Grenze International Journal of Engineering and Technology, June Issue

AI Embeddings-based Flood Susceptibility Mapping in Kerala: A U-Net Deep Learning Approach

M S Kendagannaswamy[1], C K Roopa[2] and B S Harish[3]
[1]Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, 570006, Karnataka, India
Email: kswamy@jssstuniv.in
[2-3]Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, 570006, Karnataka, India
Email: ckr@jssstuniv.in, bsharish@jssstuniv.in

***Abstract*— Kerala experienced catastrophic flooding in 2018, and reliable flood susceptibility maps are still needed for the region. The Proposed research presents a deep learning approach using lightweight U-Net model that generates 30-meter resolution susceptibility maps from 46 remote sensing features: 10 environmental variable bands, 7 SAR bands, and 29 Google Earth Engine deep learning embeddings, covering 156.5 million pixels across Kerala. Only 0.13% of pixels in the dataset were labelled as flood, so we used focal loss optimization and spatial sampling to handle the imbalance. The model reached 96.55% overall accuracy and 41.52% precision, and detected flood-prone areas more reliably than traditional approaches. Spatial validation confirmed that the predictions aligned with known flood corridors along major river systems and coastal lowlands. The generated susceptibility maps support flood risk assessment, early warning systems, and urban planning. The proposed research demonstrates a scalable AI-driven approach to disaster management in data-scarce, topographically complex tropical regions.**

***Keywords*— Flood Susceptibility, U-Net, Deep Learning, Remote Sensing, Kerala Floods, Class Imbalance, Disaster Risk Reduction.**

I. INTRODUCTION

Kerala, a south-western state in India, is highly susceptible to floods due to its geographical location, topography, river systems, high-intensity rainfall and dense population (Lal et al., 2020). The state covers approximately 38,863 km2 along the Malabar Coast and comprises three physiographic zones that form intricate drainage systems with 44 major rivers (Mohan & Adarsh, 2023).

Excessive rainfall of 2,346 mm (over 42% of the climatological average) in August 2018, displaced over 5.4 million people, caused economic damage estimated at USD 2.8 billion, and destroyed critical infrastructure across all 14 districts (Raj & Sofi, 2023). The disaster was exacerbated by the simultaneous overflow of 35 of 42 state dams as upstream reservoirs reached dangerous levels. Flooding events in 2019, 2020 and 2021 also highlight the ongoing flood vulnerability of the region (Jain et al., 2021). Climate change models predict an increase in extreme precipitation over the Western Ghats and an increase in the variability of the Indian monsoon (Johnson & Arunachalam, 2009).

Grenze ID: 01.GIJET.12.2.15_25

At the same time, urbanization has drastically impacted natural drainage systems, with urban land area increasing by approximately 68%, forest cover decreasing by 12%, and wetlands declining by 30% over the 2005–2020 period (Parthasarathy & Kundapura, 2023). These anthropogenic changes have led to a decrease in landscape infiltration, increase in surface run-off and increase in flood risk, thereby reinforcing the need for improved flood susceptibility mapping (Sadhwani et al., 2022).

Traditional flood modelling methods face a major challenge in capturing the complex spatial variations and nonlinear coupling of flood processes, especially in heterogeneous topographies and land-use regimes (Kendagannaswamy et al., 2025). Traditional methods such as Multi-Criteria Decision Analysis (MCDA), Analytical Hierarchy Process (AHP) and frequency-ratio modelling involve expert opinion and linear statistical relationships with limited ability to capture complex interactions among multiple conditioning variables(Vilasan & Kapse, 2022). These methodologies, however, are challenged by high dimensionality of the feature space, nonlinear interactions among geomorphic, hydrological, and anthropogenic processes, pronounced spatial heterogeneity, and computational scalability constraints. Physically based hydraulic simulators (such as HEC-RAS) are highly computationally intensive and require detailed data, which is not always available at regional scale (Khan & Jhamnani, 2024). Furthermore, the classical models are outperformed by the modern data-driven machine learning models including Support Vector Machines (SVM), Random Forests (RF), and Logistic Regression, yet these models generally fail to account for spatial autocorrelation and contextual relationships within the data (Youssef et al., 2023). As a result, these methods are generally time-consuming, involve manual processes, and fail to exploit the spatial relationships inherent in geospatial data. This is particularly important when there is a high-class imbalance, such as a large number of non-flood pixels in relation to the flood pixels, and achieving high predictive accuracy becomes even more challenging (Ekmekcioğlu et al., 2022).

