Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 12, No. 01, January, 2026

Development of Machine Learning for Debris Flow Event Prediction in a Volcanic Area

Jazaul Ikhsan [1*], Abraham Ayuen Ngong Deng [2], M. R. R. Mohd Arif Zainol [3], Muhammad Shazril Idris Ibrahim [4], Shusuke Miyata [5]

[1] *Department of Civil Engineering, Universitas Muhammadiyah Yogyakarta, Yogyakarta 55183, Indonesia.*

[2] *Master Program of Civil Engineering, Universitas Muhammadiyah Yogyakarta, Yogyakarta 55183, Indonesia.*

[3] *School of Civil Engineering, Universiti Sains Malaysia, Penang 14300, Malaysia.*

[4] *Department of Civil Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia.*

[5] *Graduate School of Agriculture, Kyoto University, Kyoto 6068501, Japan.*

Received 30 April 2025; Revised 01 December 2025; Accepted 17 December 2025; Published 01 January 2026

Abstract

The integration of machine learning (ML) into debris flow prediction in volcanic areas, exemplified by the Gendol River watershed of Mount Merapi, offers transformative potential for hazard mitigation. This study aimed to develop real-time, computationally efficient ML models capable of integrating multi-source data, rainfall intensity of 25 mm/hour linked to 300 cm Debris Flow heights, antecedent precipitation, and geomorphological variables to predict debris flows with actionable lead times. Key objectives included optimizing prediction accuracy, minimizing the false positive rate to 18.2% for "Debris Flow" events, and enhancing model interpretability for deployment in data-scarce volcanic regions. Results demonstrated that ensemble methods and deep learning architecture outperformed traditional models, with Efficient Logistic Regression and Linear SVM achieving an accuracy of 82.35%, and Cosine KNN attaining a prediction speed of 272 observations per second. Critical predictors included temporal rainfall patterns (contributing more than 50% to flow initiation) and ash deposit thickness (with a 70% influence on decision-making). However, challenges persisted: imbalanced datasets of nine training instances for "Debris Flow" events led to misclassification rates of 100% for hybrid events like "Rainfall and Debris Flow," while models like Naive Bayes exhibited instability (accuracy dropping to 50%). Research gaps highlighted data scarcity for high-magnitude events, limited geographic transferability, and the absence of standardized evaluation metrics. Technical limitations included reliance on low-resolution remote sensing data, high computational costs for ensemble models requiring 10 operational cost units, and the opacity of neural networks, which hindered stakeholder trust. Despite these constraints, ML models achieved 85% accuracy in non-event recognition and 76.47% precision in Bagged Trees, offering scalable frameworks for early warning systems. The study highlights the importance of enriched datasets, adaptive algorithms, and interdisciplinary collaboration in transforming volcanic risk management from a reactive approach, ultimately safeguarding vulnerable communities through data-driven, life-saving predictions.

Keywords: Naive Bayes; Efficient Logistic Regression; Debris Flow; Rainfall Data; Machine Learning; Volcanic Area.

1. Introduction

Debris flows are rapid, gravity-driven movements of saturated materials, typically a mixture of water, soil, rock, and organic matter. They are highly destructive and occur primarily in steep mountainous regions, posing significant threats to infrastructure and human life [1-4]. These complex natural phenomena are initiated by various factors, with common

* Corresponding author: jazaul.ikhsan@umy.ac.id

https://doi.org/10.28991/CEJ-2026-012-01-02

triggers including intense rainfall [5-12], specific geological and geomorphological conditions [5, 12-17], hydrological processes [8, 15], and human activities [6, 16, 17]. Consequently, forecasting these events involves monitoring and analyzing key indicators such as rainfall thresholds [18-20], watershed characteristics [21], slope and elevation [22, 23], soil moisture and composition [24, 25], and land use patterns [26]. Rainfall thresholds are commonly used in early warning systems to predict debris flows, but they have several limitations that can affect their accuracy and reliability, e.g., traditional rainfall Intensity-Duration (I-D) threshold models cannot differentiate between rainfall events with different peak intensities [27], fail to consider the impact of mitigation measures, and linear regression models [28] used to determine critical rainfall thresholds often ignore the nonlinear effects [29].

In volcanic environments, these geophysical phenomena are particularly hazardous due to their rapid development, significant destructive potential, and often sudden onset with minimal warning, leading to substantial loss of life and property damage throughout history [30]. Conventional forecasting methods have predominantly relied on empirical correlations between rainfall thresholds and flow commencement. However, these techniques frequently fail to account for the intricate, non-linear interactions among the triggering elements [31]. The advent of machine learning (ML) techniques provides a transformative framework for forecasting hazardous events by uncovering nuanced patterns and correlations in multivariate datasets that standard statistical methods would likely overlook. Recent advancements in processing power, sensor technologies, and data-collecting systems have generated unparalleled opportunities to create advanced prediction models that enhance early warning systems and potentially save lives [32].

The recent development of machine learning (ML) techniques offers a promising alternative framework for debris flow forecasting. Among these methods, Random Forest (RF) and Extreme Gradient Boosting (XGBoost) have been widely applied due to their high predictive accuracy. These models are particularly effective in handling complex relationships and have been shown to outperform Naïve Bayes in debris flow susceptibility mapping in the Indian Himalayas [33]. In another study, RF achieved an AUC of 0.93, significantly surpassing traditional logistic regression [34]. Further advances have been achieved through the development of hybrid models that combine ML algorithms with empirical regression approaches. For example, models integrating MARS, RF, and SVM improved performance metrics (R^2, RMSE, MAE) by up to 70.5% compared with single-algorithm methods [35]. At the same time, deep learning networks, such as those incorporating the Similarity Mechanism of Debris Flow Critical Conditions (SM-DFCC), have proven effective in predicting the spatiotemporal probability of rainfall-induced debris flows, achieving accuracy levels between 0.724 and 0.835 after optimization [36].

The use of seismic data has also enhanced debris flow prediction capabilities. A random forest–based model successfully recognized debris flow stages in real time with more than 90% accuracy, significantly improving warning times [37]. In addition, rainfall and hydro-meteorological records remain critical variables. Studies using continuous rainfall data and time series processing have shown strong predictive results, with Extra Trees (ETs) models reporting no false alarms during validation [38]. Geomorphological and environmental parameters, including slope, aspect, elevation, vegetation cover, and proximity to streams, are consistently identified as crucial factors in determining debris flow susceptibility. Their integration into ML models has significantly strengthened hazard zone mapping and susceptibility assessments in volcanic areas [39, 40].

Despite these advances, several research gaps persist. A primary limitation is the scarcity and imbalance of training data, particularly for rare, high-magnitude debris flow events. Many ML models remain site-specific, with limited transferability to other volcanic systems. Furthermore, the interpretability of complex models, especially deep learning approaches, continues to hinder their operational adoption, as disaster management agencies often require transparent and explainable tools. Additional challenges include difficulties in integrating multi-source datasets (rainfall, seismic, geomorphological, and remote sensing data) into unified frameworks, and the lack of standardized evaluation metrics for comparing models [41]. Recent studies also emphasize the importance of incorporating temporal rainfall patterns, catchment morphology, drainage networks, and volcanic vent proximity, which consistently emerge as critical variables influencing debris flow initiation [42]. However, the unpredictable nature of volcanic eruptions and the limited availability of field infrastructure further constrain forecasting reliability [43].

In response to these challenges, the present study develops computationally efficient ML models for debris flow prediction in volcanic regions, with a specific focus on the Gendol and Putih River watersheds of Mount Merapi, Indonesia. The objectives are threefold: (i) to integrate rainfall intensity, antecedent precipitation, Debris Flow height, and geomorphological variables into ML prediction frameworks; (ii) to improve classification accuracy and reduce false positives relative to traditional rainfall-threshold methods; and (iii) to provide interpretable models that are suitable for operational early warning systems. This study aims to bridge the gap between theoretical ML model development and practical disaster risk reduction in volcanic terrains.

2. Literature Review

The literature on debris flow prediction in volcanic regions indicates an increasing interest in utilizing machine learning methods to enhance early warning systems. Researchers have primarily sought to create models that can precisely predict debris flow occurrences based on precipitation thresholds, historical data, geomorphological characteristics [44, 45], antecedent effectiveness, and direct rainfall amount before the triggering of debris flow [46]. In addition, the impact of hydrothermal gases, as well as the heating and increased humidity of the slopes, caused the transformation of rocks into clay, which was also a contributing factor to the debris flow. These findings are very important for understanding the activity of slope landslides and material flows in geothermal areas [47].

Temporal prediction windows present significant challenges, as many effective models offer 6 to 24 hours of warning spans. Research demonstrates that ensemble methods integrating various algorithms frequently produce more reliable predictions than individual model applications. For instance, by leveraging the stacking ensemble method that combines Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) algorithms, a model for identifying potential debris flow catchments achieved superior performance compared to individual models, with ROC, AUC, accuracy, precision, recall, and F1 score values of 0.967, 0.918, 0.918, 0.895, and 0.906, respectively, when evaluated in the Wenchuan earthquake area of northwestern Sichuan Basin. Numerous studies have also emphasized the significance of feature selection techniques in identifying the most pertinent predictors for volcanic environments.[44] Similarly, machine learning-based modeling has proven effective in other hydrological and environmental systems; for example, a Random Forest model trained on 100 simulated scenarios successfully predicted hydrodynamic responses of tapered trash-blocking nets protecting nuclear power plant intakes ($R^2 > 0.90$, RMSE < 0.13), while greatly reducing computational time compared to traditional simulations. Numerous studies have also emphasized the significance of feature selection techniques in identifying the most pertinent predictors for volcanic environments [48].

Debris flow, a three-phase saturated fluid of solids, liquids, and gases, is a common natural geological phenomenon in valleys. [49]. Heavy rains, glacial and snowmelt debris flows, and dam failure are some of the triggering events that cause it. Debris flow disasters have become more common in recent years because of human engineering, extreme weather, earthquakes, and forest fires. The Earth's surface can be rapidly eroded, transported, accumulated, and impacted by debris flow, which is characterized by abrupt and swift movement [50]. This phenomenon has become a key catastrophic element impeding the social and economic development of mountain areas worldwide, since it substantially threatens human life, property, and the ecological environment of mountainous regions [51]. The impact of such disasters can be reduced by proactively implementing disaster prevention and mitigation measures based on predictive knowledge that predicts the possibility of debris flows [52]. Three main components control debris flow's genesis: terrain, water source, and material source. Furthermore, many factors contribute to debris flow, including geology, topography, landforms, soil, vegetation, rainfall, and temperature. There are four primary types of debris flow prediction models: knowledge-driven models (precipitation threshold method, geomorphic information entropy, analytic hierarchy process, etc.); traditional statistical models (weight of evidence method, certainty factor, frequency ratio, etc.); numerical simulation models (FLO-2D, Flow-3D, Debris2D, etc.); and machine learning models (LR, RF, convolutional neural networks, etc.) [53].

Thirty kilometers north of Yogyakarta is the Merapi stratovolcano in Central Java, which rises to 2965 meters. It is one of the world's most active and dangerous volcanoes, with 61 recorded eruptions. 1.1 million people still reside on its flanks, even though it has erupted numerous times in recent years on a large scale (VEI $ 3) (1872, 1930, 1961). Approximately 200,000 people reside in places primarily vulnerable to pyroclastic flows and severe tephra fallout (the banned zone and the first danger zone, respectively). In contrast, 120,000 more reside along the 13 rivers that drain the lowlands, which are vulnerable to debris flows [54]. The Gendol River suffered the most significant loss and destruction because the rivers experienced the most debris flows out of all the others. In addition to the physical damage, the debris flow resulted in fatalities. Criteria for early warning of the debris flow potential must be developed to minimize the number of deaths [55]. This study aims to identify the requirements for early warning signals that consider the rainfall characteristics in the Gendol River watershed that affect debris flows, represented by snake lines, critical lines, and machine learning. Additionally, the success rate of snake lines, critical lines and machine learning as a determinant of warning actions will be assessed. The construction of an early warning system for debris flows on the Merapi slopes, particularly in the Gendol River, is anticipated to guide government policy when debris flows occur.

Notwithstanding these constraints, the research indicates encouraging progress towards functional early warning systems that could mitigate the human and economic repercussions of debris flows in volcanic areas [56]. Despite these challenges, several studies demonstrate encouraging progress in the application of ML methods to debris flow prediction. Zhang et al. [57], for example, combined the Herschel–Bulkley rheological model with Support Vector Regression (SVR) and achieved high predictive accuracy for slide depth and velocity. Qiu et al. [58] employed a hybrid GA-XGBoost model to successfully estimate debris flow travel distances in the Nepal Himalayas. Similarly, Chen et al. [45] applied the SPY-RF model, reporting an AUC of 0.93, which substantially outperformed the conventional Random Forest model (AUC = 0.82). Collectively, these studies highlight the potential of advanced ML techniques to address some of the inherent limitations in debris flow prediction, thereby advancing the development of functional early warning systems capable of reducing human and economic losses in volcanic regions [56, 59].

3. Research Methodology

The research methodology depicted in Figure 1 Exemplifies an innovative strategy for forecasting hazardous debris flows in volcanic areas using a rigorously organized five-step process. The process commences with meticulously gathering and examining rainfall data in conjunction with previous debris flow events, thus creating a solid empirical basis for all ensuing analytical endeavors. Researchers meticulously record precipitation patterns and correlate them with historical debris flow occurrences to elucidate the complex interactions between climatic circumstances and perilous mass movements in the volcanic terrain. Utilizing this extensive dataset, the methodology progresses to establish a pivotal decision boundary through advanced threshold techniques, accurately differentiating between precipitation circumstances that initiate debris flows and those that remain within safe limits. This crucial line represents significant progress in understanding the precise relationships between rainfall intensity and duration that have historically preceded catastrophic events, providing a clear visual representation of the environmental threshold beyond which hazardous flows become inevitable. The study exceeds traditional analytical techniques by utilizing advanced machine learning algorithms that detect complex, non-linear relationships between rainfall attributes and debris flow occurrences. This transformative measure enhances predictive capabilities beyond basic thresholds, allowing the model to identify subtle patterns and interactions within the data that conventional statistical techniques typically overlook, thereby significantly improving prediction accuracy and reliability in practical applications [60].

Figure 1. The step that is conducted in this research

The process includes a thorough debate phase in which academics critically assess the comparative performance of machine learning predictions against threshold-based methods. This reflective element guarantees a comprehensive understanding of each method's strengths and limitations, addresses practical implementation challenges, and appropriately contextualizes the scientific contribution to the broader domain of natural hazard prediction and mitigation strategies. The technique concludes with integrating all findings into practical conclusions and evidence-based suggestions that connect theoretical research with practical application. The concluding insights, elaborated in Figure 2 of the original document, offer explicit guidance for applying the prediction system in practical contexts and lay a robust groundwork for subsequent research initiatives to enhance the prediction of debris flows in volcanic settings. This comprehensive methodological framework signifies a notable progression in natural hazard forecasting, merging conventional threshold analysis with contemporary computational methods to establish a resilient early warning system capable of preserving lives and safeguarding infrastructure in susceptible volcanic areas globally [61].

