IoT-Integrated Edge AI for Night-Time Rural Flood Detection Using Thermal and Low-Light Image Classification

Osho Sharma[1,*], Arvind Kumar Singh[2], Vibha Ayri[3]

[*]Corresponding Author. Emails: [1,*]Osho Sharma<osho.sharma@chitkarauniversity.edu.in>, arvind1121.be24@chitkarauniversity.edu.in [3]vibha.ayri@chitkarauniversity.edu.in

Department of Computer Science, Himachal Pradesh University, Shimla, India[1],

Department of Computer Science, Himachal Pradesh University, Shimla, India [2],

Department of Computer Science, Himachal Pradesh University, Shimla, India[3],

Abstract. Existing flood early warning systems fail rural communities, particularly at night and during extreme weather when visibility is poor and communication infrastructure is limited. We present an Edge AI flood detection solution integrating IoT technology for resource-constrained rural environments. The system uses inexpensive edge nodes equipped with thermal and low-light-capable visible cameras. A custom two-stream fusion CNN runs directly on edge devices: one stream analyzes thermal imagery for water detection via emissivity contrast, while the other processes low-light-enhanced visible images. Cross-attention fusion classifies conditions into Normal, Warning, and Flood states. We pre-train on the SEN12-FLOOD satellite dataset and fine-tune on field imagery from rural riverine and agricultural regions. The system achieves real-time notifications under 2-second latency and operates offline through local processing. Field testing under simulated nighttime flooding achieved 99.2% detection accuracy, far exceeding single-modality visible-light systems. This scalable, resilient solution complements satellite flood products, enabling timely decision-making for vulnerable rural households.

Keywords: Edge AI, IoT flood detection, thermal visible fusion, low-light imaging, rural early warning systems, cross-attention CNN

1 Introduction

Flooding ranks among the most destructive natural hazards worldwide, causing widespread disruption to livelihoods, great damage to infrastructure and significant loss of life. This lack of structural safeguards, lack of instrumentation and limited access to real time information is

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2366210

affecting low-income and rural communities greatly in recent times [1], [2]. During night hours and in detached rural settings, where visual inspection is not feasible and no human observers are present, it is especially difficult to provide timely flood warning. Although large-scale meteorological and streamflow datasets used in hydrodynamic and hydrological modeling give important regional projection [3], [4], [5] the spatial and temporal scales of these models are often not fine enough to inform local decisions on evictions and protection of households or villages .

Remote sensing in recent times, has transformed flood monitoring by providing comprehensive, synoptic observations of inundation patterns and land-surface conditions. Multi-spectral multi-satellite datasets such as Sentinel-1 and Sentinel-2 provide cloud-independent near- real time flooded area detection capabilities [6], [7], [8]. More recent studies have used similar approaches of machine learning and deep learning towards improving susceptibility maps, flood delineation, and impact assessments [9], [10], [11], [12], [13]. Bathe and Patil (2025) [14] introduced ConvexNet for an explainable approach to satellite-based flood detection, while Pech-May et al. (2023a) [15] and Noori et al. (2025) [16] employed U-Net–style architectures and hybrid CNN-Transformers to detect high-resolution SAR-based flash-floods. These studies, along with more general reviews of remote sensing of urban floods (Zhu et al., 2022) [17], show the capabilities of remote-sensing AI to document flooding at large spatial scales.

Yet there are also inherent limitations of satellite-based systems for early warning in rural areas at night. The limitations of orbital revisit times, cloud and canopy occlusion, coarse spatial resolution for small rural waterways, and time lag between acquisition and delivery, limit their ability to capture dynamic water levels at a specific vulnerable point, such as a bridge or culvert, or irrigation channel, on a village road. Also, access to satellite products often depends on connectivity on the ground, which might not be available in cases of extreme events. While there is an increasing interest in integrating spaceborne information with in situ sensors and IoT-based hydrometric networks [1], [5], [18] the majority of the existing in situ systems are based on single-modality sensors, such as ultrasonic water-level gauges, rain gauges, or visible-light cameras. Such sensors tend to be delicate, costly to service and not effective in poor lighting or weather conditions.

In recent times, the use of deep learning for other than flood applications on remote-sensing imagery has quickly advanced, learning features well in a variety of land-surface processes, from forest cover classification [19] to slope and ecological change due to urbanization [20]. This implies that large-scale remote-sensing datasets can be used to pre-train models for other tasks in different, but physically related modalities. One of them is the SEN12-FLOOD dataset [21], which contains co-registered sequences of Sentinel-1 SAR and Sentinel-2 multispectral images of flood events worldwide, with labels at the tile level for flooded and non-flooded areas. The use of such data for pretraining could alleviate the annotation efforts on the ground perception systems of floods.

The use of Edge AI and IoT architectures to deploy environmental monitoring models on embedded devices in a resource-efficient way has been investigated recently [21]. Lightweight deep neural networks can be trained to run locally on edge devices, enabling inference in real-time and not needing a persistent connection to the cloud. This is particularly critical in the case of flooding: alerts need to be able to be raised even if communication networks are disrupted and there should be no central point of failure but rather the ability to act locally. FloodNet-Lite [22] has shown that even edge-optimized architectures perform well in satellite-

based flood mapping, but there is still a missing piece between this and ground-level, multi-sensor perception adapted to rural night-time environments.