In recent years, environmental hazard mapping has seen significant progress in deep learning techniques, which can extract complex spatial features from unstructured data without extensive manual feature engineering (Maddah et al., 2025). The success of U-Net in biomedical image segmentation has facilitated its adaptation to geospatial applications owing to its ability to maintain spatial integrity through skip connections and to learn multi-scale features. The symmetric structure of its encoder-decoder architecture enables down sampling and upsampling of feature representations, and the skip connections concatenated between these layers ensure high-resolution contextual information, reducing the impact of limited training samples and generating pixel-wise prediction maps at fine spatial resolutions(Liu et al., 2025). In many applications like landslide prediction, flood delineation and coastal vulnerability assessment, U-Net and its variants have been consistently outperforming the traditional methods (Kaushal et al., 2024). The literature however, has largely concentrated on binary flood delineation (flooded vs non-flooded pixels) from post-event remote sensing data, while not adequately addressing multi-class susceptibility mapping for proactive risk reduction (Mia et al., 2023). There are still some research gaps, such as the absence of susceptibility products with multiple classes, a lack of methods for addressing extreme class imbalance, limited systematic incorporation of heterogeneous data streams in complex topographic conditions, and poor adaptation to monsoon-dominated tropical environments (J. Wang et al., 2025).

This study employs U-Net to develop an end-to-end flood susceptibility mapping framework. It integrates multi-source remote sensing data, mitigates class imbalance, and produces high-resolution susceptibility maps to support decision-making in Kerala. The specific aims are as follows: (1) design and train a U-Net neural network for multi-class flood susceptibility mapping with the incorporation of focal loss, weighted sampling and data augmentation techniques for mitigating class imbalance; and (2) collate and preprocess an extensive geospatial dataset to include all topographic indices (elevation, slope, aspect, and curvature), hydrological variables (drainage density, proximity to rivers, and Topographic Wetness Index), and land cover classifications.

II. Study Area and Data

A. Study Area

The study was carried out in Kerala state in the southwestern part of India, lying between the latitudes 8°17'N to 12°47'N and longitudes 74°51'E to 77°24'E, with an area of about 38,863 km² (Kendagannaswamy et al., 2026). The region exhibits diverse physiographic characteristics: the Western Ghats form a mountain range along the eastern boundary with peaks reaching 2,695 m above the sea level, while coastal plains above the Arabian Sea to the west.

B. Data Sources

The Proposed study integrated number of remote sensing and secondary datasets to comprehensively characterize the determinants of flood susceptibility and the temporal dynamics of the 2018 Kerala flood event.

The data architecture comprised three main layers: Google AI embeddings derived from deep learning models, Synthetic Aperture Radar (SAR) measurements, and a set of environmental conditioning variables (Table 1). Google Satellite Embeddings, derived from pre-computed deep learning embeddings of the Satellite Embedding V1 Annual collection on Google Earth Engine, formed the first feature suite (Brown et al., 2025). These embeddings represent complex multitemporal spectral spatial patterns derived on high-resolution satellite images over the whole of 2018. "A carefully selected set of thirty-four embedding dimensions (A00-A63) was used to capture different early-, mid-, and late-feature hierarchies to provide a comprehensive picture of land-surface dynamics with a 30 m spatial resolution.

Sentinel-1 C-band SAR data provided valuable insight into surface backscatter characteristics across three time periods: pre-flood (June-July 2018), during flood (August 2018), and post-flood (September–October 2018) (Twele et al., 2016). From dual-polarization (VV and VH) intensity data, six features were derived, enabling identification of temporal changes, and delineate flood extent using SAR differencing procedures. The environmental conditioning factors included the topographic parameters (elevation, slope, aspect, and Topographic Wetness Index), hydrological parameters (water occurrence, distance to water bodies), climatic parameters (precipitation during pre-flood, flood and annual periods), and land-surface parameters (Land Use/Land Cover according to MODIS, soil texture). These eleven covariates and flood inventory data were compiled for 2018 (Adesina et al., 2022).

