EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

# AquaMap-XAI: A Robustness-Aware and Explainable Transfer Learning Framework for Urban Water Issue Detection

Sridevi Bonthu[1*], Padma Jyothi Uppalapati[1] and G Sankara Rao[2]

[1]Department of Computer Science and Engineering, Vishnu Institute of Technology, Bhimavaram, India
[2]Department of Computer Science and Engineering, GVP College of Engineering for Women, Visakhapatnam, India

## Abstract

Urban water-related problems such as drainage blockage, road flooding, water leakage, potholes, and visible contamination often require quick identification before civic authorities can take appropriate action. In many cities, citizens already capture and report such issues through mobile devices, but the submitted images still need to be verified and categorized before they can support decision-making. This work presents AquaMap-XAI, an explainable and robustness-aware deep learning framework for classifying urban water issue images. The work extends the preliminary AquaMap study by shifting from a single baseline classifier to a more systematic evaluation of transfer learning models, reliability through cross-validation, robustness under degraded image conditions, and visual explanation using Grad-CAM. Four pretrained convolutional models, namely ResNet18, MobileNetV3, EfficientNet-B0, and DenseNet121, were fine-tuned on a six-class AquaMap dataset. The models were evaluated using accuracy, precision, recall, Macro-F1 score, training time, and inference time. MobileNetV3 achieved the best overall performance with an accuracy of 0.88 and a Macro-F1 score of 0.87. Stratified five-fold cross-validation supported the stability of the selected model, while robustness analysis showed that Gaussian blur produced the highest performance degradation. Grad-CAM visualizations were used to inspect whether the model attended to meaningful regions related to water issues. The results indicate that AquaMap-XAI can provide a lightweight, interpretable, and practically useful direction for AI-assisted urban water issue monitoring and smart governance support.

**Keywords:** Urban water issue detection; smart governance; transfer learning; MobileNetV3; robustness analysis; Grad-CAM; explainable AI; civic issue monitoring

Received on 24 June 2026, accepted on 09 September 2026, published on 17 September 2026



doi: 10.4108/eetiot.13710

*Corresponding author. Email: sridevi.b@vishnu.edu.in

## 1. Introduction

Some of the most visible and recurrent civic problems in growing cities are water-related issues. Blocked drainage, road flooding, water leakage, potholes filled with water and visible contamination can affect public mobility, health, safety and the quality of urban life [1]. These issues may look small, but if not addressed, they can lead to traffic disruptions, infrastructure deterioration, inconvenience to the public, and service delays. Global development priorities, particularly SDG 6, which emphasizes sustainable access to water and sanitation services [2], also reflect the need for better water and sanitation management. Recent global urban reports have also warned that cities are more exposed to climate-related risks such as flooding, heat, storms and stressed infrastructure [3], [4]. In many practical situations, citizens are the first to observe local civic issues. A resident may notice a leaking pipeline, an overflowing drain, or a flooded road much earlier than a municipal inspection team. Smartphones have made it possible for citizens to capture images and share evidence quickly. Platforms like civic reporting portals and mobile complaint applications have already shown how people-generated information can help authorities identify local issues [5], [6]. But the value of such systems depends not only on collecting complaints but also on processing them in a quick and reliable manner. When the number of complaints increases, manual verification and categorization may become slow and inconsistent.

Image-based artificial intelligence can support this process by providing an initial layer of interpretation. A trained model can classify the type of issue present in an uploaded image and help move it to the appropriate authority [7]. For urban water-related problems, this means distinguishing between categories such as drainage problems, floods, potholes, water contamination, water leakage, and non-water scenes. Such classification is not always straightforward because several categories share overlapping visual patterns. For example, a pothole may contain water, a drainage problem may appear visually similar to contamination, and low-quality citizen images may contain blur, shadows, or background clutter.

Convolutional Neural Networks (CNN) have been widely used for image classification and visual recognition tasks. Transfer learning is especially useful in domain-specific applications where large annotated datasets are not always available [8], [9]. Instead of training a model from the scratch, pretrained models can be adapted to a smaller custom dataset. This approach reduces training effort and often improves performance in practical computer vision tasks. Architectures such as ResNet, MobileNet, EfficientNet, and DenseNet have been widely used because they offer different balances between accuracy, computational cost, and representation capability [10–13].
The preliminary AquaMap work introduced an application-oriented approach for reporting and mapping water-related civic issues using a custom image dataset and a ResNet18 classifier [14]. That study established the feasibility of community-level reporting and image-based classification. A practical AI system should not be evaluated only by a single accuracy value [15]. It should also be checked for model stability, performance across classes, computational efficiency, robustness to real-world image degradation, and interpretability of predictions [16].

This paper presents AquaMap-XAI, an extended framework that addresses these aspects. The proposed work compares four pretrained transfer learning models, namely ResNet18, MobileNetV3, EfficientNet-B0, and DenseNet121, under a common experimental setting. The best model is selected using Macro-F1 score, which is more suitable for multiclass evaluation than accuracy alone. To reduce dependence on a single train-test split, stratified five-fold cross-validation is performed. To understand behaviour under realistic citizen-image quality conditions, robustness analysis is carried out using Gaussian noise, Gaussian blur, and low-brightness transformations. Finally, Grad-CAM is used to generate visual explanations that show which regions of the image influenced the model's decision.
The main contributions of this work are summarized as follows.

