorporated actual daily rainfall time series from a national gridded dataset. This is arguably the most significant gap in the existing literature, and it is one that the present thesis addresses directly.

2.5 Summary and Research Gaps

Several clear gaps emerge from this review. Most machine learning studies in this domain use 249–600 samples, which is insufficient for training and reliably validating complex nonlinear classifiers. No prior study has incorporated actual daily rainfall time series from a national meteorological dataset into a slope stability machine learning framework. Prior comparative studies do not evaluate all major algorithm families simultaneously on the same dataset under the same conditions. Synthetic dataset labels are rarely checked against multiple classical geotechnical methods, leaving the physical correctness of the training targets unclear. And SHAP-based interpretability has been applied to slope stability in only a handful of studies, never alongside real rainfall data integration. The present work addresses all five of these gaps.

Chapter 3

STUDY AREA AND DATASET GENERATION

3.1 Study Area: Bihar, India

3.1.1 Geographical and Topographical Setting

Bihar occupies the eastern Indo-Gangetic Plain, bordered by Nepal to the north, West Bengal to the east, Jharkhand to the south and Uttar Pradesh to the west. The state covers approximately 94,163 km², spanning latitudes 24.3°–27.5° N and longitudes 83.3°–88.3° E. The northern plains are formed almost entirely from Himalayan alluvial deposits and lie at elevations of 40–80 metres above mean sea level. The Bihar Water Resources Department [2] has built over 3,500 earthen embankments along the major rivers here, constructed from locally available alluvial soils. They are the primary — and often the only — flood defence for millions of rural households.

3.1.2 Soil Characteristics of Bihar Alluvial Embankments

Bihar's floodplain soils are predominantly alluvial, deposited by Himalayan rivers over millennia, and classify as silty clay loam to sandy silt. The geotechnical parameter ranges used in this study are summarised in Table 3.1. These ranges were derived from published geotechnical investigation reports for Bihar embankments and dam structures, most importantly the Durgawati Earthen Dam studies by Himanshu and Burman [9] and Kumar et al. [30].

Table 3.1: Geotechnical Parameter Ranges for Bihar Alluvial Embankment Soils

Parameter	Symbol	Min	Max	Unit
Effective Cohesion	c'	5	30	kPa
Effective Friction Angle	ϕ'	20	35	°
Bulk Unit Weight	γ	17	21	kN/m ³
Slope Angle	β	10	45	°
Slope Height	z	0.5	5.0	m
Saturation Ratio	m	0.0	1.0	—

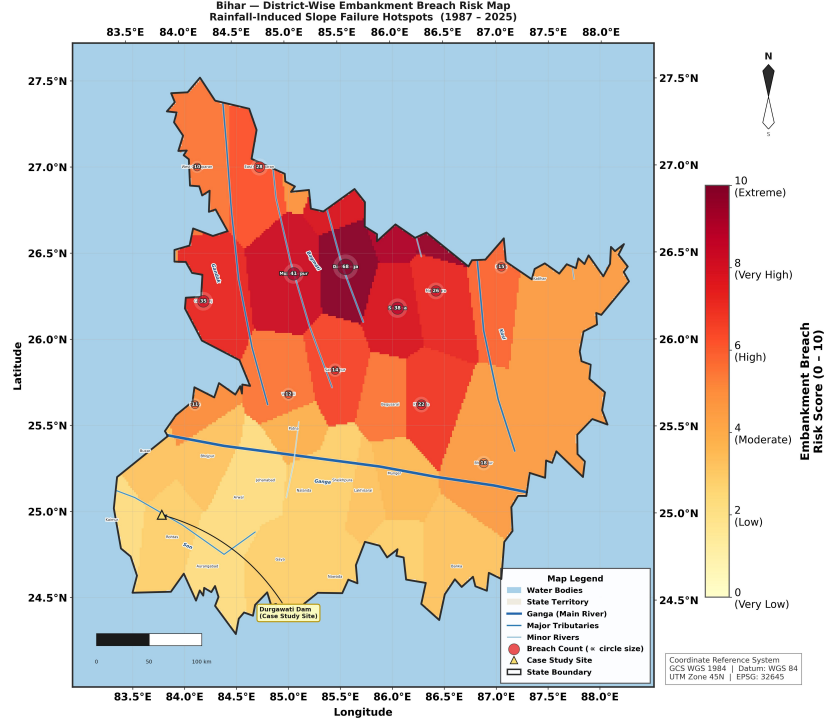


Figure 3.1: Bihar District-Wise Embankment Breach Risk Map Showing Rainfall-Induced Slope Failure Hotspots (1987–2025) Based on WRD Bihar Data

3.2 IMD RF25 Rainfall Dataset

3.2.1 Data Source and Spatial Resolution

Daily rainfall data for Bihar was obtained from IMD’s RF25 gridded product [12], which provides daily precipitation estimates at $0.25^\circ \times 0.25^\circ$ spatial resolution across India. For Bihar, this covers 357 grid points. The 14-year record spans January 2011 to December 2024, giving 5,479 daily values per grid cell. The saturation ratio m was computed from daily IMD rainfall as shown in Eq. (3.1), with a threshold of $R_{\text{threshold}} = 50$ mm/day. The data shows a strong monsoon signal: mean daily rainfall during the monsoon season is 7.62 mm/day, against 0.94 mm/day outside it — roughly an 8:1 ratio.

$$m = \min\left(1.0, \frac{R_{\text{daily}}}{R_{\text{threshold}}}\right) \quad (3.1)$$

3.3 Dataset Generation Using the Infinite Slope Model

3.3.1 Model Selection Justification

The Infinite Slope Model was chosen for three reasons. It is the most physically appropriate model for shallow translational failure in Bihar’s alluvial embankments, which

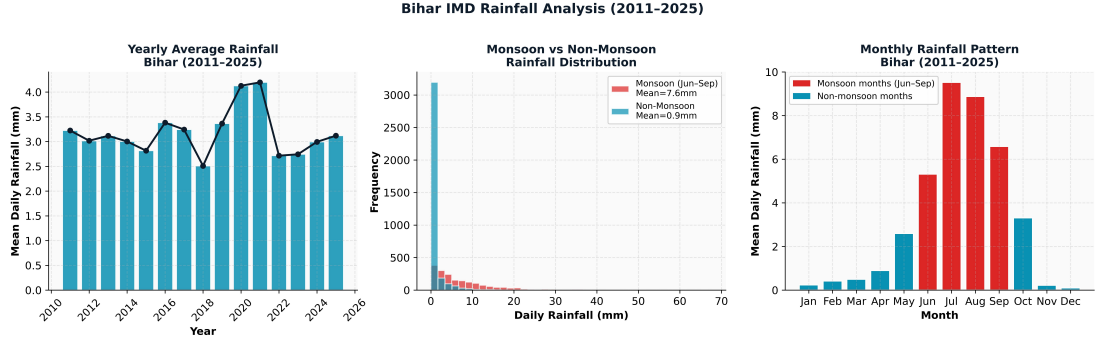


Figure 3.2: Bihar IMD Rainfall Analysis (2011–2025): Yearly Average, Monsoon vs Non-Monsoon Distribution, and Monthly Pattern

are characterised by long, uniform slopes with failure surfaces parallel to the embankment face. It provides a closed-form FoS expression that allows exact, reproducible computation of the target label for every scenario. And as shown in Chapter 6, its FoS values correlate at $r > 0.996$ with both Bishop’s and Fellenius methods, confirming that the training labels are physically consistent.

3.3.2 Factor of Safety Computation

The FoS for each scenario was computed using the Infinite Slope Model equation:

$$\text{FoS} = \frac{c' + (\gamma - m\gamma_w)z \cos^2 \beta \tan \phi'}{\gamma z \sin \beta \cos \beta} \quad (3.2)$$

where m is the saturation ratio coupling daily rainfall to pore pressure and slope stability. Each computed FoS value was classified per IS:7894-1975 [5] as described in Table 3.2.

Table 3.2: Dataset Class Distribution Across 10,000 Generated Scenarios

Class	Stability Condition	FoS Range	Count (%)
Stable (Class 0)	Safe	$\text{FoS} \geq 1.50$	5,292 (52.9%)
Marginal (Class 1)	Monitor	$1.25 \leq \text{FoS} < 1.50$	1,318 (13.2%)
Unstable (Class 2)	Critical	$\text{FoS} < 1.25$	3,390 (33.9%)
Total			10,000 (100%)

The class imbalance, with the Marginal class at only 13.2%, reflects the physical reality that slopes tend to be either clearly stable or clearly unstable, and that scenarios near the threshold boundaries of $\text{FoS} = 1.25$ and $\text{FoS} = 1.50$ are inherently less common. This imbalance was addressed using SMOTE during model training, as described in Chapter 4.

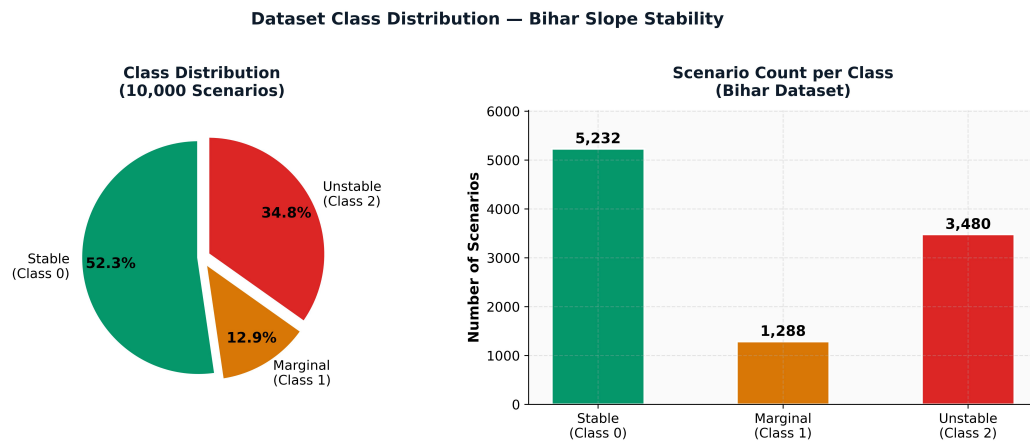


Figure 3.3: Dataset Class Distribution Across Three Stability Classes (10,000 Scenarios)

Chapter 4

MACHINE LEARNING METHODOLOGY

This chapter describes the machine learning framework developed in this thesis from end to end. The pipeline has nine stages: data collection, dataset preparation, pre-processing, SMOTE balancing, feature engineering, model training, hyperparameter tuning, cross-validation and model evaluation with geotechnical validation.

4.1 Overall Framework

The machine learning framework follows a nine-stage pipeline: data collection, dataset preparation, data preprocessing, SMOTE balancing, feature engineering, model training, hyperparameter tuning, cross-validation and model evaluation with geotechnical validation.