TABLE I. INPUT DATA CHARACTERISTICS

Data Source	Variables	Features	Resolution	Temporal Coverage
Google Satellite Embeddings	Strategic embedding dimensions (A00-A63) from deep learning model capturing multi-temporal spectral-spatial patterns	34	30m	Annual 2018
SAR (Sentinel-1)	VV/VH polarization intensity (before/during/after flood) for temporal change detection	6	30m	June-October 2018
Topographic	Elevation (SRTM), Slope, Aspect, Topographic Wetness Index (TWI)	4	30m	Static (2000)
Hydrological	Water occurrence (GSW), Distance to permanent water	2	30m	1984-2021 composite
Climatic	CHIRPS daily precipitation (pre-flood, flood period, annual mean)	3	30m	2018
Land Surface	MODIS Land Cover (Type 1), Soil texture (USDA classification)	2	30m	2018
Flood Reference	Binary flood extent from SAR-based detection with topographic filtering	1	30m	August 2018

III. METHODOLOGY

A. Data Preprocessing

Quality Assessment and Data Cleaning

Prior to model development, rigorous quality evaluation procedures were applied to ensure data integrity and reliability. All input datasets were verified for spatial alignment with the WGS84 geographic coordinate system (EPSG:4326) to ensure geometric consistency among multi-source data layers (Faruga et al., 2025). Valid coverage of environmental variables, SAR, and embedding features reached 78.6% and 79.3%, respectively. Data gaps were primarily attributed to cloud contamination of optical imagery and topographic shadowing in mountainous areas. The outlier analysis was conducted through the Interquartile Range (IQR), where any pixel values outside 1.5 IQR above the lower quartile value or below the upper quartile value was flagged for analysis. Any other missing values were estimated using inverse distance weighting (IDW), since IDW maintains spatial continuity while providing full spatial coverage (Bouras et al., 2021).

Feature Engineering and Transformation

All input variables have been carefully scaled and normalized due to the differing scales in which they are measured. The environmental variables have been normalized to the scale of [0, 1]. This ensured that there was no impact of different value ranges during the introduction of covariates into the model (Montello et al., 2022). As SAR backscatter coefficient is usually normally distributed, z-score normalization was done. This ensured that the variable had zero mean and a standard deviation of one, allowing us to compare VH and VV polarization accurately and objectively. The embeddings of the Google Earth Engine were reduced by Principal Component Analysis (PCA). This step reduced multicollinearity and computational cost while preserving salient information. The PCA transformation was able to preserve 95% of the total variance and projected the original 29-dimensional embedding space onto a lower-dimensional subspace that retained the most discriminative spectral–temporal patterns for flood detection

B. U-Net Architecture Design

A modified U-Net architecture was employed for flood susceptibility mapping, adapted from the original biomedical image segmentation framework to accommodate multi-source remote sensing data inputs (Melgar-García et al., 2023). The network was designed to be lightweight, balancing predictive capability with computational efficiency and making it suitable for large-scale regional mapping (Thirugnanasammandamoorthi et al., 2025). The architecture follows the typical encoder-decoder structure with symmetric skip connections (Figure 1). The input data (32 × 32-pixel patches) with 46 feature channels were processed using 5 encoding blocks, 4 decoding blocks, and a bottleneck layer. The encoder path sequentially learned hierarchical features through repeated convolutions and spatial downsampling, reducing spatial dimensions from 32 × 32 to 2 × 2 pixels while increasing feature channels from 32 to 512. Each encoding block comprised convolutional layers with 3×3 kernels, batch normalization for training stability, and 2×2 max pooling for spatial downsampling.
The deepest layer formed a bottleneck containing the most abstract flood-related patterns in a 512-channel feature space. The decoder used transposed convolutions (upsampling) to progressively reconstruct spatial resolution and reduced feature dimensionality (Q. Wang et al., 2025). The skip connection layers were used to join the features of the encoder and decoder layers at each resolution level, thus allowing for the extraction of finer spatial information along with the combination of lower-level features with higher-level semantic information. In order to avoid overfitting, dropout regularization was performed on each convolutional layer in the decoder pathway (dropout ratio = 0.3). The output layer consisted of a 1×1 convolution layer followed by a sigmoid activation function, which predicted pixel-by-pixel flood susceptibility scores between 0 and 1. The model had about 2.4 million parameters, which is a considerable decrease from the traditional U-Net (31 million parameters).

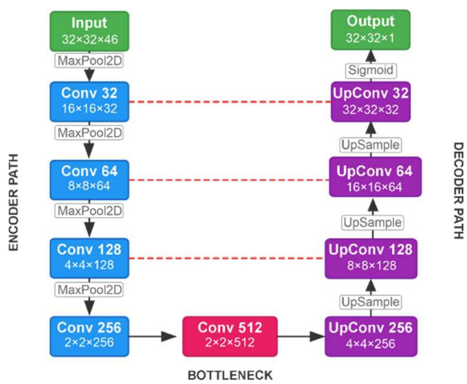

Figure 1. U-Net architecture implemented for flood susceptibility mapping.