1. The preliminary AquaMap work is extended into AquaMap-XAI, a robustness-aware and explainable transfer learning framework for urban water issue classification.
2. Four pretrained convolutional architectures are systematically compared using accuracy, precision, recall, Macro-F1 score, training time, and inference time.
3. MobileNetV3 is identified as the best model for the dataset, achieving 0.88 accuracy and 0.87 Macro-F1 score.
4. Stratified five-fold cross-validation is used to study the stability of the model across different data partitions.
5. Robustness analysis is performed under degraded image conditions to study the practical limitations of the model.
6. Grad-CAM visual explanations are generated to improve transparency and support trust in AI-assisted civic issue monitoring.

The rest of the paper is organized as follows. Section 2 reviews related work on citizen reporting, smart city monitoring, deep learning for infrastructure-related image analysis, robustness evaluation, and explainable AI. Section 3 presents the proposed methodology. Section 4 describes the experimental setup. Section 5 discusses the results and section 6 concludes the paper and outlines future directions.

## 2. Related Work

Smart city systems increasingly depend on timely information from distributed sources. While fixed sensors, surveillance systems, and municipal inspection teams are useful, they cannot always capture every local issue as soon as it appears [17]. Citizen reporting platforms help bridge this gap by allowing citizens to submit information about problems in their surroundings [18]. FixMyStreet, for example, is a well-known civic technology platform that allows people to report local issues such as potholes and broken streetlights to responsible councils [5]. Similarly, 311 service request systems have been widely studied as sources of information about urban service needs and maintenance patterns [6].

Mobile crowdsensing has also become an important concept in smart city research. In such systems, citizens use their devices as sensing units to capture information through cameras, GPS, and other mobile sensors [19–21]. This model is attractive because it can collect observations from many locations at relatively low cost. However, citizen-generated data also brings challenges. The submitted content may be incomplete, duplicated, ambiguous, poorly captured, or inconsistent in quality. As a consequence, automated support is often required to organize, classify, and prioritize incoming data.

Recent studies on smart city complaint systems and civic monitoring have explored the use of machine learning and image processing to classify reported issues [22]. However, many of these systems remain prototype-oriented and often focus mainly on reporting workflows. The present work differs by focusing on the reliability of the image classification component itself. In AquaMap-XAI, the emphasis is not only on recognizing a reported issue but also on understanding whether the model is stable, robust, and explainable enough to support real-world civic use.

## 2.1 Computer Vision for Urban Infrastructure and Water-Related Issues

Computer vision has been increasingly applied to urban infrastructure monitoring. Road damage detection, pothole recognition, and flood monitoring are among the most relevant application areas [23]. Vision-based pothole detection methods have been reviewed extensively, with recent studies showing that deep learning and computer vision methods can reduce dependence on manual road inspections [24–26]. Large road damage datasets such as RDD2022 have also supported the development and benchmarking of automated infrastructure monitoring models [27].

Flood and waterlogging detection is another related area. Deep learning has been used with satellite imagery, traffic images, and surveillance data to detect flood extent, flood depth, or water accumulation [28–30]. These studies show the growing relevance of AI-based visual analysis in urban hydrology and disaster response. However, most flood-related studies focus on satellite imagery, traffic camera feeds, or surveillance systems. Citizen-submitted ground-level images remain a more variable and less controlled source of information. AquaMap-XAI lies between these areas. It does not focus only on potholes or only on floods; instead, it addresses multiple urban water-related civic categories in a unified classification setting. This makes the task more challenging because the classes may visually overlap. For instance, water leakage, drainage problems, and water contamination may all contain water-like regions, while potholes and flooded roads may share road-surface cues. Therefore, a systematic comparison of transfer learning models is needed to understand which architecture provides the best balance for this application.

## 2.2 Transfer Learning for Practical Image Classification

Transfer learning has become a common strategy in practical computer vision because it allows models trained on large datasets to be adapted to smaller domain-specific datasets [31]. This is useful in civic and environmental applications, where collecting thousands of carefully labeled images for every local problem is often difficult. Recent surveys on transfer learning in computer vision have highlighted its value in reducing training time, improving performance, and enabling domain adaptation when labeled data are limited [8], [9]. The architectures used in this study represent different designs. ResNet introduced residual connections to improve the training of deeper networks [10]. MobileNetV3 was designed for mobile and resource-constrained settings, making it relevant for future deployment in lightweight applications [11]. EfficientNet proposed compound scaling to balance network depth, width, and resolution [12]. DenseNet introduced dense connections between layers to encourage feature reuse and better gradient flow [13]. Because these models differ in complexity and computational behaviour, comparing them under the same dataset and training conditions provides useful insight into practical model selection. In the base AquaMap work, ResNet18 was used as the primary classifier [14]. The present study expands this by comparing ResNet18 with MobileNetV3, EfficientNet-B0, and DenseNet121. This comparison allows the extended paper to move beyond baseline feasibility and toward a more informed evaluation of model performance and deployment suitability.

## 2.3 Robustness under Real-World Image Degradation

A common limitation of image classification systems is that good performance on clean test data does not always guarantee reliable behaviour in real-world conditions [32]. Images captured by citizens may be blurred, noisy, underexposed, overexposed, compressed, or partially occluded. Robustness evaluation is therefore important when a model is intended for practical use [33].

Hendrycks and Dietterich introduced benchmark ideas for evaluating neural network robustness under common image corruptions and perturbations, showing that corruption-based testing gives important information beyond standard accuracy [16]. Later studies have continued to emphasize that robustness should be considered in applied computer vision systems, especially when input quality cannot be controlled [34]. In civic issue monitoring, this concern is highly relevant because users may capture images quickly, from moving vehicles, during rain, or under poor lighting. In AquaMap-XAI, robustness is evaluated using Gaussian noise, Gaussian blur, and low-brightness transformations. These degradations are simple but meaningful because they reflect common problems in citizen-submitted images. The robustness analysis helps identify which degradation affects the model most severely and provides direction for future improvements such as blur-aware augmentation, image quality assessment, or preprocessing-based correction.

