An Image-Based Urban Flood Detection System using Deep Learning and Conformal Prediction for Edge Deployment

Guilherme Bisse Alves Rodrigues[1], **Arthur Lima Marques Rocha**[2], **Jó Ueyama**[2], **Caetano Mazzoni Ranieri**[1]

[1] Instituto de Geociências e Ciências Exatas
Universidade Estadual Paulista (UNESP) – Rio Claro – SP – Brazil

[2]Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo (USP) – São Carlos – SP – Brazil

{guilherme.bisse,cm.ranieri}@unesp.br,
arthurtuca@usp.br, joueyama@icmc.usp.br

Abstract. *Urban floods threaten infrastructure and public safety, making reliable early detection systems essential. In this context, this paper presents a solid deep learning pipeline for flood detection using real-world stream images. Due to the underrepresentation of flood images, the proposed architecture was strengthened with a semantic masking data augmentation approach, and conformal prediction was applied to set statistical confidence thresholds for model inference. Using a temporal hold-out strategy, the proposed system achieved a significant F1-score of 78.65% for the critical overflow class, successfully minimizing serious false negatives and demonstrating its operational reliability for real-time monitoring of urban flooding.*

1. Introduction

The deleterious effects of natural disasters include substantial human and material losses, and have become increasingly frequent worldwide [Cunha et al. 2023]. The economic losses resulting from natural disasters are alarming, having reached R$ 28 billion in Brazil in 2025, according to data from the website *Economic News Brasil* [Economic News 2026]. According to Tellman et al. [Tellman et al. 2021], between 2000 and 2018, a total flood area of 2.23 million square kilometers was determined, with 255 to 290 million people directly affected by the floods. It is estimated that the total population in areas with satellite-observed flooding grew between 58 and 86 million people from 2000 to 2015.

Existing flood monitoring systems often rely on physical sensors, such as water-level gauges. Although effective, these solutions may involve high installation and maintenance costs, limited spatial coverage, and be vulnerable to hardware and software failure. Hence, with the increasing availability of embedded systems and advances in computer vision, image-based monitoring has emerged as a promising alternative for environmental monitoring [Upadhyaya et al. 2025, Kalogiannidis et al. 2025]. Recent progress in deep learning, particularly convolutional neural networks (CNNs), has enabled reliable image classification in real-world scenarios. Lightweight architectures such as *MobileNet* and its variants have been widely explored and adopted to combine IoT with Machine Learning due to their reduced computational requirements [Zhan 2024].

To ensure the proposed system is truly reliable and safe when deployed in real-world scenarios, this work introduces a deep learning pipeline to address the numerous challenges associated with early water-level detection. To mitigate the severe data imbalance and prevent spatial overfitting, the pipeline introduced a novel data augmentation strategy with a dynamically segmented water mask [Jain et al. 2023]. It also applied oversampling strategies to the training samples. This forces the network to focus strictly on the visual behavior of the water flow.

In addition, our approach incorporated a *Class-Conditional Conformal Prediction* framework [Balasubramanian et al. 2014, Karimi and Samavi 2023]. This mathematical procedure allows us to quantify the model's uncertainty by calculating class-specific certainty thresholds for each water level, providing statistical guarantees even for the rarest overflow events. The pipeline combined this mathematical foundation with an anti-risk fallback policy that serves as a protective layer during periods of high uncertainty. By dynamically adjusting decision thresholds to prioritize social safeguard, this mechanism drastically reduces the occurrence of fatal false negatives when a flood is imminent.

The rest of this paper is organized as follows: Section 2 provides context by reviewing related work in vision-based environmental monitoring. Section 3 breaks down our methodological approach, from data preparation and model architecture to the conformal prediction framework. Section 4 presents and discusses the experimental results, demonstrating the system's performance. Finally, Section 5 brings the paper to a close with our conclusions and thoughts on future work.

2. Related Work

Flood monitoring has typically relied on physical sensors and high-computing tools to measure water levels and detect significant conditions. Unlike traditional warning systems, which encounter problems such as high costs for installation, maintenance and restricted coverage areas, as pointed out by Hashi et al. and Chavula et al. [Hashi et al. 2021, Chavula et al. 2025], image-based monitoring systems have emerged as a distinct alternative. These systems employ computer vision technology to automatically monitor water levels in urban areas.