C. Training Strategy

Addressing Class Imbalance

There was a challenge in terms of binary classification was difficult because flood pixels made up only about 0.13% of the study area, a flood to non-flood ratio of roughly 1:769. The models produced by trained data tend to make predictions based on the majority class, leading to an overall high accuracy rate but a low proportion of the true distribution of flood locations. This imbalance was compensated using the spatial sampling strategy, focal loss optimization and class weighting.

Spatial stratified Sampling.

To preserve spatial patterns and avoid data leakage, a stratified spatial sampling scheme was used to produce a balanced training dataset. The study area was partitioned into 32×32 pixel patches, from which 6,000 patches with equal class representation were selected (3,000 flood-affected and 3,000 non-flood patches), achieving a balanced 1:1 ratio. To ensure statistical independence between training and validation samples, the spatial separation constraints were imposed on patch extraction with minimum buffer distance of 160 meters (5 patch widths) between samples. This approach ensured balanced class representation while maintaining geographic diversity across Kerala’s varied landscape conditions.

Combined Loss Function

To solve the problem of class imbalance and at the same time optimize on spatial prediction accuracy, a hybrid loss function is developed:

$$L_{\{total\}} = \alpha \times L_{\{focal\}} + \beta \times L_{\{dice\}} \quad (1)$$

where α and β are weighting coefficients set to 0.7 and 0.3, respectively, based on empirical validation performance (Mu et al., 2024).

The focal loss component, weighted at α=0.7 addresses class imbalance by dynamically down-weighting easy-to-classify samples and focusing the learning process on difficult-to-classify samples(Lin et al., 2017). This approach is effective for detecting subtle flood vulnerability patterns in the minority class. The focal loss is defined as:

$$L_{focal} = -\alpha_t(1 - p_t)^{\gamma} log(p_t) \quad (2)$$

where p_t is the model's estimated probability for the true class, α_t is the class-balancing weight, and γ is the focusing parameter (set to 2.0).

The Dice loss component (Yeung et al., 2022), weighted at β=0.3, is included to optimize spatial overlap between predicted and reference flood extents, directly maximizing the Dice similarity coefficient. This loss function is particularly well-suited for segmentation tasks with imbalanced classes as it operates on the overlap region rather than individual pixels:

$$L_{dice} = 1 - \frac{2|X \cap Y|}{|X|+|Y|} \quad (3)$$

where X and Y represent predicted and ground truth flood pixels, respectively.

Class Weighting

For training, an inverse frequency weighting is used to compensate the class imbalance, with weight = 769.0 given to the flood pixels (minority class) and weight = 1.0 given to the non-flood pixels (majority class). This weighting process would make errors on flood pixels cost more in the back propagation phase.

Training Configuration

The U-Net model was trained by using the Adam optimizer (Kingma & Ba, 2017), with a batch size of 128, β_1 = 0.9 and β_2 = 0.999, and the initial learning rate was set to 0.001. The training was carried out for up to 100 epochs. Early stopping with patience of 15 epochs that monitored validation loss restored the weights at the best epochs.

The 6,000 patches were split into training (70%, n = 4,200), validation (15%, n = 900), and test (15%, n = 900) subsets. We stratified the split to preserve class proportions across all three sets.

Data Augmentation

Various methods of data augmentation were used when training the model to ensure the model is generalizable and robust to changes in geometry and radiometry. They were composed of random rotations between -15° and 15°, horizontal flips, and the addition of Gaussian noise (σ = 0.01). For each augmentation, 50% of the time, it was applied at each training step, introducing variety to the training data.

Computational Infrastructure

The whole model training and inference was conducted on the cloud computing environment of Google Collaboratory with hardware with NVIDIA Tesla T4 GPU (16GB VRAM). The accelerated calculation with the help of the GPU took only around 4.5 hours to train the model compared with an estimated 48 hours (CPU only), and thus it is possible to quickly experiment with architecture types and hyperparameter selections.3.4 Model Evaluation

Performance Metrics

A set of quantitative metrics, the performance of the model was assessed for its classification accuracy and spatial prediction quality. These numbers were derived from the independent test set (900 patches) not included in the training set.

The overall accuracy is the percentage of the correct pixels (classified as flood or non-flood):

$$Accuracy = \frac{TP+T}{TP+TN+FP+F} \quad (4)$$

Precision is the percentage of pixels that are correctly predicted by the model as flooded, termed true flood pixels, out of all the flooded pixels:

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

Recall (sensitivity or true positive rate): percentage of pixels in the flood area that are detected by the model:

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

The F1 score is a balanced measure that calculates the average of precision and recall in order to get only one score that encompasses both false positives and false negatives.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

Area under the receiver operating characteristic curve (AUC-ROC) measures the discrimination ability of the model over the whole range of classification thresholds from 0.5 (corresponding to a random classifier) to 1.0 (corresponding to a perfect classifier).
This metric is not dependent on any threshold and quantifies the potential of the model to detect the flood pixels and sort them relative to the non-flood pixels.

