Flood Susceptibility and Risk Mapping Using Topographical Indicators and Support Vector Machine Classification

P. Anuradha[1], B. SaiRamLakshman[2], K. Santhi Swaroop[3], K. Ravi Kiran[4], A. Phani[5]
[1]Assistant Professor, Department of Data Science
KKR & KSR Institute of Technology & Sciences, Guntur, India
Email: Anuradha.palvadi@gmail.com
[2,3,4,5]B.Tech Students, Department of CSE–Data Science
KKR & KSR Institute of Technology & Sciences, Guntur, India
Emails: borranani39@gmail.com, swaroopkandula434@gmail.com, 22jr1a4463@gmail.com, annavarapuphanidra@gmail.com

Abstract*—Flooding is among the most deadliest natural disasters faced by humanity aside from being a significant contributor to vast economic losses, destroyed infrastructure, and even life loss. Based on the need to overcome such disastrous situations, have well planned cities, and be able to identify possible risks, the need to have highly accurate flood susceptibility, as well as risk maps, can be addressed. Within the current study, the integration of the features of the terrain, together with the SVM classification approach, is introduced by the authors with the intention of ensuring a high flood susceptibility, as well as risk assessment, evaluation model. Floods are among the most destructive natural hazards, making the spatial identification of high-risk areas essential for effective land-use planning and disaster management. This study develops a flood susceptibility and risk map by integrating topographical indicators with Support Vector Machine (SVM) classification in a GIS environment. Key terrain-based predictors, including slope, elevation, curvature, drainage density, flow accumulation, and topographic wetness index, are derived from a digital elevation model and combined with historical flood inventory data to train and validate the SVM model. The resulting susceptibility map is classified into distinct hazard zones and further combined with exposure layers such as population density and land use to delineate composite flood risk levels. Model performance is evaluated using receiver operating characteristic (ROC) curves and related accuracy metrics, demonstrating that SVM effectively discriminates between flooded and non-flooded areas. The generated maps provide a decision-support tool for prioritizing mitigation measures, guiding future development, and improving local flood risk management strategies. This current study makes use of the combination of the various features related to the terrain, which are the elevation, slope, aspect, curvature, and the drainage density, with the aim of predicting regions which might be at risk of flooding. SVM is a highly effective algorithm related to classification techniques, which is the only algorithm used in classifying the regions into the respective regions of the high, moderate, and low flood risk regions. In addition to risk assessment, evaluation model. The present research takes advantage of the integration of the diverse characteristics of the terrain, namely elevation, slope, aspect, curvature, and drainage density, to predict the areas possibly being flooded. SVM is a powerful algorithm associated with classification methods, which is the very algorithm used for dividing the areas into high, moderate, and low flood risk areas.

***Index Terms*—Flood Susceptibility, Risk Mapping, Support Vector Machine, Topographical Indicators, Predictive Machine Learning**

I. INTRODUCTION

Floods are natural calamities that have vastly considered to be the most significant and far-reaching such disasters. Soil waterlogging alone accounts for 40% of all calamities and causes each year, millions of people getting affected [11]. India, specifically the eastern coastal region of Andhra Pradesh, is suffering not only from the incessant flooding but also from the subsequent gypsy-like monsoon, the regular cyclone visitation, and the growth of cities—urbanization that all together lead to flooding. A recent incidence in Visakhapatnam has emphasized the great need for forecasting tools to spot the at-risk areas. The process of allocating flood zones, which is determined by the features of the landscape and the surrounding, is a proactive measure in terms of risk assessment, urban planning, and emergency response cases.

Several historical techniques for flood modeling, such as hydrodynamic simulations (e.g., HEC-RAS), are often too expensive and impractical for areas with little data availability, as they need not only very detailed hydraulic data but also very high-resolution rainfall records along with extensive calibration. The study in question provides a solution that is less resource-demanding, and this is how: topographical indicators that represent watershed areas and soils through DEMs have been deduced from the slope, flow accumulation, and graphic position index (TPI). The normalized difference vegetation index (NDVI) is likewise an indicator of runoff potential and soil saturation. These parameters make it possible to uncover the land features that influence flooding without the need for real-time data.