With the progress in deep learning, convolutional neural networks have shown great effectiveness in image classification tasks, such as flood detection, and have achieved acceptable levels of accuracy [Ranieri et al. 2024, Roohi et al. 2025]. Recent research has shown that using lightweight CNNs in edge computing applications provides satisfactory performance along with improved efficiency, making them a practical option for edge AI [Musa et al. 2025]. However, the computational efficiency and model accuracy are not sufficient for effective disaster monitoring.

In practical applications, the reliability of present flood detection systems is a major concern. Most current techniques involve training deterministic networks, which suffer from extensive spatial overfitting and often emphasize static background characteristics rather than dynamic water properties [Atok and Chai 2025]. Furthermore, most training datasets used for natural disasters exhibit disproportionate class imbalances, making them more likely to produce biased predictions that overlook important events and instead focus on more common, typical situations.

Despite the reasonable overall accuracy of these approaches, they are ineffective

at capturing a model's uncertainty when faced with environmental noise (e.g., fog, lens reflections, or shadows) and can lead to quiet, overly confident failures. This study goes beyond conventional fine-tuning techniques to achieve the necessary degree of spatial generalization and address the identified weakness in fixed cameras regarding background memorization [EL HARIRI et al. 2023, Li et al. 2024].

Based on successful applications of segmented masks for data augmentation across various domains [Yang et al. 2022, Kebaili et al. 2023, Bosquet et al. 2023], we propose a dynamic water-mask method. This approach involves the neural network directly recognizing specific fluid characteristics from the continuous flow, enabling stable performance regardless of changes in the surrounding environment.

Even though many existing image-based flood detection systems achieve high overall accuracy [Ghosh et al. 2022, Stateczny et al. 2023], they primarily rely on deterministic point forecasts that fail to adequately consider model uncertainty [He et al. 2026, Li et al. 2026]. In critical scenarios, this overconfidence poses a major problem that can result in fatal false negatives during rare and severe flood events.

Building on the success of conformal forecasting in other risk-sensitive areas, we address this limitation by integrating a *Class-Conditional Conformal Prediction* framework directly into our pipeline. By developing prediction sets that offer statistical certainty, combined with a safety-oriented fallback policy, the system shifts the emphasis from merely maximizing standard accuracy to proactively guaranteeing public safety. It adjusts real-time decision-making criteria to prevent silent failures from occurring as a flood approaches.

3. Methodology

This section presents the dataset used in this study, the proposed architecture, and the evaluated conformal prediction framework.

3.1. Dataset Acquisition and Preparation

The public dataset presented by [Ranieri et al. 2024] was employed in the experiments made in this study. The dataset contains 68,599 images collected by a monitoring system installed in a critical flooding location at a creek located in the city of São Carlos (SP), Brazil, and was collected between 2018 and 2022. The acquisition process reflects real-world environmental conditions, including variations in luminosity, weather, and water level. The labels consider four distinct classes of images corresponding to water levels: *low*, *medium*, *high*, and *flood*.

Reflecting the nature and frequency of urban flooding, the dataset exhibits severe class imbalance, since the creek remains at a low water level when it is not raining. To ensure a reliable training and evaluation process, the dataset was divided into four subsets: *Training*, *Validation*, *Calibration* and *Test*, as shown in the model training pipeline (see Figure 2). Specifically, the test set consisted exclusively of data from 2020 in order to assess the model's temporal generalization. Following this split, an oversampling strategy was applied exclusively to the training set to artificially balance the minority classes, while keeping the validation and test sets in their original unbalanced state. This separation ensures that the model learns correctly and can generalize to real-world situations it has never encountered before.

3.2. Dynamic Masking and Data Augmentation

Images such as those of an urban stream, which feature a static background and dynamic river flow, present a unique challenge for deep learning models: spatial overfitting. Instead of learning the fluid dynamics of water, neural networks often tend to memorize static background elements, such as bridges, concrete walls, or vegetation. To counteract this tendency, we developed a custom data augmentation workflow that includes dynamic water masking, illustrated in Figure 1.