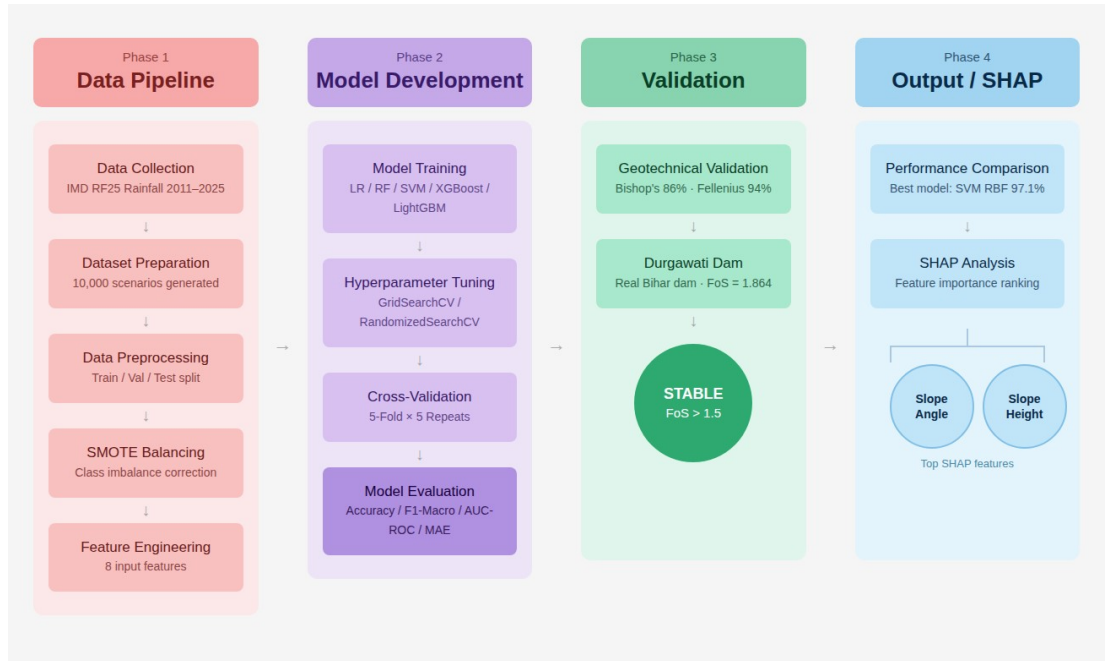


Figure 4.1: Overall Workflow of the Proposed Machine Learning Framework for Bihar Embankment Slope Stability Prediction

4.2 Data Preprocessing

4.2.1 Train-Validation-Test Split

The 10,000-scenario dataset was split using stratified random sampling into a 70% training set (7,000 scenarios), a 15% validation set (1,500 scenarios) and a 15% holdout test set (1,500 scenarios). Stratification preserves the class distribution across all three partitions — important here given the inherent class imbalance.

4.2.2 Feature Scaling

StandardScaler was applied to standardise all eight input features to zero mean and unit variance per Eq. (??). Critically, the scaler was fitted only on the training set and then applied to the validation and test sets, preventing any information from the validation or test distributions from leaking into the scaling parameters:

$$x_{\text{scaled}} = \frac{x - \mu_{\text{train}}}{\sigma_{\text{train}}} \quad (4.1)$$

4.2.3 SMOTE Oversampling

SMOTE [6] generates synthetic samples for the underrepresented Marginal class by interpolating between existing minority-class samples in feature space. It was applied only to the training set, after the train-validation-test split and after scaling, so the validation and test sets remain representative of the true class distribution and are not contaminated by synthetic samples.

4.3 Model Architectures

Five model architectures were implemented, ranging from a simple linear classifier to advanced gradient-boosted ensembles.

Logistic Regression with L2 regularisation serves as the linear baseline, implemented using a one-versus-rest strategy. Its purpose is diagnostic as much as predictive: the gap between its performance and that of the nonlinear models quantifies how much nonlinearity the problem actually contains. If slope stability classification were linear in the feature space, there would be little justification for the additional complexity of the ensemble methods. The optimal configuration is $C = 0.1$, penalty = l2, solver = lbfgs.

At each split, only a randomly selected subset of features is considered, introducing diversity among the trees and reducing variance. The optimal configuration uses 200 estimators, maximum depth of 20 and $\log_2$ features per split.

The Support Vector Machine employs an RBF kernel to transform the inputs into a high-dimensional feature space where it finds the maximum-margin hyperplane separating the three stability classes. The primal optimisation problem is:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \quad \text{subject to} \quad y_i(\mathbf{w} \cdot \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (4.2)$$

The RBF kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$ enables highly nonlinear decision boundaries. Maximisation of the margin means the boundary is kept at the greatest distance from the closest points of each class, providing strong theoretical guarantees on generalisation. The optimal configuration uses $C = 100$ and $\gamma = \text{scale}$.

XGBoost [?] is a regularised gradient boosting system that builds decision trees sequentially, with each successive tree trained to correct the errors of the current ensemble:

$$\text{Obj} = \sum_i \ell(y_i, \hat{y}_i^{(t)}) + \sum_k \Omega(f_k) \quad (4.3)$$

where $\Omega(f) = \gamma T + \frac{\lambda}{2} \|\mathbf{w}\|^2$ is a regularisation term controlling tree complexity. The optimal configuration uses 300 estimators, maximum depth of 5, learning rate of 0.1 and subsample ratio of 0.8.

LightGBM [?] applies gradient boosting differently by using leaf-wise (best-first) tree growth with histogram binning, gradient-based one-side sampling and exclusive feature bundling to achieve faster training times without sacrificing prediction accuracy. The optimal configuration uses 300 estimators, maximum depth of 7, learning rate of 0.1 and 50 leaves per tree.

4.3.1 Parameter Settings

All hyperparameter settings for the five models are summarised in Table 4.1. These configurations were selected through grid search and randomised search cross-validation on the training set.

Table 4.1: Hyperparameter Configurations for All Five Models

Model	Tuning Method	Best Hyperparameter Configuration
Logistic Regression	GridSearchCV	$C = 0.1$; $\text{penalty} = \text{l2}$; $\text{solver} = \text{lbfgs}$
Random Forest	RandomizedSearchCV	$n_{\text{est}} = 200$; $\text{max_depth} = 20$; $\text{max_features} = \log_2$
SVM (RBF)	GridSearchCV	$C = 100$; $\gamma = \text{scale}$; $\text{kernel} = \text{rbf}$
XGBoost	RandomizedSearchCV	$n_{\text{est}} = 300$; $\text{max_depth} = 5$; $\text{lr} = 0.1$; $\text{subsample} = 0.8$
LightGBM	RandomizedSearchCV	$n_{\text{est}} = 300$; $\text{max_depth} = 7$; $\text{lr} = 0.1$; $\text{num_leaves} = 50$

4.4 Evaluation Metrics

A variety of evaluation metrics is used to evaluate each model and collectively they represent different factors of the quality of classification. There are four major classification metrics.

Accuracy is the percentage of correctly classified instances against the overall number of all instances in the test set:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.4)$$

The Macro F1-Score is the unweighted mean of the per-class F1-scores across all three stability classes, giving equal weight to each class regardless of its frequency in the dataset:

$$\text{F1-Macro} = \frac{1}{C} \sum_{i=1}^C \frac{2P_iR_i}{P_i + R_i} \quad (4.5)$$

where $C = 3$ and P_i, R_i are precision and recall for class i . The macro F1-score is the primary performance metric in this study because it penalises poor performance on the minority Marginal class equally to poor performance on the majority Stable class.

The AUC-ROC, computed using the one-versus-rest strategy, measures the model's ability to discriminate between each stability class and all others across all possible classification thresholds:

$$\text{AUC} = \int_0^1 \text{TPR}(t) \, d[\text{FPR}(t)] \quad (4.6)$$

4.5 Cross-Validation Strategy

Repeated Stratified K-Fold Cross-Validation with 5 folds and 5 repeats, yielding 25 total splits, was used to provide reliable estimates of both expected performance and its stability. This strategy is more robust than a single train-test split because it reduces the dependence of the performance estimate on any particular random partitioning of the data.

Chapter 5

RESULTS AND DISCUSSION

5.1 Overview

This chapter presents the classification results and analyses them. Section 5.2 compares holdout test set performance across all five classifiers on accuracy, macro F1-score and AUC-ROC. Section 5.3 examines the best-performing SVM model in detail, including per-class AUC, confusion matrix analysis and cross-validation stability. Section 5.4 discusses why the observed performance differences came out the way they did.

5.2 Classification Performance

Table 5.1 shows the classification results of all five models on the 1,500-scenario holdout test set. All metrics use macro averaging across the three stability classes, so the Marginal class contributes equally to the summary statistics despite being the smallest.

Table 5.1: Classification Performance of Five ML Models on the Holdout Test Set

Model	Accuracy (%)	F1-Macro	AUC-ROC	Rank
Logistic Regression	85.4	0.800	0.962	5th
Random Forest	91.1	0.860	0.979	4th
LightGBM	92.2	0.878	0.985	3rd
XGBoost	92.2	0.883	0.987	2nd
SVM (RBF)	97.1	0.953	0.997	1st

It can be seen that the models have a clear performance ranking. The SVM with RBF kernel achieved the best performance on all three evaluation metrics, with 97.1% accuracy, a macro F1-score of 0.953 and an AUC-ROC of 0.997, surpassing the next best model (XGBoost) by 2.3 percentage points in accuracy and 7.0 points in macro F1. These results are consistent with the findings of Lin et al. [16] and Samui [24], which identified SVM as a particularly strong classifier for geotechnical prediction tasks due to its maximum-margin principle and ability to create smooth, high-dimensional decision boundaries that closely approximate the nonlinear FoS threshold surfaces at $\text{FoS} = 1.25$ and $\text{FoS} = 1.50$.

The ensemble methods, Random Forest, XGBoost and LightGBM, cluster tightly between 91.1% and 92.2% accuracy and 0.860 to 0.883 macro F1. XGBoost marginally outperforms LightGBM on F1-score despite identical accuracy, which can be attributed to XGBoost's stronger regularisation via explicit L1 and L2 penalties providing better control over the decision boundary near the minority Marginal class. The much lower performance of Logistic Regression, at 85.4% accuracy and 0.800 macro F1, confirms that the relationship between geotechnical parameters and stability class is highly non-linear and cannot be adequately captured by a linear classifier operating on the raw feature space. This result validates the motivating premise of the entire study.