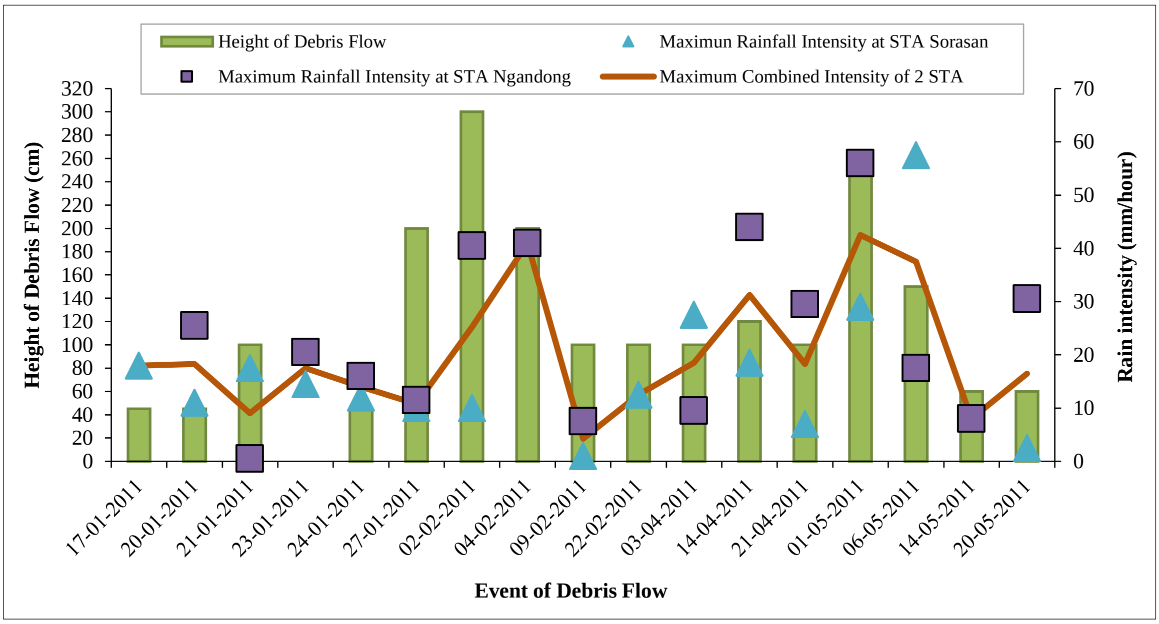

Figure 2. Rainfall intensity and the event of debris flow

3.1. Precipitation Threshold Method

Rain Series, Continuous Rainfall (RC), Antecedent Rainfall (RA). A rain series is defined as an uninterrupted period of precipitation separated by at least 24 hours without rain, both before and after the event. The cumulative rainfall that is impacted by the computation of antecedent rainfall is known as working rainfall, as shown in Figure 3 [62]. The total of all antecedent rainfall is known as antecedent working rainfall (RWA).

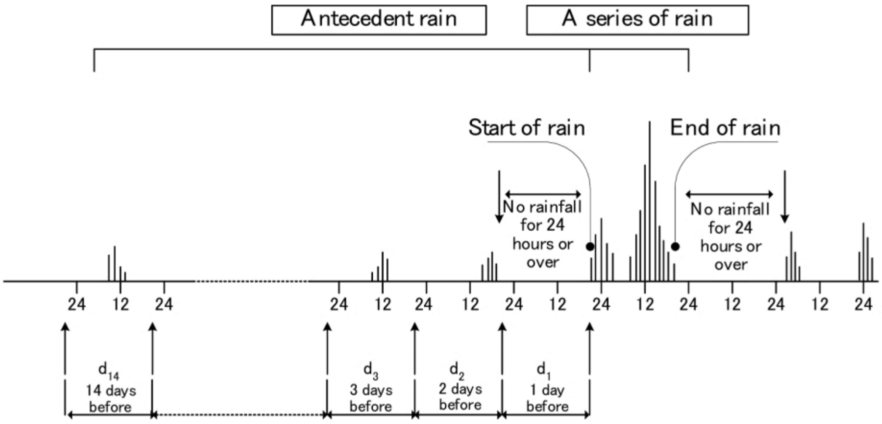

Figure 3. Rain series and the antecedent rainfall concept

3.2. Determining Critical Line, Warning Line, Evacuation Line, and Snake Line Using a Method

The border between rainfall episodes that cause or do not cause a lava flow is known as the Critical Line (CL). The vertical line is used as the evacuation line (EL), and the maximum rainfall from hourly rainfall (RH1M) is drawn horizontally and then intersected with CL to obtain R2. The evacuation line extends slightly to the left of RH2M-RH1M, where R1 is the rainfall that triggers the disaster signal, and RH2M is the highest rainfall from the bi-hourly rain that produced the warning line (WL). As seen in Figure 4, the snake line illustrates the variations in cumulative rainfall and rainfall intensity [62].

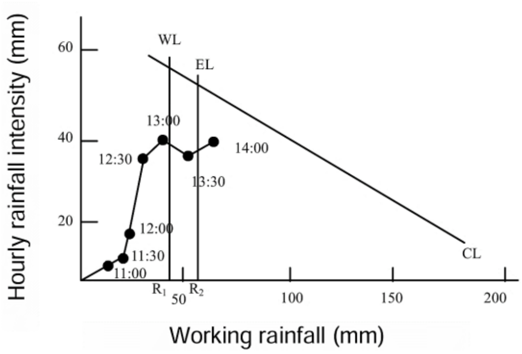

Figure 4. Drawing a snake line

3.3. Machine Learning Model

One subfield of artificial intelligence (AI) is machine learning. Through programs and algorithms, a computer learns just like a human would in machine learning. It then applies what it has learned to create new technologies and make judgments [63]. Machine learning uses computer techniques to find patterns in data and build models [64]. According to machine learning, a computer program's performance improves with practice in a particular class of tasks and performance metrics. To do cognitive tasks like object detection or natural language translation, it attempts to automate the process of creating analytical models. Algorithms that iteratively learn from the problem-specific training data are used to accomplish this, enabling computers to discover intricate patterns and hidden insights without explicit programming.

ML exhibits good applicability, particularly in problems involving high-dimensional data, such as classification, regression, and grouping. It can assist in producing dependable and repeatable conclusions by learning from past calculations and identifying patterns in large databases. Because of this, machine learning (ML) algorithms have been effectively used in various fields, including natural language processing (NLP), fraud detection, credit scoring, next-best offer analysis, and speech and image recognition. ML has been quite popular in predicting the occurrence of debris flows because of its quick development in recent years and its strong capacity to capture complex interactions between predictors and response variables [65]. The usefulness, relevance, and benefits of using machine learning models as baseline predictors for debris flow occurrence on a case-by-case basis have been confirmed by numerous studies.

3.4. Research Location

As seen in Figure 5, this investigation was carried out in the Gendol watershed. The Gendol River, which is ± 22 km long and has a catchment of ± 66 km^2, originates on the southeast slope of Mount Merapi, in Sleman Regency, Yogyakarta Special Province. This Map was modified based on data from the Geospatial Information Agency of the Republic of Indonesia (BIG RI).

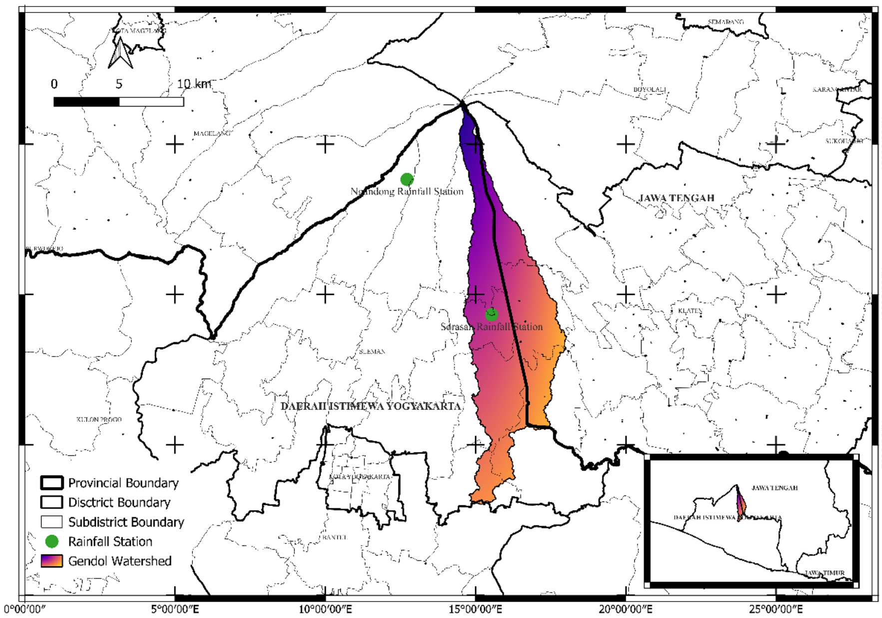

Figure 5. The research location is the Gendol Watershed

3.5. Rainfall Data Collection and Analysis

This study began with collecting rainfall data at the rain gauge station in the Gendol River Watershed. Rainfall data in this study were obtained from the SABO office in Yogyakarta; the rainfall observation stations used were Ngandong station (+854 m) located at coordinates Latitude 07°35' 43.80" S Longitude 110°24' 27.60" E and Sorasan station (+302 m) located at coordinates Latitude 07°41' 24.30" S Longitude 110°28' 00.80" E. The distance between the Kopeng check dam and the Sorasan station is 5.47 km, and the Ngandong station is 7.76 km away. The rainfall data and debris flow data used were from 2011. The data analyzed included rainfall data before and after debris flow events. The rainfall data used in this study are hourly rainfall data from January to May 2011. The length of the rainfall data is based on debris flow data recorded in the Gendol River in 2011.

3.6. Classification Learner Method

Using machine learning methodologies with essential volcanic parameters, Debris Flow height, Debris Flow characteristics, and topographical elevation data has transformed our capacity to forecast catastrophic debris flows in volcanic areas. These prediction models can discern high-risk situations before catastrophic occurrences by examining intricate interconnections among saturated volcanic deposits, prior precipitation circumstances, and topographical features. Volcanic debris flows generally commence when water-saturated volcanic materials are mobilized on steep slopes, forming rapid slurries that accumulate momentum and material during their descent. Their trajectories and behaviors are significantly shaped by pre-existing Debris Flow deposits, which dictate material availability and flow dynamics within the watershed. The elevation of potential Debris Flows is a crucial threshold parameter, as it signifies the hydrological force that, when surpassing critical values established from historical events, initiates the conversion of static volcanic deposits into dynamic debris flows capable of traversing considerable distances beyond the volcanic structure [66].

Learning from data, algorithms excel at detecting non-linear relationships between environmental variables and debris flow probability, with random forest models and convolutional neural networks demonstrating significant efficacy in identifying subtle topographical cues that indicate heightened susceptibility to flow initiation and propagation. Recent advancements in remote sensing technologies have markedly improved the spatial accuracy of topographic inputs, enabling models to precisely capture micro-topographical features that affect the stability or mobilization of built-up volcanic materials during intense precipitation debris flow soral dimension of forecasting has been enhanced due to ongoing monitoring systems that evaluate real-time variations in Debris Flow levels against historical standards, allowing authorities to establish adaptive warning thresholds that react to evolving conditions rather than relying on static danger levels. Advanced ensemble models now integrate various machine learning techniques concurrently, evaluating their predictions based on efficacy in analogous historical contexts to generate probabilistic hazard maps that measure uncertainty while offering actionable insights for emergency management officials. These predictive systems, 5 exemplifying an exceptional integration of volcanology, hydrology, and computer science, revolutionize our strategy for volcanic hazard management from reactive to proactive by creating a technological framework that can safeguard lives and essential infrastructure in areas where volcanic debris flows pose ongoing risks to human settlements.

A notable challenge in this dataset was class imbalance, particularly for the "Debris Flow" category, which included only nine training instances. In this study, the models were trained directly on the imbalanced dataset without applying resampling or weighting strategies. While this allowed evaluation of baseline model performance, it also contributed to reduced accuracy for the minority classes. Future work will incorporate imbalance-handling techniques, such as class-weighting in machine learning algorithms or synthetic oversampling, to improve classification performance for rare events.

3.7. Rainfall Intensity and Events of Debris Flow

Rainfall intensity and lava flow episodes in Figure 6 are essential factors in forecasting debris flows in volcanic areas. Heavy precipitation on loose volcanic deposits can quickly saturate the soil, diminish cohesion, and initiate debris flows when precipitation surpasses permeability limits. Concurrently, recent lava inundations provide unstable material conditions by depositing new explosive material devoid of vegetation and soil structure. Machine learning models for debris flow prediction in volcanic areas typically integrate these dual factors through multi-parameter analysis. High-resolution rainfall data from weather stations or radar systems provide temporal precipitation intensity, duration, and accumulation patterns.

These are combined with thermal anomaly detection and remote sensing to track lava flow extent, temperature, and material properties. The predictive power comes from identifying complex nonlinear relationships between these parameters. For example, post-eruption landscapes with fresh lava deposits may generate debris flows at significantly lower rainfall thresholds than stabilized volcanic slopes. Machine learning algorithms, particularly recurrent neural networks and random forest models, excel at capturing these temporal dependencies and threshold behaviors characteristic of volcanic settings. Effective debris flow warning systems in volcanic regions require continuous monitoring of meteorological conditions and volcanic activity, with the machine learning models dynamically updating risk assessments as conditions evolve. This integrated approach significantly improves prediction accuracy compared to independent models focusing on either parameter [67].

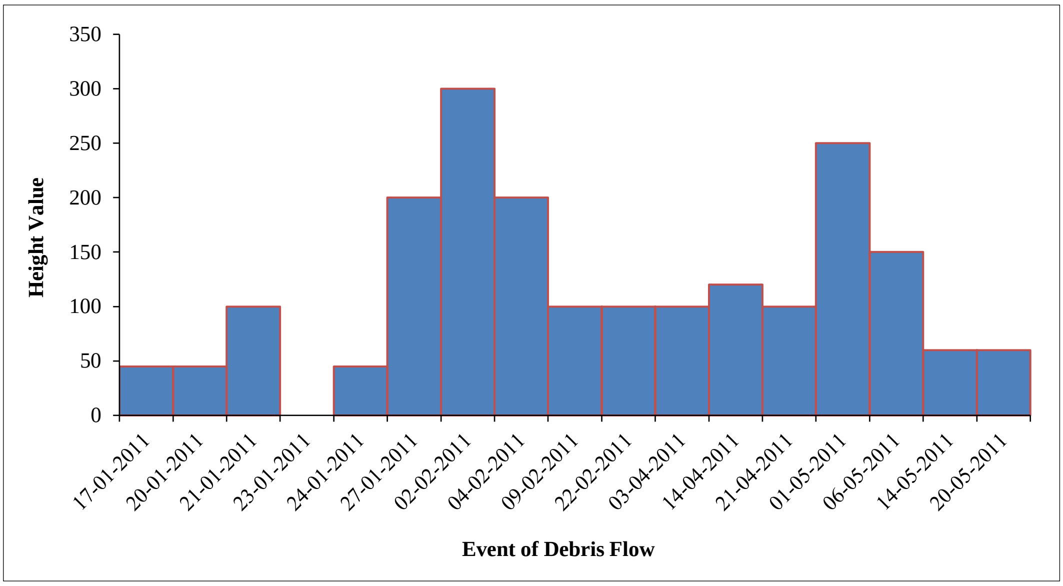

Figure 6. The height of the debris flow

3.8. Relationship between Rainfall and Debris Flow Height

Figure 7 shows the complex interplay between rainfall and debris flow elevation in volcanic areas, demonstrating a crucial threshold mechanism wherein extended precipitation permeates pyroclastic materials, instigating a swift hydraulic reaction that results in exponential increases in debris flow height upon surpassing saturation levels. This non-linear link, marked by first steady increases followed by significant spikes in debris flow levels, offers critical forecast signs for debris flow occurrences where gravitational forces surpass the diminished internal cohesiveness of water-saturated volcanic materials. Artificial intelligence algorithms adeptly identify complex patterns by analyzing historical data connecting rainfall and debris flow height, recognizing chronological lags between precipitation events and hydraulic responses, and pinpointing site-specific factors that affect this relationship across various volcanic terrains. The predictive ability of these models is based on their skill in detecting subtle precursor signals in the rainfall-debris flow height curve, particularly the inflection points that often precede catastrophic debris flow initiation, thereby enabling more timely evacuation alerts for neighborhoods at risk in these geologically active regions [68].