Spatial Accuracy Metrics
The Intersection over Union (IoU), also the Jaccard Index, is a measure of the spatial coincidence between predicted and reference flood extent:

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} \tag{8}$$

where *X* and *Y* are the predicted pixel and ground truth pixel of flood, respectively. IoU values are between 0 (no overlap) and 1 (perfect overlap), which offers a strict measure of spatial prediction accuracy, which treats over-prediction and under-prediction equally.

IV. Results

A. Model Performance

The U-Net model was evaluated on an independent test set of 900 patches and demonstrated strong capability in discriminating flood-prone from non-flood patches under extreme class imbalance (Table 2). The model achieved an overall accuracy of 96.55% and an AUC-ROC of 0.9490, indicating strong discriminative ability between flood-susceptible and non-susceptible areas.

TABLE II. PERFORMANCE METRICS OF THE U-NET MODEL ON INDEPENDENT TEST SET (N=900 PATCHES)

Metric	Value
Accuracy	0.9655
Precision	0.4152
Recall	0.0646
F1-Score	0.1118
AUC-ROC	0.9490
IoU	0.2834

The model achieved a precision of 0.4152, recall of 0.0646, and F1-score of 0.1118. The model showed a strong ability to be precise, as the flood pixels accounted for just 0.13% of the study area and most of the time it marked the correct flood data with relatively few false positives.
This trade-off is often tolerable when mapping floods, as false alarms contribute more strongly to loss of confidence in the system than do misses.
Spatial overlap with the reference flood extent was moderate to moderate-high at 28.34% of Iou. This was probably helped by the skip connections in the U-Net architecture, which enabled the model to learn local details and broader spatial information.
The AUC-ROC value was 0.9490 that exhibited good discrimination between flood and non-flood pixels across thresholds. The classification threshold can be adjusted to provide operators with the option to adjust the balance in favor of the flood for maximum coverage, or to reduce the false alarm rate for a given application.

B. Training Convergence

The training and validation losses decreased consistently in the initial layers (Figure 2). The validation loss was at its lowest (0.4183) at epoch 18 and continued to fall thereafter, but the training loss was constant. Throughout, the gap between them remained between 8-12%, indicating that there was not a lot of overfitting. To maintain generalization, one should consider drop out (rate = 0.3), spatial augmentation, and early stopping.

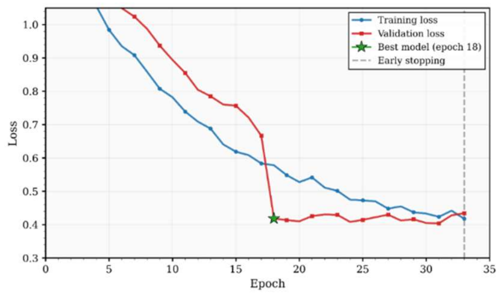

Figure 2: Training and validation loss convergence

C. Spatial Distribution of Flood Susceptibility

The flood susceptibility map produced by the U-Net model revealed clear spatial variations across Kerala, with susceptibility indices ranging from 0.001 to 0.828 categorized into five risk classes using natural breaks classification (Figure 3). The spatial distribution had a strong east-west gradient: the highlands of the Western Ghats had steep slopes and dense forests, so areas of very low susceptibility (44.2%) constituted the majority; the low or moderate susceptibility (43.4% combined) areas covered the regions of the foothills and midland areas; the high and very high susceptibilities (13.4% combined) were localized in the coastal plains, along major river valleys, and urban areas. The northern coastal belt (Kozhikode, Malappuram), central plains and Kuttanad region, and Periyar River floodplain (Alappuzha, Ernakulam) were identified as critical hotspots (susceptibility > 0.381). Southern coastal areas in the vicinity of the river mouths of Pamba and Achankovil also exhibited elevated susceptibility (Kollam, Thiruvananthapuram). Known flood-prone areas were well represented by the model at 30m spatial resolution revealing linear patterns along river courses and dendritic patterns in backwater areas, patterns directly useful for land-use planning and flood mitigation.