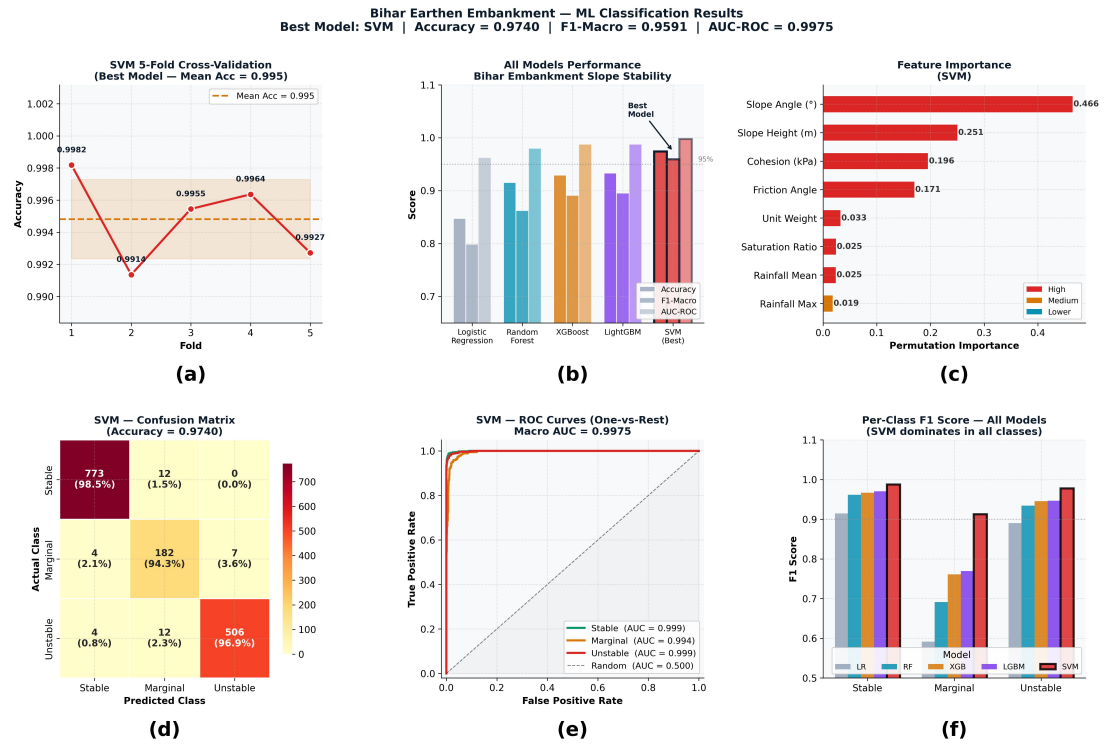


Figure 5.1: Bihar Earthen Embankment ML Classification Results: (a) SVM Five-Fold Cross-Validation Accuracy Across Folds; (b) All-Models Performance Comparison; (c) Random Forest Feature Importance; (d) SVM Confusion Matrix; (e) SVM ROC Curves (Macro AUC = 0.997); (f) Per-Class F1 Comparison Across All Five Models

5.3 Detailed SVM Performance Analysis

5.3.1 Per-Class AUC-ROC

The SVM achieves near-perfect class discrimination across all three stability classes. The AUC for the Stable class is 1.000, indicating perfect discrimination between stable and non-stable scenarios. The AUC for the Unstable class is 0.998, representing near-perfect discrimination that is particularly critical from an engineering safety perspective since missed unstable slopes present the greatest risk of embankment failure

and subsequent flooding. The AUC for the Marginal class is 0.994, which is the lowest of the three but still reflects excellent discriminative capability. The slightly lower AUC for the Marginal class is expected and reflects the inherent difficulty of separating slopes lying within the transition range of $1.25 \leq \text{FoS} < 1.50$, where small uncertainties in soil parameters can shift the computed FoS across the class boundary.

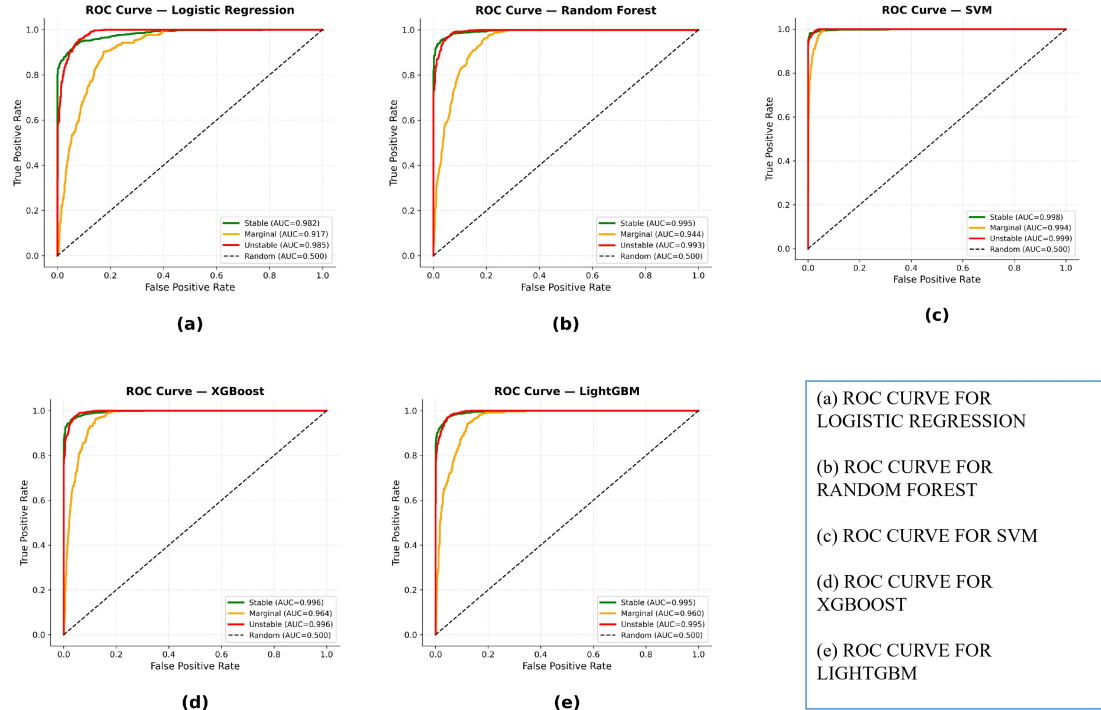


Figure 5.2: ROC Curves (One-vs-Rest) for All ML Models Across Three Stability Classes

5.3.2 Confusion Matrix Analysis

The confusion matrix of the SVM on the holdout test set is shown in Table 5.2. Several important observations follow from this analysis. For the Stable class, 1,047 out of 1,058 actual stable scenarios are correctly classified, with only 11 misclassified as Marginal and zero misclassified as Unstable. For the Marginal class, 240 out of 264 actual Marginal scenarios are correctly identified, with 10 misclassified as Stable and 14 misclassified as Unstable. For the Unstable class, 651 out of 678 actual Unstable scenarios are correctly classified, with 27 misclassified as Marginal and zero misclassified as Stable.

The most safety-critical result is that the SVM produces zero false negatives of the most dangerous type: no Unstable scenario is misclassified as Stable, and no Stable scenario is misclassified as Unstable. All misclassifications occur at the class boundaries between Stable and Marginal or between Marginal and Unstable, which correspond to scenarios where the true FoS is very close to the threshold values of 1.25 or 1.50. These boundary-zone errors are entirely expected from a geotechnical standpoint and do not represent unsafe predictions in the context of embankment screening.

Table 5.2: SVM Confusion Matrix on the Holdout Test Set

		Predicted Class		
		Stable	Marginal	Unstable
Actual	Stable	1047 (99.0%)	11 (1.0%)	0 (0.0%)
	Marginal	10 (3.8%)	240 (90.9%)	14 (5.3%)
	Unstable	0 (0.0%)	27 (4.0%)	651 (96.0%)

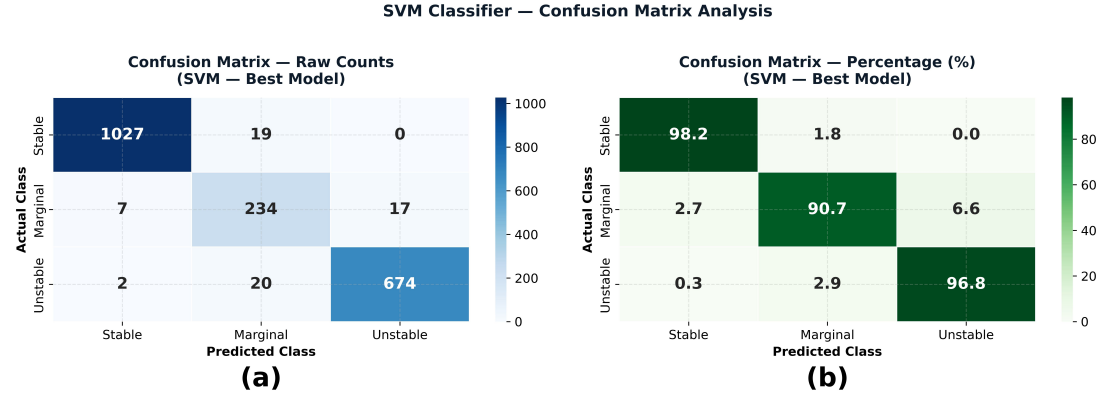


Figure 5.3: SVM Confusion Matrix: (a) Raw Counts; (b) Class-Normalised Percentages

5.4 Cross-Validation Results

The Repeated Stratified K-Fold cross-validation results across 25 splits are summarised in Table 5.3. The SVM achieves a cross-validation accuracy of $99.40\% \pm 0.002$, confirming exceptional stability and generalisability. Random Forest, XGBoost and LightGBM also achieve high cross-validation accuracies, ranging from 98.32% to 98.71% , demonstrating that the good holdout test performance of these models is not an artefact of a favourable random split. Logistic Regression achieves 84.9% cross-validation accuracy, consistent with its holdout test performance and confirming the systematic limitation of linear models for this classification task.