## 2.4 Explainable AI for Visual Decision Support

Deep learning models are often accurate but difficult to interpret. In applications that support public services, this lack of transparency can reduce trust [35]. Municipal users may hesitate to rely on an automated prediction if the system cannot provide any explanation for its decision. Explainable AI methods attempt to address this issue by making model predictions more understandable. Grad-CAM is a widely used method for visualizing the regions of an image that contribute to a CNN prediction [36]. It has been used in many computer vision tasks to inspect whether a model focuses on relevant image regions or irrelevant background patterns. Recent surveys on explainable AI in computer vision also emphasize the importance of visual explanation methods for improving interpretability, debugging model behaviour, and supporting user trust [37], [38].

In AquaMap-XAI, Grad-CAM is used to visualize the regions that influence predictions for urban water issue images. This is useful because the model should ideally focus on visible civic issue regions such as blocked drains, accumulated water, leakage areas, potholes, or contamination patterns. If the model attends to irrelevant backgrounds, that would indicate a need for dataset refinement or improved training. Thus, explainability is not treated as an optional addition but as part of the model evaluation process. Existing work provides useful foundations in smart city reporting, road damage detection, flood monitoring, transfer learning, robustness testing, and explainable AI. However, there is still a gap in developing a unified, practical, and interpretable framework for classifying multiple urban water-related civic issues from citizen-submitted images. Many civic reporting systems focus on complaint collection but not on robust image interpretation. Many computer vision studies focus on single categories such as potholes or floods. Similarly, the preliminary AquaMap work demonstrated feasibility but did not examine multiple pretrained models, cross-validation reliability, degradation robustness, or visual explainability in depth.

The present study addresses this gap through AquaMap-XAI. It combines comparative transfer learning, reliability assessment, robustness analysis, and Grad-CAM explanation within a single experimental framework. This makes the work more suitable for smart governance applications where accuracy, efficiency, reliability, and transparency must be considered together.

# 3. Methodology

This section presents the framework of AquaMap-XAI, which is an extension of AquaMap framework [14]. It incorporates a systematic model comparison, robustness analysis, reliability assessment based on cross-validation, and visual explainability through Grad-CAM. The framework follows a pipeline with multiple stages consisting of dataset preparation, classification based on transfer learning, robustness evaluation and explainability analysis. Subsequently, multiple pretrained CNN architectures are employed using transfer learning to perform multiclass classification of urban water issues. To ensure fair comparison, all models are trained under identical experimental settings and evaluated using multiple performance metrics, including accuracy, precision, recall, and macro-F1 score. In addition, stratified five-fold cross-validation is performed to assess model stability and reduce dependence on a single train-test split. The overall workflow of the proposed AquaMap-XAI framework is illustrated in Fig. 1. The framework aims to provide a practical and scalable AI-assisted solution for smart urban water issue monitoring and governance applications.

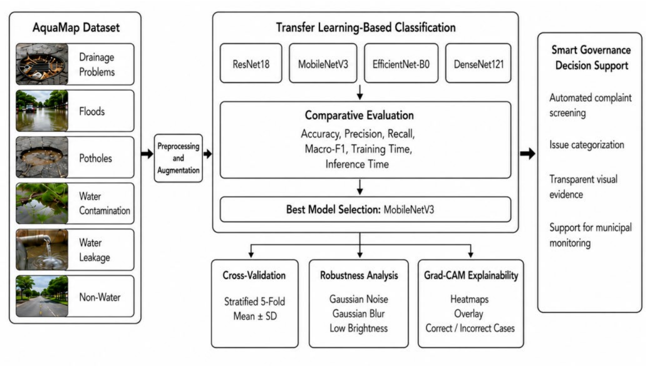



**Figure 1.** Architecture of proposed AquaMap-XAI framework

## 3.1 Dataset Preparation

This work used AquaMap dataset which consists of urban water-related civic issue images collected to support image-based monitoring and classification of municipal infrastructure problems [14]. The dataset was originally collected as part of the preliminary AquaMap framework. To ensure robust represenation, 100 images were collected for each class, culminating in an extensive dataset of 600 images.It has been further refined and reorganized in the present study to support robustness-aware and explainable deep learning experiments. The images in the dataset are categorized into six distinct classes corresponding to commonly observed urban water-related issues. These classes are drainage problems, floods, potholes, water contamination, water leakage, and non-water scenes. Figure 2 presents representative class samples. The images exhibit variability and diversity for improving the generalization capability of the trained models, but this feature also exhibits challenges in modeling. Prior to model training, the dataset was examined to remove unreadable and corrupted image samples. Dataset consistency was improved by inspecting for duplicate images and poor quality samples. Subsequently, all valid images were organized into class-specific directories to facilitate preprocessing and transfer learning experiments. Substantial variations exist in the images present in the dataset as they were collected from diverse environments. Therefore, a preprocessing approach was adopted to standardize the input data before model training and evaluation. Initially, all images were converted into RGB format and resized to a uniform spatial resolution of 224 × 224 pixels. Next, pixel normalization was performed using the standard ImageNet mean and standard deviation values as the selected models were pretrained on ImageNet data.

**Figure 2.** Representative samples from AquaMap dataset. The classes include floods, potholes, water leakage, water contamination,drainage problems, and non-water scenes

Data Augmentation techniques were employed during training to improve the generalization capability of the models and to reduce the risk of overfitting. The augmentation techniques employed in this study are random horizontal flipping, random rotation, affine transformations, scaling, translation, and brightness variation. In addition, brightness and contrast variations were introduced using color jitter operations to simulate changes in illumination conditions caused by weather, shadows, or device-specific camera characteristics. The validation and testing datasets were maintained without augmentation to ensure unbiased performance evaluation.