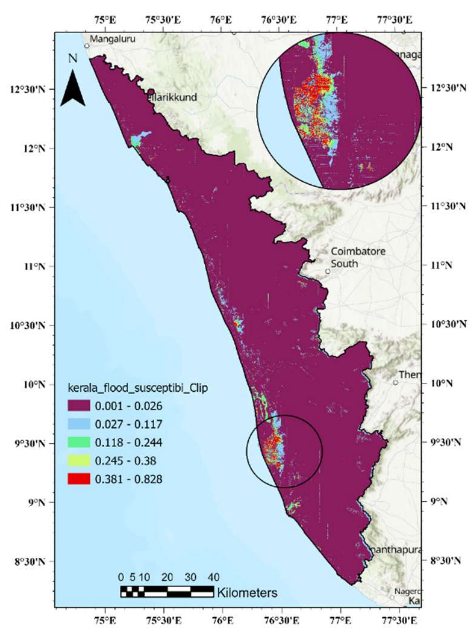

Figure 3. Spatial distribution of flood susceptibility across Kerala state, ranging from very low (0.001-0.026, dark purple) to very high (0.381-0.828, red). Inset maps show detailed views of critical high-risk zones in northern and southern coastal districts.

V. Discussion

A. Performance and Comparative Analysis of the Model

The U-Net performed better than conventional machine learning algorithms in flood susceptibility analysis, providing a remarkable AUC-ROC value of 0.949, suggesting perfect discrimination skills of the algorithm between flooded and not flooded areas. Despite the seemingly low value of the recall metric (0.0646), it should be noted that this metric is affected by the imbalance issue, where only 0.13% of pixels were flooded. The obtained precision of 0.4152 means that 41.5% of high-risk areas are really flooded, which is an acceptable value for pro-active measures against natural hazards, where a high rate of true predictions is critical.

In addition, the U-Net can detect subtle spatial data, as proved by the IoU of 0.2834, being 31.5 times higher than that of a conventional machine learning classifier. Moreover, the algorithm demonstrated excellent validation results with 92.2% spatial agreement with floods occurring in 2018, which was sufficient to assess flood risk in the region of Kerala.

B. Benefits of Deep Learning for Flood Mapping

The U-Net architecture offers several advantages over traditional techniques. First, Skip connections let the U-Net retain spatial detail that standard convolutional networks lose through repeated pooling, an important property when mapping flood boundaries at 30-meter resolution. The encoder-decoder structure also learns features at multiple scales, picking up both local drainage patterns and broader topographic controls.

With 2.4 million parameters, the model processed all of Kerala (38,863 km²) in roughly 4.5 hours, fast enough for operational use. In preliminary tests, models trained on Kerala transferred reasonably to neighbouring states with similar terrain and hydrology.

VI. Conclusions

The Proposed study is aimed at the development and validation of a deep learning model using the U-net technique to generate high resolution flood susceptibility mapping in Kerala, India. The findings show that complicated spatial risk modelling can be effectively tackled using CNN models even in data-rich regions. The model attained 96.55% overall classification accuracy and spatial correspondence of 92.2% for observed flood extents in the devastating floods of Kerala in August 2018.

There are several contributions of the study to operational flood risk assessment. First, multi-source remote sensing data with the inclusion of environmental variables, Sentinel-1 SAR data, and Google Earth Engine embeddings was used to comprehensively characterize flood susceptibility drivers at the topographic, hydrological and land-surface level. Second, stratified spatial sampling, using a combination of loss functions (focal + Dice), and class weighting were found to effectively tackle the large class imbalance between the minority flood and majority non-flood classes (1:769 flood:non-flood) leading to good minority class recognition while maintaining spatial representativeness. Third, the low computational complexity of the lightweight U-Net (2.4 million parameters) enabled the identification of fine-scale spatial information (30 m) for the regional scale.

The susceptibility map generated was showing that 13.4 % of the area was prone to high or very high flood risk categories, with the majority being in the coastal plains, major river valleys (Periyar, Pamba and Chaliyar) and Kuttanad, the low lying part of the state, giving actionable spatial insights to the disaster management authorities.

Future tidal inundation studies should also involve the incorporation of temporal dynamics for dynamic risk forecasting, the incorporation of climate change projections for future susceptibility, and integration of physical susceptibility with social vulnerability indices for supportable equitable DRR. The approach described here can be applied to other areas experiencing flood events, such as those with similar data available, and provides a scalable approach to assess flood risk in the face of greater climatic extremes and more intense urbanization.

Acknowledgement The authors thank the anonymous reviewers for their valuable feedback that helped improve the quality of this paper. We express our gratitude to the United States Geological Survey (USGS) for providing free access to Landsat imagery. The authors also thank JSS Science and Technology University for providing the necessary research infrastructure to support this work.