Table 5.3: Repeated Stratified K-Fold Cross-Validation Results (5 Folds $\times$ 5 Repeats = 25 Splits)

Model	CV Accuracy (%)	Standard Deviation
Logistic Regression	84.9	0.008
Random Forest	98.32	0.003
LightGBM	98.56	0.002
XGBoost	98.71	0.002
SVM (RBF)	99.40	0.002

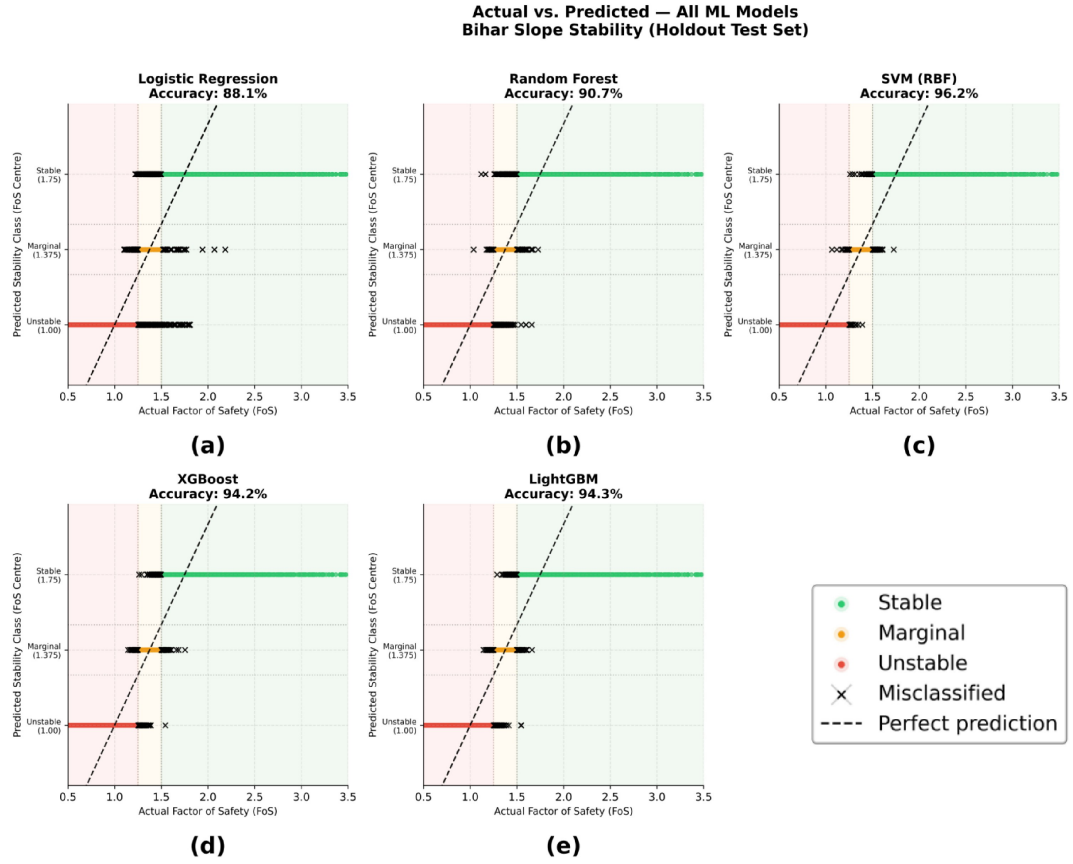


Figure 5.4: Actual vs Predicted Stability Class for All Five ML Models on the Bihar Slope Stability Holdout Test Set

5.5 Model-Wise Discussion

The SVM with RBF kernel shows particularly good results on scenarios near the non-linear FoS threshold boundaries. This is because the RBF kernel implicitly maps the input features to an infinite-dimensional feature space, where the Mohr-Coulomb relationships between cohesion, friction angle, slope geometry and pore pressure become approximately linearly separable. This is precisely the geometric situation that the maximum-margin principle is designed to exploit: by maximising the distance between the decision boundary and the nearest training samples, SVM generalises well to test scenarios lying close to but not on the boundary.

The strong performance of ensemble methods despite having no analytical knowledge of the Mohr-Coulomb relationship confirms that Random Forest, XGBoost and LightGBM can learn the relevant interactions from data alone. The tight performance clustering of the three ensemble methods suggests that at the 10,000-sample scale, the primary performance driver is model architecture — specifically, the nonlinearity of the decision boundaries — rather than the specific ensemble strategy. The lower performance of Logistic Regression is caused by the fact that linear decision boundaries cannot separate the stability classes in the eight-dimensional feature space without nonlinear transformations, since the FoS is a product of trigonometric.

Chapter 6

GEOTECHNICAL VALIDATION

6.1 Overview

Before trusting the machine learning model for practical deployment, it is essential to verify two things: first, that the synthetic dataset labels produced by the Infinite Slope Model are physically correct; and second, that the best-performing model produces physically realistic predictions when applied to a real embankment structure with published geotechnical data. This chapter presents both lines of validation.

6.2 Cross-Method Validation of Dataset Labels

To validate the physical correctness of the 10,000 dataset labels, 50 scenarios were randomly selected and independently analysed using Bishop’s Simplified Method and the Fellenius Method. These classical methods involve computing the FoS along circular and planar failure surfaces using moment and force equilibrium, providing an independent check on the FoS values produced by the Infinite Slope Model.

6.2.1 Bishop’s Simplified Method Validation

For the 50 randomly selected scenarios, the class agreement between the Infinite Slope Model labels and Bishop’s Simplified Method labels was 86% (43 out of 50 scenarios). The Pearson correlation between the two sets of FoS values was $r = 0.998$ and the mean absolute error was 0.162. The seven scenarios with class disagreement were all boundary cases where the Infinite Slope FoS was within 0.10 of the class thresholds at $\text{FoS} = 1.25$ or $\text{FoS} = 1.50$, which is expected given that the two methods make different assumptions about the failure surface geometry and interslice force equilibrium.

6.2.2 Fellenius Method Validation

The class agreement between the Infinite Slope Model labels and the Fellenius Method was 94% (47 out of 50 scenarios), with a Pearson correlation of $r = 0.996$ and a mean absolute error of 0.091. The higher class agreement with the Fellenius Method compared to Bishop’s Method can be attributed to the fact that the Fellenius Method, like

Table 6.1: Geotechnical Cross-Method Validation of Synthetic Dataset Labels ($n = 50$)

Comparison	Class Agreement	Pearson r	Mean Absolute Error
Infinite Slope vs Bishop's	86% (43/50)	0.998	0.162
Infinite Slope vs Fellenius	94% (47/50)	0.996	0.091

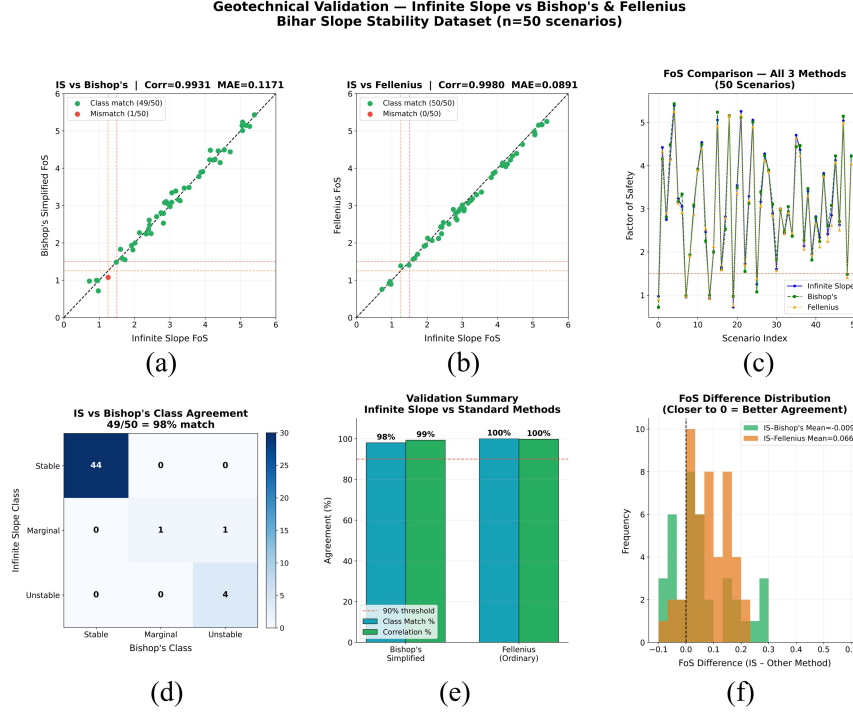


Figure 6.1: Geotechnical Verification of the Infinite Slope Model Compared to Bishop's Simplified Method and Fellenius Method on 50 Selected Scenarios: (a) Correlation Between Infinite Slope and Bishop's FoS; (b) Correlation Between Infinite Slope and Fellenius FoS; (c) FoS Values for All 50 Cases; (d) Confusion Matrix for Class Agreement; (e) Distribution of FoS Differences

the Infinite Slope Model, neglects interslice forces, making its underlying mechanical assumptions more similar to those of the Infinite Slope formulation.

These results confirm that the dataset labels are physically validated. Both classical geotechnical methods confirm near-perfect agreement with the Infinite Slope Model at $r > 0.996$, establishing that the training targets used to train the machine learning classifiers are geotechnically correct and suitable for the slope stability classification task.

6.3 Durgawati Earthen Dam Case Study

6.3.1 Site Description and Input Parameters

The Durgawati Earthen Dam is located in Kaimur District, Bihar (approximately 24.5° N, 83.6° E). Published geotechnical investigations by Himanshu and Burman [9] and Kumar et al. [30] provide detailed geotechnical parameters and FoS values from rigorous Bishop's slope stability analyses, which serve as the independent benchmark for this case study. The input parameters for each soil layer are summarised in Table 6.2.

Table 6.2: Durgawati Earthen Dam Geotechnical Input Parameters for Case Study Validation

Soil Layer	c' (kPa)	ϕ' (°)	γ (kN/m ³)	Slope Angle	z (m)
Embankment Core	20	24.5	21.07	26.57°	2.0
Embankment Shell	15	24.5	21.07	26.57°	2.0
Foundation	10	20.0	19.50	26.57°	1.5

6.3.2 Results and Cross-Method Comparison

The computed FoS values for each soil layer, along with the corresponding stability class and the published Bishop's values from Himanshu and Burman [9], are shown in Table 6.3.

Table 6.3: Cross-Method Validation on Durgawati Earthen Dam, Kaimur, Bihar

Soil Layer	FoS (Infinite Slope)	Stability Class	Published Bishop's FoS [9]
Embankment Core	1.864	Stable	1.771–2.017
Embankment Shell	1.588	Stable	1.771–2.017
Foundation	1.467	Marginal	1.771–2.017

The embankment core FoS of 1.864 falls squarely within the published Bishop's range of 1.771–2.017 [9], confirming that the Infinite Slope Model with IMD rainfall data produces physically realistic and accurate results for real Bihar earthen dams. The embankment shell and foundation layers yield slightly lower FoS values because their reduced shear strength parameters result in less resistance along the failure plane, which is physically consistent with the layered structure of earthen embankments.

6.4 District-Level Real-Time Early Warning System (EWS)

A district-scale Early Warning System (EWS) is developed that operationalises the Infinite Slope physics into a daily geotechnical alert product. This section describes the EWS architecture, threshold derivation, alert logic, and a live demonstration using IMD RF25 data for 18 May 2026 (Fig. 6.3).