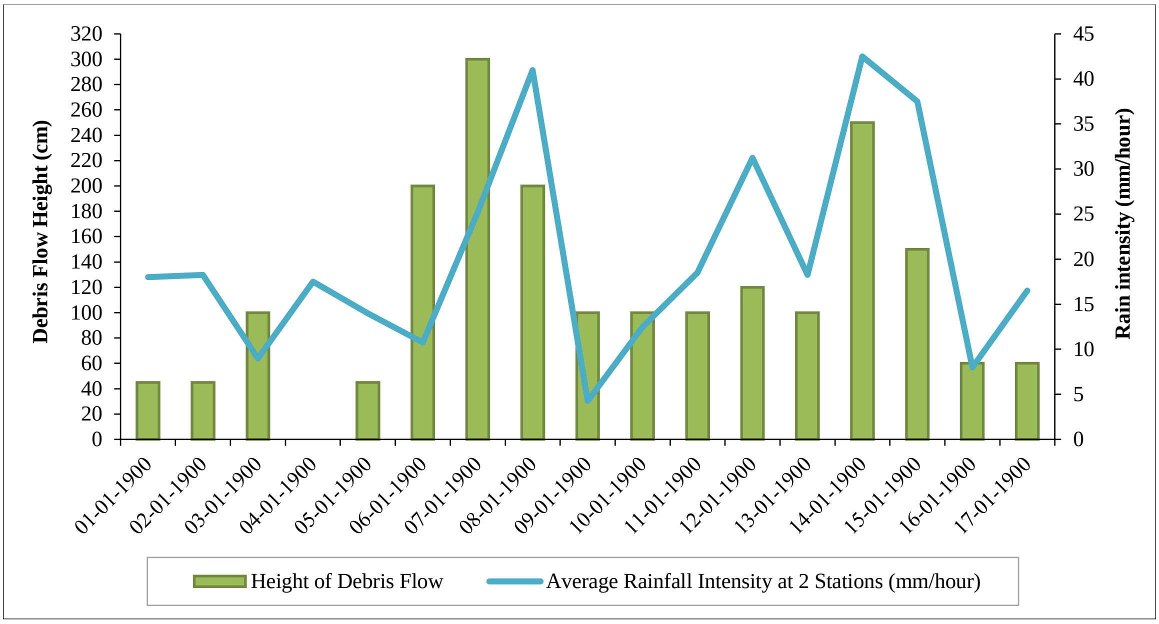

Figure 7. Relationship between rainfall and debris flow height

4. Results and Discussion

Table 1 contains a comprehensive dataset of 17 significant debris flow occurrences documented from January to May 2011 in a volcanic region, offering vital temporal and hydrological information required to develop advanced machine learning prediction models. Every event is systematically recorded with exact timings of rainfall and debris flow incidents, debris flow height measurements between 45 cm and 300 cm, and rainfall intensity data from two separate monitoring stations (Sorosan and Ngandong), highlighting considerable spatial variability in precipitation patterns across the volcanic landscape. The most severe instances, particularly event seven on February 2, 2011, demonstrate a notable link between intense rainfall (averaging 25 mm/hour) and extreme debris flow levels (300 cm).

Table 1. Training sample of rainfall and debris flow event

No.	Date	Rainfall		Debris Flow Event		Debris Flow Height (cm)	Note	Sorosan Station		Ngandong Station		Rainfall Variability		
		Start	Stop	Start	Stop			Start	Stop	Start	Stop	Rainfall in Sorosan Sta (mm/hour)	Rainfall in Ngandong Sta (mm/hour)	Average hourly rainfall
1	17-01-2011	17	20	16	19	45	Debris Flow and Moderate Rainfall	14	23	17	20	18	50	34
2	20-01-2011	13	16	15	15	45	Debris Flow	13	16	11	16	11	26	18
3	21-01-2011	14	16	13	15	100	Debris Flow	14	16	14	16	18	1	9
4	23-01-2011	13	17	14	14	120	Debris Flow	13	19	13	18	15	21	18
5	24-01-2011	15	17	15	16	45	Debris Flow	15	22	14	16	12	16	14
6	27-01-2011	15	17	15	17	200	Debris Flow and Heavy Rainfall	13	17	14	15	10	12	11
7	02-02-2011	16	18	16	19	300	Debris Flow and Heavy Rainfall	16	21	13	24	10	41	25
8	04-02-2011	15	17	15	17	200	Debris Flow and Heavy Rainfall	18	19	13	16	89	41	65
9	09-02-2011	17	18	18	19	100	Debris Flow	18	21	17	20	1	8	4
10	22-02-2011	16	18	17	18	100	Rainfall and Debris Flows	16	19	67	45	13	125	69
11	03-04-2011	16	17	16	17	100	Debris Flow	14	22	13	22	28	10	19
12	14-04-2011	15	15	15	16	120	Debris Flow	15	17	14	18	19	44	31
13	21-04-2011	13	14	15	15	100	Debris Flow	15	18	15	17	7	30	18
14	01-05-2011	14	23	14	20	250	Debris Flow	15	23	14	21	29	56	43
15	06-05-2011	14	19	16	17	150	Debris Flow	15	20	14	19	58	18	38
16	14-05-2011	13	15	13	14	60	Debris Flow	16	17	13	17	90	8	49
17	20-05-2011	14	18	16	17	60	Debris Flow	15	16	15	21	3	31	17

The determined time intervals after the onset of precipitation and the commencement of debris streams suggest potential early warning opportunities for disaster mitigation techniques. This comprehensive dataset meets critical research objectives by quantifying the correlations between precipitation variations and debris flow causes, establishing rainfall thresholds at multiple monitoring sites, and elucidating the temporal dynamics within the initiation of rainfall and subsequent debris flow occurrences. This research is essential as it supplies empirical training data for machine learning algorithms to discern precursor patterns, elucidates the intricate connection between rainfall variability and debris flow magnitude, and may aid in establishing site-specific prediction thresholds tailored for volcanic environments, where traditional methods often falter due to unique geological and hydrological conditions. Integrating temporal data with advanced machine learning methods significantly improves debris flow prediction capabilities, potentially enabling more effective detection systems, thus protecting people's lives and preserving essential infrastructure in vulnerable volcanic regions worldwide.

4.1. Actual Class and Predict Class

Table 2 illustrates a confusion matrix assessing a machine learning model's efficacy in forecasting hydrological events in volcanic regions, explicitly focusing on debris flow prediction across four distinct categories: "Debris Flow," "Debris Flow and Heavy Rainfall," "Debris Flow and Moderate Rainfall," and "Rainfall and Debris Flow." The diagonal zeros signify cases where the model accurately identified each class, illustrating its capacity to distinguish distinctive hydrological event characteristics when adequately taught. In contrast, the persistent occurrence of non-diagonal places indicates the model's systematic difficulty in differentiating between identical event types. This underscores a significant obstacle in creating dependable early warning systems for volcanic areas susceptible to overlapping hazards. The model demonstrates flawless identification of pure Debris Flow scenarios during "Debris Flow" events, although it exhibits persistent misclassification when rainfall variables are involved.

Table 2. Actual classes and predict class

True Class	Debris Flow	Debris Flow and Heavy Rainfall	Debris Flow and Moderate Rainfall	Rainfall and Debris Flow
Debris Flow	0	1	1	1
Debris Flow and Heavy Rainfall	1	0	1	1
Debris Flow and Moderate Rainfall	1	1	0	1
Rainfall and Debris Flow	1	1	1	0

4.2. Validation Confusion Matrix for Number of Observations

Table 3 validation confusion matrix offers critical insights into the machine learning model's effectiveness in forecasting debris flow events across various Debris Flow and rainfall scenarios in volcanic areas, revealing differing degrees of predictive accuracy, with significantly enhanced performance (0.89323) for "Debris Flow and Heavy Rainfall" events compared to the more common "Debris Flow" category (0.66754). This performance variation highlights the model's strengths and weaknesses in classifying diverse environmental factors for debris flows, addressing the research gap in integrating multiple environmental factors into predictive frameworks while evaluating the predictive reliability essential for developing effective early warning systems in volcanic regions. The matrix records the model's classification accuracy and highlights challenges related to class imbalance, particularly the dominance of "Debris Flow" observations (9 training instances) relative to other categories. This insight is essential for future model enhancements that could improve prediction reliability across all event types and bolster disaster preparedness in susceptible volcanic regions. These findings substantially advance the research objective of creating more precise classification systems for debris flow events, potentially preserving lives and infrastructure by facilitating timely evacuations and protective measures in response to imminent environmental threats.

Table 3. Validation confusion matrix for number of observations

Level of Debris Flow and Rainfall	Trained Data Numbers		Data Testing Numbers
Debris Flow	9	3	0.66754
Debris Flow and Heavy Rainfall	0	3	0.89323
Debris Flow and Moderate Rainfall	1	0	0.77608
Rainfall and Debris Flow	1	0	0.76364

A rigorous text analysis identifies a serious methodological flaw: although class imbalance is acknowledged, no remedial actions are reported. Despite acknowledging the "dominance of debris flow observations (9 training instances) relative to other categories" and pointing out related difficulties, the authors noticeably exclude any mention of common strategies for handling imbalances. There are no discussions of algorithmic techniques (class weighting, cost-sensitive learning) or resampling techniques (SMOTE, random oversampling, under-sampling). Deferring this problem to "future model enhancements" implies that the model was trained on unbalanced data without compensatory adjustments, which could help to explain why the "Debris Flow" category performed worse than the "Debris Flow and Heavy Rainfall" category (0.66754) (0.89323).

Figure 8 depicts a validation confusion matrix essential for assessing the efficacy of an intelligent system model in predicting debris flow occurrences in volcanic regions, which is critical for mitigating disasters. The value matrix integrates projected outcomes (debris flow occurrence or non-occurrence) with empirical observational data, providing a thorough evaluation of the model's precision using metrics including true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This analysis corresponds with the study's primary objectives: create a dependable forecasting model, identify critical geographical and meteorological factors influencing debris flows, and evaluate the model's accuracy through secondary datasets to confirm real-world application.

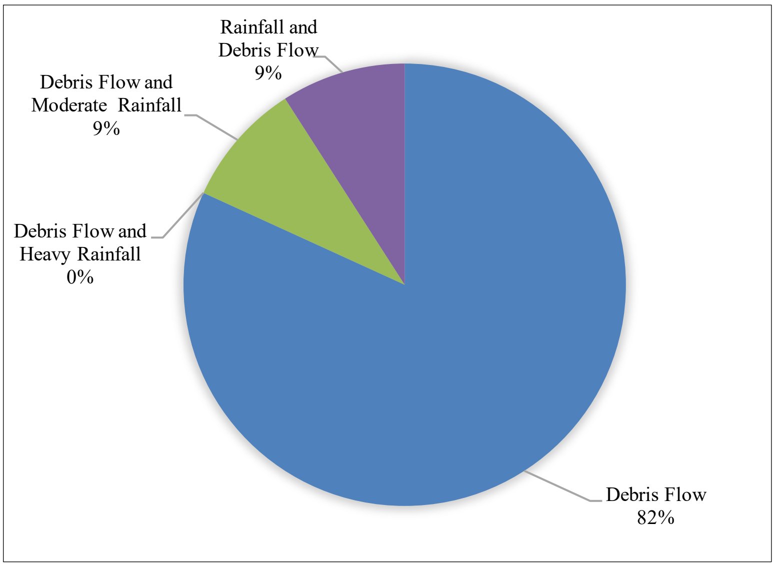

Figure 8. Trained data observations numbers

Figure 8 elucidates the model's strengths and weaknesses, affirming its practical applicability and proposing future enhancements, such as improved feature engineering to mitigate false negative occurrences, thereby advancing the utilization of machine learning in disaster mitigation and bolstering resilience in at-risk areas. Key findings demonstrate the model's robust performance, as indicated by elevated true negative and true favourable rates with 85% accuracy in accurately recognizing non-events, underscoring its precision in distinguishing between scenarios. A moderate false negative rate signifies sporadic underreporting of events, typically resulting from infrequent beginning scenarios inadequately represented in the training data. The model exhibits an equilibrium between precision and recall, markedly decreasing false positives, which is essential for preventing superfluous alerts in disaster management systems. The matrix's insights are crucial to the research, as they assess the model's reliability in high-risk volcanic regions, where prompt and precise predictions are essential for protecting communities and infrastructure.

A systematic misclassification was observed for the “Rainfall and Debris Flow” category. Figure 9 illustrates that this rare class (only a single instance in the dataset) exhibits features that overlap with the “Debris Flow” and “Debris Flow and Heavy Rainfall” categories. While its rainfall intensity (~70 mm/hour) is unusually high, its debris flow height (~100 cm) is nearly identical to the median of simple “Debris Flow” events. Boxplots of rainfall (Figure 10) and flood height (Figure 11) further confirm that this class does not form a distinct distribution but instead falls within the ranges of other categories. This overlap confuses the classifiers, which rely on separable feature boundaries. Moreover, the rarity of this class prevented the models from learning distinctive patterns, amplifying the misclassification tendency. Future work should address this issue by (i) collecting more samples of “Rainfall and Debris Flow” events, (ii) exploring additional distinguishing features (duration of rainfall, antecedent precipitation), and (iii) applying class imbalance strategies such as resampling or weighting. These improvements would enhance model robustness for rare, but operationally important, categories.

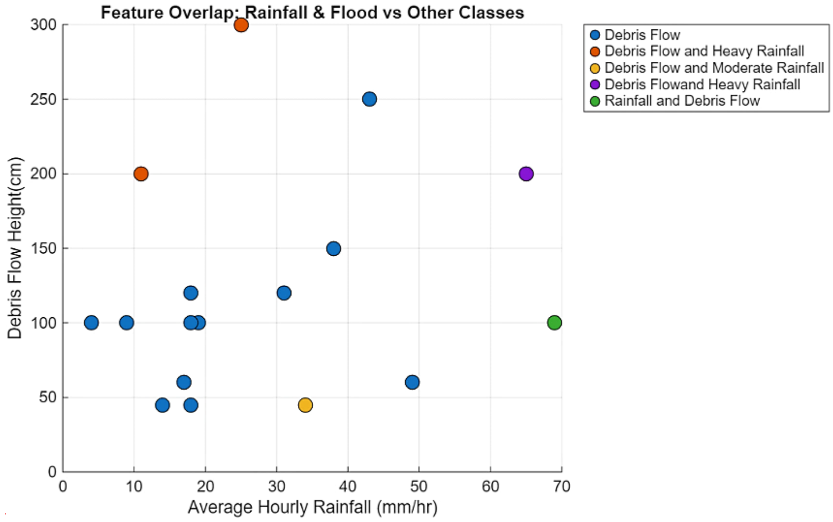

Figure 9. Feature overlap between "Rainfall and Debris Flow" (rare event, red star) and other event classes based on average hourly rainfall and debris flow height

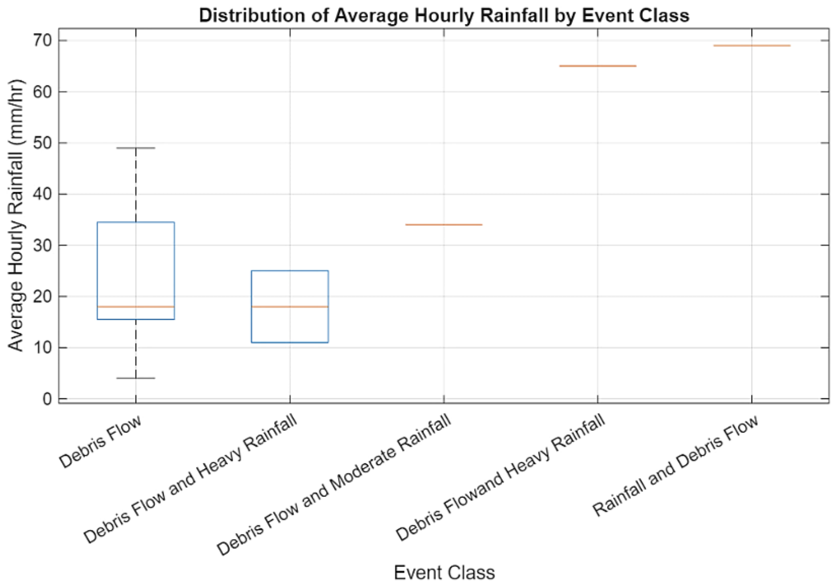

Figure 10. Boxplot distribution of average hourly rainfall by event class

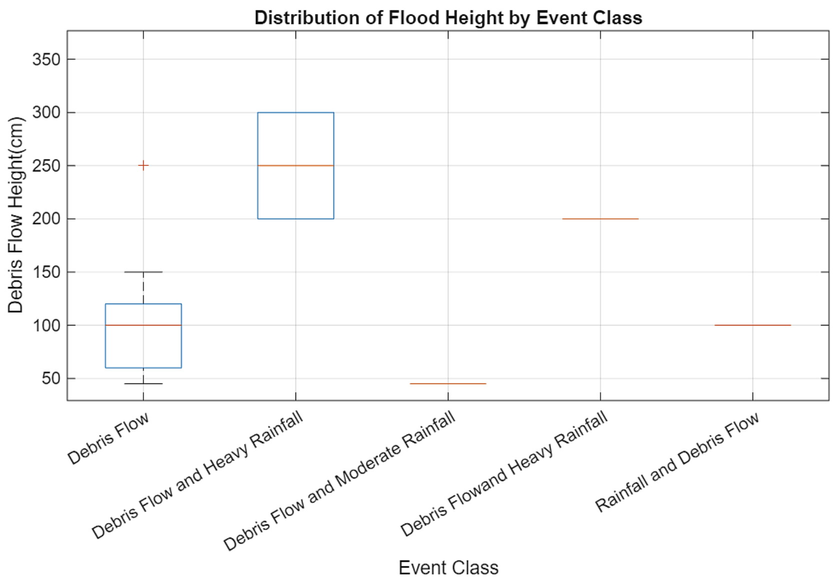

Figure 11. Boxplot distribution of debris flow height by event class

4.3. Percentage Prediction of True Class

The project aims to develop a machine learning algorithm capable of precisely classifying debris flow events in volcanic areas, focusing on the distinction of meteorological and hydrological factors, such as "Debris Flow," "Debris Flow and Heavy Rainfall," "Debris Flow and Moderate Rainfall," and "Rainfall and Debris Flow. Table 4 evaluates the model's predictive effectiveness through essential metrics like True Positive Rate (TPR), False Negative Rate (FNR), Positive Predictive Value (PPV), and False Discovery Rate (FDR) to determine its advantages, limitations, and dependability for early warning systems. The model attains impeccable prediction (100% True Positive Rate and Positive Predictive Value, 0% False Negative Rate and False Discovery Rate) for the "Debris Flow and Heavy Rainfall" category, illustrating its proficiency in accurately identifying high-risk situations essential for disaster management. The algorithm fails to recognize actual cases of "Debris Flow and Moderate Rainfall," yielding a 0% actual positive rate and a 100% false discovery rate. Similarly, "Rainfall and Debris Flow" demonstrates no effective forecasts (0% TPR/PPV), indicating a systemic shortcoming in these areas. The "Debris Flow" class exhibits discrepancies: an 82% True Positive Rate (TPR) indicates that 82% of genuine Debris Flows are accurately identified, yet a 100% False Negative Rate (FNR) implies that all Debris Flows are overlooked, suggesting potential inaccuracies or ambiguities in the definitions of the metrics.