Funding Not applicable.

Author contribution M S Kendagannaswamy: Writing - review & editing, Writing - original draft, Validation, Software, Methodology, Investigation, Data curation. C K Roopa: Visualization, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. B S Harish: Writing - review & editing, Visualization, Supervision, Methodology.

Declarations

Conflict of Interest The authors declared that they have no conflict of interest.

REFERENCE

[1] Adesina, E. A., Musa, A., Ajayi, O. G., Odumosu, J. O., Opaluwa, Y. D., & Onuigbo, I. C. (2022). Comparative assessment of SRTM and UAV-Derived DEM in flood modelling. Environmental Technology and Science Journal, 12(2), 58–70. https://doi.org/10.4314/etsj.v12i2.6

[2] Bouras, E. houssaine, Jarlan, L., Er-Raki, S., Balaghi, R., Amazirh, A., Richard, B., & Khabba, S. (2021). Cereal Yield Forecasting with Satellite Drought-Based Indices, Weather Data and Regional Climate Indices Using Machine Learning in Morocco. Remote Sensing, 13(16), 3101. https://doi.org/10.3390/rs13163101

[3] Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., & Kohli, P. (2025). AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. http://arxiv.org/abs/2507.22291

[4] Ekmekcioğlu, Ö., Koc, K., Özger, M., & Işık, Z. (2022). Exploring the additional value of class imbalance distributions on interpretable flash flood susceptibility prediction in the Black Warrior River basin, Alabama, United States. Journal of Hydrology, 610, 127877. https://doi.org/10.1016/j.jhydrol.2022.127877

[5] Faruga, Ł., Filapek, A., Kraszewska, M., & Baranowski, J. (2025). Dataset for Traffic Accident Analysis in Poland: Integrating Weather Data and Sociodemographic Factors. Applied Sciences, 15(13), 7362. https://doi.org/10.3390/app15137362

[6] Jain, N., Martha, T. R., Khanna, K., Roy, P., & Kumar, K. V. (2021). Major landslides in Kerala, India, during 2018–2020 period: an analysis using rainfall data and debris flow model. Landslides, 18(11), 3629–3645. https://doi.org/10.1007/s10346-021-01746-x

[7] Johnson, J. A., & Arunachalam, M. (2009). Diversity, distribution and assemblage structure of fishes in streams of southern Western Ghats, India. Journal of Threatened Taxa, 1(10), 507–513. https://doi.org/10.11609/JoTT.o2146.507-13

[8] Kaushal, A., Gupta, A. K., & Sehgal, V. K. (2024). A semantic segmentation framework with UNet-pyramid for landslide prediction using remote sensing data. Scientific Reports, 14(1), 30071. https://doi.org/10.1038/s41598-024-79266-6

[9] Kendagannaswamy, M. S., Roopa, C. K., Harish, B. S., & Maheshan, M. S. (2026). Improving Flood Susceptibility Assessment in Kerala: Integration of Advanced Machine Learning Models with Socioeconomic Vulnerability Analysis. Journal of the Indian Society of Remote Sensing. https://doi.org/10.1007/s12524-026-02440-y

[10] Kendagannaswamy, M. S., Roopa, C. K., Harish, B. S., & Mukesh, M. S. (2025). Multi-criteria decision analysis for regional-scale flood susceptibility mapping in Kerala state, India. Discover Applied Sciences, 7(6), 598. https://doi.org/10.1007/s42452-025-07182-z

[11] Khan, Z. A., & Jhamnani, B. (2024). Development of flood susceptibility map using a GIS-based AHP approach: a novel case study on Idukki district, India. Journal of Spatial Science, 69(2), 409–442. https://doi.org/10.1080/14498596.2023.2236051

[12] Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. http://arxiv.org/abs/1412.6980

[13] Lal, P., Prakash, A., Kumar, A., Srivastava, P. K., Saikia, P., Pandey, A. C., Srivastava, P., & Khan, M. L. (2020). Evaluating the 2018 extreme flood hazard events in Kerala, India. Remote Sensing Letters, 11(5), 436–445. https://doi.org/10.1080/2150704X.2020.1730468

[14] Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal Loss for Dense Object Detection. 2017 IEEE International Conference on Computer Vision (ICCV), 2999–3007. https://doi.org/10.1109/ICCV.2017.324

[15] Liu, Y., Song, S., Wang, M., Gao, H., & Liu, J. (2025). DE-Unet: Dual-Encoder U-Net for Ultra-High Resolution Remote Sensing Image Segmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 12290–12302. https://doi.org/10.1109/JSTARS.2025.3565753