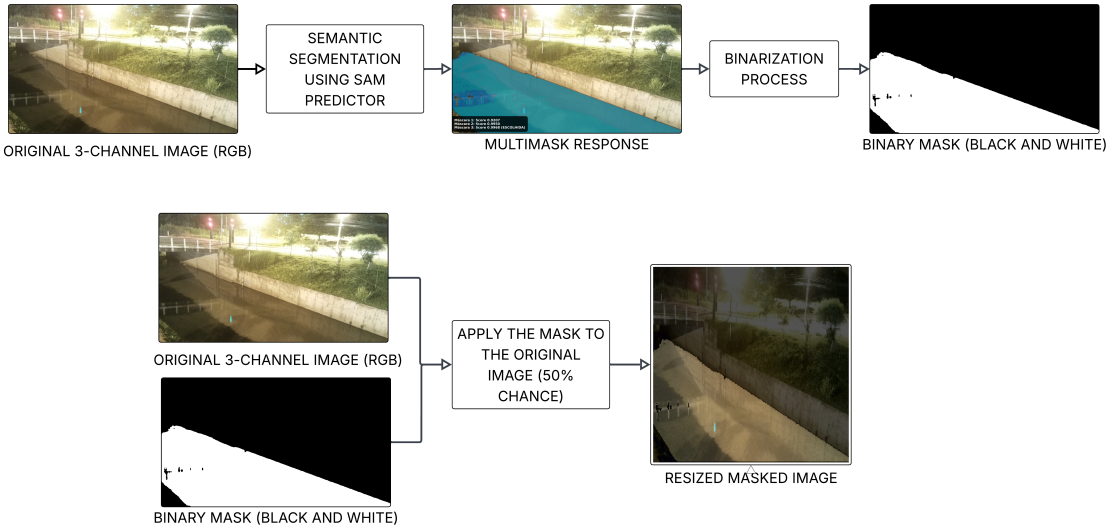


Figure 1. Mask generation pipeline using the SAM predictor.

Initially, the Segment Anything Model (SAM) [Kirillov et al. 2023] was used to generate a binary mask of the main flow area of the watercourse. Instead of applying this mask statically to all training data, we integrated it probabilistically. During the batch loading process, each image has a 50% probability of being masked — darkening the background and isolating the water — and a 50% probability of remaining unchanged. This approach forces the network to extract generalized textural features directly from the water's surface, while maintaining sufficient global context to understand the scene. Additional real-world transformations, such as slight rotations and brightness adjustments, were also applied to simulate variable weather and camera conditions.

3.3. Model Architecture and Imbalance Handling

Given the strict hardware constraints of edge computing environments, such as IoT cameras, we selected MobileNetV3 Large as our primary architecture. Unlike traditional models, which require significant computational power, MobileNetV3 uses depth-separable convolutions, inverted residual blocks, and lightweight attention mechanisms to maximize feature extraction while keeping computational overhead to a minimum [Howard et al. 2019]. To promote transparency and full reproducibility, the complete implementation of this framework, including the training pipeline, preprocessing sequence, customized datasets, and inference protocols, is publicly available in our open-source repository[1].