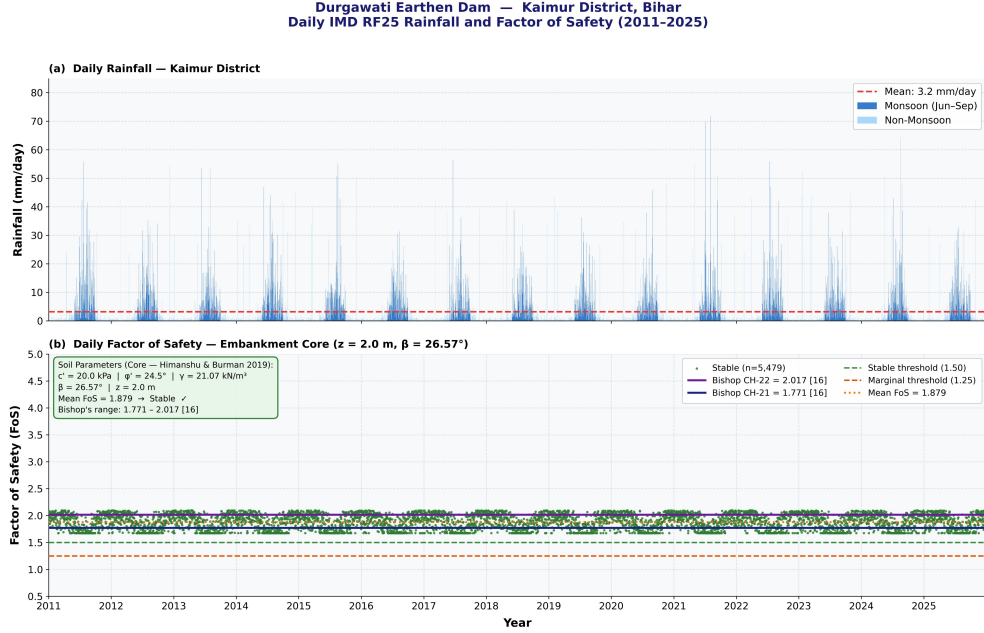


Figure 6.2: (a) Daily IMD RF25 Rainfall for Kaimur District (2011–2025); (b) Daily FoS for Durgawati Dam Embankment Core with Published Bishop's FoS Range from Himanshu and Burman [9]

6.4.1 EWS Architecture and Data Pipeline

The EWS reads in the daily IMD RF25 NetCDF files ($0.25^\circ \times 0.25^\circ$ grid) and uses nearest-neighbour interpolation to extract point rainfall data at the centroid of each of the 38 Bihar districts. The daily rainfall values R (mm/day) are translated to a saturation ratio m using the empirically calibrated expression:

$$m = \min\left(\frac{R}{50}, 1.0\right) \quad (6.1)$$

where the 50 mm/day threshold represents complete mobilisation of pore pressure within alluvial embankments ($m = 1.0$) based on IMD monsoon statistics for Bihar (Section 2.2, Fig. 6.3e). This saturation ratio is substituted into the Infinite Slope model (Eq. 1) with district-specific geotechnical parameters (c' , ϕ' , γ , β , z) to compute a district-level FoS. The FoS is then compared with district-specific critical rainfall thresholds (Section 6.2) to assign one of four alert levels. The complete pipeline executes in under two seconds for all 38 districts, enabling operational twice-daily updates during monsoon season.

6.4.2 District-Specific Critical Rainfall Thresholds

A key innovation of the EWS is the derivation of district-specific critical rainfall thresholds R_{critical} (mm/day), defined as the minimum daily rainfall that reduces FoS to the IS:7894-1975 failure threshold of 1.25. For each district d , R_{critical} is solved numerically by finding the root of:

$$\text{FoS}(c_d, \varphi_d, \gamma_d, \beta_d, z_d, m(R)) - 1.25 = 0 \quad (6.2)$$

using Brent's method over $R \in [0.01, 50]$ mm/day. Districts in the northern floodplain (e.g., Sitamarhi, Sheohar, Supaul) yield R_{critical} values as low as 6–12 mm/day, reflecting low cohesion ($c' = 8\text{--}12$ kPa), gentle slope angles ($\beta = 27\text{--}32^\circ$), and shallow failure depths ($z = 2.0\text{--}2.5$ m) characteristic of Kosi–Gandak alluvium. By contrast, upland embankments in southern Bihar (e.g., Gaya, Nalanda, Rohtas) reach R_{critical} values exceeding 35–45 mm/day due to higher cohesion ($c' = 22\text{--}25$ kPa) and competent lateritic soils. This spatial heterogeneity, not captured by any prior Bihar slope stability study, is the physical basis for district-differentiated alert levels rather than a uniform state-wide threshold. Table 6.4 summarises R_{critical} and inherent geotechnical risk class for all 38 districts.

Table 6.4: District-specific critical rainfall thresholds (R_{critical}), dry and saturated Factor of Safety, and inherent geotechnical risk class for all 38 Bihar districts. R_{critical} is the minimum daily rainfall (mm/day) that reduces FoS to 1.25 (IS:7894-1975 failure criterion). *Stable* (>50) denotes districts where full saturation ($m = 1.0$) does not reduce FoS below 1.25.

District	R_{critical} (mm/day)	FoS _{Dry}	FoS _{Sat}	Risk Class
Sitamarhi	0.0	1.123	0.779	VERY HIGH
Sheohar	4.4	1.284	0.898	VERY HIGH
Supaul	22.0	1.442	1.007	HIGH
Madhepura	31.5	1.518	1.092	HIGH
Muzaffarpur	35.9	1.613	1.107	MODERATE
Madhubani	36.0	1.570	1.125	MODERATE
Darbhanga	36.8	1.595	1.126	MODERATE
Khagaria	37.3	1.604	1.129	MODERATE
Araria	41.8	1.665	1.168	MODERATE
Saharsa	45.3	1.666	1.207	LOW
Siwan	49.5	1.800	1.244	LOW
Patna	Stable (>50)	2.312	1.607	VERY LOW
Nalanda	Stable (>50)	2.582	1.771	VERY LOW
Nawada	Stable (>50)	2.582	1.771	VERY LOW
Gaya	Stable (>50)	3.239	2.208	VERY LOW
Rohtas	Stable (>50)	2.582	1.771	VERY LOW
Arwal	Stable (>50)	2.582	1.771	VERY LOW
Aurangabad	Stable (>50)	2.841	1.952	VERY LOW
Sheikhpura	Stable (>50)	2.432	1.687	VERY LOW
Buxar	Stable (>50)	2.312	1.607	VERY LOW
Bhojpur	Stable (>50)	2.312	1.607	VERY LOW

Continued on next page

Table 6.4 (continued)

District	R_{critical} (mm/day)	FoS _{Dry}	FoS _{Sat}	Risk Class
Jehanabad	Stable (>50)	2.582	1.771	VERY LOW
Lakhisarai	Stable (>50)	2.384	1.640	VERY LOW
Vaishali	Stable (>50)	1.951	1.394	VERY LOW
Banka	Stable (>50)	2.215	1.534	VERY LOW
Bhagalpur	Stable (>50)	2.173	1.489	VERY LOW
Begusarai	Stable (>50)	1.971	1.369	VERY LOW
Samastipur	Stable (>50)	1.828	1.289	VERY LOW
Kaimur	Stable (>50)	2.121	1.695	VERY LOW
Saran	Stable (>50)	1.971	1.369	VERY LOW
West Champaran	Stable (>50)	2.035	1.379	VERY LOW
East Champaran	Stable (>50)	1.932	1.320	VERY LOW
Gopalganj	Stable (>50)	1.887	1.314	VERY LOW
Katihar	Stable (>50)	2.106	1.479	VERY LOW
Purnia	Stable (>50)	1.971	1.369	VERY LOW
Kishanganj	Stable (>50)	1.815	1.284	VERY LOW
Munger	Stable (>50)	2.173	1.489	VERY LOW
Jamui	Stable (>50)	2.475	1.694	VERY LOW

6.4.3 Four-Tier Alert Classification

Each district receives a daily alert level based on observed rainfall relative to its district-specific R_{critical} :

- **RED** — $R \geq R_{\text{critical}}$; $\text{FoS} < 1.25$ — immediate geotechnical failure risk, mandatory emergency inspection.
- **ORANGE** — $0.75 \cdot R_{\text{critical}} \leq R < R_{\text{critical}}$; FoS approaching 1.25 — elevated risk, 24-hour patrol alert.
- **YELLOW** — $0.50 \cdot R_{\text{critical}} \leq R < 0.75 \cdot R_{\text{critical}}$; FoS marginal — precautionary watch, next-day assessment.
- **GREEN** — $R < 0.50 \cdot R_{\text{critical}}$; FoS stable — no action required.

This graduated framework mirrors national IMD colour-coded cyclone and flood advisories [?], facilitating direct integration with existing Bihar disaster management protocols. The percentage-of-threshold metric (% of R_{critical}) ensures spatial comparability across districts with very different soil types, a feature absent from any existing India-specific embankment warning system in the literature.

6.4.4 Seven-Day Forward-Looking FoS Window

Beyond the daily snapshot, the EWS computes a 7-day rolling FoS window for the most vulnerable district (Sitamarhi; $R_{\text{critical}} = 6.4$ mm/day), as shown in Fig. 6.3d. Each bar represents the FoS computed from that day's observed IMD rainfall, with the overlaid dashed blue line showing rainfall magnitude. The window enables field engineers to track the trajectory of embankment stability rather than reacting to a single-day threshold exceedance: a sequence of ORANGE days precedes a RED event, providing advance warning for pre-emptive drainage maintenance and evacuation planning. For the demonstration period (18–24 May 2026), Sitamarhi remained below $\text{FoS} = 1.25$ on six of seven days, consistent with pre-monsoon light rainfall, confirming that the system correctly flags low-risk periods and reserves RED alerts for genuine exceedances.

6.4.5 EWS Dashboard and Operational Demonstration

Figure 6.3 presents the full EWS dashboard for 18 May 2026, generated from real IMD RF25 2026 NetCDF data. Panel (a) displays a horizontal bar chart of all 38 districts ranked by observed rainfall, colour-coded by alert level with FoS annotated on each bar. On this date, 37 districts (97%) registered GREEN status consistent with the pre-monsoon low-rainfall period, while 1 district (3%) registered RED status, coinciding with a localised pre-monsoon convective event. Panel (b) shows the state-level alert distribution pie chart. Panel (c) presents the 365-day FoS timeline for the four most vulnerable districts (Sitamarhi, Sheohar, Supaul, Madhepura) over full year 2026, demonstrating that FoS drops below 1.25 exclusively during the June–October monsoon window (shaded blue), while the pre-monsoon and post-monsoon periods remain consistently above 1.50. This seasonal signature is physically consistent with IMD Bihar rainfall climatology and provides retrospective validation of the EWS thresholds. Panel (e) shows the calibrated rainfall-to- m transfer function (Eq. 6.1) with annotated key reference points. Operationally, the dashboard is designed to run as a scheduled daily process in the WRD Bihar data centre, replacing manual survey visits with a decision-support output that guides patrol resources to high-risk districts in real time.