## 3.2 Classification Models

When compared with benchmark datasets, the real-world domain specific datasets are relatively small. In these cases, transfer learning can be leveraged because it uses pretrained feature representations learned from large datasets. This approach is well-proven in many scenarios. This study examines multiple pretrained models such as ResNet18, MobileNetV3, EfficientNet-B0 and DenseNet121 to investigate their suitability for urban water issue classification. All the selected models were initialized using pretrained ImageNet weights, and the final classification layers were modified to perform multiclass classification corresponding to the six AquaMap categories of our dataset. Identical preprocessing, augmentation, training, evaluation settings, and hyperparameters were used across all architectures to ensure fair comparison.

### 3.2.1 ResNet18

ResNet18 was used as baseline model in our study due to its effectiveness in image classification tasks. Through skip connections, ResNet architecture addresses the degradation problem of deep neural networks. The residual learning mechanism can be expressed as $\mathrm{y} = \mathrm{F}(\mathrm{x}, \mathrm{W_i}) + \mathrm{x}$, where x represents the input feature map, and the function F denotes the residual mapping learned by convolution layers. This model serves as a strong baseline for evaluating the effectiveness of the other pretrained models employed in this study.

### 3.2.2 MobileNetV3

MobileNetV3 model has a lightweight architecture and suitable for resource constrained environments. This architecture is suitable for urban monitoring systems, as they require deployment on edge devices, or mobile platforms. This model leverages depthwise separable convolutions and maintains good classification performance. The architecture of this model uses squeeze-and-excitation mechanisms and optimized activation functions. The lightweight nature of MobileNetV3 makes it particularly suitable for real-time urban monitoring scenarios involving mobile image acquisition and rapid inference requirements.

### 3.2.3 EfficientNet-B0

EfficientNet-B0 has balanced scaling strategy and strong feature representation capability. Unlike conventional CNN architectures that independently scale network depth or width, EfficientNet employs compound scaling to simultaneously balance network depth, width, and input resolution in a systematic manner. This strategy enables improved performance while maintaining computational efficiency. The compound scaling principle adopted in EfficientNet can be represented as $depth = \alpha^{\phi}, width = \beta^{\phi}, resolution = \gamma^{\phi}$, where α, β and γ denote scaling coefficients corresponding to network depth, width, and input resolution, respectively, while $\phi$ controls the overall scaling factor.

### 3.2.4 DenseNet121

DenseNet121 model uses dense connectivity mechanism and efficient feature reuse capability. In DenseNet architectures, each layer receives feature information from all preceding layers through direct connections. This dense feature propagation strategy improves gradient flow, encourages feature reuse, and reduces redundancy in learned representations. The dense connectivity mechanism can be mathematically expressed as $\mathrm{x_l} = \mathrm{H_l}([\mathrm{x_0}, \mathrm{x_1}, \dots, \mathrm{x_{l-1}}])$ where $x_l$ represents the output of the

$l^{th}$ layer. The characteristics of this model are beneficial for handling complex urban image scenes involving varying textures, structural patterns, and environmental conditions.

## 3.3 Model Training and Evaluation

All the adopted models employed in the AquaMap-XAI framework were trained using a unified training strategy to ensure fair and consistent performance comparison. All experiments were conducted using the PyTorch deep learning framework with GPU acceleration support to improve computational efficiency and reduce training time. The adopted training pipeline was designed to provide stable optimization behaviour while maintaining reproducibility across all evaluated architectures. During training, the pretrained CNN models were initialized using ImageNet weights, and the final classification head was modified to suit classification corresponding to the six AquaMap categories. Fine-tuning was then performed using the processed AquaMap dataset to adapt the pretrained feature representations to the urban water issue detection task. The categorical cross-entropy loss function was employed to optimize the multiclass classification objective. The cross-entropy loss is defined as:

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

where C denotes the total number of classes (in this case 6), $y_i$ represents the actual class, and $\hat{y}_i$ denotes the predicted probability corresponding to class $i$. Adam optimization algorithm was employed due to its stable convergence characteristics. The initial learning rate is $1 \times 10^{-4}$ and it is used along with weight decay regularization. All the models are trained for a fixed number of epochs (20) with a batch size of 32. While training, the models checkpoints to the best validation Macro-F1 score to ensure best model selection for further evaluation. The training pipeline was designed to support robust performance evaluation, cross-validation analysis, robustness testing, and explainable AI interpretation within the AquaMap-XAI framework.

To evaluate the performance of the models, multiple evaluation metrics like accuracy, precision, recall and Macro-F1 were used as the study involves multiclass classification under real environmental conditions. Accuracy indicates the percentage of correctly classified samples among the total number of evaluation samples. To adequately capture the class-wise behavior precision and recall were employed. F1-score is employed for balanced evaluation. All the four metrics are mathematically expressed as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The class wise prediction behavior was performed by confusion matrix analysis. Computational efficiency was evaluated by using training and inference times.

## 3.4 Reliability and Robustness Analysis

Stratified 5-fold cross-validation was done on the best model to ensure reliable and unbiased evaluation. It enables a systematic investigation of the model's consistency across the data partitions and at the same times also preserves the class distributions. It also enables reliable estimation of model generalization capability with different distributions of the data. The same preprocessing, augmentation, optimization and hyperparameters were retained during cross-validation experiments also. For every fold, all the evaluation metrics are recomputed and from them the mean and standard deviation of the obtained metrics was calculated to summarize model stability and performance variability. This strategy improves the rigor of the proposed framework.