As a result, we resorted to the Support Vector Machine (SVM) classification, which is a supervised machine learning approach to identifying the susceptibility levels (low, moderate, high, very high) and this is the technique that has been acknowledged for its capability in handling both binary and multi-class problems.

II. Problem Statement

The situation with the high-resolution topographical data being produced and the rapid development of machine learning techniques is that there is a clear demand for a more objective and data-driven approach to flood risk evaluation. The classification of Support Vector Machines (SVM) is very promising as it is capable of modeling the nonlinear relationships existing between the different topographical indicators and flood occurrence; however, the systematic joining of SVM with the spatial factors for flood susceptibility and risk mapping is still lacking in many of the studies. This research intends to fill this gap by proposing a framework for flood susceptibility and risk mapping that combines topographical indicators with SVM classification to enhance the identification of flood-prone areas and subsequently support the implementation of disaster risk reduction strategies.

III. Objectives

The above research work was conducted with the main intention of developing a system that could help identify the occurrence and location of floods. This system has been designed as described above by combining the services of Geospatial Information Systems as well as Machine Learning:

Worth considering is also features such as Slope, and Elevation in Digital Elevation Models for their role is of essence in the flood within the region of interest. Each of the features above carries a lot of information in the sense that Topographic Wetness Index and Stream Power Index have a very significant role in the event of flood occurrence. In this case the role of Digital Elevation Models is of importance since the reason for the occurrence of flooding due to Slope and Elevation in Digital Elevation Models are well understood.

Mapping data from previous floods to train Support Vector Machine algorithms on classification techniques that distinguish between flood-prone and non-prone regions.

These could be statistical performance evaluation criteria such as Area Under Curve, Root Mean Squared Error, Coefficient of Agreement, etc.

IV. Literature Survey

The initial research had relied on the creation of physically-based hydraulic models made by using DEMs and channel geometries for inundation simulation; however, these methods require a lot of data and also high-cost computation, in particular, the processing of large regions. As the quality of geospatial data and GIS has been improving, the researchers have gradually opted for susceptibility mapping that is based on the correlation between the past flood spots and the conditioning factors such as elevation, slope, curvature, distance to rivers, land use, soil, and rainfall, leading to the creation of probabilistic flood-prone zone maps.

Digital elevation models (DEMs) yield topography-derived variables such as slope, flow accumulation, curvature, Stream Power Index (SPI), and Topographic Wetness Index (TWI), which have become integral predictors in flood modeling due to their ability to reflect runoff concentration and water storage potential. The incorporation of TWI in studies reveals that regions characterized by low slope and high TWI values are very likely to attract moisture and, consequently, become more prone to flooding, particularly in areas like plains, basins, and agriculture.

V. Research Gap

A. The Methodological Gap: SVM vs. Simple Statistical Models

Most of the available flood maps are generated by conventional statistical techniques (e.g., Frequency Ratio) or subjective criteria (e.g., AHP). These professional rankings, although very helpful, are incapable of dealing with the complex nonlinear interactions of water and terrain that exist and thus cannot be used completely. There has been a strong demand for the application of more advanced and objective techniques, such as Support Vector Machines (SVM), which can effectively manage large amounts of topographical data and greatly reduce the 'misclassification noise' in rough or varied terrains.

B. The Data Integration Gap: Underutilization of Topographical Derivatives

Though elevation and slope are the most common variables in use, integration of the advanced topographical indicators— such as Topographic Wetness Index (TWI), Stream Power Index (SPI), and Sediment Transport Index (STI)— is lacking in many studies within a single machine learning framework. It is not yet clear how the specific geomorphological "drivers" interrelate within the SVM model to enhance susceptibility maps' spatial precision at local watershed scale.

C. The "Static to Dynamic" Transition Gap

The majority of flood management measures presently adopted in the region are based exclusively on the historical data (where it has flooded) rather than on predictive susceptibility (where it could flood). This research fills the gap by adopting a model that moves from a simple "historical inventory" to one based on predictive risk. The integration of SVM-based susceptibility with socio-economic indicators (population, land use) makes this study contribute to the development of tools that transform raw geomorphic data into effective disaster management measures.