[1] https://github.com/GuilhermeBisse/Flood-Detector

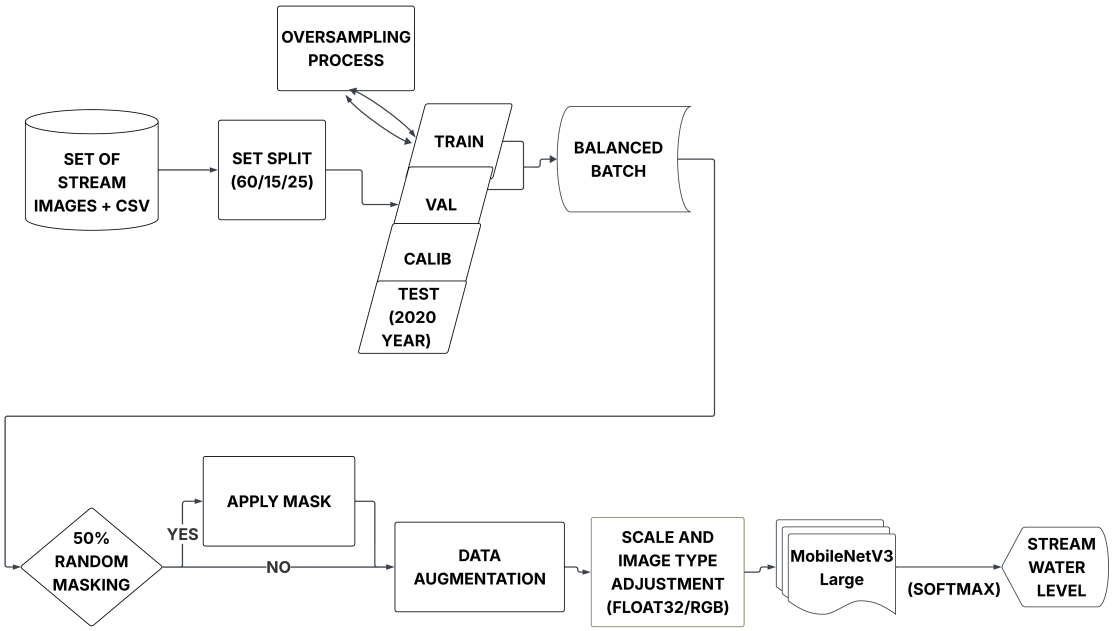


Figure 2. Main Workflow

For our deep learning pipeline, the most important architectural adaptation addresses the extreme class imbalance inherent to natural disaster datasets. Under standard Categorical Cross-Entropy (CE), the loss is computed proportionally to sample frequency. In our scenario, where normal baseline conditions (*low*) outnumber critical events (*flood*) by orders of magnitude, the accumulated gradients are overwhelmingly dominated by the easily classified majority class. Consequently, a network optimizing CE can achieve artificially high global accuracy by essentially ignoring the minority class—a fatal operational flaw for an early warning system.

To mathematically correct this asymmetry, we replaced the standard loss function with Sparse Categorical Focal Loss [Lin et al. 2017]. Focal Loss extends standard CE by introducing a modulating factor $(1 - p_t)^\gamma$ that dynamically scales the loss based on prediction confidence rather than mere class frequency:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \tag{1}$$

The focusing parameter γ smoothly adjusts the rate at which easy examples are down-weighted. Following the empirical foundations established by Lin et al. [Lin et al. 2017], which demonstrated optimal hard-example mining in severe imbalance scenarios, we set $\gamma = 2.0$. In the context of flood detection, this value aggressively suppresses the gradient contribution of the tens of thousands of highly confident *low*-level stream images (driving their loss near zero) without rendering the network hyper-reactive to random visual noise. Coupled with the weighting factor α_t, dynamically computed for each class via Scikit-Learn [Pedregosa et al. 2011] to establish a baseline frequency inversion, this nonlinear mechanism forces the network to allocate its learning capacity entirely to the infrequent flood events. This ensures that critical risk states drive the weight updates, effectively preventing the silent failures typical of traditional CE optimizations.

The simultaneous use of techniques such as oversampling, class-wise α weighting and focal loss is an aggressive strategy, but extremely important when dealing with severe dataset imbalance. Empirical tests conducted throughout the study demonstrated that applying these techniques in isolation failed to guarantee operational safety: removing oversampling led to unacceptable spatial overfitting on minority classes, while omitting focal loss caused the model to predict many false negatives during critical events. Therefore, the over-corrective behavior was a deliberate and strategic trade-off to prioritize public safety.

3.4. Conformal Prediction and Risk-Averse Fallback

In environmental monitoring, model overconfidence constitutes a critical vulnerability. To transform our model from a simple image classifier into a reliable early warning system, we implemented a Class-Conditional Conformal Prediction framework at the end of the pipeline.

Given a calibration dataset, the framework computes a non-compliance score $s(x, y)$ for each sample, which measures the degree of deviation of a true label y from the model's prediction $\hat{f}(x)$. In this work, we adopt the standard hinge loss or probability-based score:

$$s(x, y) = 1 - \hat{f}(x)_y \tag{2}$$

where $\hat{f}(x)_y$ represents the Softmax probability assigned to the true class y. Based on the distribution of these scores within the calibration set, a quantile $\hat{q}$ is determined to ensure a user-defined error rate α (set to $\alpha = 0.05$ in our experiments). For any new, unseen test image X, the system constructs a prediction set $\hat{C}(X)$ containing all classes whose scores fall below the calculated threshold. This prediction set satisfies the formal marginal coverage guarantee:

$$P(Y \in \hat{C}(X)) \geq 1 - \alpha \tag{3}$$

To ensure operational safety, this mathematical framework is combined with a risk-averse contingency policy. If a captured image is highly ambiguous, distorted by bad weather, or affected by nighttime glare, the conformational predictor may produce an empty set ($\hat{C}(X) = \emptyset$). Instead of halting the operation, an empty set triggers a safety protocol that bypasses the statistical threshold and evaluates the raw probability distribution, automatically assuming the highest detected risk level. This design ensures that the edge system fails safely, strongly favoring false alarms over fatal false negatives during imminent overflow events.

