Advanced Ensemble Deep Learning Framework for Enhanced River Water Level Detection: Integrating Transfer Learning

Nisreen Tawfeeq [1*], Jameelah Harbi [2]
[1, 2] Mustansiriyah University College of Science, Computer Department, Baghdad, Iraq
Email: [1] nsrntawfk@gmail.com, [2] dr.jameelahharbi@uomustansiriyah.edu.iq
*Corresponding Author

***Abstract*—The precise monitoring and prediction of river water levels are crucial for effective environmental management, flood prevention, and ensuring water security. This paper introduces an advanced deep learning framework that utilizes an ensemble of state-of-the-art neural networks, namely InceptionV3, VGG16, Xception, MobileNet, and ResNet152, to enhance the accuracy of water level detection from river imagery. The proposed system integrates these models through a robust ensemble methodology that leverages hard voting to improve predictive performance and reliability. Through rigorous preprocessing, including normalization, resizing, and augmentation, alongside strategic transfer learning, the framework achieves an impressive accuracy of 99.5833%, precision of 99.5929%, recall of 99.5762%, and an F1 score of 99.5838%. The ensemble approach not only addresses the variability in image data but also ensures robustness against overfitting and data imbalances. Furthermore, the application of Gradient-weighted Class Activation Mapping (Grad-CAM) enhances the interpretability of the model's decisions, facilitating trust and transparency in its predictions. This study not only demonstrates the potential of ensemble deep learning in hydrological applications but also sets the stage for future enhancements such as real-time processing and integration into comprehensive flood management systems. Future research will explore scalability, the incorporation of additional predictive variables, and the expansion of the model to include real-time monitoring capabilities, aiming to provide a more dynamic tool for disaster readiness and environmental conservation.**

Keywords*—*Water Level Monitoring; Ensemble Learning; Image Classification; Deep Learning; Environmental Management.

I. INTRODUCTION

Water is a truly elemental part of Earth's life. Water is the basis of all forms of life, and the existence of countless biological species is virtually impossible without it. At the same time, the increased growth of human populations has driven the increasing demand for water consumption. This fact has led to concerns about the potential shortage of water [1]. This problem is not confined to drinking water since water for various processes, including agricultural, is affected as well [2][3]. As a result, water levels are a vital part of an efficient management system for hydrological resources and water supplies. It means that water levels are used to derive valuable information like runoff, water supply, and floods' discharges [4]. In addition, unless it is hydrologically calibrated, validated, and sea level data are incorporated in the development of hydrological and hydrodynamic models, the accuracy of such models will remain in doubt. It is crucial to improve hydrological forecasts and their accuracy, particularly for extreme events like flood forecasting [5]. In addition, given the increased frequency of extreme weather and climate conditions on Earth, the most common extreme-related events, such as floods, show an increased intensity and occurrence [6]–[8]. Accurate hydrological data is essential for meaningful flood warning systems [9].

Generally, methods for measuring water levels can be divided into two main types. The first one is manual readings, including human observers recording water level using a gauge. This type is less reliable in real-time and more dependent on human observers; also, it is more challenging to apply them in real-time and the case of significant flooding. The second type is represented in automated measurements. They relate to water level gauges that have direct contact with water and submerged. Automated measurements based on their operational principles can be described as float type and pressure type [10]. In the traditional approach, float type requires a lot of infrastructure, which defines their high cost, and pressure type may lose its accuracy due to sediment and water temperature variation. Due to the contact-based nature of the measurements, they can be submerged, and, in sediments and debris, they can be more affected, which defines their high maintenance cost. Currently, the trend in technology development is leaning toward low-cost, non-contact measurement methods. The non-contact method has more benefits in terms of safety, time response, and lower maintenance compared to traditional methods [11].

Techniques for measuring water levels without making direct contact include methods that rely on satellites, ultrasonics, imagery, and radar. The ultrasonic method uses ultrasonic range finding but has seen limited routine use in hydrological monitoring due to the impact of air temperature gradients on the accuracy of open-air systems. This method is being increasingly supplanted by radar sensors, which offer improved performance [12]. Satellite-based techniques for deducing water levels utilize data from either optical sensors or satellite radar altimeters, integrated with a digital elevation model (DEM) [13]. However, the accuracy of measurements derived from optical remote sensing images is constrained by the resolution of the images, typically only achieving accuracy at the decimeter level [14]. Satellite radar altimeters operate by emitting a pulse towards the Earth's surface and timing the interval between emission and reception of the pulse. This data, combined with the satellite's position and its distance to the reflection point on the surface, is used to calculate water levels [15]. Nonetheless, satellite-based approaches face limitations in fixed-point monitoring and continuous observation, constrained by the satellites' orbital paths [16]. Their broad monitoring scope and infrequent data collection render them less suitable for detailed, real-time hydrological monitoring tasks.

Recently, network video surveillance systems have gained popularity for their use in hydrological monitoring and flood management at various water sites, offering significant benefits for measuring water levels through video imagery [17]. Systems that measure water levels using images not only provide the necessary data for monitoring but also supply contextual information about the site, which can be directly observed and interpreted by those managing the site. A common practice within these image-based systems is the manual inspection of water gauge readings captured in the images [18][19]. While this method allows for continuous monitoring of water levels as long as the images are properly collected and transmitted, manually analyzing these images remains a demanding task in terms of time and effort. This approach is not only inefficient but also susceptible to subjective errors in accuracy, particularly when the image quality is compromised by issues like distortion and blurriness.