The real-world images contain noise, blur, poor lighting and motion artifacts which affect the performance of the models. To understand the practical deployment capability of the AquaMap-XAI, it is necessary to evaluate the model under degraded image conditions. Gaussian noise, Gaussian blur, and low-brightness conditions are used to examine model reliability. These are introduced during evaluation to simulate the image acquisition issues in reality. The robustness evaluation was performed by using the best model identified during the comparative analysis stage. The trained model was evaluated independently under each degradation condition while preserving identical testing configurations. All the performance metrics for each condition were computed to quantify performance degradation relative to the clean-image baseline.

## 3.5 Explainable AI

Interpretability and transparency play crucial role in applications like urban civic issue monitoring as end users may require justification for automated predictions. Therefore, Explainable AI is incorporated into the framework to improve interpretability and thereby trustworthiness of the classification. Gradient-weighted Class Activation Mapping (Grad-CAM) was used to generate visual explanations corresponding to the predictions produced by the best model. Grad-CAM identifies discriminative image regions contributing most significantly to a particular prediction. Grad-CAM enables better understanding of the internal behaviour of CNN models by highlighting relevant areas in the image. It computes importance weight with the help of the gradients of the target class. In this work, Grad-CAM visualizations were generated for a representative set of test images. The generated heatmaps were then overlaid on the original images to visually highlight the regions that influenced the model predictions to know whether the model is focusing on meaningful water related structures or not.

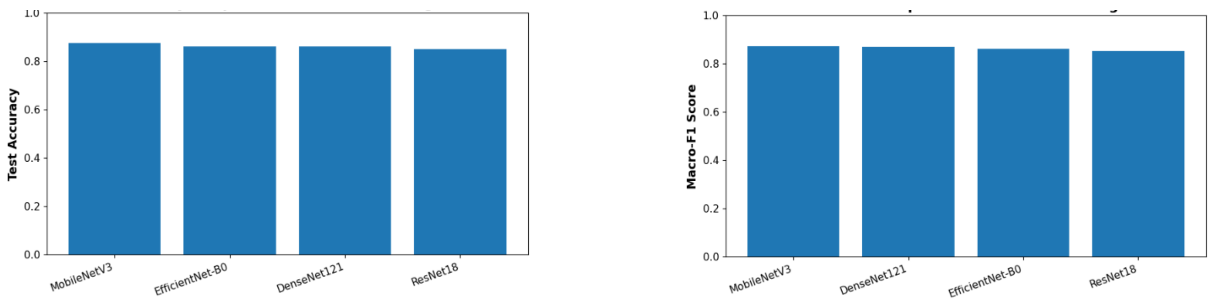
## 4. Results and Discussion

This section presents the experimental results of the AquaMap-XAI framework. The major aspects used for the analysis are comparison of model performance, computational efficiency, cross-validation reliability, robustness under image degradation, and explainability.

Table 1. Comparative performance of transfer learning models on the AquaMap test set

| Model | Accuracy | Precision | Recall | Macro-F1 |
|---|---|---|---|---|
| ResNet18 | 0.85 | 0.85 | 0.86 | 0.85 |
| MobileNetV3 | 0.88 | 0.87 | 0.89 | 0.87 |
| EfficientNet-B0 | 0.86 | 0.85 | 0.87 | 0.86 |
| DenseNet121 | 0.86 | 0.88 | 0.87 | 0.87 |

From the performance of the evaluated models shown at table 1, the best performance with an accuracy of 0.88 and a Macro-F1 score of 0.87 is produced by MobileNetV3. DenseNet121 also produced a competitive result, particularly in terms of precision, but MobileNetV3 provided the best balance across accuracy, recall, and Macro-F1 score. Figure 3 visually presents the comparison. MobileNetV3 was able to learn discriminative features for the six AquaMap classes while maintaining balanced class-wise performance. The comparative analysis is important for this work, as the task involves visually diverse urban images captured under uncontrolled conditions. The strong performance of MobileNetV3 suggests that lightweight architectures can be effective for practical smart governance applications.




(a) (b)

**Figure 3**. Comparison of (a) test accuracy and (b) F1-Score of the studied models

To understand the practical deployment suitability of the models, computational efficiency was also analyzed using training and average inference time per image. Table 2 presents the computational efficiency of the four models under study. RestNet18 and EfficientNet-B0 has shown faster inference time when compared with other two. MobileNetV3 and ResNet18 have taken least training times. DenseNet121 requires higher training time due to its dense connectivity structure.

Table 2. Computational efficiency comparison of transfer learning models

| Model | Inference Time / Image (s) | Training Time (s) |
|---|---|---|
| ResNet18 | 0.0104 | 164.8562 |
| MobileNetV3 | 0.0167 | 163.6342 |
| EfficientNet-B0 | 0.0108 | 170.1593 |
| DenseNet121 | 0.0129 | 188.7033 |

From the comparative results of performance, and computational efficiency, MobileNetV3 provides a suitable balance. This balancing nature is appropriate for future deployment in lightweight or mobile-assisted civic issue reporting systems. F1-score is an appropriate metric for unbalanced datasets. Based on the F1-score also MobileNetV3 can be selected as the best performing model among the four.

The confusion matrix of the MobileNetV3 model is presented in Fig. 4. This analysis provides class-wise insight into the prediction behavior of the model. This matrix enables us to identify the categories which were classified reliably and which produced misclassifications. In urban water issue detection, confusion may occur between visually related categories such as drainage problems, water leakage, and water contamination, especially when images contain overlapping visual patterns such as stagnant water, damaged surfaces, or unclear boundaries. Similarly, potholes and flood-related images may show visual similarity when water accumulation is present on road surfaces.