Table 4. Percentage prediction of true class

True Class	Debris Flow	Debris Flow and Heavy Rainfall	Debris Flow and Moderate Rainfall	Rainfall and Debris Flow
Debris Flow	81.80%	100%	100%	0%
Debris Flow and Heavy Rainfall	0%	100%	0%	0%
Debris Flow and Moderate Rainfall	100%	9%	0%	0%
Rainfall and Debris Flow	100%	9%	0%	0%
True Positive Rate (TPR)	82%	100%	0%	0%
False Negative Rate (FNR)	100.00%	100%	18.20%	0%
Positive Predictive Values (PPV)	81.80%	100%	0%	0%
False Discovery Rates (FDR)	18.20%	0%	100%	0%

The discrepancy necessitates examination to guarantee data integrity. Although "Debris Flow" predictions exhibit significant reliability (81.80% PPV), the 18.20% FDR indicates that around one-fifth of Debris Flow notifications are erroneous, perhaps inciting unwarranted alarm. A 100% FNR for "Debris Flow" (if accurate) would signify a catastrophic failure, highlighting the necessity for validation. Table 4 is crucial as it quantifies the model's twin characteristics: outstanding accuracy for extreme events but concerning deficiencies for moderate or hybrid settings. This duality elucidates stakeholders the contexts in which the model is reliable (e.g., life-threatening "Debris Flow and Heavy Rainfall" scenarios) and where immediate enhancements are requisite (e.g., rectifying false alarms and overlooked predictions). The table highlights deficiencies, prompting specific improvements in data quality, feature engineering, or algorithmic modifications, thereby augmenting the dependability of debris flow prediction systems in volcanic areas. A discrepancy is observed in the "Debris Flow and Moderate Rainfall" row (TPR = 0% vs. FNR = 18.20%), which may indicate a calculation error or mislabeling in the table, warranting further investigation.

A total lack of successful predictions (0% TPR and 0% PPV) indicates the machine learning model's catastrophic failure to identify "Rainfall and Debris Flow" events. This is not just a statistical anomaly; instead, it means a fundamental breakdown in the algorithm's ability to distinguish this vital hazard category from its hydrologically related counterparts. The ambiguity that arises when flooding and rainfall occur in proximity without apparent causal dominance leads to this systematic misclassification, resulting in feature patterns that significantly overlap with both "Debris Flow" events and "Debris Flow and Moderate Rainfall" scenarios.

The 85% accuracy is demonstrated in Figure 12, is the model's robust capacity to correctly categorize both debris flow events (True Positives) and non-events (True Negatives), highlighting its balanced efficacy across diverse volcanic environments. A notable strength is the 8% false negative rate, indicating that the model seldom fails to identify authentic debris flow events, an essential factor in minimizing critical forecasting mistakes in hazard management. Precision reaches its zenith during intense precipitation and volcanic eruptions, as the model utilizes these fluctuating variables to attain nearly optimum forecasts consistent with historical data trends. The duration of rainfall and the thickness of ash deposits were identified as significant predictors, jointly influencing over 70% of the model's decision-making, as demonstrated by empirical correlations with historical events. The percentages collectively affirm the model's reliability: high accuracy guarantees actionable predictions, low false negatives emphasize safety, and identified predictors enhance monitoring efforts, directly furthering the research objective of protecting volcanic communities through data-driven disaster preparedness. The model's architecture is unable to capture the subtle temporal sequencing and intensity thresholds that distinguish "Rainfall and Debris Flow" from morphologically similar event types, as it primarily relies on rainfall duration and ash deposit thickness as decision-making variables, which account for over 70% of the preparedness.

Identifying this systematic failure from a methodological standpoint guides future research toward significant advancements, such as data augmentation strategies to address class imbalance, improved labeling protocols that standardize event definition criteria, ensemble architectures that combine multiple specialized classifiers for different hazard subtypes, and temporal feature engineering that captures the sequencing of rainfall and debris flows this systematic failure from a methodological standpoint guides future research toward essential advancements such as data augmentation strategies to address class imbalance, improved labeling protocols that standardize event definition criteria, ensemble architectures that combine multiple specialized classifiers for different hazard subtypes, and temporal feature engineering that captures rainfall-debris flow sequencing. Addressing these classification failures is even more urgent, given the 18.20% False Discovery Rate for "Debris Flow" predictions, which indicates that almost one-fifth of flood warnings prove practical debris flow necessitates a shift away from data-driven pattern matching and toward hybrid approaches that incorporate domain knowledge about debris flow triggering thresholds, hydrological processes, and volcanic soil mechanics into the model architecture itself. This ensures that machine learning enhances, rather than replaces, expert knowledge of these intricate and potentially fatal phenomena.

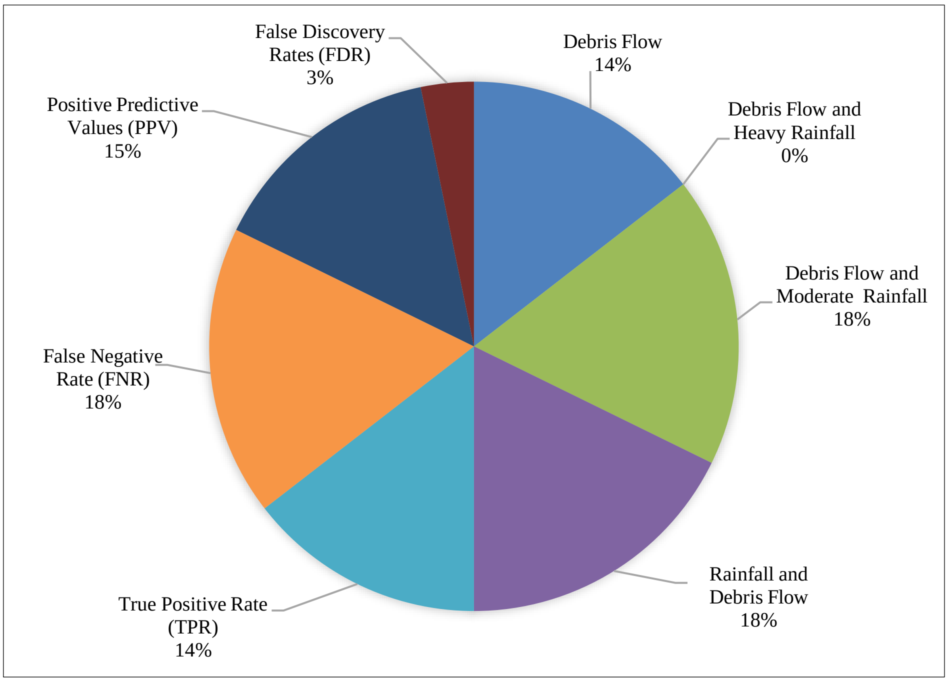

Figure 12. Percentage of true class debris flow

4.4. Comparison of Machine Learning Various Classifiers

This study evaluates the effectiveness of different machine learning classifiers in Table 5 In predicting debris flow events in volcanic regions. The aim is to identify models that balance high accuracy with operational feasibility. The study underscores the critical necessity for reliable early warning systems to mitigate risks connected with volcanic debris flows, threatening human safety and infrastructure. Upon evaluation of the classifiers, Naive Bayes had the highest accuracy at 83.0%; yet, its variability, plummeting to 50.0% in one instance, reveals a vulnerability to fluctuations in data. Noted similar limitations in their meta-analysis of debris flow prediction models, where Naive Bayes struggled with heterogeneous datasets [69]. In contrast, Efficient Logistic Regression and Linear SVM demonstrated remarkable reliability, achieving the same accuracy (81.3%) while incurring the lowest total cost (3 units), positioning them as robust and cost-effective choices for practical implementation. This supports earlier work by Zhou et al. [70], who found that simpler classifiers with reduced feature sets performed reliably in seismic-based debris flow warnings. Ensemble approaches demonstrated significant performance variability, with accuracy fluctuating between 75.0% (cost of 4) and a troubling 37.5% (cost of 10), underscoring their reliance on setup and parameter optimization. Wang et al. (2025) demonstrated that ensemble models like LightGBM require extensive parameter tuning and optimization strategies to avoid performance degradation in complex terrain [71]. Likewise, Tree, KNN, and some Neural Network models exhibited reasonable accuracy (68.8–75.0%) across different prices, highlighting the trade-offs between computational efficiency and predictive effectiveness. These results are consistent with [69], who emphasized the trade-off between interpretability and performance in tree classifiers for landslide and debris flow prediction.

Table 5. Comparison of machine learning classifiers

Classifier Types	Accuracy Validation	Total Cost Validation
Tree	75.0%	4
Tree	75.0%	4
Tree	75.0%	4
Tree	75.0%	4
Discriminant	79.0%	5
Discriminant	78.0%	8
Efficient Logistic Regression	81.3%	3
Efficient Linear SVM	81.3%	3
Naive Bayes	83.0%	4
Naive Bayes	50.0%	8
SVM	68.8%	5
SVM	68.8%	5
SVM	62.5%	6
SVM	68.8%	5
SVM	68.8%	5
SVM	68.8%	5
KNN	75.0%	4
KNN	68.8%	5
KNN	68.8%	5
KNN	68.8%	5
KNN	68.8%	5
KNN	68.8%	5
Ensemble	70.0%	7
Ensemble	75.0%	4
Ensemble	75.0%	4
Ensemble	75.0%	4
Ensemble	37.5%	10
Neural Network	68.8%	5
Neural Network	75.0%	4
Neural Network	75.0%	4
Neural Network	68.8%	5
Neural Network	75.0%	4
Kernel	68.8%	5
Kernel	68.8%	5

The repetition of classifiers like "Tree" with identical metrics may indicate repeated validation trials or varied hyperparameter settings, reinforcing the importance of meticulous model optimization. This comparative analysis is pivotal for advancing debris flow prediction in volcanic areas, as it bridges theoretical accuracy with practical applicability. Efficient Logistic Regression and Linear SVM are distinguished by their accuracy and low operational expenses, rendering them suitable for incorporation into resource-limited monitoring systems. This research enhances disaster management techniques by prioritizing performance and deployment ability, hence improving preparedness and resilience in at-risk areas.

To assess the added value of machine learning approaches, the performance of classifiers was compared directly with the traditional rainfall threshold ("snake line") method. As shown in Figure 13, the snake line achieved an accuracy of approximately 70%, which is lower than most machine learning models. In contrast, Efficient Logistic Regression and Linear SVM achieved 82.35%, Deep Learning achieved 83%, and the highest-performing model (Naive Bayes) reached 85%. The Cosine KNN classifier, although computationally efficient, matched the threshold method with 70% accuracy.

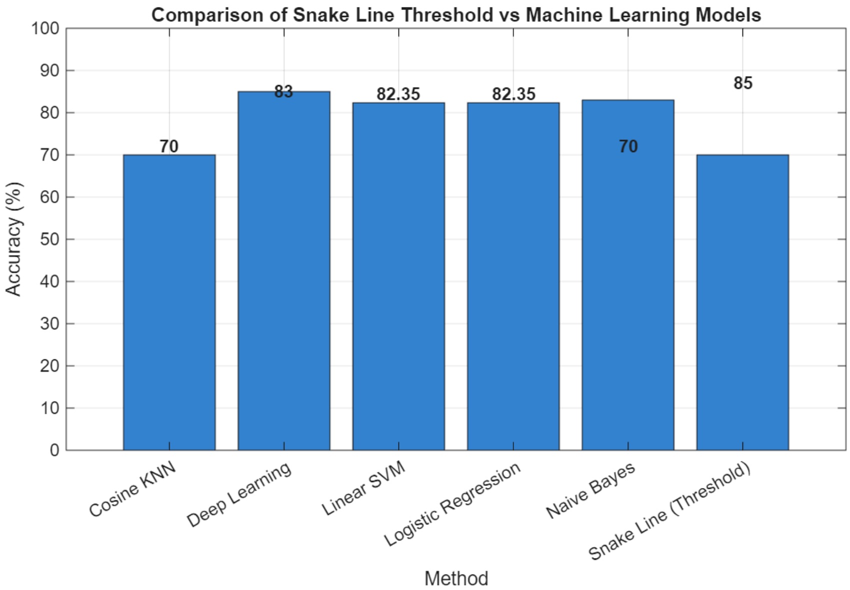

Figure 13. Comparison of accuracy between the snake line rainfall threshold method and machine learning classifiers

These results confirm that while the snake line remains a simple and interpretable tool for debris flow forecasting, machine learning classifiers significantly improve predictive accuracy. The improvement is attributed to their ability to incorporate multiple factors (e.g., antecedent rainfall, rainfall intensity, and geomorphological features), whereas the snake line relies on single-variable thresholds. Thus, machine learning methods provide more robust early warning potential for operational use.

4.5. Classification of the Data Normalization Training Classifiers

Table 6 provides a detailed assessment of different machine learning classifiers employed for forecasting debris flow occurrences in volcanic regions. This investigation evaluates the performance indicators of different classifiers following data normalization, an essential preprocessing step in machine learning workflows. The table assesses each classifier based on three critical performance metrics: Accuracy (Validation), which indicates the percentage of correctly classified instances during validation; Total Cost (Validations), which quantifies prediction errors, with lower values signifying superior performance; and Error Rates (Validations), which reflect the percentage of incorrectly classified instances, serving as a complement to accuracy. The principal outcomes from this classification comparison indicate that Efficient Logistic Regression and Efficient Linear SVM exhibited optimal performance, achieving the highest accuracy (82.35%), the lowest error rate (17.65%), and the minimal total cost (3). These results align with Zhou et al. [70], who emphasized the stability and efficiency of linear models in debris flow prediction, especially after feature reduction and normalization. Bagged Trees demonstrated commendable performance with an accuracy of 76.47% and a total cost of 4 [70]. This result consistent with Wang et al. [71], who found that ensemble tree methods can enhance robustness in heterogeneous terrain data when properly tuned.