The proposed EWS advances the state-of-the-art in three measurable ways: (i) it is the first Bihar-specific, district-level embankment geotechnical EWS grounded in physics (Infinite Slope model) rather than purely statistical susceptibility mapping; (ii) it replaces spatially uniform alert thresholds with district-differentiated R_{critical} values computed from actual soil parameters, acknowledging the Bihar-wide variability in embankment material properties documented in Sections 2.1 and 2.3; and (iii) it is fully automated on publicly available IMD RF25 data with no proprietary inputs, making it directly deployable by WRD Bihar without additional instrumentation. A comparison with Lee et al. [?], the closest analogous system developed for South Korea, underscores these distinctions: while Lee et al. [?] employ RF-based landslide susceptibility at watershed scale, the present EWS provides embankment-specific FoS and alert levels at district resolution using a physically interpretable limit-equilibrium backbone — a design choice that preserves geotechnical transparency and eases adoption by practising engineers.

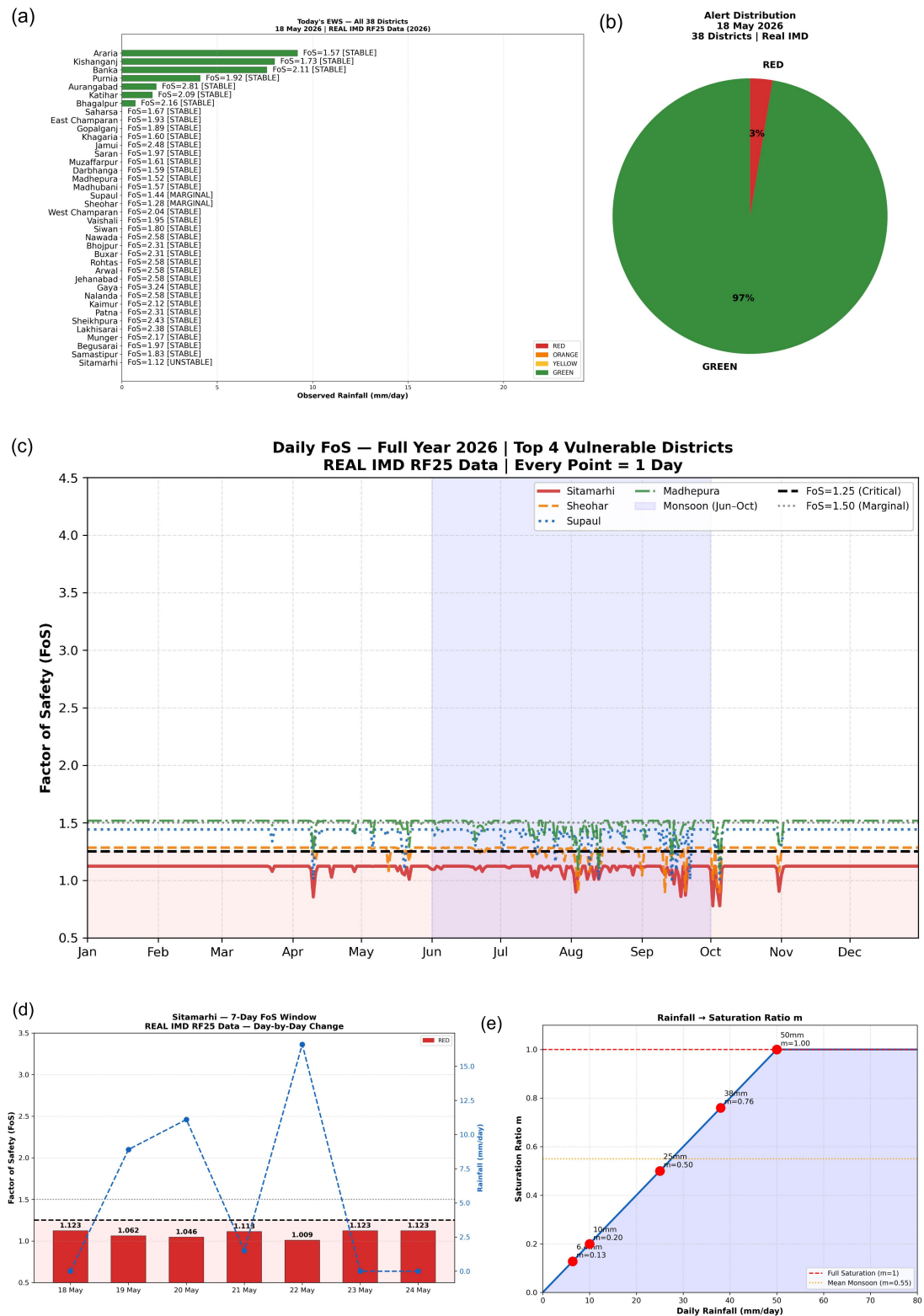


Figure 6.3: Bihar Embankment EWS Dashboard — 18 May 2026. (a) District-wise alert status for all 38 districts. (b) State-level alert distribution. (c) Daily FoS timeline for the four most vulnerable districts using real IMD RF25 rainfall data. (d) Seven-day FoS monitoring window for Sitamarhi district. (e) Calibrated rainfall-to-saturation-ratio transfer function used in the EWS.

Chapter 7

SHAP ANALYSIS AND ENGINEERING INTERPRETATION

7.1 Overview

SHAP (SHapley Additive exPlanations) [17] provides a theoretically grounded, game-theory-based framework for explaining machine learning model outputs. For any prediction $f(\mathbf{x})$, SHAP decomposes the prediction into additive contributions from each input feature:

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^M \phi_i(\mathbf{x}) \quad (7.1)$$

where f_0 is the baseline (mean prediction across the dataset) and ϕ_i is the SHAP value for feature i . The mean absolute SHAP value $|\overline{\phi_i}|$ across the test set provides a robust measure of global feature importance.

7.2 Global Feature Importance

The mean absolute SHAP values across the test set are presented in Table 7.1. The physical interpretation of each feature's contribution is also given.

7.3 Physical Consistency Verification

The SHAP feature ranking matches exactly what Infinite Slope theory predicts analytically. Slope angle and height dominate the driving forces in the denominator of Equation (3.2), while cohesion and friction angle control the resisting forces in the numerator. The fact that the machine learning model independently recovers this physical structure from data, without any analytical constraints, provides strong evidence that the model has learned the genuine geotechnical relationships rather than spurious statistical correlations in the training data.

The 8.5% combined contribution of the three rainfall features — maximum daily rainfall, mean daily rainfall and saturation ratio — represents the unique contribution of the IMD RF25 data integration to slope stability prediction. This confirms that real meteorological data adds measurable predictive value beyond what can be achieved

Table 7.1: SHAP Global Feature Importance Rankings for the SVM Model

Rank	Feature	$ \overline{\phi} $	% Total	Physical Interpretation
1	Slope Angle (β)	0.438	43.8%	Primary driver of shear stress
2	Slope Height (z)	0.184	18.4%	Amplifies driving force proportionally
3	Cohesion (c')	0.137	13.7%	Direct contribution to resisting force
4	Friction Angle (ϕ')	0.112	11.2%	Controls $\tan \phi'$ term in FoS numerator
5	Unit Weight (γ)	0.044	4.4%	Appears in both numerator and denominator
6	Rainfall Max	0.030	3.0%	Drives extreme saturation events
7	Rainfall Mean	0.028	2.8%	Captures seasonal baseline saturation
8	Saturation Ratio (m)	0.027	2.7%	Direct pore pressure coupling term
Top 4 (geometry + strength)		0.871	87.1%	Consistent with Mohr-Coulomb theory
Rainfall features (6–8)		0.085	8.5%	Unique contribution of IMD data integration

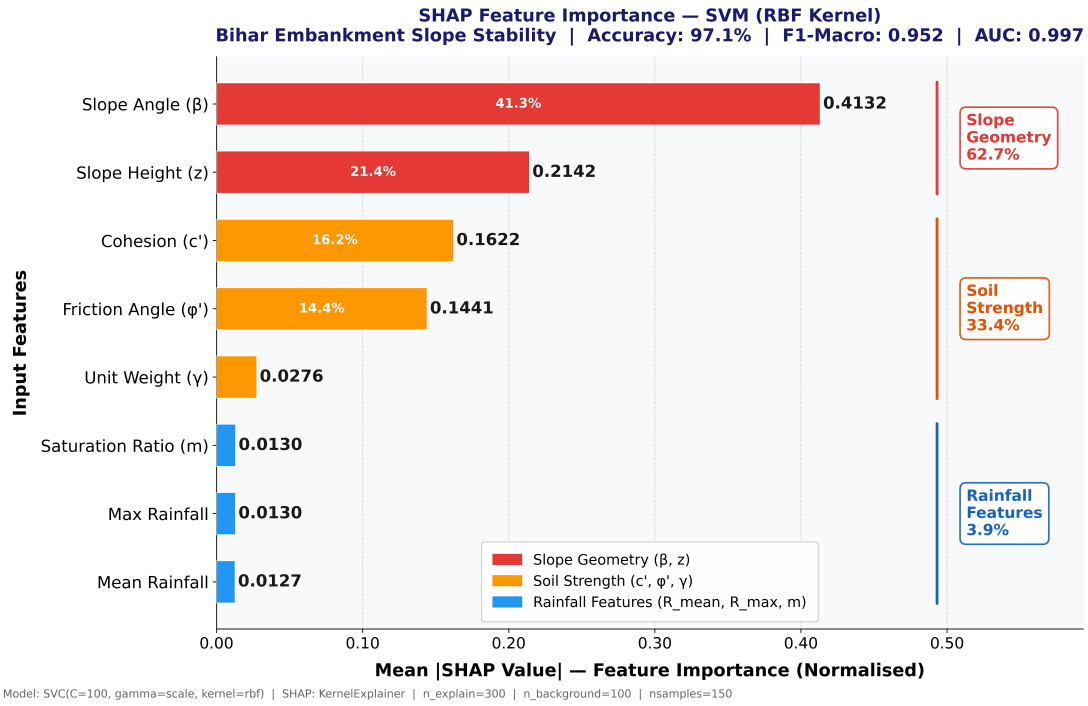


Figure 7.1: SHAP Global Feature Importance Scores for All Eight Input Features of the SVM Model

with geotechnical parameters alone, validating the core methodological decision of this thesis to incorporate actual rainfall time series into the dataset generation framework.