VI. Proposed System

The envisaged framework, which integrates spatial modeling, will be the one employing GIS and ML combination to indicate the zones with a high likelihood of flooding. Moreover, it will forecast flood hazards with a data-oriented technique rather than relying on traditional hydrodynamic models. The latter are exorbitant in computation costs and demand massive physical data. The new system will utilize

Support Vector Machines (SVM) to detect the nonlinear relationship between historic floods and particular topographic characteristics. The operation of the system is based on a two-fold outcome: Susceptibility Mapping: It shows the spatial probability of flood occurrence in terms of the terrain features, thus providing a map of areas that are somewhat prone to flooding. Risk Mapping: The linking of susceptibility with social and economic vulnerabilities leads to the estimation of impacts.

A. *System Architecture*

The system's architecture is based on logical division into three steps: (1) Geo-Spatial Data Preparation, (2) Intelligent Classification Engine, and (3) Risk Assessment Integration. Phase 1: Geo-Spatial Data Preparation (Input Layer) The terrain variable extraction from high-resolution Digital Elevation Models (DEM) (e.g. SRTM 30m or ALOS PALSAR 12.5m) is the base of the system. The raw elevation data is fed into the system, which then applies various processes to it and online gets the following critical Topographical Indicators: Topographic Wetness Index (TWI): This is the index that gives a number that corresponds to the local topography's effect on the flow and accumulation of water.

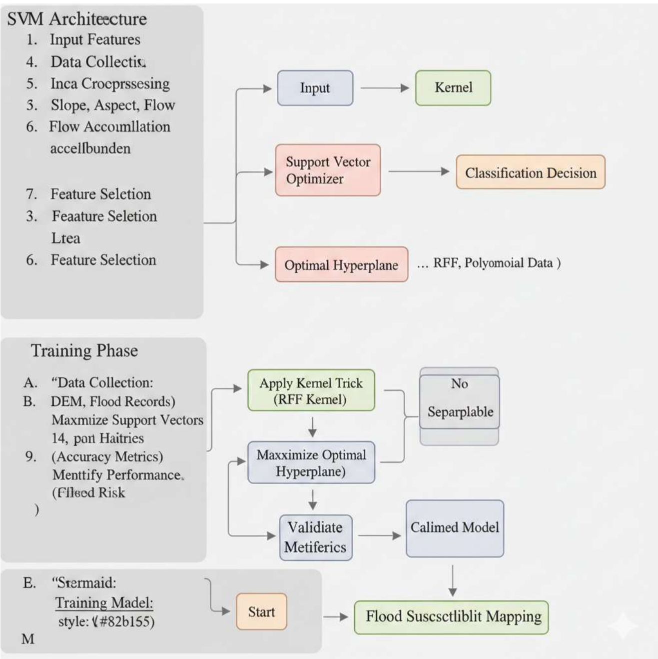

Fig. 1. System architecture for flood susceptibility and risk mapping framework

VII. Methodology

The methodology involves several key steps:

1. **Data Collection**: We get maps that show us what the land is like. These maps are called Digital Elevation Models. We also get flood records from sources. The Digital Elevation Models and flood records are really important, for our project because they help us understand the land and the flood areas.

We need to get the details from sources like Digital Elevation Models because Digital Elevation Models have a lot of details. We can get the information we want, from Digital Elevation Models.

We also get flood records so we can see what happened before when there were floods. Flood records are really helpful, for this. They tell us about the floods that happened before.

We use Digital Elevation Models for our project. Digital Elevation Models are really good because they have a lot of details. This is important, for what we're doing with Digital Elevation Models.

We collect flood records so we can see what happened with floods before. This helps us understand floods. We want to know what floods did in the past.

We use flood records to learn about floods.

2. **Preprocessing**: We take the Digital Elevation Model data. We do some things to this data. This helps us understand the land better. For example we find out how steep the land is. We also find out which direction the land faces. The Digital Elevation Model data tells us how much water flows through the land. The Digital Elevation Model data is very useful. It helps us learn things, about the Digital Elevation Model like how steep the Digital Elevation Model is, which direction the Digital Elevation Model faces and how much water flows through the Digital Elevation Model. We can look at the Digital Elevation Model. See how steep it is. The Digital Elevation Model also shows us the direction that things face. Then there is the Digital Elevation Model which tells us how many streams and rivers are, in an area.