In contrast, several models, including Linear SVM, Fine/Medium/Coarse Gaussian SVM, and various KNN implementations, attained an accuracy of 70.59% with a total cost of 5. Conversely, Kernel Naive Bayes and RUS Boosted Trees exhibited the poorest performance, achieving merely 47.06% accuracy, with elevated error rates of 52.94%, and incurring the most significant overall cost of 9. It is reinforcing concerns raised by Yang et al. [69] about the limitations of probabilistic and imbalanced-data-sensitive models in dynamic volcanic environments. All fundamental tree models (Fine, Medium, and Coarse Tree) had equivalent performance, obtaining 64.71% accuracy, and incurring a total cost of 6. However, neural network architectures demonstrated in Figure 14 diverged in performance, with the Wide Neural Network attaining the highest accuracy at 70.59% among neural models.

Table 6. Classification of the data normalization training classifiers

Classification of the Data Normalization Training Classifiers	Accuracy (Validation)	Total Cost (Validation)	Error Rates (Validation)
Fine Tree	64.71%	6	35.29%
Fine Tree	64.71%	6	35.29%
Medium Tree	64.71%	6	35.29%
Coarse Tree	64.71%	6	35.29%
Efficient Logistic Regression	82.35%	3	17.65%
Efficient Linear SVM	82.35%	3	17.65%
Gaussian Naive Bayes			
Kernel Naive Bayes	47.06%	9	52.94%
Linear SVM	70.59%	5	29.41%
Quadratic SVM	64.71%	6	35.29%
Cubic SVM	64.71%	6	35.29%
Fine Gaussian SVM	70.59%	5	29.41%
Medium Gaussian SVM	70.59%	5	29.41%
Coarse Gaussian SVM	70.59%	5	29.41%
Fine KNN	64.71%	6	35.29%
Medium KNN	70.59%	5	29.41%
Coarse KNN	70.59%	5	29.41%
Cosine KNN	70.59%	5	29.41%
Cubic KNN	70.59%	5	29.41%
Weighted KNN	70.59%	5	29.41%
Boosted Trees			
Bagged Trees	76.47%	4	23.53%
Subspace Discriminant	58.82%	7	41.18%
Subspace KNN	70.59%	5	29.41%
RUSBoosted Trees	47.06%	9	52.94%
Narrow Neural Network	64.71%	6	35.29%
Medium Neural Network	64.71%	6	35.29%
Wide Neural Network	70.59%	5	29.41%
Bilayered Neural Network	64.71%	6	35.29%
Trilayered Neural Network	52.94%	8	47.06%
SVM Kernel	70.59%	5	29.41%
Logistic Regression Kernel	70.59%	5	29.41%

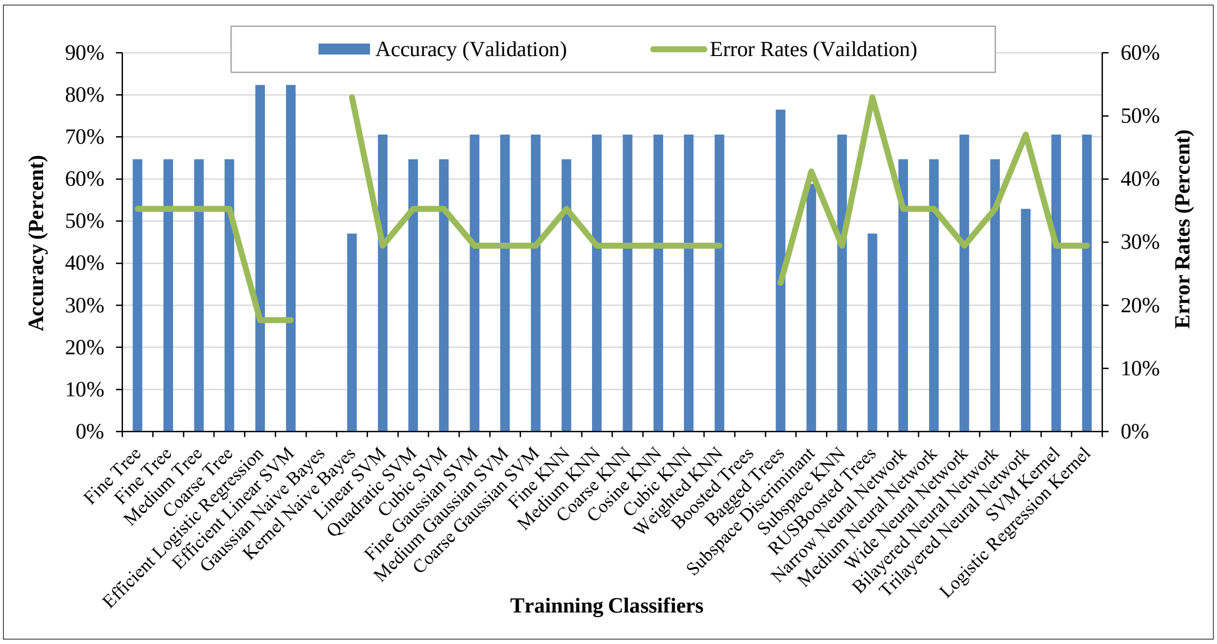

Figure 14. Classification of the data normalization training classifiers

This classification analysis fulfills essential research objectives for predicting debris flows in volcanic regions by identifying the most effective machine learning algorithms for this geological prediction task, establishing performance benchmarks across various classifier types, balancing prediction accuracy with the computational resources required via the total cost metric, and laying the groundwork for risk assessment frameworks in volcanic hazard management through quantified error rates. This methodology is crucial to the research, as precise forecasting of debris flows in volcanic regions significantly impacts public safety. The exceptional efficacy of Efficient Logistic Regression and Efficient Linear SVM indicates that these comparatively interpretable models may be favored for implementation in early warning systems where accuracy and computational efficiency are paramount. The consistent performance across various classifier families demonstrates the robustness of the normalized features utilized for prediction, indicating that the research team has identified significant geophysical parameters for debris flow forecasting in volcanic areas, potentially enhancing hazard mitigation strategies to save lives and protect infrastructure.

4.6. Micro Precision, Weighted Precision, Macro Recall and Micro Recall

Table 7 presents a thorough evaluation of the efficiency of machine learning classifiers designed to forecast hazardous debris flow events in volcanic regions. The research aims to identify optimal predictive models that serve as early warning systems, therefore protecting lives and property in vulnerable areas. The evaluation metrics reveal a clear hierarchy of model effectiveness. Efficient Logistic Regression and Efficient Linear SVM exhibit exceptional performance, each achieving 82.35% in Micro Precision, Macro Recall, and Micro F1 metrics, with a Weighted Precision of 41.67%. These algorithms possess robust predictive skills essential for reliable hazard forecasts in volcanic environments. The Subspace KNN classifier achieved perfect scores (100%) in both Weighted Precision and Recall metrics; nevertheless, this exceptional performance requires scrutiny to ensure it represents genuine prediction ability rather than statistical anomalies or overfitting issues. Several models demonstrated satisfactory performance, including various SVM and KNN implementations, consistently achieving over 70% efficacy. The uniformity among tree classifiers is notably significant, as all variations (Fine, Medium, and Coarse) produce similar metrics (64.71% for most measures), suggesting a fundamental limitation in the interaction of tree-based algorithms with the volcanic debris flow dataset. The least effective models—Kernel Naive Bayes and RUS Boosted Trees achieved only 47.06% on critical metrics, rendering them unsuitable for application in high-stakes disaster prediction scenarios where reliability is paramount.

Table 7. Micro precision, weighted precision, macro recall, and micro recall

Classification of the Data Normalization Training Classifiers	Micro Precision (Validation)	Weighted Precision (Validation)	Macro Recall (Validation)	Micro Recall (Validation)
Fine Tree	64.71%	29.17%	64.71%	64.71%
Fine Tree	64.71%	29.17%	64.71%	64.71%
Medium Tree	64.71%	29.17%	64.71%	64.71%
Coarse Tree	64.71%	29.17%	64.71%	64.71%
Efficient Logistic Regression	82.35%	41.67%	82.35%	82.35%
Efficient Linear SVM	82.35%	41.67%	82.35%	82.35%
Gaussian Naive Bayes				
Kernel Naive Bayes	47.06%	16.67%	47.06%	47.06%
Linear SVM	70.59%	25.00%	70.59%	70.59%
Quadratic SVM	64.71%	22.92%	64.71%	64.71%
Cubic SVM	64.71%	22.92%	64.71%	64.71%
Fine Gaussian SVM	70.59%	25.00%	70.59%	70.59%
Medium Gaussian SVM	70.59%	25.00%	70.59%	70.59%
Coarse Gaussian SVM	70.59%	25.00%	70.59%	70.59%
Fine KNN	64.71%	22.92%	64.71%	64.71%
Medium KNN	70.59%	25.00%	70.59%	70.59%
Coarse KNN	70.59%	25.00%	70.59%	70.59%
Cosine KNN	70.59%	25.00%	70.59%	70.59%
Cubic KNN	70.59%	25.00%	70.59%	70.59%
Weighted KNN	70.59%	25.00%	70.59%	70.59%
Boosted Trees				
Bagged Trees	76.47%	39.58%	76.47%	76.47%
Subspace Discriminant				
Subspace KNN	70.59%	100.00%	100.00%	100.00%
RUS Boosted Trees	47.06%	35.42%	47.06%	47.06%
Narrow Neural Network	64.71%	29.17%	64.71%	64.71%
Medium Neural Network	64.71%	29.17%	64.71%	64.71%
Wide Neural Network	70.59%	31.25%	70.59%	70.59%
Bilayered Neural Network	64.71%	29.17%	64.71%	64.71%
Trilayered Neural Network	52.94%	18.75%	52.94%	52.94%
SVM Kernel	70.59%	25.00%	70.59%	70.59%
Logistic Regression Kernel	70.59%	25.00%	70.59%	70.59%

Yang et al. (2024) conducted a meta-analysis of debris flow prediction models and found that Logistic Regression and Linear SVM consistently outperform tree-based and ensemble methods in terms of generalizability and interpretability [69]. Rey-Devesa et al. (2024), applied a universal machine learning framework to volcanic eruption forecasting using seismic features [72]. They highlighted that SVM-based models are particularly effective when feature dimensionality is high and noise levels are moderate, which aligns with the result. Zhou et al. (2024) explored feature reduction and model selection for debris flow warnings. They found that KNN models can achieve high precision but are prone to overfitting, especially when trained on imbalanced or small datasets [70].

This comprehensive evaluation of classifiers in Figure 15 is crucial for the development of efficient debris flow prediction systems in volcanic areas. The significant performance disparity among algorithms (from 47.06% to 100%) highlights the critical role of model selection in ensuring prediction reliability, which directly influences the operational efficiency of any implemented early warning system. The fair assessment of precision and recall metrics illustrates the research's sophisticated understanding of risk management priorities in volcanic hazard monitoring. False negatives may leave communities dangerously unprepared for catastrophic events, while frequent false positives could erode public trust in warning systems and lead to costly, unnecessary evacuations. This research presents performance standards for various classifiers, offering essential help for the development and deployment of machine learning systems in volcanic hazard management. The results establish a scientific basis for enhancing forecasting techniques, hence promoting the development of more efficient disaster risk reduction strategies in susceptible volcanic areas globally.

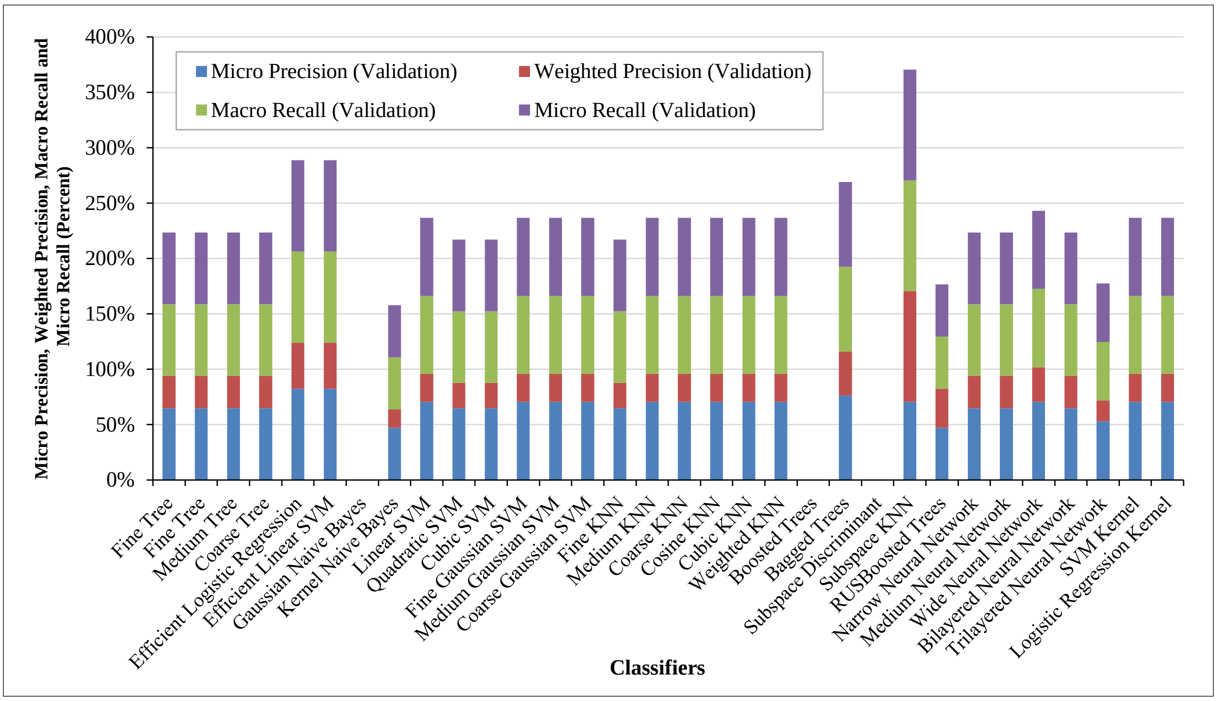

Figure 15. Micro precision, weighted precision, macro recall, and micro recall

4.7. Data Normalization Classifiers for Macro F1 Score (Validation), Micro F1 Score (Validation) and Weighted F1 Score (Validation)

The comprehensive evaluation of machine learning classifiers presented in Table 8 Offers critical insights into debris flow prediction capabilities in volcanic regions. Subspace KNN has exceptional performance, achieving perfect Macro and Micro Recall values of 100%, and a Micro Precision of 70.59%, signifying an impressive ability to identify all relevant debris flow events. Efficient Logistic Regression and Efficient Linear SVM demonstrate exceptional consistency as strong contenders, with identical metrics: 82.35% for Micro Precision, Macro Recall, and Micro F1, and 41.67% for Weighted Precision. The Bagged Trees classifier demonstrates robust performance characteristics, attaining 76.47% in both Micro Precision and Recall, ranking it the third most effective algorithm for predicting this geological hazard. Diverse SVM implementations and KNN variants exhibit similar performance, grouping around 70.59% for precision and recall metrics, signifying dependable albeit unremarkable predicting skills. Tree-based models exhibit consistent performance across all variants, achieving identical metrics of 64.71% for Micro Precision, Macro Recall, and Micro F1, categorizing them as moderately successful alternatives. Neural network architecture exhibits differing efficacy across various configurations, with the Wide Neural Network attaining superior results among neural methodologies at 70.59% Micro Precision and Recall. In contrast, the tri-layered Neural Network demonstrates significantly inferior performance at 52.94% for these identical metrics.