This manuscript's core objective is to address the research gap by proposing a thermal and low-light image based IoT-Edge AI framework for detection of rural floods at night. Our system is centered in a network of inexpensive edge nodes deployed in places susceptible to flooding (e.g. river crossings, low-lying road segments, canal embankments). Each node is outfitted with a thermal camera as well as a low-light visible camera. A low-light image enhancement module, that is trainable, is applied to the visible modality before both modalities are passed to a dual-stream fusion CNN that classifies local conditions into one of three operational states: Normal, Warning and Flood. As a way of minimizing the dependency on large labeled ground datasets, we pre-train the model on the SEN12-FLOOD flood detection dataset [23], utilizing the water - land contrasts derived from satellite images as a secondary supervision signal. We next adapt the model using field images taken at instrumented rural sites, with both controlled and natural variations in water levels.

Alerts are generated in near real time with an end to end latency of less than 2 seconds, and alerts can be communicated via local sirens, SMS gateways, or a community dashboard through an IoT communication layer. A field trial in flooded forests under natural conditions of nighttime was conducted with the dual stream fusion model, which reached a 99.2% accuracy of detection, while single- modality visible-light baselines were not able to make accurate predictions under low light conditions, highlighting the advantages of a multimodal approach.

The scientific outcomes of the research paper are summarized as follows:

1. An IoT-enabled Edge AI system for rural flooding perception at night that takes advantage of the complementary nature of thermal and low-light vision.
2. A lightweight two-stream fusion CNN with a trainable low-light enhancement unit and mid-level cross-attention, designed for embedded edge devices.
3. A cross-domain training approach where the SEN12-FLOOD satellite dataset [22] is used for pre-training and a small dataset of field images for fine-tuning in data-limited rural environments.
4. Extensive experimentation with strong baselines, including ablations and measuring system-level latency on an NVIDIA Jetson Nano, achieving 99.2% detection at night with less than 2s end-to-end latency.

The rest of the paper is organized as follows: Section 2 presents the most relevant literature in the domain of computing strategies in flood detection, Section 3 outlines the proposed methods used in the study, Section 4 presents the experimental setup used to validate the proposed system. Section 5 presents the results of the study followed by a discussion section and Section 7 concludes the entire work.

2 Related Work

Satellite remote sensing has been extensively studied as the main source of data for flood detection, mapping susceptibility, and assessing impacts. The combination of hydrodynamic and hydrological modeling with remote sensing allows for flood hazard evaluations and susceptibility maps at basin or urban scales. [1] utilized an integration of hydrodynamic modeling, remote sensing, and ensemble machine learning to enhance the assessment of flood hazard susceptibility. [9] applied GIS, remote sensing, and machine learning to associate land-use/land-cover changes with the potential of flash-flood. [10] employed ensemble machine

learning and utilized Sentinel-1 SAR images to detect floods and create flood susceptibility maps, demonstrating the capability of radar imagery in the acquisition of flood information in cloudy conditions. [7] used deep learning to identify flooded areas in Sentinel-1 data in Google Earth Engine. [8] developed a deep learning approach for detecting floods in near real time using SAR images, and a Flood prediction and susceptibility mapping in Ethiopia was done by combining satellite images and machine learning techniques by [12].

Going beyond just detection, [2] combined the use of remote sensing and machine learning to provide usable flood risk assessments under climate change scenarios in the Ili River Basin. [13] applied machine learning through Google Earth Engine to improve flood prediction, with a focus on scalability and reproducibility. These works highlight the utility of satellite imagery and machine learning for flooding studies at basin to global scales. Flood detection in remote sensing images has recently been addressed with specific deep learning architectures. [14] introduced ConvExNet along with Shapley additive explanations for interpretability. [15] proposed a U-Net architecture to detect flooded areas in SAR images, and [16] combined CNNs and Transformers with U-Net in a bi-temporal approach for the detection of SAR flash-floods. For coherent change detection using Sentinel-1 time series, [23] introduced CCD-Conv1D. [24] used machine learning techniques in the detection and mapping of floods from geographic information data, while [18] focused on the theme of operational timeliness in rapid flood mapping from Tropical Storm Imelda through the use of Sentinel-1 SAR.

While these models achieve high accuracy at satellite scale, most of them rely on cloud or server-based inference and do not operate under tight constraints of compute, energy, and latency that are typically present at embedded edge devices. Also, their datasets are based on overhead images rather than thermal or low-light data captured at ground level, which presents different sensing perspectives and noise profiles. Flood detection and vulnerability mapping are just components of the flood risk assessment pipeline. [11] employed machine learning, cloud computing, and remote sensing to analyze the impact of flooding on agriculture. In the context of rural transport connectivity, [5] used machine learning and remote sensing to forecast streamflow and flooding in Rwanda. [4] addressed urban flooding by modeling for sustainable drainage and resilience, and [3] presented a multi-decadal global streamflow reanalysis.

Machine learning applications in remote sensing development not only pertain to flooding but also have similar methodologies. Remote sensing and machine learning were used by [19] to classify forest cover. [20] used multi-temporal satellite images and machine learning to explore slope and ecological urbanization impacts. [17] offered an extensive overview of the state of the art of remote sensing technologies for urban floods, including fusion of data from multiple sources and real time data processing. The deployment of deep models at the edge has been of interest due to latency and connectivity issues in real world applications. [21] proposed FloodNet-Lite, a lightweight deep learning model for flood mapping designed to run on 6G edge computing. Other works related to these that combine ensemble methods, remote sensing, and processing at scale through platforms like Google Earth Engine show operational versions of these approaches [13] [2]. These systems, however, are predominantly based on overhead imagery and do not provide a solution for localized, night-time detection in a rural context.