4. Results and Discussion

The experiments performed were evaluated according to the model performance, the impacts of the dynamic data augmentation strategy, and the conformal prediction framework.

4.1. Model Performance

The performance of the proposed architecture was evaluated using an out-of-time temporal split with the 2020 test dataset, which was separated from the remaining years to

ensure that the model is capable of generalizing to images and unexpected noises it has never seen before. The classification results are presented in the confusion matrix (Figure 4). The training curves are shown in the Figure 3. In the context of early warning systems, the global training accuracy (96.48%) is an inadequate metric, as it evaluates overall performance but overlooks significant details, such as severe class imbalance and asymmetric importance levels. In such situations, the model's classification performance should be measured by its ability to correctly identify critical levels, despite the overwhelming number of images under normal conditions.

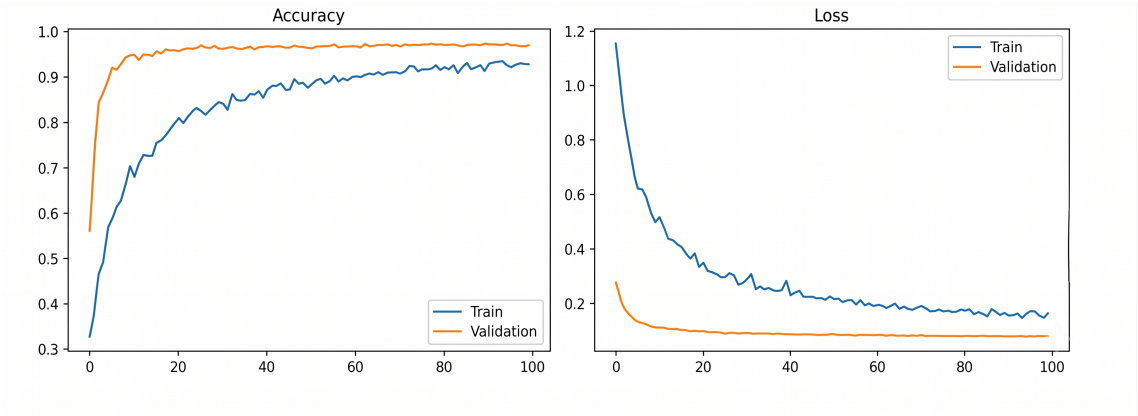


Figure 3. Training Curves

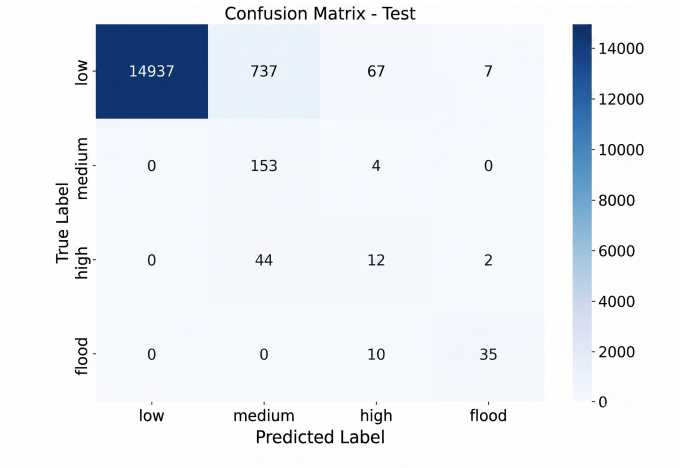


Figure 4. Confusion Matrix

It is essential to clarify the transition from the set-valued outputs of the Conformal Prediction to the single-label deterministic metrics reported in the confusion matrix and F1-scores. The metrics presented reflect the final operational state of the pipeline after the application of the risk-averse fallback logic. When the CP framework generates a multi-class set due to high uncertainty (e.g., $\{high, flood\}$), the fallback logic forces a deterministic selection of the most critical state to prioritize safety. Conversely, when severe visual noise prevents any class from reaching its confidence threshold, the framework produces an empty set ($\emptyset$), yielding an average ambiguity size of 0.98. In cases of empty sets, the fallback logic then selects the argmax prediction for the test image,

ensuring that the model's predictions align with those shown in the confusion matrix. Finally, the reported performance metrics evaluate the entire integrated system—classifier, CP calibration, and fallback logic—rather than just the raw neural network output.