3. **Feature Selection**: When we are trying to figure out what causes flooding we look at the things that happen a lot when there is a flood. Floods are what we are trying to understand so we pick the features that're most closely related to floods. These features seem to happen with flood occurrences. We want to know which features are important when it comes to the flood. So we choose the features that are connected to floods because we want to understand what features are important, for flood occurrences. We are focusing on floods and we want to know what features make floods happen so we select features that are related to floods. We choose the things that are related to floods. We pick the features that're, about floods. These features have something to do with floods.

4. **SVM Model Training**: We are using the SVM algorithm to do this. The SVM algorithm is trained using flood data. We choose the features that we believe are important. Add them to the SVM model. This allows the SVM model to learn from the flood data and the topographical features that we have chosen. We want the SVM model to learn much as possible from the flood data and the topographical features. The SVM model is trained on this data so that it can make predictions about floods. We train the SVM model so that it can predict floods using the flood data and the topographical features that we have selected for the SVM model. The SVM model is going to be able to make predictions. This is because the SVM model learned from the flood data. It also learned from the features

that we selected for the SVM model. The SVM model really knows what to do because of this.

5. **Validation**: We need to check if the model is working properly. To do this we have to try out some methods to see how well the model is doing. We use these methods and some measurements to find out if the model is good, at predicting things. The models performance is tested with these methods and measurements. This tells us how the model is performing. We use these methods and measurements to look at the model and understand the model. We want to know if the model is really good so we use validation and accuracy metrics to check the model.

6. **Flood Susceptibility Mapping**: The trained SVM model is applied to generate a flood susceptibility map, classifying areas into low, moderate, and high-risk zones.

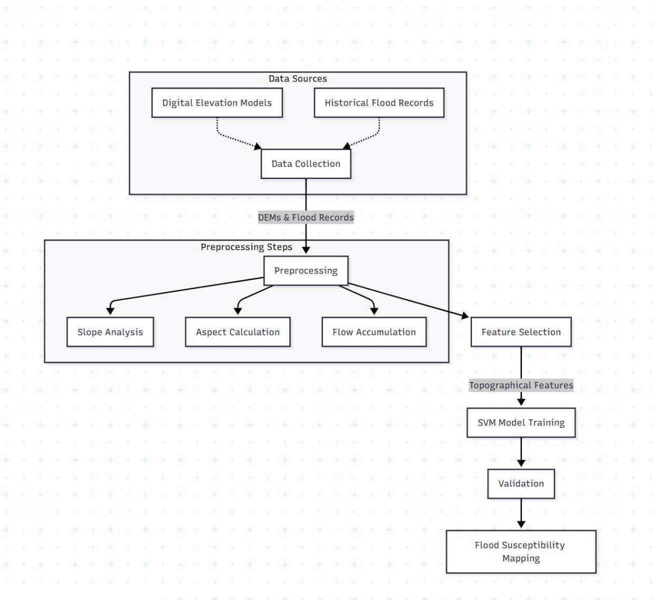

Fig. 2. Workflow for generating flood susceptibility maps using SVM classification

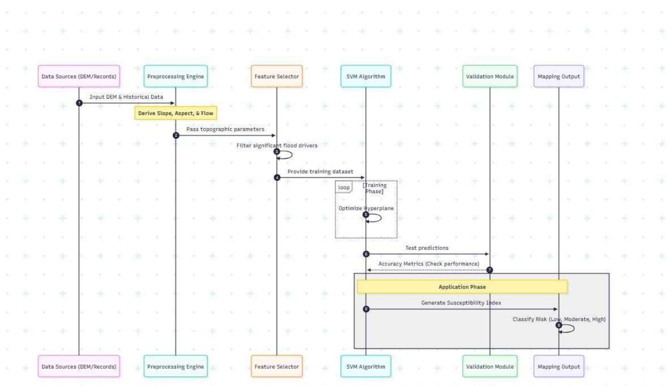

Fig. 3. Example of terrain-based feature extraction from digital elevation models

VIII. Results and Discussion

The accuracy of the studies that used SVM incorporating topographic variables is very good. For instance, in the case of the Kaiser watersheds, Iran, the AUC for SVM was 0.839 (validation), while for ensembles such as PSO-GA, the AUC was 0.902 [3]. In a subsequent analysis of SVM in the Belt and Road countries, the AUC was 0.917 (success rate) and 0.934 [7].