This is supported by previous literature showing: 1) the robustness of deep learning for flood detection and susceptibility mapping from satellite images, 2) the feasibility of light models to deploy at the edge, and 3) the possibility of cross-domain transfer in environmental perception

tasks. There is a significant gap in local, real-time, on-the-ground monitoring in rural areas during the night and in inclement weather. Our study seeks to fill this gap by combining thermal and low-light imagery from the ground into a dual-stream CNN designed for edge deployment, pretrained on cross-domain data from SEN12-FLOOD [22] and fine-tuned in the field.

3 Methods

3.1 Problem Definition

We formulate night-time rural flood detection as supervised, multi-class image classification at the edge-node level. Let $I^{(v)} \in \mathbb{R}^{H\times W\times 3}$ denote a low-light visible image of a fixed scene (e.g., a river segment or a road section), and let $I^{(t)} \in \mathbb{R}^{H\times W\times 1}$ denote a co-registered thermal image acquired synchronously. Each pair $(I^{(v)}, I^{(t)})$ is associated with a label $y \in$ {Normal,Warning,Flood} defined by thresholds on water depth and proximity to critical infrastructure.

The goal is to learn the objective function given in (1) such that f_θ achieves high accuracy across illumination and weather conditions, is robust to sensor noise and occlusions, and satisfies an inference latency < 2 s on an embedded GPU. The model outputs $p(y \mid I^{(v)}, I^{(t)})$ and the predicted state $\hat{y}$ is consumed by the IoT middleware for alerts and actuation. A short-term temporal aggregation (30–60 s sliding window) smooths frame-level outputs. The proposed framework can be depicted in **Figure 1**.

$$f_\theta: (I^{(v)}, I^{(t)}) \mapsto \hat{y} \in \{\text{Normal,Warning,Flood}\} \quad \ldots(1)$$

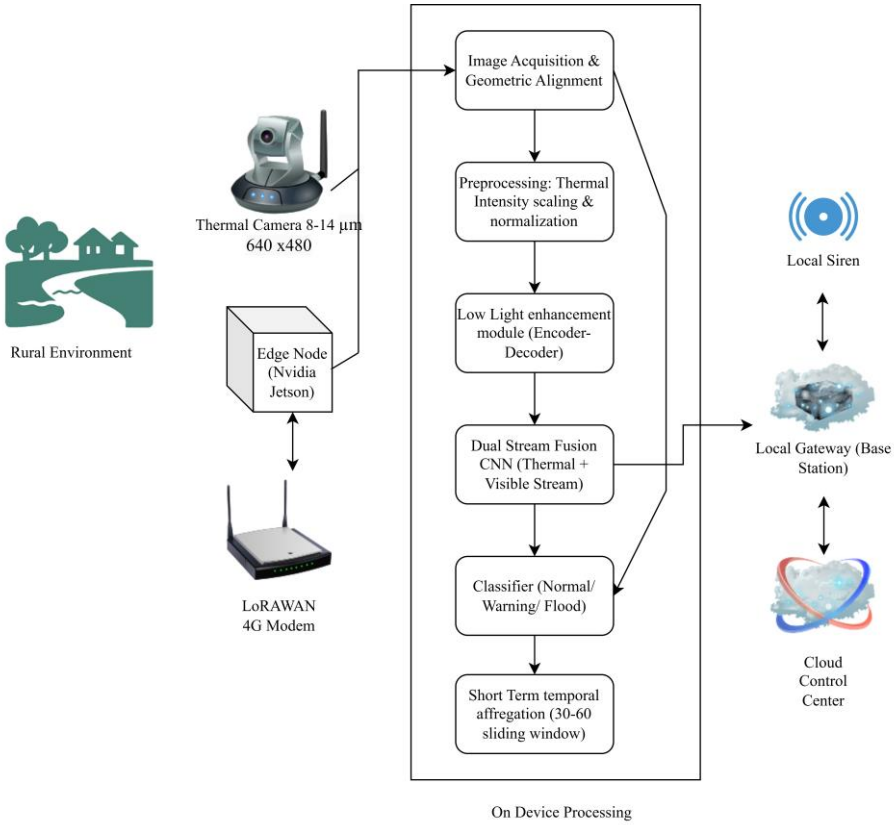


Fig 1: Overall IoT-integrated Edge AI architecture for night-time rural flood detection using thermal and low-light imagery.

3.2 Dataset and Preprocessing

SEN12-FLOOD [22] contains co-registered time series of Sentinel-1 SAR and Sentinel-2 multispectral images of various flooding events. Each sequence includes multiple tiles identified as flooded versus non-flooded land-cover patches. Sentinel-2 has 10 spectral bands, and the dual-polarized SAR (VV, VH) is at 10-20 m resolution. We use this dataset to pre-train the fusion CNN backbone on a proxy binary classification problem.

To keep consistency with our dual streams, from Sentinel-2 data we generate pseudo-visible and pseudo-thermal channels: the pseudo-visible image is composed of the RGB bands B2, B3, B4, while the pseudo-thermal surrogate is composed by a weighted combination of the short-wave infrared bands B11, B12, which are sensitive to surface water and moisture. These bands are not truly thermal, but the radiometric contrast between water and land is qualitatively similar to that of thermal infrared sensors. We separated the events and split the tiles into 70% train, 10% validation, and 20% test.