The confusion matrix reflects the system's operational behavior after the application of the conformal fallback logic, revealing a highly conservative performance: the strict minimization of fatal false negatives. Among all critical High and Flood events in the test dataset, the classifier produced 0 predictions for the *Low* class (normal water level). For flood events, it correctly classified 35 of the 45 images (77.78% Recall), and was able to classify the remaining 10 images as *High*, triggering an orange alert rather than ruling out the possibility of a disaster. Consequently, the proposed architecture achieved a strong F1-Score of 78.65% for the *Flood* class, despite having only 45 samples from that class in 2020, demonstrating the model's high capability to reliably and accurately identify anomalies and disasters in real-world scenarios.

Table 1. Predictive Performance Metrics

Metric	Value
Global Accuracy	96.48%
Balanced Accuracy	72.69%
Test Loss	0.0846
Average Inference Time	96.78 ms

Table 2. Class-wise Performance Evaluation

Class	Precision	Recall	F1-Score	Support
Low	1.0000	0.9485	0.9736	15,748
Medium	0.1638	0.9745	0.2806	157
High	0.1290	0.2069	0.1589	58
Flood	0.7954	0.7778	0.7865	45

As anticipated by the risk-averse design, the pipeline exhibits a deliberate trade-off when making predictions involving water level transition classes. For the *Medium* class, it achieved an extremely high recall (97.45%), with an F1-score of 28.05%. The architecture classified 737 normal (*Low*) samples as Medium, which explains the extremely low Precision rate for this class (16.38%). For the *High* class, it exhibited very low levels of Precision and Recall, at 12.90% and 20.69%, respectively, achieving an F1-Score of only 15.89%. Considering also that the Normal level (*Low*) class achieved 100% Precision, these results reflect the model's conservative behavior, avoiding classifying them as low as much as possible and prioritizing the intermediate classes.

4.2. Impacts of Dynamic Augmentation

To assist the neural network during training, probabilistic semantic masking and class-conditional augmentations were applied. This strategy proved extremely effective in mitigating spatial overfitting by helping the network focus on water flow dynamics and adverse environmental conditions. Nevertheless, it introduced a bias that controls the false positive rate.

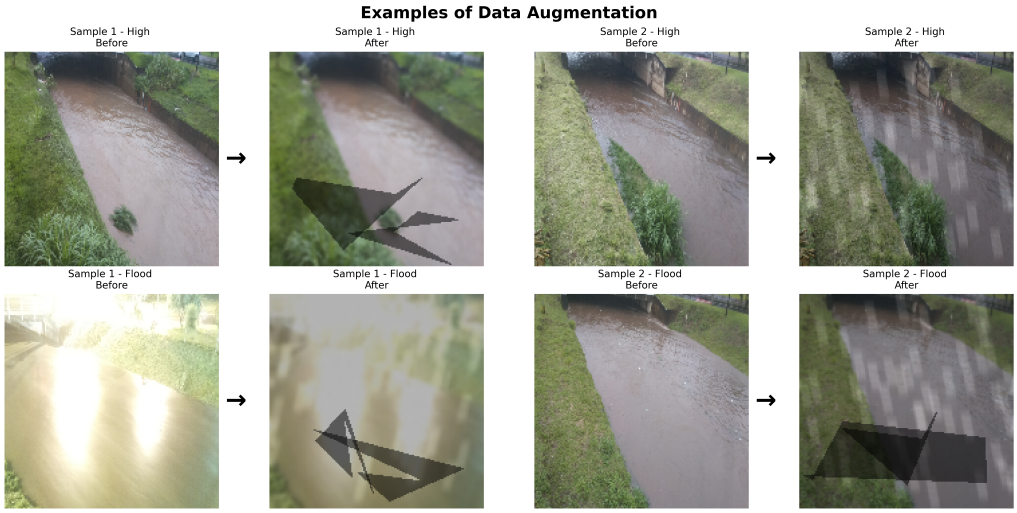


Figure 5. Dynamic Augmentation - Before and After Samples

Since the *Low* and *Medium* classes go through the same transformation pipeline, in cases where normal images are blurred or have altered contrast, the model may interpret them as being more saturated than they actually are. That explains once again why those 737 *low* samples were actually classified as *medium*.