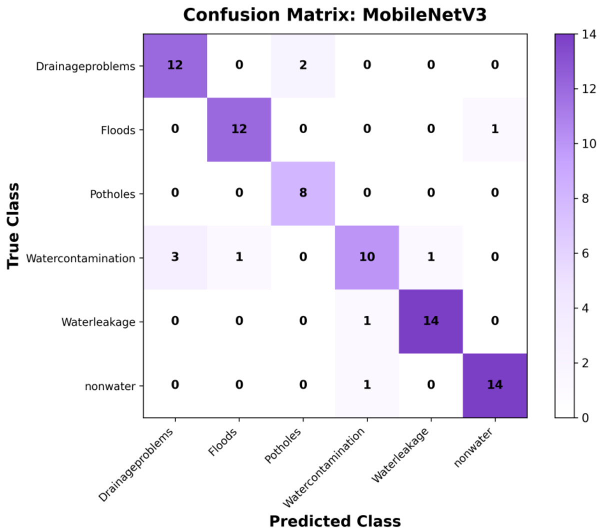



**Figure 4**. Confusion Matrix of the MobileNetV3 Model

To verify the reliability of the selected model, stratified five-fold cross-validation was performed using MobileNetV3. The cross-validation results are summarized in Table 3. The cross-validation results show that MobileNetV3 achieved a mean accuracy of 0.87 and a mean Macro-F1 score of 0.88 across 5-folds. The relatively small standard deviation values indicate stable performance across different data partitions. These results confirm that the model's performance is not dependent on a single train-test split. This analysis is useful because it moves beyond aggregate accuracy and provides a more detailed understanding of model behaviour. Such class-level interpretation is important for practical deployment, where different water-related issues may require different municipal responses.

Table 3. Stratified 5-fold cross-validation summary of MobileNetV3 model

| | Mean | Standard Deviation |
|---|---|---|
| Accuracy | 0.87 | 0.03 |
| Precision | 0.88 | 0.02 |
| Recall | 0.88 | 0.04 |
| Macro-F1 | 0.88 | 0.03 |

The images generated by citizens often contain light issues, blur, noise etc. Robustness analysis can evaluate the model under these degraded image conditions. The performance drop in this analysis is shown using heatmap in figure 5. A reasonable performance of 0.84 accuracy, which is only 0.03 drop. A substantial performance drop is observed using Gaussian blur. It indicates that blur introduces a drop in performance. An interesting observation through this analysis is low-brightness condition did not reduce performance, instead in this setting, it raised the by 0.02. This is because the image features are discriminative even after the reduction of brightness. The performance degradation under blur suggests that future versions of the framework will benefit from blur-specific augmentation, image deblurring preprocessing, or quality-aware input filtering.