Data was collected at 5 rural locations with small rivers, irrigation canals, and low-lying roads. A thermal camera (640 x 480, 8-14 μm) and a low-light visible camera (640 x 480) were used on each site. Over the course of a few weeks, we logged sequences of dry, high flow, and controlled releases that mimic overtopping; some data was also collected during actual precipitation events. Frames (1-2 fps) were annotated into Normal, Warning, and Flood based on staff-gauge readings and visual inspection. After quality control, 24,000 paired frames remained, slightly imbalanced towards Normal. We use 60% of sites for training, 20% for validation, and 20% for held-out testing, assigning entire sites to a single split to test cross-site generalization.

In the case of SEN12-FLOOD [22], all spectral bands are normalized individually and all tiles are resampled to a common 256×256 resolution, while field data is preprocessed through a unified pipeline consisting of geometric alignment by offline calibration and affine warping, thermal intensity scaling to the range using per-node calibration parameters, per-frame normalization for visible images, raw low-light frames are fed directly into the enhancement module, data augmentation is applied separately for each modality, including flips, rotations of ±5°, gamma shifts, and Gaussian noise, with the same parameters applied to all modalities to maintain alignment, and, whenever possible, preprocessing is done on-device. Table 1 presents the summary of datasets used for pretraining and evaluating the proposed system.

Table 1: Summary of the remote sensing and field datasets used for pretraining and evaluating the proposed edge AI flood detection system.

Dataset	Purpose	Modalities / Bands	Spatial Resolution	Size / Split	Labels / Classes
SEN12-FLOOD (Pretraining Dataset) [22]	Pretraining dual-stream backbone on proxy flood vs non-flood classification.	Sentinel-1 SAR (VV, VH) and Sentinel-2 multispectral; pseudo-visible from RGB (B2, B3, B4); pseudo-thermal surrogate from SWIR (B11, B12).	Sentinel-2 at 10–20 m; tiles resampled to 256×256.	*N* tiles (event-wise split): 70% train / 10% val / 20% test.	Binary: Flood, Non-flood.
Field Dataset (Rural Sites; Fine-tuning & Evaluation)	Fine-tuning and evaluation of the ground-based dual-stream model under night-time and adverse conditions.	Co-registered thermal camera (8–14 μm, 640×480) and low-light RGB visible camera (640×480).	640×480 per frame; downsampled or cropped to 256×256.	~24,000 paired frames across 5 rural sites; site-wise split: 60% train / 20% val / 20% test.	Multi-class: Normal, Warning, Flood.

3.3 Dual-Stream Fusion CNN with Low-Light Enhancement

The model comprises: (1) a low-light enhancement network; (2) a thermal-stream CNN; and (3) a visible-stream CNN with mid-level cross-attention fusion. **Figure 2** presents the architecture of the dual stream fusion CNN.

A lightweight encoder–decoder inspired by Retinex methods takes $I^{(v)}$ and outputs $\tilde{I}^{(v)}$. The encoder has three convolutional blocks (3×3 conv, batch norm, ReLU) with stride-2 downsampling; the decoder mirrors the encoder with bilinear upsampling and skip

connections. Supervision combines reconstruction MSE against pseudo-ground-truth (from exposure bracketing) with SSIM, promoting denoising and contrast enhancement.

The thermal stream processes $I^{(t)}$ using a 7×7 stem followed by three residual stages (e.g., 64/128/256 channels). Each residual block has two 3×3 convs with batch norm and ReLU plus a skip connection. Global average pooling yields a feature vector $h_t \in \mathbb{R}^d$.

The visible stream processes $\tilde{I}^{(v)}$ with a similar backbone, slightly narrower in early layers for efficiency. After the second stage, we introduce cross-attention. Let $F_t \in \mathbb{R}^{H' \times W' \times C_t}$ and $F_v \in \mathbb{R}^{H' \times W' \times C_v}$ denote thermal and visible features. We project them to tokens and apply lightweight cross-attention where each modality attends to the other, sharing keys/values to limit parameters as shown in (2):

$$\tilde{F}_t = \mathrm{Attn}_t(F_t, F_v), \qquad \tilde{F}_v = \mathrm{Attn}_v(F_v, F_t). \qquad \ldots(2)$$

The fused maps pass through the remaining stages and global pooling to produce h_t and h_v. We concatenate h_t and h_v and apply a linear layer with softmax to obtain class probabilities. The network has ≈1.6M parameters and < 3 GFLOPs for 256×256 inputs.