It also triggered a red alert for 7 normal samples, classifying them as *Flood*, which represents an error rate of just 0.04% among the 15,748 *Low* samples. These false positives are often produced by exaggerated distortions in the image, allowing the model to make mistakes and push original samples from the *Low* class into adjacent classes.

From a deployment perspective, this trade-off is optimal. In the context of real-world edge deployment, this simple problem could be resolved by establishing an alert persistence strategy across multiple consecutive frames, preventing a false positive from triggering a response from emergency control agencies.

4.3. Conformal Prediction

The operational robustness of the proposed pipeline is empirically bounded using the Class-Conditional Conformal Prediction Framework. Tested using the calibration dataset, the outputs of the deterministic neural network were adjusted using an error tolerance of $\alpha = 0.05$. This framework established dynamic confidence thresholds for each of the four classes: **71.85% for Low, 38.49% for Medium, 10.07% for High, and 52.79% for Flood**.

The results obtained from the calibration process reflect the physical reality of the urban stream. The model must exceed a high threshold to actually classify a sample as a normal water level (*Low*), whereas the confidence thresholds for classifying an image as *Medium* and, especially, *High* are significantly lower. These values are also useful for understanding the results displayed by the confusion matrix.

The Conformal Prediction framework was calibrated to guarantee a marginal coverage of $1 - \alpha = 0.95$, yet empirical evaluation on the 2020 test set yielded 90.84%. This deviation results from the fundamental limitations of applying standard CP to an out-of-time temporal split. CP's mathematical guarantees rely strictly on exchangeabil-

ity between calibration and test distributions, violated by temporal domain shifts (e.g., seasonal lighting changes and sensor degradation) in the subsequent year. Furthermore, the extreme scarcity of calibration samples for minority classes (*high* and *flood*) constrained the uncertainty bounding for rare events. Ultimately, this highlights that while CP provides robust theoretical bounds under i.i.d. conditions, real-world IoT deployments require conservative operational safeguards to counteract temporal degradation.

Acknowledgements

This work received funding from the São Paulo Research Foundation (FAPESP), grant 2025/20968-7.

5. Conclusion

This paper introduced a comprehensive and reliable computer vision pipeline that goes beyond the standard methodology of image classification models and aims to address common challenges in water flow and natural disaster scenarios, such as extreme dataset imbalance, spatial overfitting, and overconfidence.

The pipeline was developed by integrating the MobileNetV3 backbone with class-conditional data augmentation techniques, probabilistic semantic masks, and conformal prediction. The proposed architecture achieved an overall accuracy of 96.48% and a satisfactory F1-score of 78.65% for the critical *Flood* class, and produced 0 fatal false negatives, pushing difficult samples into the adjacent High class, thereby triggering an attentiveness state.

The most significant contribution of this article lies in the implementation of the class-based conditional conformational prediction framework, which constructs sets of predictions based on a predefined level of certainty ($\alpha = 0.05$) instead of making deterministic and biased predictions. This approach resulted in statistical coverage of over 90% of the images, making the model much more reliable. It also achieved an average ambiguity of 0.98 classes per set. The incorporation of risk-averse fallback logic provides an effective filter for this ambiguous predictions and empty sets, prioritizing public safety without overloading the early warning system with excessive false alarms.

A limitation of the projected model is the difficulty in correctly classifying the *Medium* and *High* transition classes, which had F1 scores of 28.0% and 15.89%, respectively, as reflected in the number of false positives. Another problem concerns a specific behavior of the classifier, which tends to generate false positives in cases of extreme environmental noise. A follow-up to this work will involve improving the model's capacity on correctly generalize to transition classes and deploying this pipeline on a micro controller device for the purpose of conducting a qualitative analysis of the model's performance at the edge. In addition, temporal continuous inference strategies will be useful for ensuring reliability as frames change, avoiding false alarms caused by highly noisy data, and contributing to the creation of an autonomous and trustworthy early warning system for urban flooding.

References

Atok, D. A. M. and Chai, S. S. (2025). Addressing overfitting and overestimation challenges in landslide susceptibility modeling: a case study of penang island, malaysia. *Natural Hazards*, 121(11):13577–13604.

Balasubramanian, V., Ho, S.-S., and Vovk, V. (2014). *Conformal prediction for reliable machine learning: theory, adaptations and applications*. Newnes.

Bosquet, B., Cores, D., Seidenari, L., Brea, V. M., Mucientes, M., and Del Bimbo, A. (2023). A full data augmentation pipeline for small object detection based on generative adversarial networks. *Pattern Recognition*, 133:108998.