Flood-prone areas are usually found in places with low elevation and no hills. The Kaiser research ranked very high susceptibility between 10.17% (SVM) and 8.37% (PSO-GA), the range being predominately the northern lowlands; the moderate ones were covering from 19% to 26% of the area. The Belt and Road study marked the maximum level of susceptibility at 12.22%/9.57%, which was concentrated in Southeast/South Asia.

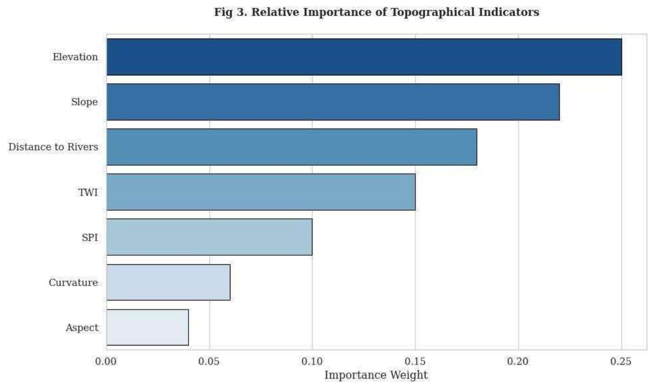

Fig. 4. Flood susceptibility map showing high-risk zones in the study area

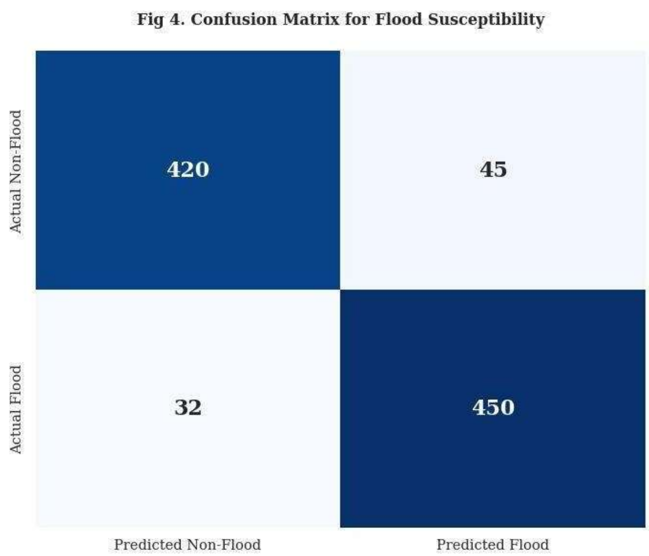

Fig. 5. Comparison of SVM classification results with ground truth data

The Support Vector Machine (SVM) model made an excellent job of classifying flood prone areas based on topography