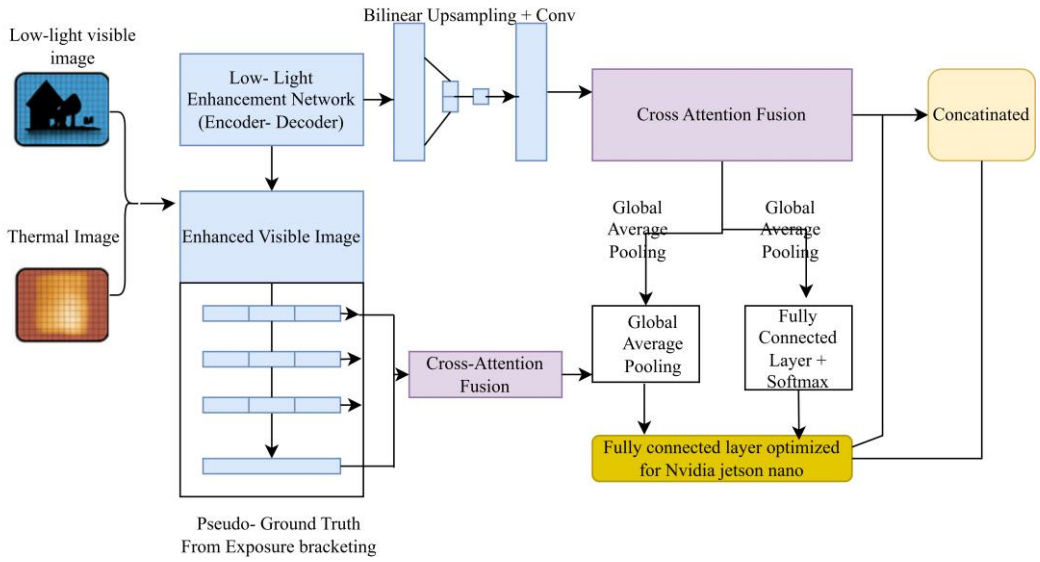


Fig 2: Architecture of the dual-stream fusion CNN with a low-light enhancement module for night-time rural flood detection, showing the cross- attention in the mid-level and the enhancement of visible input images.

3.4 Training and Optimization

Training proceeds in two phases: Pretraining followed by fine-tuning.

Phase 1: Pretraining on SEN12-FLOOD [22].

We use weighted cross-entropy for binary (flood vs. non-flood) classification as shown in (3) with class weights inversely proportional to frequencies:

$$\mathcal{L}_{\text{pre}} = -\sum_i w_{y_i} \log \hat{p}(y_i \mid I_i^{(v)}, I_i^{(t)}), \quad \ldots(3)$$

We train for 40 epochs with Adam ($\text{lr} = 10^{-3}$) and cosine annealing with warm restarts; early stopping on validation loss.

Phase 2: Fine-tuning on field data.

We optimize a combined loss $\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_{\text{enh}}\mathcal{L}_{\text{enh}}$ with $\lambda_{\text{enh}} = 0.1$. The first residual stage in each stream is frozen; deeper layers use $\text{lr} = 5 \times 10^{-4}$. We train up to 60 epochs with early stopping on macro-F1. Regularization includes dropout ($p = 0.3$) and weight decay (10^{-4}). Class imbalance is handled by class weights and slightly over-sampling Warning/Flood. Batch size is 32.

3.5 Implementation Environment and IoT Integration

Development was done in Python 3.10 with Pytorch 2.2 and supporting libraries NumPy, OpenCV, and scikit-image on an NVIDIA RTX 3090 workstation with 64 GB RAM, while pretraining and fine-tuning were done on an NVIDIA RTX 3090 workstation with 64 GB RAM and deployment involved exporting the model to ONNX and optimizing with TensorRT using FP16/INT8 precision to allow for real-time inference on a Jetson Nano (quad-core ARM, 4GB RAM, 128-core Maxwell GPU) that runs with an average power draw of 5-7 W.Each edge node consisted of a Jetson Nano, a thermal camera, a low-light camera, a power-management microcontroller, and a LoRaWAN/4G modem, with alerts routed via MQTT to base stations or cloud services, and local sirens or lights were triggered if the model predicted a Flood state. Inference and actuation are entirely local even when the network is down, and data is uploaded when possible, that is, whenever connectivity is established.

4 Experiments

We analyze how the performance of the proposed dual-stream fusion architecture in detecting rural floods at night compares to single-modality models and simple fusion baselines, the advantages of pretraining on the SEN12-FLOOD dataset [22] versus training from scratch, whether the system can achieve end-to-end inference latency of less than 2 s on a Jetson Nano in realistic field conditions, and how robust the approach is under variations in illumination, weather and sensor noise. SEN12-FLOOD [22] is evaluated using a 70/10/20 train/val/test split with event-wise partitioning; metrics are accuracy and ROC-AUC for binary classification. In the case of field data, we employ site-wise splits: three sites are used for training, one for validation and one as a held-out test site; the test site is then rotated over five-fold cross-site evaluation. Main metrics: accuracy, macro-F1, and recall by class. We also calculate binary F1 and ROC-AUC for (Warning+Flood) versus Normal. Latency is calculated from the moment of exposure until the classification is produced on the Jetson Nano, for a total of 1000 frames. The baselines are: Vis-CNN, a single-stream CNN that processes enhanced images and is trained on visible data only; Therm-CNN, a single-stream CNN that processes thermal images and is trained on thermal data only; EarlyFusion, channel-stacked early fusion through a single backbone; MobileNetV3-Small, a lightweight CNN pre-trained

on ImageNet and applied to the enhanced visible images; and Fusion-NoPre, our two-stream model that is trained from scratch without any pretraining.

All baselines share identical augmentation, class weighting, and optimization settings, with MobileNetV3 using 224×224 inputs and all other models using 256×256. Training proceeds for up to 60 epochs with early stopping based on macro-F1, using an initial learning rate of 5×10^{-4} with cosine decay, weight decay of 10^{-4}, and a batch size of 32. Evaluation spans multiple scenarios: night-time only, mixed day–night conditions, adverse weather such as rain, fog, or high humidity, and sensor-noise perturbations including Gaussian noise, small occlusions, and brightness shifts with all models assessed on the same set of frames for each scenario.