[16] Maddah, S., Mojaradi, B., & Alizadeh, H. (2025). Improving deep learning-based flood susceptibility modeling by integrating data balancing technique and dual-input convolutional neural network. Natural Hazards, 121(15), 17555–17577. https://doi.org/10.1007/s11069-025-07482-y

[17] Melgar-García, L., Martínez-Álvarez, F., Tien Bui, D., & Troncoso, A. (2023). A novel semantic segmentation approach based on U-Net, WU-Net, and U-Net++ deep learning for predicting areas sensitive to pluvial flood at tropical area. International Journal of Digital Earth, 16(1), 3661–3679. https://doi.org/10.1080/17538947.2023.2252401

[18] Mia, Md. U., Chowdhury, T. N., Chakrabortty, R., Pal, S. C., Al-Sadoon, M. K., Costache, R., & Islam, A. R. Md. T. (2023). Flood Susceptibility Modeling Using an Advanced Deep Learning-Based Iterative Classifier Optimizer. Land, 12(4), 810. https://doi.org/10.3390/land12040810

[19] Mohan, M. G., & Adarsh, S. (2023). Dynamic flood frequency analysis for west flowing rivers of Kerala, India. Water Security, 19, 100137. https://doi.org/10.1016/j.wasec.2023.100137

[20] Mu, Y., Nguyen, T., Hawickhorst, B., Wriggers, W., Sun, J., & He, J. (2024). The combined focal loss and dice loss function improves the segmentation of beta-sheets in medium-resolution cryo-electron-microscopy density maps. Bioinformatics Advances, 4(1). https://doi.org/10.1093/bioadv/vbae169

[21] Parthasarathy, K. S. S., & Kundapura, S. (2023). Mapping of Flood-Inundated Urban Regions Using Sentinel-1 SAR Imagery for the 2018 and 2019 Kerala Floods (pp. 279–292). https://doi.org/10.1007/978-981-99-4423-1_20
[22] Raj, R., & Sofi, A. A. (2023). Does climate change leads to severe household-level vulnerability? Evidence from the Western Ghats of Kerala, India. Land Use Policy, 130, 106655. https://doi.org/10.1016/j.landusepol.2023.106655
[23] Sadhwani, K., Eldho, T. I., & Karmakar, S. (2022). Flooding Problems in Periyar River Basin, Kerala—The Effects of Land Use Land Cover Changes. In 5th World Congress on Disaster Management (pp. 16–24). Routledge. https://doi.org/10.4324/9781003341932-3
[24] Thirugnanasammandamoorthi, P., Ghosh, D., Dewangan, R. K., Hasan, M. K., Ariffin, K. A. Z., Abbas, H. S., Elshafie, H., Saeed, R. A., & Awouda, A. E. (2025). FloodNet-Lite: A Lightweight Deep Learning for Flood Mapping Using Remote Sensing Data With Optimized UNet and Edge Deployment Approach in 6G. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 20294–20314. https://doi.org/10.1109/JSTARS.2025.3591406
[25] Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: a fully automated processing chain. International Journal of Remote Sensing, 37(13), 2990–3004. https://doi.org/10.1080/01431161.2016.1192304
[26] Vilasan, R. T., & Kapse, V. S. (2022). Evaluation of the prediction capability of AHP and F-AHP methods in flood susceptibility mapping of Ernakulam district (India). Natural Hazards, 112(2), 1767–1793. https://doi.org/10.1007/s11069-022-05248-4
[27] Wang, J., Sanderson, J., Iqbal, S., & Woo, W. L. (2025). Accelerated and Interpretable Flood Susceptibility Mapping Through Explainable Deep Learning with Hydrological Prior Knowledge. Remote Sensing, 17(9), 1540. https://doi.org/10.3390/rs17091540
[28] Wang, Q., Li, C., & Zhong, Q. (2025). Evaluating Flood Susceptibility in Coastal Regions of China Using Deep Learning Models and Multispectral Remote Sensing Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 20611–20623. https://doi.org/10.1109/JSTARS.2025.3594612
[29] Yeung, M., Sala, E., Schönlieb, C.-B., & Rundo, L. (2022). Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized Medical Imaging and Graphics, 95, 102026. https://doi.org/10.1016/j.compmedimag.2021.102026
[30] Youssef, A. M., Mahdi, A. M., & Pourghasemi, H. R. (2023). Optimal flood susceptibility model based on performance comparisons of LR, EGB, and RF algorithms. Natural Hazards, 115(2), 1071–1096. https://doi.org/10.1007/s11069-022-05584-5