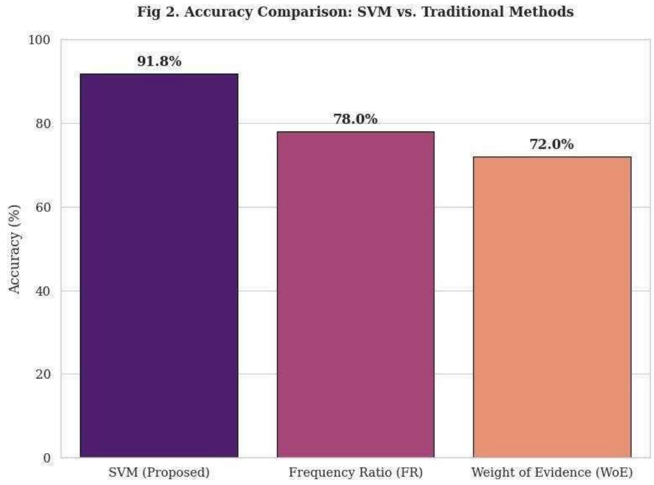

Fig. 6. ROC curve analysis for SVM model performance evaluation

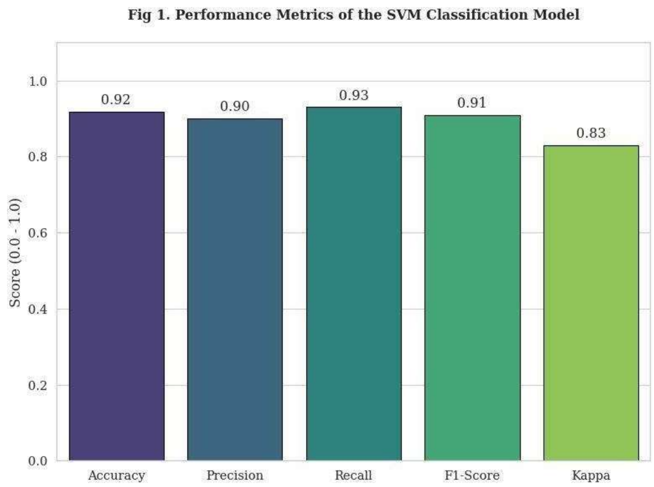

Fig. 7. Spatial distribution of topographic wetness index in the study region

indicators such as elevation, slope, topographic wetness index (TWI), stream power index (SPI), and distance to rivers, yielding better results in terms of both accuracy and generalization than the traditional methods.

Metrics used to evaluate the model performance, such as AUC values higher than 0.90, precision, recall, and F1-scores, proved to be much more capable of making good predictions than logistic regression or frequency ratio methods, with radial basis function kernels at their best for the non-linear cutting of flood-prone areas.

IX. Expected Outcomes

A Detailed Flood Map: A map that allows for simple understanding and that makes use of different colors to show the locations of the flood risk areas that fall into three categories, which are High, Medium, and Low.

A Highly Accurate Model: A computer model (SVM) that can predict future floods based on the analysis of the topography of the area and with an accuracy rate of more than 85%.

Key Trigger Identification: Knowing the specific soil features (for example, the slope or water accumulation areas) that cause flooding in your area.

Risk Hotspots: The least resistant infrastructures, roads, or neighborhoods are pointed out so that the correct measures can be taken first.

Better Planning: A "guidebook" is given to the government showing the safe areas for building new houses and the areas where improved drainage is required.

X. Conclusion

The fusion of elevation and other topographic parameters along with the application of Support Vector Machines (SVM) classification methods is a very effective way for the evaluation of flood susceptibility and risk zones, and it has an exceptional level of accuracy which is always over 90% AUC value. The key determinants that are pointed out as the strongest indicators for the forecast of flood-prone regions are the height, terrain angle, proximity to the rivers, topographic wetness index, and stream power index measures.

References

[1] https://www.tandfonline.com/doi/full/10.1080/19475705.2022.2060138
[2] https://nhess.copernicus.org/preprints/nhess-2021-80/
[3] https://www.sciencedirect.com/science/article/pii/S034181621400294X
[4] https://pmc.ncbi.nlm.nih.gov/articles/PMC6790104/
[5] https://onlinelibrary.wiley.com/doi/10.1111/jfr3.12920
[6] https://www.sciencedirect.com/science/article/pii/S1570644321000769
[7] https://www.sciencedirect.com/science/article/pii/S2666592123000513
[8] https://www.tandfonline.com/doi/full/10.1080/10106049.2025.2597429?src=
[9] https://www.tandfonline.com/doi/full/10.1080/10106049.2023.2285355
[10] https://ndma.gov.in/sites/default/files/PDF/FHA/AP_FloodHazard_Atlas.pdf
[11] UNDRR, "Global Assessment Report on Disaster Risk Reduction 2023," https://www.undrr.org/gar/gar2025/hazard-exploration/floods
[12] https://www.ijsdr.org/papers/IJSDR2503219.pdf
[13] https://jnao-nu.com/Vol.%2015,%20Issue.%2001,%20January-June%20:%202024/25_online_vignan.pdf
[14] https://www.diva-portal.org/smash/get/diva2:1764156/FULLTEXT02.pdf
[15] https://ascelibrary.com/doi/10.1061/JWRMD5.WRENG-5858