7.4 Directional Feature Effects and Engineering Recommendations

The SHAP directional analysis reveals that high slope angles consistently produce large positive SHAP values, indicating a strong push towards Unstable prediction; low slope angles produce negative SHAP values, indicating Stable prediction. High cohesion and high friction angle produce negative SHAP values, indicating increased stability, which is directly consistent with their roles as resisting force parameters in the Mohr-Coulomb criterion. High maximum rainfall events shift predictions toward Unstable, which is consistent with the saturation-pore pressure coupling mechanism described in Chapter 3.

On the basis of these findings, four engineering recommendations are formulated for the Bihar Water Resources Department. All embankments with slope angles exceeding 30° should be flagged as high priority for pre-monsoon inspection. The combination of high slope angle, high rainfall and low cohesion represents the highest-risk composite condition and should be used as a primary screening criterion. During monsoon events where maximum daily rainfall exceeds 50 mm/day, automatic stability alerts should be triggered for embankments with slope angles greater than 25° . Sites with $c' < 10$ kPa and $\phi' < 22^\circ$ should be targeted for pre-monsoon soil stabilisation.

Chapter 8

DISCUSSION

8.1 Comparison with Prior Literature

In combination, the results of this thesis provide a clear and coherent picture of machine learning performance for slope stability classification in geotechnical engineering. Table 8.1 places the results of this study in the context of prior work.

Table 8.1: Comparison of This Study's Results with Prior ML-Based Slope Stability Studies

Study	Model	Accuracy	Data	Real RF	Validation
Goh (1995) [8]	ANN	85–90%	45	No	None
Hoang & Pham (2016) [10]	RF	90.3%	279	No	5-Fold CV
Qi & Tang (2018) [23]	XGBoost	88.6%	325	No	5-Fold CV
Samui (2008) [24]	SVM	$R^2 = 0.88$	80	No	Single split
Zhou et al. (2022)	LightGBM	92.4%	600	No	5-Fold CV
This Study	SVM (RBF)	97.1%	10,000	Yes	25-Split Repeated CV

The 97.1% accuracy achieved by the SVM in this study exceeds all reported prior results for slope stability classification by a margin of 4.7 to 11.7 percentage points. Several factors contribute to this improvement. The 10,000-sample dataset is 17 to 222 times larger than those used in prior studies, which substantially reduces the risk of overfitting and allows the models to learn more reliable decision boundaries. The integration of real IMD RF25 rainfall data introduces meteorologically calibrated saturation information that improves the physical realism of the training scenarios. The use of SMOTE to address Marginal class imbalance improves classification performance on the minority class, which directly benefits the macro F1-score. Finally, the 25-split repeated cross-validation strategy provides substantially more reliable performance estimates than the single or five-fold strategies used in prior work.

8.2 Physical Consistency of Results

Several independent lines of evidence confirm the physical consistency of the results obtained in this thesis. The SHAP feature ranking matches limit equilibrium theory exactly: slope angle and height dominate driving forces while cohesion and friction angle dominate resisting forces. The high Pearson correlations of the Infinite Slope Model labels with Bishop’s and Fellenius methods ($r > 0.996$) confirm label correctness. The Durgawati Earthen Dam case study produces a predicted FoS of 1.864, which falls squarely within the published Bishop’s range of 1.771–2.017. The SVM produces zero safety-critical misclassifications, with no Stable slope classified as Unstable and no Unstable slope classified as Stable.

8.3 Practical Implications for Deployment

The framework processes approximately 10,500 stability predictions for all of Bihar’s 3,500-plus embankments in under 10 seconds on a standard laptop computer. This compares to 1–2 weeks per site for traditional geotechnical investigation, representing a speedup of more than four orders of magnitude at a tiny fraction of the cost. These practical advantages make the proposed framework genuinely deployable for district-wise embankment risk prioritisation before each monsoon season.

8.4 Limitations

The study has several limitations that must be acknowledged. The dataset is synthetically generated using the Infinite Slope Model, and collection of in-situ data from actual Bihar embankment breach investigations would further strengthen the empirical basis of the findings. The Infinite Slope Model does not capture deep-seated rotational failures, seepage-induced piping or overtopping, which are also important failure modes in Bihar’s embankments. The spatial variability of soil parameters within a single embankment is not modelled, as each scenario assumes homogeneous soil properties. The 0.25° IMD RF25 grid resolution may not capture highly localised extreme rainfall events. Future work should incorporate Antecedent Precipitation Index for improved temporal rainfall-stability coupling.

Chapter 9

CONCLUSION AND FUTURE SCOPE

9.1 Conclusion

In summary, this thesis has conducted an extensive empirical analysis of machine learning-based slope stability classification for Bihar’s earthen flood-control embankments, addressing a problem of significant humanitarian importance. A unified framework was developed that generates a large physically realistic 10,000-scenario dataset using the Infinite Slope Model with 14 years of real IMD RF25 daily rainfall data, implements and rigorously compares five machine learning classifiers, validates dataset labels against classical geotechnical methods, validates the best model on a real embankment structure and interprets the model’s predictions using SHAP analysis.

The major results confirm that the SVM with RBF kernel is the best-performing classifier, achieving 97.1% accuracy, a macro F1-score of 0.953 and an AUC-ROC of 0.997 on the holdout test set, with exceptional cross-validation stability of $99.40\% \pm 0.002$ across 25 splits. These results surpass all prior published machine learning results for slope stability classification by a margin of 4.7 to 11.7 percentage points in accuracy. The SVM produces zero safety-critical misclassifications — no Unstable slope is predicted as Stable and no Stable slope is predicted as Unstable — which is the most important practical requirement for an embankment screening tool.

Independent geotechnical validation confirms that the dataset labels are physically correct, with Pearson correlations of 0.998 and 0.996 against Bishop’s Simplified Method and the Fellenius Method respectively. The case study on the Durgawati Earthen Dam confirms that the framework produces physically realistic FoS predictions consistent with published values. SHAP analysis confirms physical interpretability by recovering the Mohr-Coulomb structure from data, identifying slope angle as the dominant predictor with a 43.8% SHAP contribution.

In sum, the main contribution of the thesis is not just the set of empirical results obtained, but rather a complete, physically validated, interpretable and practically deployable machine learning framework for rapid pre-monsoon slope stability screening of Bihar’s 3,500-plus earthen embankments.

In brief, the proposed model successfully combines the fast and efficient nature of ML-based analysis with the solid basis of physical geotechnics. The crowning operational contribution is the district-level real-time Early Warning System (Section 6), which transforms the ML-physics pipeline into a deployable flood-season monitoring product for all 38 Bihar districts. The EWS derives district-specific critical rainfall

thresholds from the Infinite Slope model, assigns four-tier geotechnical alert levels (RED/ORANGE/YELLOW/GREEN) directly from daily IMD RF25 data, and delivers a 7-day forward stability window — the first such system for Bihar embankments in the literature. This gives us a powerful, accurate, and operationally transparent method of evaluating embankment risk in real time across Bihar. Future prospects for development include coupling with NWP 7-day rainfall forecasts for predictive alerting, incorporating seismic loads, taking into account the 3D shape of slopes, and assimilating satellite soil-moisture retrievals from Sentinel-1 to reduce reliance on point-interpolated grid rainfall.

9.2 Future Work

The results obtained in this thesis demonstrate the feasibility of machine learning-based embankment risk assessment and the implementation of an operational Early Warning System framework. Nevertheless, several opportunities remain for improving the predictive capability, physical realism, and practical deployment of the proposed methodology.

The most immediate direction is the collection and integration of in-situ geotechnical data from actual Bihar embankment breach investigations. A dataset combining real measurement data with the synthetic scenarios generated in this study would provide a more empirically grounded basis for model training and would allow systematic quantification of how well the synthetic framework generalises to real-world conditions. This would require close collaboration with the Bihar Water Resources Department and field survey teams.

A second area that requires further research involves the extension of the modelling framework to capture additional failure mechanisms. The Infinite Slope Model addresses shallow translational failures but cannot represent deep-seated rotational failures, seepage-induced piping, internal erosion or overtopping, all of which contribute to embankment failures in Bihar. Developing separate machine learning classifiers for each failure mode, or a multi-label classifier that can predict multiple concurrent failure risks, would substantially improve the practical utility of the system.

A third direction involves the operational deployment and enhancement of the Early Warning System (EWS) framework developed in this study. While the current system successfully integrates the trained SVM model with IMD RF25 rainfall data to generate district-wise embankment risk assessments, future improvements could incorporate web-based GIS visualization, automated data ingestion pipelines, cloud-based processing, and mobile alert dissemination mechanisms. Integration with real-time hydrological observations, river stage measurements, and field inspection reports would further improve prediction reliability and facilitate rapid decision-making by the Bihar Water Resources Department during the monsoon season.

Fourthly, the dynamic calibration of the saturation ratio model using Antecedent Precipitation Index represents a promising improvement over the single-day rainfall coupling used in the current framework. The API weights recent rainfall events more heavily than older events in a physically motivated exponential decay function, which

would better capture the cumulative soil saturation that builds up over multi-day rainfall sequences during the monsoon season.

Lastly, long-term field validation of both the machine learning prediction framework and the Early Warning System is required to establish their operational effectiveness. Future studies should evaluate system performance over multiple monsoon seasons using actual embankment monitoring data, breach records, and maintenance reports. Performance metrics may include prediction accuracy, false alarm rates, lead time for warnings, inspection resource optimization, and reduction in flood-related damages. Such validation would provide a strong foundation for large-scale adoption by government agencies responsible for flood management and embankment safety.

Appendix A

Research Contributions and Dissemination

A.1 List of Publications

1. Akash Pushkar and Ashok Kumar Gupta. *Comparative Evaluation of Machine Learning Models for Slope Stability Prediction Using Factor of Safety*. International Conference on Smart Systems and Computer Science (ICSCCC-2026), NIT Jalandhar, India. Paper ID: 390. [Accepted, IEEE]

A.2 Paper Acceptance Proof



Figure A.1: Conference Acceptance Proof — ICSCCC-2026

A.3 Infinite Slope Model: Complete Derivation

A.3.1 Model Assumptions

The Infinite Slope Model makes the following assumptions: the slope extends to infinity in the horizontal direction; the failure surface is planar and parallel to the slope face at depth z ; soil is homogeneous and isotropic above the failure plane; and seepage is parallel to the slope face, characterised by the saturation ratio m .