Chavula, P., Kayusi, F., Lungu, G., and Uwimbabazi, A. (2025). The current landscape of early warning systems and traditional approaches to disaster detection. *LatIA*, (3):77.

Cunha, A. d. S. et al. (2023). A gestão de risco para redução de desastres naturais e a construção de cidades resilientes.

Economic News (2026). Desastres naturais no brasil geram r$ 28 bi em perdas. `https://economicnewsbrasil.com.br/2026/02/16/desastres-naturais-no-brasil-2025-prejuizos/`. Acessado em 21/02/2026.

EL HARIRI, A., MOUITI, M., HABIBI, O., and LAZAAR, M. (2023). Improving deep learning performance using sampling techniques for iot imbalanced data. *Procedia Computer Science*, 224:180–187.

Ghosh, B., Garg, S., and Motagh, M. (2022). Automatic flood detection from sentinel-1 data using deep learning architectures. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:201–208.

Hashi, A. O., Abdirahman, A. A., Elmi, M. A., Hashi, S. Z. M., and Rodriguez, O. E. R. (2021). A real-time flood detection system based on machine learning algorithms with emphasis on deep learning. *International Journal of Engineering Trends and Technology*, 69(5):249–256.

He, W., Jiang, Z., Xiao, T., Xu, Z., and Li, Y. (2026). A survey on uncertainty quantification methods for deep learning. *ACM Computing Surveys*, 58(7):1–35.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., and Adam, H. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Jain, J., Singh, A., Orlov, N., Huang, Z., Li, J., Walton, S., and Shi, H. (2023). Semask: Semantically masked transformers for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 752–761.

Kalogiannidis, S., Kalfas, D., Kontsas, S., Papaevangelou, O., and Chatzitheodoridis, F. (2025). Evaluating the effectiveness of early warning systems in reducing loss of life in natural disasters: a case study of greece. *Journal of Risk Analysis and Crisis Response*, 15(1):33–33.

Karimi, H. and Samavi, R. (2023). Quantifying deep learning model uncertainty in conformal prediction. In *Proceedings of the AAAI Symposium Series*, volume 1, pages 142–148.

Kebaili, A., Lapuyade-Lahorgue, J., and Ruan, S. (2023). Deep learning approaches for data augmentation in medical imaging: a review. *Journal of imaging*, 9(4):81.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollar, P., and Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4015–4026.

Li, H., Rajbahadur, G. K., Lin, D., Bezemer, C.-P., and Jiang, Z. M. (2024). Keeping deep learning models in check: A history-based approach to mitigate overfitting. *IEEE Access*, 12:70676–70689.

Li, Y., Matgen, P., and Chini, M. (2026). Quantifying and communicating uncertainty in sar-based flood mapping via density-aware neural networks and conformal risk control. *IEEE Transactions on Geoscience and Remote Sensing*.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollar, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Musa, A., Kakudi, H. A., Hassan, M., Hamada, M., Umar, U., and Salisu, M. L. (2025). Lightweight deep learning models for edge devices—a survey. *International Journal of Computer Information Systems and Industrial Management Applications*, 17:18–18.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Ranieri, C. M., Souza, T. L. D. e., Nishijima, M., Krishnamachari, B., and Ueyama, J. (2024). A deep learning workflow enhanced with optical flow fields for flood risk estimation. *Applied Intelligence*, 54(7):5536–5557.

Roohi, M., Ghafouri, H. R., and Ashrafi, S. M. (2025). Advancing flood disaster management: leveraging deep learning and remote sensing technologies. *Acta Geophysica*, 73(1):557–575.

Stateczny, A., Praveena, H. D., Krishnappa, R. H., Chythanya, K. R., and Babysarojam, B. B. (2023). Optimized deep learning model for flood detection using satellite images. *Remote Sensing*, 15(20):5037.

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., and Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, 596(7870):80–86.

Upadhyaya, A. N., Debnath, P., Rani, P., Selvakumar, P., Manjunath, T., and Bhattacharya, S. (2025). Ai-based disaster prediction and early warning systems. In *AI and emerging technologies for emergency response and smart cities*, pages 55–74. IGI Global Scientific Publishing.

Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., and Shen, F. (2022). Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610*.

Zhan, J. (2024). Mobilenet compression and edge computing strategy for low-latency monitoring. *Journal of Computer Science and Software Applications*, 4(4).