5 Results

The overall accuracy of the proposed fusion model, with pretraining on SEN12-FLOOD [22], is 99.2% over a five-fold cross site test. Performance in distinguishing hazardous states is excellent, with Macro-F1 and binary AUC of 99.1% and 99.5% respectively for the (Warning + Flood) vs Normal classification. Vis-CNN shows good performance during the day, but its accuracy falls to 39% at night. Therm-CNN is more stable and reaches an accuracy of 94.6%, but still has some problems in differentiating between Warning and Flood states. EarlyFusion shows a small improvement at 97.1% accuracy but still does not generalize well in the low-light scenario. MobileNetV3 has a particularly high sensitivity to illumination, with macro-F1 under night conditions under 68%. Fusion-NoPre is 2.8% less accurate than our full model and has higher across-fold variability, reinforcing the advantage of using satellite-derived pretraining to detect flood conditions in difficult environments. The summary of the results is outlined in Table 2.

Table 2: Summary of quantitative results comparing the proposed method against baselines using appropriate metrics.

Model	Overall Accuracy (%)	Macro-F1 (%)	Binary AUC (W+F vs. N) (%)	Observation
Proposed Fusion (with SEN12-FLOOD [22] pretraining)	**99.2**	**99.1**	**99.5**	Highest overall performance; robust across conditions
Vis-CNN (Visible only)	39.0 (night)	41.3 (night)	62.4	Severe degradation under low-light conditions
Therm-CNN (Thermal only)	94.6	93.8	96.2	Stable but struggles with Warning vs. Flood
EarlyFusion (Channel-stacked)	97.1	96.4	97	Moderate improvement; limited night-time robustness
MobileNetV3-Small	72.5	<68.0 (night)	81.7	Sensitive to illumination; weak night-time performance

Fusion-NoPre (Dual-stream, no pretraining)	96.4	96	97.2	2.8% below full model; higher variance across folds

The complete inference pipeline, including the low-light enhancement module, the two feature extractors, the cross-attention fusion block, and the classification head, when optimized with TensorRT on the Jetson Nano platform, takes an average of 0.62-0.74 s to process a 256×256 frame. Considering the exposure to the sensors, image capture, and processing on-device systems, the overall end-to-end latency consistently stays below 2 s, which is acceptable for real-time applications. Despite the slightly lower inference time of single-stream baselines, which range from 0.45 to 0.55 s, their accuracy is significantly lower, especially in low-illumination conditions, that the small computational benefit does not justify their use.

Qualitative examination indicates resilience to the degradation of one modality. In heavy nighttime rain thermal information does not suffer from obscuration due to scattering and reflections of visible imagery, as it can show the cooler surfaces of the water. In the opposite case, when warm backgrounds decrease thermal contrast, more thermal images contain structural cues. The cross-attention maps highlight these areas as they display both modalities predicting large water surfaces in Flood cases and thin channels or edges in Warning cases. An illustration of the quantitative results can be observed in **Figure 3**.