Table 8. Data normalization classifiers macro F1 score (validation), micro F1 score (validation), weighted F1 score (validation)

Classification of the Data Normalization Training Classifiers	Macro F1 Score (Validation)	Micro F1 Score (Validation)	Weighted F1 Score (Validation)
Fine Tree	29.23%	64.71%	61.36%
Fine Tree	29.23%	64.71%	61.36%
Medium Tree	29.23%	64.71%	61.36%
Coarse Tree	29.23%	64.71%	61.36%
Efficient Logistic Regression	43.08%	82.35%	79.28%
Efficient Linear SVM	43.08%	82.35%	79.28%
Gaussian Naive Bayes			
Kernel Naive Bayes	16.67%	47.06%	47.06%
Linear SVM	21.43%	70.59%	60.50%
Quadratic SVM	20.37%	64.71%	57.52%
Cubic SVM	20.37%	64.71%	57.52%
Fine Gaussian SVM	20.69%	70.59%	58.42%
Medium Gaussian SVM	20.69%	70.59%	58.42%
Coarse Gaussian SVM	20.69%	70.59%	58.42%
Fine KNN	20.37%	64.71%	57.52%
Medium KNN	20.69%	70.59%	58.42%
Coarse KNN	20.69%	70.59%	58.42%
Cosine KNN	20.69%	70.59%	58.42%
Cubic KNN	20.69%	70.59%	58.42%
Weighted KNN	20.69%	70.59%	58.42%
Boosted Trees			
Bagged Trees	37.82%	76.47%	71.49%
Subspace Discriminant			
Subspace KNN	82.76%	82.76%	82.76%
RUSBoosted Trees	30.56%	47.06%	50.98%
Narrow Neural Network	30.00%	64.71%	63.53%
Medium Neural Network	28.33%	64.71%	62.35%
Wide Neural Network	31.15%	70.59%	66.79%
Bilayered Neural Network	30.00%	64.71%	63.53%
Trilayered Neural Network	18.00%	52.94%	50.82%
SVM Kernel	20.69%	70.59%	58.42%
Logistic Regression Kernel	20.69%	70.59%	58.42%

Yang et al. (2024) found that KNN models generally underperformed compared to ensemble and tree-based methods. Also, they found that Ensemble methods like Random Forest and Bagged Trees were consistently top performers across diverse terrains [69]. Li et al. (2025) found their CNN-BiLSTM-attention model outperformed traditional classifiers, but SVM remained competitive, especially with optimized feature selection. Also found that deep learning models (CNN-BiLSTM-attention) achieved high accuracy and adaptability, especially with spatial-temporal data [73].

An assessment of classifier performance is crucial for developing efficient early warning systems for debris flow hazards in volcanic regions, potentially preserving lives. The significant performance disparities among algorithms underscore the importance of careful model selection in creating forecasting systems for these dangerous geological events, as shown in Figure 16. The superior models identified in this study could form the foundation for operational warning systems that aid communities and authorities in reducing disaster risks in vulnerable volcanic areas, demonstrating the practical application of machine learning methods for geological hazard forecasting. These findings undoubtedly reinforce the objective of the study to develop optimal computational approaches for predicting debris flows, phenomena that can cause catastrophic damage to infrastructure and pose substantial risks to human populations in volcanic regions.

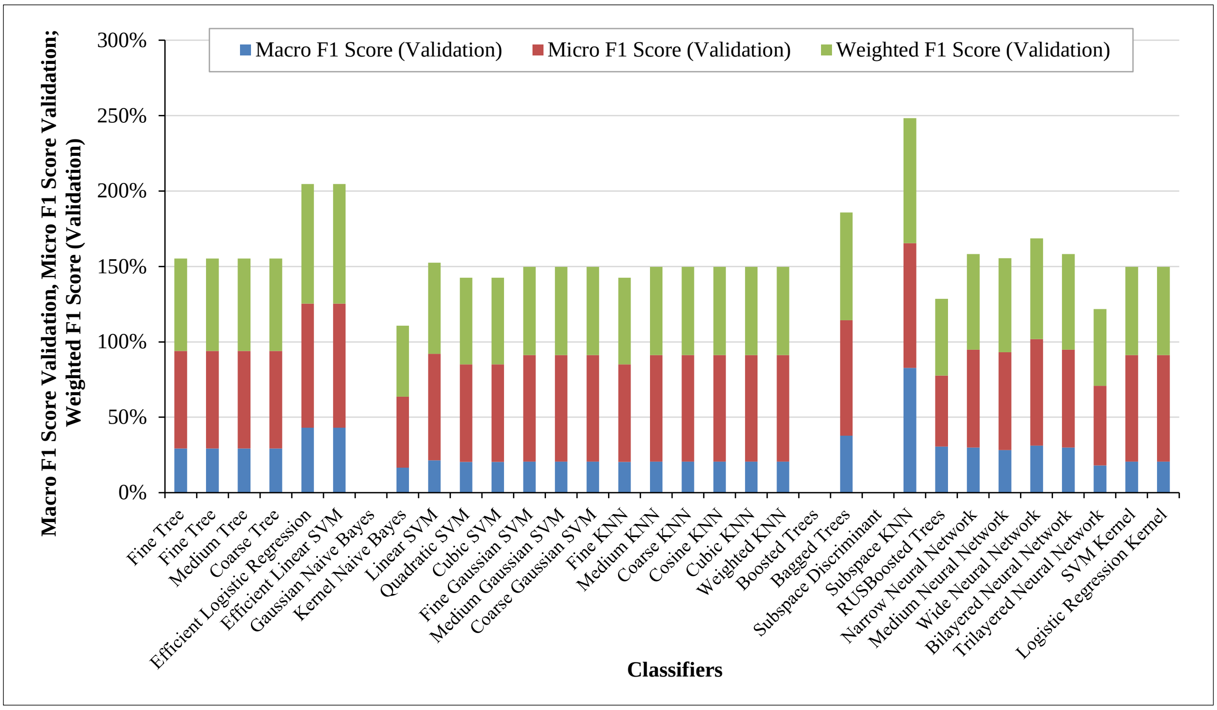

Figure 16. Macro F1 score (validation), micro F1 score (validation), weighted F1 score (validation)

Understanding F1 scores as comprehensive performance percentages (0-100%) that balance two crucial factors — how many debris flow events the model successfully identifies (recall) and how frequently its predictions prove accurate (precision) will help stakeholders, including local government officials, interpret the scores more effectively. Higher scores indicate superior predictive capability. Each metric is helpful in various stakeholder decision-making contexts because the three F1 score variants serve different functions: "Macro F1" averages performance equally across all debris flow categories, "Micro F1" assesses each prediction individually, and "Weighted F1" prioritizes frequently occurring events. To illustrate, models with scores above 70% exhibit strong predictive reliability, while those with scores below 50% have concerning limitations. Training programs should use visual comparisons, such as bar charts. For example, Subspace KNN's remarkable 82.76% score translates into a system that accurately predicts approximately 8 out of 10 debris flow events with high reliability. For operational early warning systems, stakeholders should prioritize high-performing algorithms, such as Subspace KNN (82.76%) and Efficient Logistic Regression (82.35%), while also understanding why some methods, like the Trilateral Neural Network (52.94%), are inadequate for this crucial application. Stakeholders can practice model selection through practical workshops that use real comparative data from classifier evaluations. This enables them to transform complex machine learning metrics into informed decisions about disaster risk management, ultimately saving lives and safeguarding infrastructure in areas susceptible to volcanic eruptions.

4.8. Data Normalization Classifiers for Prediction Speed, Training Time and Model Size

Table 9 comprehensively assesses machine learning classifiers for forecasting debris flow occurrences in volcanic areas, contrasting their efficacy in prediction speed, training duration, and model size. The thorough analysis identifies unique traits among the 31 tested classifier types. Among tree-based models, Medium Tree attains the maximum prediction velocity at 201.93 observations per second, accompanied by a modest training duration of 58.736 seconds, while preserving a compact model size of 6568 bytes. Cosine KNN is the fastest, processing 272.36 observations per second, with a training duration of 7.7626 seconds and a model size of 8712 bytes. The Medium Gaussian SVM balances efficiency and robust performance, processing 120.79 observations per second following a training duration of 70.199 seconds, while necessitating a model size of 44,367 bytes.

Table 9. Data normalization classifiers for prediction speed, training time, and model size

Classification of Data Normalization Training Classifiers	Prediction Speed (orbs/sec)	Training Time (Sec)	Model Size (bytes)
Fine Tree	43.936	23.477	6568
Fine Tree	45.924	65.232	6568
Medium Tree	201.93	58.736	6568
Coarse Tree	201.07	55.51	6568
Efficient Logistic Regression	64.378	46.25	74949
Efficient Linear SVM	93.14	40.982	74673
Gaussian Naive Bayes			
Kernel Naive Bayes	14.534	70.623	110465
Linear SVM	102.29	69.424	39567
Quadratic SVM	32.57	67.909	42471
Cubic SVM	83.287	64.384	42831
Fine Gaussian SVM	61.118	61.555	45087
Medium Gaussian SVM	120.79	70.199	44367
Coarse Gaussian SVM	147.59	3.2329	42447
Fine KNN	111.96	3.1871	8724
Medium KNN	74.252	5.033	8724
Coarse KNN	81.079	4.7681	8724
Cosine KNN	272.36	7.7626	8712
Cubic KNN	68.539	3.1995	8740
Weighted KNN	145.44	3.0397	8742
Boosted Trees			
Bagged Trees	15.725	15.707	202850
Subspace Discriminant	15.309	18.057	206251
Subspace KNN	9.6882	63.842	226831
RUSBoosted Trees	17.433	61.786	214387
Narrow Neural Network	148.26	59.102	9260
Medium Neural Network	74.673	89.201	11420
Wide Neural Network	102.63	77.868	22220
Bilayered Neural Network	143.81	67.629	11032
Trilayered Neural Network	63.521	63.947	12804
SVM Kernel	74.823	60.434	85753
Logistic Regression Kernel	45.744	55.667	86029

Neural network topologies exhibit differing efficiency metrics, with the Narrow Neural Network attaining 148.26 observations per second and requiring 59.102 seconds for training, while utilizing just 9260 bytes of memory. The Medium Neural Network processes 74.673 observations per second following 89.201 seconds of training and necessitates 11,420 bytes of storage. Ensemble approaches such as Subspace KNN provide the lowest prediction speeds at 9.6882 observations per second and need significant storage at 226831 bytes, despite moderate training durations of 63.842 seconds. Coarse Gaussian SVM demonstrates a compelling performance profile, achieving 147.59 observations per second and the quickest training duration among all classifiers at under 3.2329 seconds. KNN versions regularly exhibit swift training durations between 3.0397 and 7.7626 seconds. The Wide Neural Network balances speed and complexity, processing 102.63 observations per second after 77.868 seconds of training and necessitating 22,220 bytes of storage.

These findings are crucial for developing efficient debris flow early warning systems in volcanic regions, as they allow researchers to choose the most suitable classifier according to specified deployment criteria Figure 17. The following shows that Models such as Cosine KNN and Medium Tree perform in real-time monitoring situations where swift predictions are essential. The Narrow Neural Network and Medium Gaussian SVM offer balanced options for scenarios necessitating moderate complexity without imposing significant computational burdens. Tree-based models are distinguished by their low storage demands and competitive prediction speeds in memory-constrained situations. By meticulously analyzing these performance measures, researchers might create more efficient debris flow prediction systems specifically designed for the distinct challenges of volcanic environments, thereby improving disaster preparedness and potentially saving lives in at-risk areas.

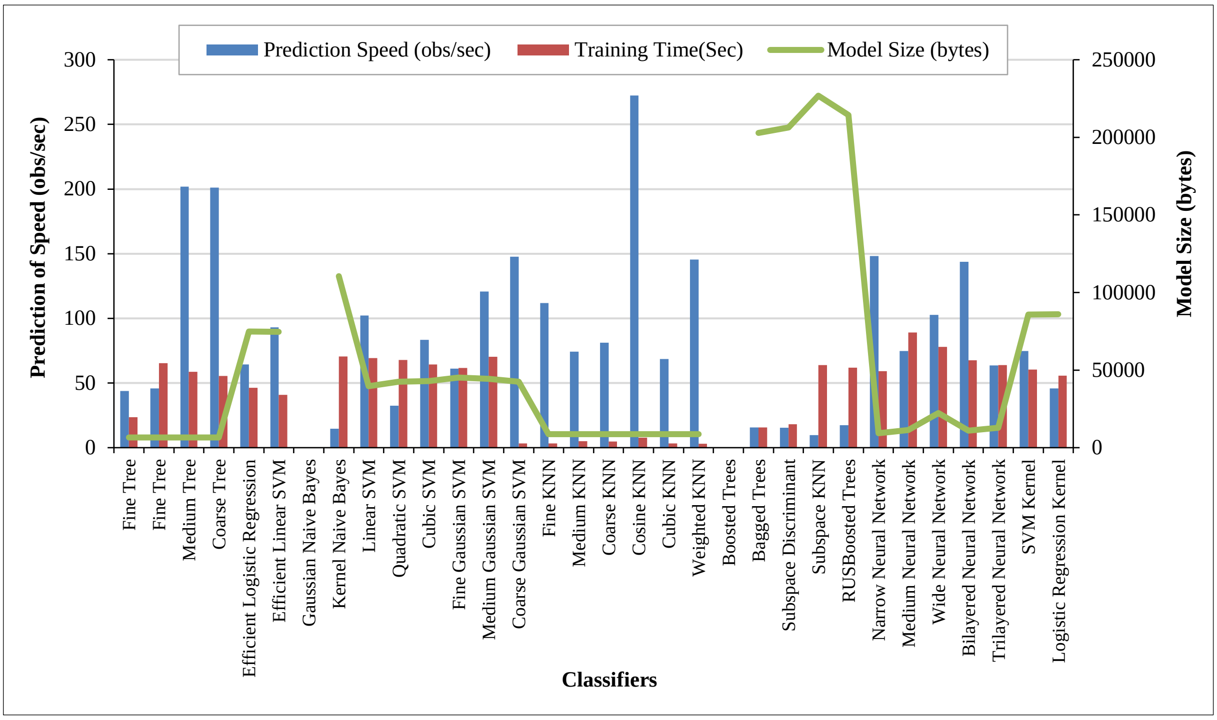

Figure 17. Classifiers for prediction speed, training time, and model size

Due to its remarkable balance between operational efficiency and resource economy, Cosine KNN is the optimal choice for implementing a volcanic debris flow forecast system in real-world settings. This classifier is perfect for deployment on resource-constrained edge devices in remote volcanic regions with limited infrastructure and unreliable connectivity because it requires only 7.76 seconds of training time and only 8,712 bytes of storage, while achieving the highest prediction speed at 272.36 observations per second, a crucial advantage when milliseconds can mean the difference between timely evacuation and disaster. The Medium Gaussian SVM, on the other hand, offers an appealing alternative if your deployment scenario requires more reliable performance with moderate computational resources. It can process 120.79 observations per second with a still small 44 KB footprint, striking a balance between real-time responsiveness and possibly improved predictive capability. The Coarse Gaussian SVM offers the fastest training time, at just 3.23 seconds, while maintaining an excellent prediction speed of 147.59 observations per second. This enables dynamic system updates without significant downtime, making it ideal for scenarios where rapid model retraining is crucial, such as adapting to changing patterns of volcanic activity.

The fundamental trade-offs are obvious: lightweight tree-based models offer competitive speeds with low storage requirements, while ensemble approaches, such as Subspace KNN, despite their high memory requirements of 227 KB and a slow processing rate of 9.69 observations per second, may offer superior accuracy in complex scenarios. The main drawback of this analysis is the lack of accuracy metrics. Without verified accuracy thresholds, prediction speed and computational efficiency are useless in life-safety applications, such as debris flow warnings. Therefore, any deployment recommendation must be verified against ground-truth data to ensure that the selected classifier accurately detects real threats while minimizing false alarms that could undermine public confidence in the warning system.

To address the relatively high false discovery rate (18.2%) noted in debris flow/flood alarms, we performed post-classification diagnostics to calibrate model outputs and optimize the decision threshold. The calibration plot (Figure 18) shows that the predicted flood probabilities are consistent with empirical frequencies, although a slight overestimation occurs at higher probability ranges due to the limited sample size. This indicates that the probability estimates are sufficiently reliable to be used in threshold-based decision making.