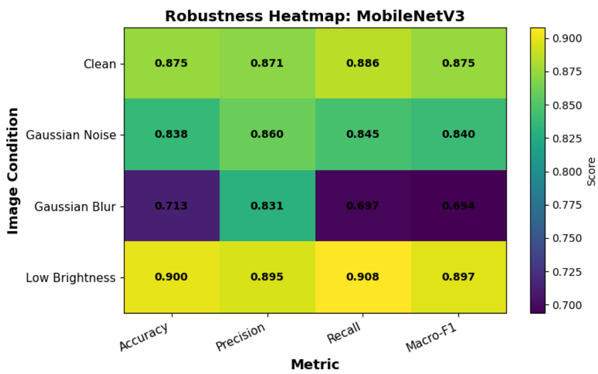



**Figure 5.** Robustness Analysis under degraded image conditions

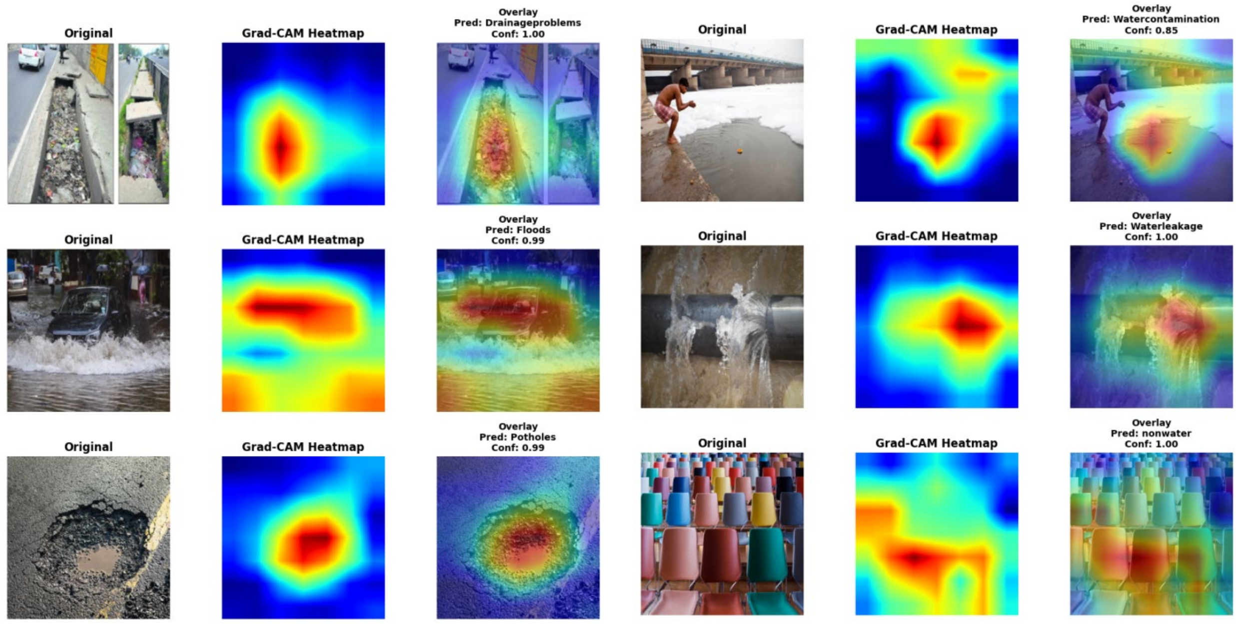



**Figure 6.** Grad-CAM visual explanations for MobileNetV3

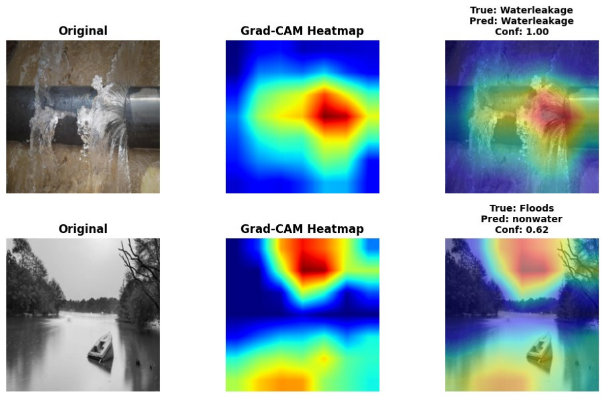



**Figure 7.** Grad-CAM analysis of correct and incorrect predictions

To verify whether the selected MobileNetV3 model focused on meaningful regions related to urban water issues, Grad-CAM visualizations were generated and interpreted. Figure 6 presents Grad-CAM visual explanations for representative samples from AquaMap classes. These outputs provide visual evidence of the regions that influence the model predictions. In the correctly classified samples, the highlighted regions are expected to correspond to relevant visual cues such as accumulated water, road damage, drainage blockage, leakage areas, or contamination patterns. Fig 7 provides analysis of correct and incorrect predictions and their visual cues. In misclassified examples, Grad-CAM can help reveal whether the model attended to irrelevant background regions or ambiguous visual patterns. This is useful for identifying dataset limitations and improving future model training. These visual explanations improve the interpretability of the proposed method and make the predictions more transparent for governance-oriented applications, thereby improving trust in AI-assisted civic issue monitoring.

The above results clearly indicate the AquaMap-XAI incorporating MobileNetV3 is suitable for smart governance oriented urban water issue monitoring, as it have better classification performance and lightweight. The cross validation results confirmed the reliability of the model across data partitions. The practical value of the work is further strengthened with explainability analysis. When compared with base Aquamap version, the proposed AquamapXAI provides a stronger experimental foundation.

## Conclusion

The extended AquaMap-XAI study examined how far the original AquaMap idea can be strengthened through a more careful model evaluation. Instead of relying on a single classifier, four pretrained models were compared under the same experimental setting. Among them, MobileNetV3 gave the most balanced result, achieving an accuracy of 0.88 and a Macro-F1 score of 0.87. Its performance, along with its lightweight nature, makes it a practical choice for image-based civic issue reporting systems. The additional experiments also provided useful insights. Cross-validation showed that the selected model was reasonably stable across different data partitions. The robustness study revealed that blurred images affect performance more seriously than noise or reduced brightness, which is an important point for citizen-captured images. Grad-CAM visualizations helped verify that the model was generally attending to relevant water-related regions rather than only background patterns. Overall, the study shows that AquaMap can be extended into a more reliable and interpretable framework for urban water issue classification. Future work should focus on adding more field-level images, improving the handling of blurred or poor-quality inputs, and testing the framework with real municipal complaint workflows.

## References


[1] Rammile S, Bhanye JI, Lehobo T, Hugo M. Building resilient urban infrastructure and innovation for urban public health: A case study of Bloemfontein, South Africa. In Sustainable development goals and urban health: Strides, challenges and way forward for poor neighborhoods 2024 Sep 29 (pp. 69-96). Cham: Springer Nature Switzerland.

[2] Sharma R, Katoch OR. Analysis of the targets and progress toward meeting the 2030 sustainable development goal-6 on clean water and sanitation: evidence from India. South Asian Journal of Social Studies and Economics. 2022 Sep 13;15(3):16-26.

[3] UN-HABITAT. World cities report 2024: Cities and climate action. Stylus Publishing, LLC; 2024 Nov 5.

[4] Parmesan C, Morecroft MD, Trisurat Y, Mezzi D. Climate change 2022: Impacts, adaptation and vulnerability (Doctoral dissertation, GIEC; IPCC).

[5] Wittemyer R, Bailur S, Anand N, Park KR, Gigler BS. New routes to governance: a review of cases in participation, transparency, and accountability. Closing the Feedback Loop. 2014:43.

[6] Wang L, Qian C, Kats P, Kontokosta C, Sobolevsky S. Structure of 311 service requests as a signature of urban location. PloS one. 2017 Oct 17;12(10):e0186314.

[7] Obuchowicz R, Strzelecki M, Piórkowski A. Clinical applications of artificial intelligence in medical imaging and image processing—A review. Cancers. 2024 May 14;16(10):1870.

[8] Yu F, Xiu X, Li Y. A survey on deep transfer learning and beyond. Mathematics. 2022 Oct 3;10(19):3619.

[9] Panda A, Panigrahi D, Mitra S, Mittal S, Rahimi S. Transfer learning applied to computer vision problems: survey on current Progress, limitations, and opportunities. arXiv preprint arXiv:2409.07736. 2024 Sep 12.

[10] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[11] Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV. Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision 2019 (pp. 1314-1324).

[12] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. InInternational conference on machine learning 2019 May 24 (pp. 6105-6114). PMLR.

[13] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4700-4708).