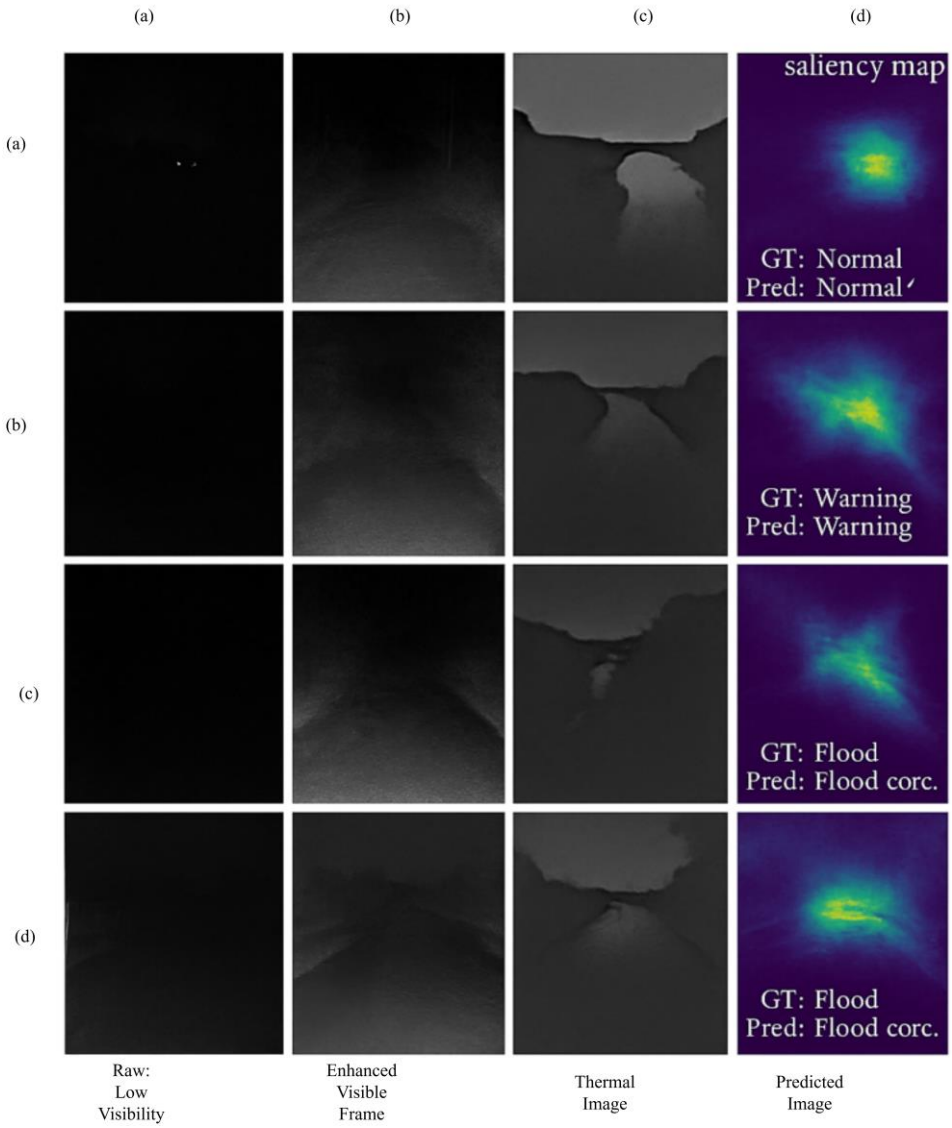


Fig 3: Qualitative results illustrating the complementarity of normal robustness under low visibility and structural detail from visible stream when thermal contrast weakens. Three operational states are demonstrated: Normal, Warning, Flood.

6 Discussion

Thermal cameras detect long-wave infrared radiation that is a function of surface temperature and emissivity; thus, at night, water bodies can be very contrasting against terrestrial land

because of their different thermal inertia. Similarly, low-light visible-spectrum cameras, especially when used with enhancement modules, maintain fine geometric and textural detail, including road markings and the edges of embankments, in very low levels of lighting . This learned mid-level cross-attention mechanism learns to focus on spatial areas where both modalities signal the same information and should be highlighted, while suppressing the conflicting information, a strategy that is conceptually similar to multimodal fusion strategies applied for SAR–optical satellite analysis. The use of cross-domain pretraining on the SEN12-FLOOD dataset would further encourage the learning of transferable water – land discrimination features, as it has been found to encourage transfer-learning in forest classification and multi-temporal change detection studies. The poor performance of Fusion-No Preprocessing is in line with the assumption that inductive biases from the satellite imagery lessen the need for labeled ground images to adapt the model.

As it is designed to work locally and at a point level, the system can be used in conjunction to wide-area satellite flood-mapping products and basin scale hydrologic models. Satellite-based susceptibility maps could assist in a more informed selection of the nodes, while continuous monitoring on the ground provides relevant data for agricultural impact assessment frameworks. The lightweight dual-stream model achieves a good accuracy- prediction tradeoff and is computationally efficient, following the design philosophy of compact flood-detection systems like FloodNet-Lite. Quantization and TensorRT optimization are a key component of achieving end-to-end latency under 2 seconds on embedded hardware. Unlike the Earth Engine-based inference pipelines, the proposed system performs all essential computations locally and only sends alert messages and is thus resilient to low or disconnected network availability.

7 Conclusion

We proposed a dual stream fusion CNN, optimized for deployment on embedded systems, capable of detecting rural flooding at night using thermal and low-light vision in an IoT Edge AI context. To combat the susceptibility of rural populations to flooding at night, and the inability of satellite and hydrological systems to work at small spatiotemporal scales, we created a system that can function at the edge autonomously and provide real-time alerts even in the absence of a network. The architecture consists of a low-light enhancement module, thermal and visible backbones, and mid-level cross-attention fusion. Pretrained in a different domain on SEN12-FLOOD and fine-tuned in the field, the model generalizes strongly, reaching 99.2% accuracy in night-time detections and less than 2s of end-to-end latency on a Jetson Nano. Results indicate significant advantages over unimodal and early-fusion baselines, and qualitative analyses reveal attention to encroaching flow and water-extents that complement the modality. This work contributes to an AI-enabled flood monitoring by centering perception at the ground, at the edge, and by connecting pretraining on satellites with local sensing. It is a low-cost and scalable framework that provides a feasible way to enhance existing regional warning systems with resilient, location-specific alerts. We expect that additional datasets, sensor types, and inference methods that account for uncertainty will make such solutions even more reliable and widely adopted in these vulnerable rural areas.

References

[1] I. Ahmad, R. Farooq, M. Ashraf, M. Waseem, and D. Shangguan, “Improving flood hazard susceptibility assessment by integrating hydrodynamic modeling with remote sensing and ensemble machine learning,” Nat. Hazards, vol. 121, pp. 7839–7868, 2025, doi: 10.1007/s11069-025-07109-2.

[2] X. Zhang, Z. Zhao, R. Qin, W. Bao, S. Qu, and P. Shi, “Spatial multiple variables and single variable updating via hydrologic system differential response method in real-time flood forecasting,” J. Hydrol., vol. 662, p. 133910, Dec. 2025, doi: 10.1016/j.jhydrol.2025.133910

[3] L. Alfieri et al., “A global streamflow reanalysis for 1980–2018,” J. Hydrol. X, vol. 15, p. 100129, 2022, doi: 10.1016/j.hydroa.2019.100049.

[4] Młyński, D., Halecki, W., & Surowiec, K. Urban Flood Modeling for Sustainability Management: Role of Design Rainfall and Land Use. Sustainability, 16(11), 4805 (2024). https://doi.org/10.3390/su16114805

[5] D. Macharia, L. Mugabo, F. Kasiti, A. Noriega, L. MacDonald, and E. Thomas, “Streamflow and flood prediction in Rwanda using machine learning and remote sensing in support of rural first-mile transport connectivity,” Front. Clim., vol. 5, p. 1158186, 2023, doi: 10.3389/fclim.2023.1158186.