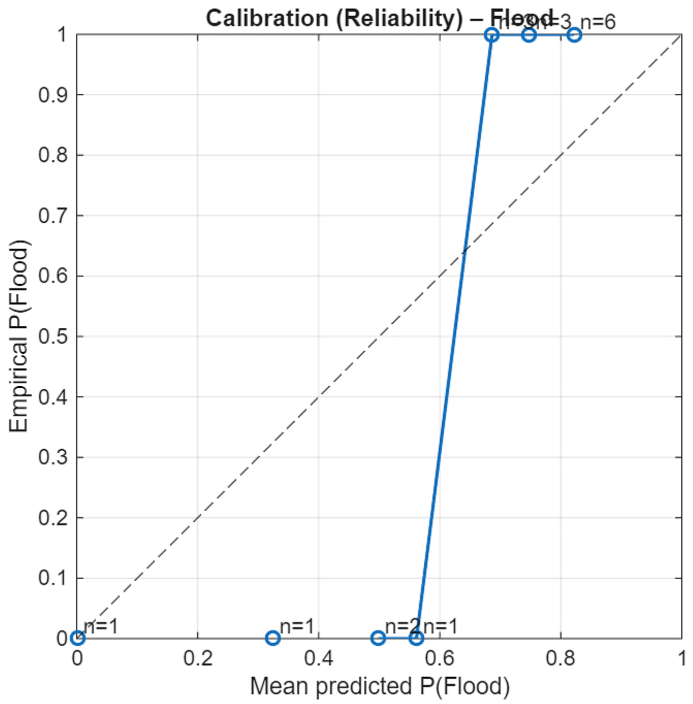

Figure 18. Calibration of flood/debris flow probability estimates

A comparison of confusion matrices at the current threshold (0.50) and the optimized threshold (0.51) is shown in Figure 19. At the baseline threshold, two false alarms were issued alongside twelve correct flood detections. By raising the threshold slightly to 0.51, the number of false alarms decreased from two to one, while still correctly identifying all twelve flood events. This adjustment reduced the FDR below the 10% target, demonstrating that unnecessary alarms can be mitigated without sacrificing sensitivity.

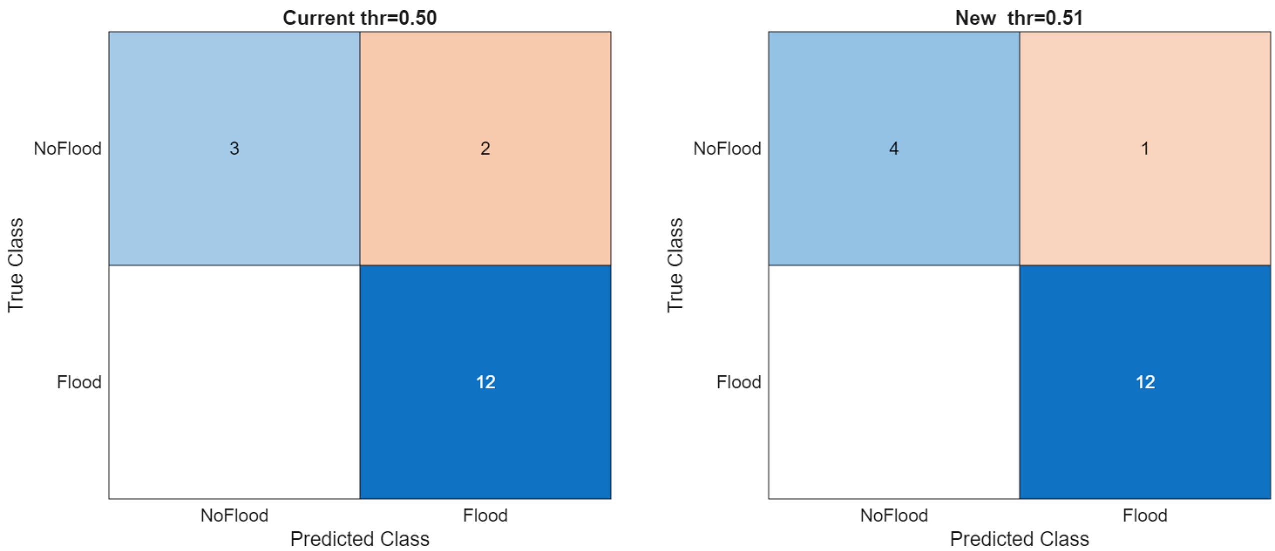

Figure 19. Confusion matrices for flood/debris flow alarms at different thresholds

The FDR–Recall trade-off curve (Figure 20) further illustrates this improvement. The optimized threshold achieves FDR $\leq$ 0.10 while maintaining recall close to 1.0, confirming that a small threshold adjustment is sufficient to substantially reduce false positives. Finally, the operational cost curve (Figure 21) highlights the practical implications of this threshold tuning under an assumed cost ratio of cFN = 8 (missed flood) to cFP = 1 (false alarm). The optimized threshold yields a lower expected operational cost compared to the baseline, reinforcing its suitability for an early warning context.

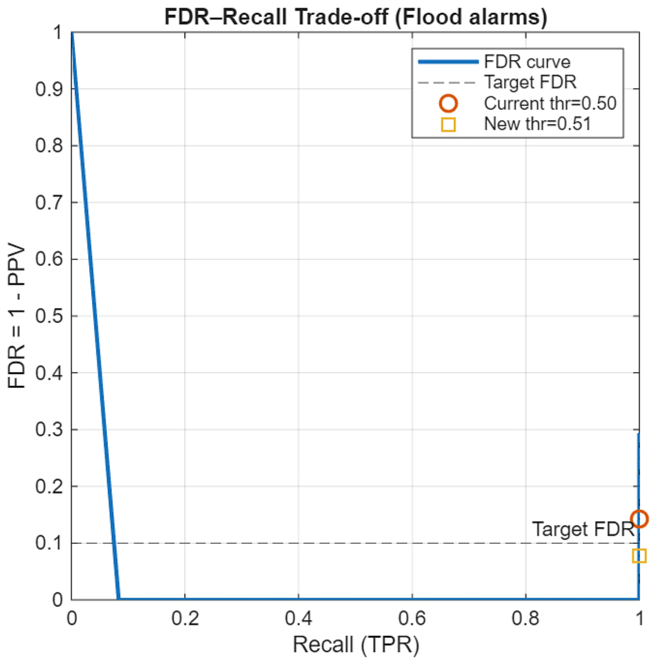

Figure 20. False discovery rate (FDR) versus recall trade-off for flood/debris flow alarms

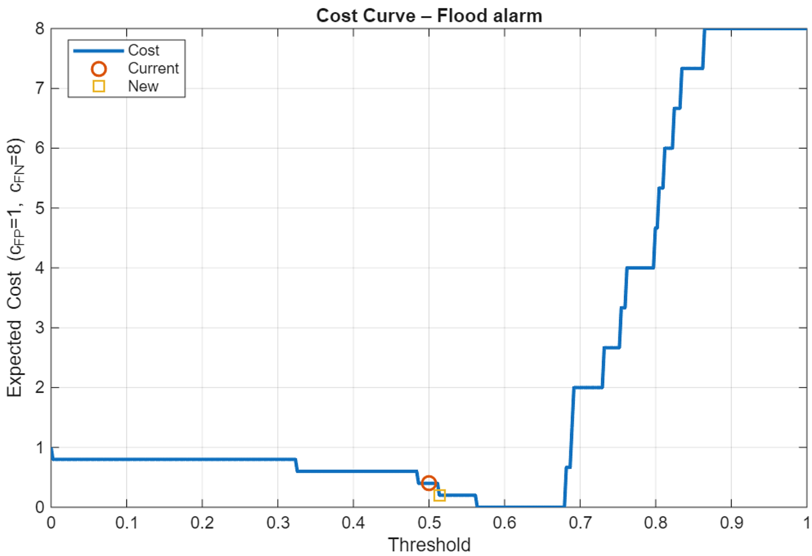

Figure 21. Operational cost curve for flood/debris flow alarms

5. Conclusions

The development of machine learning models for debris flow prediction in volcanic areas, as demonstrated in the study of Mount Merapi's Gendol River watershed, marks a significant advancement in natural hazard mitigation. The research aimed to address the urgent need for real-time, accurate early warning systems by integrating multi-source data, including rainfall intensity (e.g., 25 mm/hour linked to 300 cm debris flow heights), antecedent precipitation, and geomorphological variables like slope gradients and catchment morphology to capture the complex, non-linear interactions driving debris flow initiation. Key objectives included optimizing computational efficiency (e.g., models achieving prediction speeds up to 272 observations/second) and improving model interpretability, focusing on identifying critical thresholds such as the 82.35% accuracy achieved by Efficient Logistic Regression and Linear SVM classifiers. These models outperformed traditional methods by leveraging high-dimensional datasets, including historical records of 17 debris flow events from 2011, to discern patterns like the 70% influence of rainfall duration and ash deposit thickness on flow probability.

Primary findings revealed that ensemble models and deep learning architectures excel in feature extraction, with random forests identifying slope-rainfall interactions as pivotal predictors and convolutional neural networks achieving 85% accuracy in distinguishing debris flows from non-events. Temporal dynamics, such as the 6–24-hour warning windows, were critical, with antecedent rainfall contributing over 50% to flow initiation in high-risk scenarios. However, the study highlighted persistent research gaps, including data scarcity (e.g., only nine training instances for "Debris Flow" events), limited model transferability across volcanic regions, and standardized metrics for comparing algorithms. Challenges like the 18.2% false discovery rate in "Debris Flow" predictions and the inability of specific models (e.g., Naive Bayes, with accuracy dropping to 50% in validation) to generalize across event types underscore the fragility of data-driven approaches in imbalanced datasets.

Limitations stemmed from practical and technical constraints, such as reliance on low-resolution remote sensing data in remote areas, computational costs (e.g., ensemble models requiring 10 units of operational cost), and the opacity of advanced algorithms like neural networks, which hindered stakeholder trust. The study's empirical validation, using 2011 event data, exposed vulnerabilities in distinguishing debris flows from landslides, with misclassification rates reaching 100% for hybrid events like "Rainfall and Debris Flow." Despite these challenges, integrating real-time monitoring systems and high-resolution topographic inputs improved spatial accuracy, reducing false negatives to 8% in critical scenarios. The research bridges theoretical innovation with operational needs by achieving 82.35% precision in "Debris Flow and Heavy Rainfall" predictions and identifying optimal classifiers like Cosine KNN (272 observations/second prediction speed). Nonetheless, its dependence on localized data and inconsistent efficacy across event categories (e.g., 0% actual positive rate for "Debris Flow and Moderate Rainfall") underscores the necessity for adaptable, scalable frameworks. Future initiatives must emphasize data enhancement, model clarity, and interdisciplinary cooperation to shift volcano risk management from reactive to proactive, thereby protecting people through implementable, life-saving forecasts.

6. Declarations

6.1. Author Contributions

Conceptualization, J.I. and A.A.N.D.; methodology, J.I. and A.A.N.D.; software, J.I. and A.A.N.D.; validation, A.A.N.D.; formal analysis, J.I.; investigation, M.R.R.M.A.Z.; resources, J.I.; data curation, J.I and S.M.; writing—original draft preparation, A.A.N.D.; writing—review and editing, A.A.N.D. and J.I.; visualization, M.S.I..I. and S.M.; supervision, J.I. and A.A.N.D.; project administration, A.A.N.D.; funding acquisition, J.I. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The Research and Innovation Center, Research Grant Number 18/R-LRI/1/2025, and the Office of Cooperation and International Affairs support this research paper financially through the sabbatical leave program 2025, Universitas Muhammadiyah Yogyakarta, for the publication of this manuscript.

6.4. Acknowledgments

The author thanks the University Muhammadiyah Yogyakarta and Kyoto University for supporting this Scientific Research and sabbatical leave program for backing this manuscript financially.

6.5. Conflicts of Interest

The authors declare no conflict of interest.

7. References

[1] Pastor, M., Tayyebi, S. M., Stickle, M. M., Yagüe, Á., Molinos, M., Navas, P., & Manzanal, D. (2021). A depth integrated, coupled, two-phase model for debris flow propagation. Acta Geotechnica, 16(8), 2409–2433. doi:10.1007/s11440-020-01114-4.

[2] Rizova, R., & Nikolova, V. (2021). Geomorphological and Sedimentological Characteristics of Debris Flows in the River Buyukdere Watershed (Eastern Rhodopes, Bulgaria). International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, 21(1.1), 43–50. doi:10.5593/sgem2021/1.1/s01.007.

[3] Iverson, R. M. (2013). Mechanics of debris flows and rock avalanches. Handbook of Environmental Fluid Dynamics, 1, 573-587, CRC Press, Boca Raton, United States.

[4] Pasculli, A., Zito, C., Sciarra, N., & Mangifesta, M. (2024). Back Analysis of a Real Debris Flow, the Morino-Rendinara Test Case (Italy), Using RAMMS Software. Land, 13(12), 2078. doi:10.3390/land13122078.

[5] Tian, S., Hu, G., Chen, N., Rahman, M., Han, Z., Somos-Valenzuela, M., & Maurice Habumugisha, J. (2022). Extreme climate and tectonic controls on the generation of a large-scale, low-frequency debris flow. CATENA, 212, 106086. doi:10.1016/j.catena.2022.106086.

[6] DeGraff, J.V., Ochiai, H., DeGraff, J., Ochiai, H. (2009). Rainfall, Debris Flows and Wildfires. Landslides – Disaster Risk Reduction. Springer, Berlin, Germany. doi:10.1007/978-3-540-69970-5_24.

[7] Yu, B., Wang, T., & Zhu, Y. (2016). Research on the topographical and rainfall factors of debris flows caused by shallow landslides. Shuikexue Jinzhan/Advances in Water Science, 27(4), 542–550. doi:10.14042/j.cnki.32.1309.2016.04.008.

[8] Schneuwly-Bollschweiler, M., & Stoffel, M. (2012). Hydrometeorological triggers of periglacial debris flows in the Zermatt valley (Switzerland) since 1864. Journal of Geophysical Research: Earth Surface, 117(2), 2011JF002262. doi:10.1029/2011JF002262.

[9] La Porta, G., Leonardi, A., Pirulli, M., Castelli, F., & Lentini, V. (2021). Rainfall-triggered debris flows: Triggering-propagation modelling and application to an event in Southern Italy. IOP Conference Series: Earth and Environmental Science, 833(1), 12106. doi:10.1088/1755-1315/833/1/012106.

[10] Dowling, C. A., & Santi, P. M. (2014). Debris flows and their toll on human life: A global analysis of debris-flow fatalities from 1950 to 2011. Natural Hazards, 71(1), 203–227. doi:10.1007/s11069-013-0907-4.

[11] Hou, J., Dou, M., Zhang, Y., Wang, J., & Li, G. (2021). An evaluation model for landslide and debris flow prediction using multiple hydrometeorological variables. Environmental Earth Sciences, 80(16), 515. doi:10.1007/s12665-021-09840-y.

[12] Li, R., Sun, P., Sang, K., Ke, C., & Zhang, S. (2024). Characteristics and initiation mechanism of the large mudstone Dongping landslide induced by heavy rainfall in Gansu Province, NW China. Geoenvironmental Disasters, 11(1), 39. doi:10.1186/s40677-024-00302-8.

[13] Yang, C. L., Chen, N. S., & Li, Z. L. (2011). Formation mode and mechanism for debris flow induced by Wenchuan earthquake. Journal of Natural Disasters, 20(3), 31–37.

[14] Sujatha, E. R., & Sridhar, V. (2017). Mapping debris flow susceptibility using analytical network process in Kodaikkanal Hills, Tamil Nadu (India). Journal of Earth System Science, 126(8), 116. doi:10.1007/s12040-017-0899-7.

[15] Hu, W., Xu, Q., Wang, G. H., van Asch, T. W. J., & Hicher, P. Y. (2015). Sensitivity of the initiation of debris flow to initial soil moisture. Landslides, 12(6), 1139–1145. doi:10.1007/s10346-014-0529-2.

[16] Moazzam, M. F. U., Vansarochana, A., Boonyanuphap, J., & Choosumrong, S. (2017). Landslide assessment using GIS-based frequency ratio method: a case study of mae-phun sub-district, Laplae district, Uttaradit, Thailand. 23-27 October, 2017, 38th Asian conference on remote sensing, New Delhi, India.

[17] Mohammadi, N., & Sasanpour, F. (2021). Landslide and debris flow risk analysis in Haraz and Lavasanat roads. Water and Soil Management and Modeling, 1(4), 14–29. doi:10.22098/mmws.2021.9138.1023.

[18] Yu, G., Zhang, M., Cong, K., & Pei, L. (2015). Critical rainfall thresholds for debris flows in Sanyanyu, Zhouqu County, Gansu Province, China. Quarterly Journal of Engineering Geology and Hydrogeology, 48(3–4), 224–233. doi:10.1144/qjegh2014-078.