[14] Nalla HL, Bhuchupalli A, Addala T, Sabbineni Y, Geetha KS, Aasha G, Bonthu S. AquaMap: Empowering Communities to Report and Map Water-Related Issues in Real-Time with Deep Learning. InInternational Conference on Cognitive Computing and Cyber Physical Systems 2024 Apr 4 (pp. 296-305). Cham: Springer Nature Switzerland.

[15] Van Calster B, Collins GS, Vickers AJ, Wynants L, Kerr KF, Barreñada L, Varoquaux G, Singh K, Moons KG, Hernandez-Boussard T, Timmerman D. Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance. The Lancet Digital Health. 2025 Dec 13.

[16] Liu C, Dong Y, Xiang W, Yang X, Su H, Zhu J, Chen Y, He Y, Xue H, Zheng S. A comprehensive study on robustness of image classification models: Benchmarking and rethinking. International Journal of Computer Vision. 2025 Feb;133(2):567-89.

[17] Almulhim AI. Building urban resilience through smart city planning: a systematic literature review. Smart Cities. 2025 Jan 31;8(1):22.

[18] Asimakopoulos G, Antonopoulou H, Giotopoulos K, Halkiopoulos C. Impact of information and communication technologies on democratic processes and citizen participation. Societies. 2025 Feb 18;15(2):40.

[19] Chen LW, Liao CW, Liu JX. Proactive crowdsourced monitoring and sensing with expansible activity recognition based on Internet of Things Localization. IEEE Internet of Things Journal. 2025 Feb 5;12(11):17674-86.

[20] Yorucu V, Bekun FV, Yitmen I. The Anatomy of Smart Cities: A Global Perspective. In Modern Approach of Resilient and Sustainable Smart Cities 2026 Jan 24 (pp. 7-19). Cham: Springer Nature Switzerland.

[21] Hou Y, Chen X, Dong Q, Zhang F, Sun Y, Wu L, Liu Y. Enhancing drive-by sensing power in urban hotspots through multi-objective vehicle selection optimization. International Journal of Geographical Information Science. 2026 Jan 2;40(1):49-77.

[22] Sarika S, Hanan A, Afna VN, Johnson J, Nair KV. Agentic AI-Driven Grievance Redressal: Automating Categorization and Routing in Smart Governance. In2026 International Conference on Innovative Trends in Information Technology (ICITIIT) 2026 Mar 27 (pp. 1-6). IEEE.

[23] Darmawan MA, Mukti SN, Tahar KN. Pothole Detection Based on UAV Photogrammetry. Revue Internationale de Geomatique. 2025 Jan 1;34(1).

[24] Bhatt AK, Raj H, Sharma VB, Biswas S, Singh A, Silori R, Pandey M. Advancements in pothole detection techniques: a comprehensive review and comparative analysis. Discover Artificial Intelligence. 2025 Oct 3;5(1):255.

[25] Alam MI. Dual-Task Real-Time Low-Light Lane and Pothole Detection for Resource-Constrained Environments (Doctoral dissertation, IUB).

[26] Safyari Y. Pothole detection using machine learning and computer vision techniques (Doctoral dissertation, Memorial University of Newfoundland).

[27] Jayhind Pardeshi P, Shailja, Chaudhary A, Arya, Yadav M. Patch-Based Self-Supervised Learning for Road Damage Classification: A Case Study on the RDD2022 dataset of Indian test site. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2025 Dec 19; 10:459-66.

[28] Awasthi YY, Chinzvende J. Deep learning-based flood forecasting using satellite imagery and IoT sensor fusion.

[29] Shi G, Chen Z, Lu R. Improved lightweight water body recognition model using Mobile-Former network. Hydrology Research. 2025 Sep 1;56(9):937-50.

[30] Zhao Z, Lu B, Zhang S, Wang D, Wan J, Liu R, Li H, Wang Z. Evaluating the performance of automatic detection for urban flood levels using different deep learning approaches. Water Resources Management. 2026 May;40(7):280.

[31] Gholizade M, Soltanizadeh H, Rahmanimanesh M, Sana SS. A review of recent advances and strategies in transfer learning. International Journal of System Assurance Engineering and Management. 2025 Mar;16(3):1123-62.

[32] Liu M, Wei J, Liu Y, Davis J. Human and ai perceptual differences in image classification errors. In Proceedings of the AAAI Conference on Artificial Intelligence 2025 Apr 11 (Vol. 39, No. 13, pp. 14318-14326).

[33] Khan D, Khan N, Ullah S. Harnessing hybrid intelligence and explainable AI for urban growth prediction: A Data-Driven framework for sustainable cities. Environment, Development and Sustainability. 2025 Sep 20:1-40.

[34] Zheng S, Cui X, Sun Y, Li J, Li H, Zhang Y, Chen P, Jing X, Ye Z, Yang L. Benchmarking pathclip for pathology image analysis. Journal of Imaging Informatics in Medicine. 2025 Feb;38(1):422-38.

[35] Xu B, Yang G. Interpretability research of deep learning: A literature survey. Information Fusion. 2025 Mar 1; 115:102721.

[36] Mothkur R, Soubhagyalakshmi P. Grad-CAM Based Visualization for Interpretable Lung Cancer Categorization Using Deep CNN Models. Journal of Electronics, Electromedical Engineering, and Medical Informatics. 2025 May 3;7(3):567-80.

[37] Andrae S. Explainable Artificial Intelligence. In Encyclopedia of Modern Artificial Intelligence 2026 (pp. 1-21). IGI Global Scientific Publishing.

[38] Mathew DE, Ebem DU, Ikegwu AC, Ukeoma PE, Dibiaezue NF. Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human. Neural Processing Letters. 2025 Feb 7;57(1):16.