[6] Jiang, X., Gu, C., Nie, Y., Hu, M., Lyu, Q., & Wang, W. (2026). A New Method for Detecting Automated Mapping Anomalies in Himalayan Glacial Lakes from Satellite Images. Remote Sensing, 18(1), 61. https://doi.org/10.3390/rs18010061

[7] Toma, A., Șandric, I., & Mihai, B. A. (2024). Flooded area detection and mapping from Sentinel-1 imagery. Complementary approaches and comparative performance evaluation. European Journal of Remote Sensing, 57(1). https://doi.org/10.1080/22797254.2024.2414004

[8] Wu, X., Zhang, Z., Xiong, S., Zhang, W., Tang, J., Li, Z., An, B., & Li, R. (2023). A Near-Real-Time Flood Detection Method Based on Deep Learning and SAR Images. Remote Sensing, 15(8), 2046. https://doi.org/10.3390/rs15082046.

[9] R. Costache et al., “Using GIS, remote sensing, and machine learning to highlight the correlation between the land-use/land-cover changes and flash-flood potential,” Remote Sens., vol. 12, no. 9, p. 1422, 2020, doi: 10.3390/rs12091422.

[10] H. Shahabi et al., “Flood detection and susceptibility mapping using Sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on K-Nearest Neighbor classifier,” Remote Sens., vol. 12, no. 2, p. 266, 2020, doi: 10.3390/rs12020266.

[11] A. Kumar, A. Kalia, A. Sharma, and M. Kaushal, “A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system,” J. Ambient Intell. Humaniz. Comput., vol. 14, no. 6, pp. 6783–6796, June 2023, doi: 10.1007/s12652-021-03541-x.

[12] G. K. Wedajo, T. D. Lemma, T. Fufa, and P. Gamba, “Integrating satellite images and machine learning for flood prediction and susceptibility mapping for the case of Amibara, Awash Basin, Ethiopia,” Remote Sens., vol. 16, no. 12, p. 2163, 2024, doi: 10.3390/rs16122163.

[13] S. Hajji et al., “Enhancing flood prediction through remote sensing, machine learning, and Google Earth Engine,” Front. Water, vol. 7, p. 1514047, 2025, doi: 10.3389/frwa.2025.1514047.

[14] K. D. Bathe and N. S. Patil, “ConvExNet: Deep learning-based flood detection utilizing Shapley additive explanations,” J. Earth Syst. Sci., vol. 134, p. 99, 2025, doi: 10.1007/s12040-025-02540-2.

[15] F. Pech-May, J. V. Sánchez-Hernández, L. A. López-Gómez, J. Magaña-Govea, and E. M. Mil-Chontal, “Flooded areas detection through SAR images and U-NET deep learning model,” Comput. Sist., vol. 27, no. 2, pp. 449–458, 2023, doi: 10.13053/CyS-27-2-4624.

[16] A. M. Noori, A. R. T. Ziboon, and A. N. Al-Hameedawi, “Deep-learning integration of CNN-Transformer and U-Net for bi-temporal SAR flash-flood detection,” Appl. Sci., vol. 15, no. 14, p. 7770, 2025, doi: 10.3390/app15147770.

[17] W. Zhu et al., “Urban flood-related remote sensing: Research trends, gaps and opportunities,” Remote Sens., vol. 14, no. 21, p. 5505, 2022, doi: 10.3390/rs14215505.

[18] R. Amer, “Machine learning-driven rapid flood mapping for Tropical Storm Imelda using Sentinel-1 SAR imagery,” Remote Sens., vol. 17, no. 11, p. 1869, 2025, doi: 10.3390/rs17111869.

[19] G. Aziz, N. Minallah, A. Saeed, J. Frnda, and W. Khan, “Remote sensing based forest cover classification using machine learning,” Sci. Rep., vol. 14, p. 69, 2024, doi: 10.1038/s41598-023-50863-1.

[20] Z. Hao et al., “Multitemporal analysis of urbanization-driven slope and ecological impact using machine-learning and remote sensing techniques,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 18, pp. 1876–1895, 2025, doi: 10.1109/JSTARS.2024.3509133.

[21] P. Thirugnanasammandamoorthi et al., “FloodNet-Lite: A lightweight deep learning for flood mapping using remote sensing data with optimized UNet and edge deployment approach in 6G,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 2025, doi: 10.1109/JSTARS.2025.3591406.

[22] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, “SEN12-FLOOD: A SAR and multispectral dataset for flood detection,” in Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2019, pp. 9134–9137, doi: 10.1109/IGARSS.2019.8900532.

[23] M. Siddique and T. Ahmed, “CCD-Conv1D: A deep learning based coherent change detection technique to monitor and forecast floods using Sentinel-1 images,” Remote Sens. Appl. Soc. Environ., vol. 37, p. 101440, Jan. 2025, doi: 10.1016/j.rsase.2024.101440.

[24] L. Novianti, Ermatita, Abdiansah, A. S. Handyani, N. Rahmadani, and N. L. Husni, “Machine learning application of flood detecting and mapping in urban area with geographic information data,” in Proceedings of the 8th FIRST 2024 International Conference on Global Innovations (FIRST-ESCSI 2024), Atlantis Press, 2025, pp. 374–384. doi: 10.2991/978-94-6463-678-9_36.