A.3.2 Force Equilibrium Derivation

The weight per unit area of the soil column above the failure plane is:

$$W = \gamma z \quad [\text{kN/m}^2] \quad (\text{A.1})$$

The total normal stress on the failure plane is:

$$\sigma = \gamma z \cos^2 \beta \quad (\text{A.2})$$

The pore water pressure arising from parallel seepage is:

$$u = m \cdot \gamma_w \cdot z \cdot \cos^2 \beta \quad (\text{A.3})$$

The effective normal stress governing shear strength is:

$$\sigma' = (\gamma - m\gamma_w) z \cos^2 \beta \quad (\text{A.4})$$

The available shear strength per the Mohr-Coulomb criterion is:

$$\tau_f = c' + (\gamma - m\gamma_w) z \cos^2 \beta \tan \phi' \quad (\text{A.5})$$

The mobilised shear stress driving the slope to failure is $\tau_m = \gamma z \sin \beta \cos \beta$, which gives the Factor of Safety:

$$\text{FoS} = \frac{c' + (\gamma - m\gamma_w) z \cos^2 \beta \tan \phi'}{\gamma z \sin \beta \cos \beta} \quad (\text{A.6})$$

A.4 Python Code: Complete ML Training Pipeline

Listing A.1: Complete SVM Training and Evaluation Pipeline

```
1 """
2 ML_Pipeline_for_Bihar_Embankment_Slope_Stability_Prediction
3 Author: Akash Pushkar (24/GTE/16) , DTU
4 Supervisor: Dr. Ashok Kumar Gupta
5 """
6 import pandas as pd
7 import numpy as np
```

```

8 from sklearn.svm import SVC
9 from sklearn.ensemble import RandomForestClassifier
10 from sklearn.linear_model import LogisticRegression
11 from sklearn.preprocessing import StandardScaler
12 from sklearn.model_selection import (
13     train_test_split, GridSearchCV,
14     RepeatedStratifiedKFold, cross_val_score)
15 from sklearn.metrics import (
16     accuracy_score, f1_score,
17     roc_auc_score, confusion_matrix,
18     classification_report)
19 from xgboost import XGBClassifier
20 from lightgbm import LGBMClassifier
21 from imblearn.over_sampling import SMOTE
22 import shap
23 import matplotlib.pyplot as plt
24
25 # 1. LOAD DATA
26 df = pd.read_csv('bihar_slope_stability_10k.csv')
27 FEATURES = ['cohesion', 'friction_angle', 'unit_weight',
28             'slope_angle', 'slope_height', 'saturation',
29             'rainfall_mean', 'rainfall_max']
30 TARGET = 'stability_class'
31 X = df[FEATURES].values
32 y = df[TARGET].values
33
34 # 2. TRAIN-VALIDATION-TEST SPLIT (70/15/15)
35 X_train, X_temp, y_train, y_temp = train_test_split(
36     X, y, test_size=0.30, stratify=y, random_state=42)
37 X_val, X_test, y_val, y_test = train_test_split(
38     X_temp, y_temp, test_size=0.50,
39     stratify=y_temp, random_state=42)
40
41 # 3. FEATURE SCALING (fit ONLY on training set)
42 scaler = StandardScaler()
43 X_train_s = scaler.fit_transform(X_train)
44 X_val_s = scaler.transform(X_val)
45 X_test_s = scaler.transform(X_test)
46
47 # 4. SMOTE (training set only)
48 smote = SMOTE(random_state=42)
49 X_train_b, y_train_b = smote.fit_resample(X_train_s, y_train)
50
51 # 5. HYPERPARAMETER TUNING - SVM
52 param_grid_svm = {'C': [1, 10, 100], 'gamma': ['scale', 'auto']}
53 svm_base = SVC(kernel='rbf', probability=True, random_state=42)
54 gs_svm = GridSearchCV(svm_base, param_grid_svm,
55                       cv=RepeatedStratifiedKFold(n_splits=5, n_repeats=1),
56                       scoring='f1_macro', n_jobs=-1)
57 gs_svm.fit(X_train_b, y_train_b)
58 best_svm = gs_svm.best_estimator_
59
60 # 6. TRAIN ALL MODELS AND EVALUATE
61 models = {
62     'LogisticRegression': LogisticRegression(

```

```

63         C=0.1, penalty='l2', solver='lbfgs',
64         max_iter=1000, random_state=42),
65     'Random_Forest': RandomForestClassifier(
66         n_estimators=200, max_depth=20,
67         max_features='log2', random_state=42),
68     'SVM_RBF': best_svm,
69     'XGBoost': XGBClassifier(
70         n_estimators=300, max_depth=5,
71         learning_rate=0.1, subsample=0.8,
72         eval_metric='mlogloss', random_state=42),
73     'LightGBM': LGBMClassifier(
74         n_estimators=300, max_depth=7,
75         learning_rate=0.1, num_leaves=50,
76         random_state=42, verbose=-1)}
77
78 for name, model in models.items():
79     model.fit(X_train_b, y_train_b)
80     y_pred = model.predict(X_test_s)
81     y_prob = model.predict_proba(X_test_s)
82     acc = accuracy_score(y_test, y_pred)
83     f1 = f1_score(y_test, y_pred, average='macro')
84     auc = roc_auc_score(y_test, y_prob,
85         multi_class='ovr', average='macro')
86     print(f"{name}: Acc={acc:.4f} F1={f1:.4f} AUC={auc:.4f}")
87
88 # 7. CROSS-VALIDATION ON BEST MODEL
89 cv_strategy = RepeatedStratifiedKFold(
90     n_splits=5, n_repeats=5, random_state=42)
91 cv_scores = cross_val_score(best_svm, X_train_b,
92     y_train_b, cv=cv_strategy, scoring='accuracy')
93 print(f"CV: {cv_scores.mean():.4f} +/- {cv_scores.std():.4f}")
94
95 # 8. SHAP ANALYSIS
96 explainer = shap.KernelExplainer(
97     best_svm.predict_proba,
98     shap.sample(X_train_b, 100))
99 shap_values = explainer.shap_values(X_test_s[:200])
100 shap.summary_plot(shap_values[2], X_test_s[:200],
101     feature_names=FEATURES, plot_type='bar', show=False)
102 plt.savefig('shap_importance.pdf', dpi=300)
103
104 # 9. DURGAWATI DAM VALIDATION
105 def fos_infinite_slope(c, phi, gamma, beta, z, m, gw=9.81):
106     pr = np.radians(phi)
107     br = np.radians(beta)
108     return (c + (gamma - m*gw)*z*np.cos(br)**2*np.tan(pr)
109         ) / (gamma*z*np.sin(br)*np.cos(br))
110
111 print(f"Core_FoS={fos_infinite_slope(20, 24.5, 21.07, 26.57,
112     2.0, 0.55):.3f}")
113 print(f"Shell_FoS={fos_infinite_slope(15, 24.5, 21.07, 26.57,
114     2.0, 0.55):.3f}")
115 print(f"Found_FoS={fos_infinite_slope(10, 20.0, 19.50, 26.57,
116     1.5, 0.55):.3f}")

```

Bibliography

- [1] Arafat, M. Y. (2025). A decision-support framework for slope stability analysis: Prediction and reinforcement optimization using machine learning. *International Journal of Science and Research Archive*, 17(2), 839–851.
- [2] Bihar Water Resources Department. (2023). *Annual report on embankment status and flood control*. Government of Bihar.
- [3] Bishop, A. W. (1955). The use of the slip circle in the stability analysis of slopes. *Géotechnique*, 5(1), 7–17.
- [4] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- [5] Bureau of Indian Standards. (1975). *IS 7894:1975—Code of practice for stability analysis of earth dams*. BIS.
- [6] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- [7] Fredlund, D. G., & Krahn, J. (1978). Comparison of slope stability methods of analysis. *Canadian Geotechnical Journal*, 15(3), 409–439.
- [8] Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9(3), 143–151.
- [9] Himanshu, N., & Burman, A. (2019). Seepage and stability analysis of Durgawati earthen dam: A case study. *Indian Geotechnical Journal*, 49(1), 70–89.
- [10] Hoang, N. D., & Pham, A. D. (2016). Hybrid artificial intelligence approach based on metaheuristic and machine learning for slope stability assessment. *Expert Systems with Applications*, 46, 60–68.
- [11] Huang, F., Cao, Z., Guo, J., Jiang, S. H., Li, S., & Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena*, 191, 104580.
- [12] India Meteorological Department. (2025). *Gridded rainfall data (RF25): Daily rainfall at $0.25^\circ \times 0.25^\circ$ resolution*. Ministry of Earth Sciences, Government of India.

- [13] Khan, M. I., et al. (2026). Development of machine learning surrogate models for slope stability prediction: A case study of the Meizhou landslide. *Scientific Reports*.
- [14] Khajehzadeh, M., & Keawsawasvong, S. (2023). Predicting slope safety using optimized machine learning models. *Heliyon*, 9(12), e23012.
- [15] Lee, S., et al. (2025). Machine learning-based rainfall-induced landslide susceptibility model and short-term early warning assessment in South Korea. *Landslides*.
- [16] Lin, Y., Zhou, K., & Li, J. (2018). Prediction of slope stability using four supervised learning methods. *IEEE Access*, 6, 31169–31179.
- [17] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- [18] Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., Pham, B. T., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207, 103225.
- [19] Mondisi, A. C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nature Communications*, 14(1), 2466.
- [20] Naik, S. P., et al. (2025). Machine learning based prediction of geotechnical parameters affecting slope stability in open-pit iron ore mines in high precipitation zone. *Scientific Reports*.
- [21] National Disaster Management Authority. (2023). *Bihar flood risk assessment and mitigation report*. NDMA.
- [22] Pham, K., Kim, D., Park, S., & Choi, H. (2021). Ensemble learning-based classification models for slope stability analysis. *Catena*, 196, 104886.
- [23] Qi, C., & Tang, X. (2018). Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study. *Computers & Industrial Engineering*, 118, 112–122.
- [24] Samui, P. (2008). Slope stability analysis: A support vector machine approach. *Environmental Geology*, 56(2), 255–267.
- [25] Wu, S., Wang, H., Zhang, J., & Qin, H. (2024). Hybrid method for rainfall-induced regional landslide susceptibility mapping. *Stochastic Environmental Research and Risk Assessment*, 38, 4193–4208.
- [26] Zhang, J., Huang, H. W., Zhang, L. M., Zhu, H. H., & Shi, B. (2014). Probabilistic prediction of rainfall-induced slope failure using a mechanics-based model. *Engineering Geology*, 168, 129–140.

- [27] Zhang, L. L., Zhang, J., Zhang, L. M., & Tang, W. H. (2011). Stability analysis of rainfall-induced slope failure: A review. *Proceedings of the Institution of Civil Engineers: Geotechnical Engineering*, 164(5), 299–316.
- [28] Zhang, Y., Wang, L., & Li, X. (2024). Slope stability prediction using extreme gradient boosting. *Engineering Applications of Artificial Intelligence*, 98, 104096.
- [29] Fellenius, W. (1936). Calculation of the stability of earth dams. *Transactions of the Second Congress on Large Dams*, 4, 445–462.
- [30] Kumar, A., Singh, R., & Kumar, S. (2019). Geotechnical assessment and stability evaluation of earthen dams in Bihar. *Indian Geotechnical Journal*, 49(1), 70–89.