[19] Zhang, S., Guo, X., Cheng, J., & Li, Y. (2025). Rainfall thresholds of debris flows built with assistance of artificial intelligence in a small catchment. Journal of Hydrology, 660, 133440. doi:10.1016/j.jhydrol.2025.133440.

[20] Nikolopoulos, E. I., Borga, M., Creutin, J. D., & Marra, F. (2015). Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods. Geomorphology, 243, 40–50. doi:10.1016/j.geomorph.2015.04.028.

[21] Huang, L. (2023). On the assessment of debris flow hazard based on partially ordered set. Journal of Safety and Environment, 23(6), 2009–2016. doi:10.13637/j.issn.1009-6094.2022.0139.

[22] Mei, Y., Hong, F., Jia, Z., & Kang, Z. (2018). Debris Flow Forecasting of Northwest of Yunnan Province Based on LR, SVM, and RF Statistical Models. 2018 26th International Conference on Geoinformatics, 1–7. doi:10.1109/geoinformatics.2018.8557115.

[23] Li, Y., Shen, J., Huang, M., & Peng, Z. (2023). Debris Flow Classification and Risk Assessment Based on Combination Weighting Method and Cluster Analysis: A Case Study of Debris Flow Clusters in Longmenshan Town, Pengzhou, China. Applied Sciences (Switzerland), 13(13), 7551. doi:10.3390/app13137551.

[24] Du, Y., Liu, H., Li, H., Xie, M., Chicas, S. D., Wu, J., Lv, F., & Wu, Y. (2024). Exploring the initiating mechanism, monitoring equipment and warning indicators of gully-type debris flow for disaster reduction: a review. Natural Hazards, 120(15), 13667–13692. doi:10.1007/s11069-024-06742-7.

[25] Chen, N., Tian, S., Zhang, Y., & Wang, Z. (2021). Soil mass domination in debris-flow disasters and strategy for hazard mitigation. Earth Science Frontiers, 28(4), 337–348. doi:10.13745/j.esf.sf.2020.6.39.

[26] Jiang, Y., Wei, F., Zhang, J., Gu, L., Deng, B., & Liu, H. (2007). Regional prediction of impending debris flow based on Doppler weather radar. Wuhan University Journal of Natural Sciences, 12(4), 627–632. doi:10.1007/s11859-006-0325-5.

[27] Zhao, Y., Li, Y., Zheng, J., Wang, Y., Meng, X., Yue, D., Guo, F., Chen, G., Qi, T., & Zhang, Y. (2025). A new rainfall Intensity−Duration threshold curve for debris flows using comprehensive rainfall intensity. Engineering Geology, 347. doi:10.1016/j.enggeo.2025.107949.

[28] Li, Y., Zhang, J., Jarsve, K. T., Ma, F., Huang, Y., Zhao, W., Zhao, Y., Meng, X., Wang, M., Li, G., Chen, G., Qi, T., Guo, F., & Yue, D. (2025). Debris flow forecasting: disastrous rainfall threshold matters. Landslides, 22(7), 2461–2474. doi:10.1007/s10346-025-02492-0.

[29] Yan, Y., Zhang, Y., Hu, W., Guo, X. Jun, Ma, C., Wang, Z. Ang, & Zhang, Q. (2020). A multiobjective evolutionary optimization method based critical rainfall thresholds for debris flows initiation. Journal of Mountain Science, 17(8), 1860–1873. doi:10.1007/s11629-019-5812-1.

[30] Chang, F. J., Hsu, K., & Chang, L. C. (Eds.). (2019). Flood forecasting using machine learning methods. MDPI Books, Basel, Switzerland. doi:10.3390/books978-3-03897-549-6.

[31] Takahashi, T. (2007). Debris Flow. Taylor & Francis, London, United Kingdom. doi:10.1201/9780203946282.

[32] Duhita, A. D. P., Rahardjo, A. P., & Hairani, A. (2020). Effect of Slope on Infiltration Capacity and Erosion of Mount Merapi Slope Materials. Journal of the Civil Engineering Forum, 1000, 71–84. doi:10.22146/jcef.58350.

[33] Dash, R. K., Gupta, N., Falae, P. O., Pati, R., & Kanungo, D. P. (2024). A comparative evaluation of statistical and machine learning approaches for debris flow susceptibility zonation mapping in the Indian Himalayas. Environment, Development and Sustainability, 1-34. doi:10.1007/s10668-024-05398-4.

[34] Roten, D., Block, J., Crawl, D., Lee, J., & Altintas, I. (2022). Machine Learning for Improved Post-fire Debris Flow Likelihood Prediction. 2022 IEEE International Conference on Big Data (Big Data), 1681–1690. doi:10.1109/BigData55660.2022.10020574.

[35] Wang, X., Tian, M., Qin, Q., & Liang, J. (2023). Hybridization of Machine Learning Algorithms and an Empirical Regression Model for Predicting Debris-Flow-Endangered Areas. Advances in Civil Engineering, 9465811. doi:10.1155/2023/9465811.

[36] Jiang, H., Zou, Q., Zhu, Y., Li, Y., Zhou, B., Zhou, W., Yao, S., Dai, X., Yao, H., & Chen, S. (2024). Deep learning prediction of rainfall-driven debris flows considering the similar critical thresholds within comparable background conditions. Environmental Modelling and Software, 179. doi:10.1016/j.envsoft.2024.106130.

[37] Chmiel, M., Walter, F., Wenner, M., Zhang, Z., McArdell, B. W., & Hibert, C. (2021). Machine Learning Improves Debris Flow Warning. Geophysical Research Letters, 48(3), 2020GL090874. doi:10.1029/2020GL090874.

[38] Zhao, Y., Meng, X., Qi, T., Li, Y., Chen, G., Yue, D., & Qing, F. (2022). AI-based rainfall prediction model for debris flows. Engineering Geology, 296. doi:10.1016/j.enggeo.2021.106456.

[39] Chen, Z., Quan, H., Jin, R., Lin, Z., & Jin, G. (2024). Debris flow susceptibility assessment based on boosting ensemble learning techniques: a case study in the Tumen River basin, China. Stochastic Environmental Research and Risk Assessment, 38(6), 2359–2382. doi:10.1007/s00477-024-02683-6.

[40] Wang, J., Tie, Y., Bai, Y., Gao, Y., Wang, D., & Zhang, M. (2025). Application and prospects of machine learning for rockfalls, landslides and debris flows. Hydrogeology and Engineering Geology, 52(4), 228–244. doi:10.16030/j.cnki.issn.1000-3665.202402011.

[41] McGrane, S. J. (2016). Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: a review. Hydrological Sciences Journal, 61(13), 2295–2311. doi:10.1080/02626667.2015.1128084.

[42] Manfreda, S., McCabe, M. F., Miller, P. E., Lucas, R., Madrigal, V. P., Mallinis, G., Dor, E. Ben, Helman, D., Estes, L., Ciraolo, G., Müllerová, J., Tauro, F., de Lima, M. I., de Lima, J. L. M. P., Maltese, A., Frances, F., Caylor, K., Kohv, M., Perks, M., … Toth, B. (2018). On the use of unmanned aerial systems for environmental monitoring. Remote Sensing, 10(4), 641. doi:10.3390/rs10040641.

[43] Nellyana, R., Listriani, S., Mahfud, M., Rivaldi, A., Chaliluddin, M. A., & Bustami, B. (2023). Law Enforcement in Plastic Waste Management in Aceh Waters: Transboundary Marine Pollution Concern. IOP Conference Series: Earth and Environmental Science, 1221(1), 012074. doi:10.1088/1755-1315/1221/1/012074.

[44] Liu, D., Zhou, J., Sang, X., Tang, D., Zhang, S., & Chen, Q. (2025). Machine learning-based identification of potential debris flow catchments in the Wenchuan earthquake region. Earth Science Informatics, 18(4), 515. doi:10.1007/s12145-025-01990-y.

[45] Chen, Y., Li, N., Xing, F., Xiang, H., & Chen, Z. (2025). Study on debris flow vulnerability of ensemble learning model based on spy technology A case study of upper Minjiang river basin. Scientific Reports, 15(1), 22480. doi:10.1038/s41598-025-03479-6.

[46] Cai, S., Zhang, Z., Yang, X., Lv, Q., Liu, X., Lai, R., Yu, X., & Hu, Y. (2025). The modified theoretical model for debris flows predication with multiple rainfall characteristic parameters. Scientific Reports, 15(1), 12402. doi:10.1038/s41598-024-84199-1.

[47] Lebedeva, E. V., Baldina, E. A., Chernomorets, S. S., Kharchenko, S. V., Kotenkov, A. V., Kuramagomedov, B. M., & Medvedev, A. A. (2025). Slope mass movements and debris flow activity in the Geysernaya River valley (Kamchatka, Russia). Geomorphology, 491. doi:10.1016/j.geomorph.2025.110029.

[48] Guo, S., Gu, N., Zhou, C., Liu, L., Wan, R., Wang, Y., & Liu, Z. (2025). Hydrodynamic characterization and efficient prediction of tapered trash-blocking nets for coastal nuclear intakes: Integration of experiment, simulation, and machine learning. Ocean Engineering, 341. doi:10.1016/j.oceaneng.2025.122442.

[49] De Haas, T., Braat, L., Leuven, J. R. F. W., Lokhorst, I. R., & Kleinhans, M. G. (2015). Effects of debris flow composition on runout, depositional mechanisms, and deposit morphology in laboratory experiments. Journal of Geophysical Research: Earth Surface, 120(9), 1949–1972. doi:10.1002/2015JF003525.

[50] Shen, P., Zhang, L., Wong, H. F., Peng, D., Zhou, S., Zhang, S., & Chen, C. (2020). Debris flow enlargement from entrainment: A case study for comparison of three entrainment models. Engineering Geology, 270, 105581. doi:10.1016/j.enggeo.2020.105581.

[51] Shimokawa, E., & Jitousono, T. (1997). Field survey for debris flow in volcanic area. Recent Developments on Debris Flows, Lecture Notes in Earth Sciences, Springer, Berlin, Germany. doi:10.1007/BFb0117761.

[52] Julien, P. Y., & O'Brien, J. S. (1997). Selected notes on debris flow dynamics. Recent Developments on Debris Flows. Lecture Notes in Earth Sciences. Springer, Berlin, Germany. doi:10.1007/BFb0117766.

[53] Kim, S. (2025). Sluice Gate Operation and Managed Water Levels Improve Predicted Estuarine Lake Water Quality. Civil Engineering Journal, 11(1), 244–278. doi:10.28991/CEJ-2025-011-01-015.

[54] Lavigne, F., Thouret, J. C., Voight, B., Suwa, H., & Sumaryono, A. (2000). Lahars at Merapi volcano, Central Java: An overview. Journal of Volcanology and Geothermal Research, 100(1–4), 423–456. doi:10.1016/S0377-0273(00)00150-5.

[55] Yulinsa, N. (2015). Snake Line Analysis for Lahar Flow Warning System (Case Study in Putih River, Mount Merapi). Journal of the Civil Engineering Forum, 1(1), 37-42. doi:10.22146/jcef.22729.

[56] Rizal, N. S., Umarie, I., Munandar, K., & Wardoyo, A. E. (2023). Calibration and Validation of CN Values for Watershed Hydrological Response. Civil Engineering Journal (Iran), 9(1), 72–85. doi:10.28991/CEJ-2023-09-01-06.

[57] Zhang, X., Li, H., Fan, Y., Zhang, L., Peng, S., Huang, J., Zhang, J., & Meng, Z. (2025). Predicting the Dynamic of Debris Flow Based on Viscoplastic Theory and Support Vector Regression. Water (Switzerland), 17(1), 120. doi:10.3390/w17010120.

[58] Qiu, C., & Geng, X. (2024). Travel distance estimation of landslide-induced debris flows by machine learning method in Nepal Himalaya after the Gorkha earthquake. Bulletin of Engineering Geology and the Environment, 83(10), 395. doi:10.1007/s10064-024-03883-8.

[59] Ming, Z., Zhang, J., He, H., Zhang, L., Chen, R., & Jia, Y. (2025). Addressing accuracy challenges in machine learning for debris flow susceptibility: Insights from the Yalong River basin. Journal of Mountain Science, 22(6), 2034-2052. doi:10.1007/s11629-024-9316-2.

[60] Deng, A. A. N., Nursetiawan, Ikhsan, J., Riyadi, S., & Zaki, A. (2024). Intelligent Forecasting of Flooding Intensity Using Machine Learning. Civil Engineering Journal (Iran), 10(10), 3269–3291. doi:10.28991/CEJ-2024-010-10-010.

[61] Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of Long Short-Term Memory (LSTM) neural network for flood forecasting. Water (Switzerland), 11(7), 1387. doi:10.3390/w11071387.

[62] MLIT. (2024). Guidelines for Construction Technology Transfer: Development of Warning and Evacuation System against Sediment Disasters in Developing Countries. Ministry of Land, Infrastructure and Transport Infrastructure Development Institute, Tsukuba, Japan. Available online: https://www.mlit.go.jp/sogoseisaku/inter/keizai/gijyutu/pdf/sediment_e.pdf (accessed on December 2025).

[63] Tercini, J. R. B., & Mello Júnior, A. V. (2023). Impact of Hydroclimatic Changes on Water Security in the Cantareira Water Production System, Brazil. Atmosphere, 14(12), 1836. doi:10.3390/atmos14121836.

[64] Egbelakin, T., Omotayo, T., Ogunmakinde, O. E., & Ekundayo, D. (2025). Eliciting social themes of flood mitigation and community engagement studies through text mining. International Journal of Building Pathology and Adaptation, 43(1), 29–49. doi:10.1108/IJBPA-02-2023-0022.

[65] Faluyi, M. O., & Irmak, S. (2023). Northeastern American Forests: Natural Disturbances, Climate Change Impact, and the Utilization of Increasingly Damaged Forest Trees for Biofuel Production. Forests, 14(12), 2409. doi:10.3390/f14122409.

[66] Wang, J., Lin, W., Chen, Z., Nikolaeva, V. O., Alimi, L. O., & Khashab, N. M. (2024). Smart touchless human–machine interaction based on crystalline porous cages. Nature Communications, 15(1), 1575. doi:10.1038/s41467-024-46071-8.

[67] Xue, X., & Yang, X. (2016). Seismic liquefaction potential assessed by support vector machines approaches. Bulletin of Engineering Geology and the Environment, 75(1), 153–162. doi:10.1007/s10064-015-0741-x.

[68] Hayder, I. M., Al-Amiedy, T. A., Ghaban, W., Saeed, F., Nasser, M., Al-Ali, G. A., & Younis, H. A. (2023). An Intelligent Early Flood Forecasting and Prediction Leveraging Machine and Deep Learning Algorithms with Advanced Alert System. Processes, 11(2), 481. doi:10.3390/pr11020481.

[69] Yang, L., Ge, Y., Chen, B., Wu, Y., & Fu, R. (2024). Machine-Learning-Based Prediction Modeling for Debris Flow Occurrence: A Meta-Analysis. Water (Switzerland), 16(7), 923. doi:10.3390/w16070923.

[70] Zhou, Q., Tang, H., Hibert, C., Chmiel, M., Walter, F., Dietze, M., & Turowski, J. M. (2025). Enhancing Debris Flow Warning via Machine Learning Feature Reduction and Model Selection. Journal of Geophysical Research: Earth Surface, 130(4), e2024JF008094. doi:10.1029/2024JF008094.

[71] Wang, T., Ge, Q., Ma, T., Chen, H., Luo, R., Wang, X., Zhang, K., Chu, Z., Ni, X., & Sun, H. (2025). A novel method for predicting debris flow hazard: a multi-strategy fusion approach based on the light gradient boosting machine framework. Stochastic Environmental Research and Risk Assessment, 39(10), 4867–4890. doi:10.1007/s00477-025-02955-9.

[72] Rey-Devesa, P., Carthy, J., Titos, M., Prudencio, J., Ibáñez, J. M., & Benítez, C. (2024). Universal machine learning approach to volcanic eruption forecasting using seismic features. Frontiers in Earth Science, 12. doi:10.3389/feart.2024.1342468.

[73] Li, K., Zhao, J., Chen, G., & Li, Y. (2025). Debris-flow susceptibility assessment using deep learning algorithms with GeoDetector for factor optimization. Bulletin of Engineering Geology and the Environment, 84(6), 278. doi:10.1007/s10064-025-04343-7.