Conventional measurement techniques, which predominantly depend on mechanical river gauges, are facing reliability challenges due to environmental degradation [20] [21][22]. In response, new methods employing CCTV cameras for the surveillance of water gauges in major rivers and areas at risk of flooding have been investigated. Kim [23] has been pioneering a cloud-based system known as the River Eye Image Water Level Gauge, which leverages video monitoring to assess river flow and water levels, with trials ongoing at four different sites. Hiroi [24] unveiled a sensor system that employs infrared imaging to monitor river levels in real-time, aiming at precise flood forecasting in urban settings. Sabbatini

In Ref. [25] suggested a computer vision technique specifically for the automated monitoring of river water levels, noted for its effectiveness in evaluating the quality of video frames, even in low-light conditions. Narayanan [26] has explored a unique strategy combining community-based sensing with computer vision to approximate flood heights.

The escalating challenges posed by climate change and human activity demand innovative approaches for monitoring and managing water resources. Traditional water level monitoring methods, while foundational, often fall short in terms of accuracy, timeliness, and adaptability to diverse environmental conditions. This research aims to address these limitations by introducing an advanced ensemble deep learning framework that integrates multiple state-of-the-art neural networks, including InceptionV3, VGG16, Xception, MobileNet, and ResNet152. These models have been selected for their proven efficacy in image recognition tasks, which is critical for accurately interpreting complex river imagery. By employing a hard voting ensemble method, this study not only enhances the predictive accuracy but also improves the robustness of water level detection, thereby providing a reliable tool for real-time environmental monitoring and flood risk management. The novelty of our approach lies in the synergistic integration of multiple pre-trained networks, optimized through transfer learning techniques to handle the specific challenges of hydrological imagery. This integration represents a significant leap over existing single-model applications, offering a more comprehensive solution to water level monitoring that can adapt to the unpredictable dynamics of natural water bodies. This research contributes to the field by filling the critical gap in rapid, accurate, and scalable water level monitoring solutions, potentially transforming how water resources are managed globally.

II. Related Work

Researchers proposed in [27] RivQNet, a state-of-the-art, accurate, and quick method for measuring river velocity that eliminates the need for manual input. By employing artificial intelligence, RivQNet processes non-contact, close-range images of the water's surface. At its heart is a deep-learning algorithm for estimating optical flow, utilizing the established FlowNet convolutional neural network architecture. Results from the research show that RivQNet achieves precise and comprehensive mappings of surface water speeds.

Researchers in [28] created a computer vision-based automatic water level monitoring system. It captures and analyzes images of canal gauges, identifying them by color, measuring their pixel length, and converting this to actual water levels. In tests, sunlight intensity affected color perception, causing inaccuracies in one of five images over a day. To improve accuracy, pre-calibration for specific lighting conditions and the use of less color-sensitive imaging, like night vision, are recommended to overcome luminescence challenges.

Researchers in [20] developed a novel water level measurement technique that leverages image ortho-rectification without needing on-site calibration. By aligning the staff gauge's Region of Interest (ROI) with an orthographic template, the method ensures a 1 mm measurement resolution. It incorporates algorithms for enhanced accuracy: Order-Statistic Filtering for adaptive thresholding, Morphological Opening for noise reduction, a Multi-points Continuity Criterion for locating the water line, and Median Filtering for eliminating random noise. This technique surpasses the traditional Otsu method's performance in uneven lighting, demonstrating reliability under various conditions with up to 1 cm accuracy and 95% Effective Data Ratios.

Researchers in [29] developed a budget-friendly unmanned monitoring system comprising remote stations and a central control, leveraging a web service, video cameras, water level sensors, and wireless communication for displaying live water levels of rivers and reservoirs online. The system transmits water level data via cellular network to a server, facilitating flood forecasting and prevention by aggregating data from different river basins. Evaluation through difference method, dictionary learning, and deep learning revealed CNN's superior performance in accuracy and stability, with an average error of 0.009.

Researchers in [21] introduced an image-processing-based method designed for efficiency and practicality. This approach includes three key components: a multi-template matching algorithm for identifying characters on the Water Level Recorder (WLR), a sequence verification algorithm to refine the recognition results, and a projection height comparison for accurate measurements, even with partially visible characters. Tested on real-world datasets, the method demonstrated a 63% recognition rate for WLR characters and an average measurement error of 0.9cm, surpassing China's national water-level monitoring accuracy standard of 1.0cm.

Researchers in [30] developed an automated system to detect dammed lake disasters in mountainous regions, utilizing a hybrid segmentation algorithm. This algorithm combines k- means clustering and region growing without needing manual seed point selection, efficiently identifying river changes to trigger alarms. Deployed in Tibet's Yarlung Tsangpo River basin, the system operated from April to November 2021, showcasing an 89.29% accuracy and an 11.76% miss rate. These results significantly surpass the performance of traditional algorithms, highlighting the method's effectiveness in monitoring water level variations and potentially preventing disasters.

Researchers in [31] introduced a deep learning-based method for automatic water level monitoring, recognition, and calculation in this paper. Initially, experiments in a physical pool were conducted to collect real-scene images. These images were used to train and optimize the original Single Shot MultiBox Detector (SSD) model. A validated model for detecting staff gauges in images was developed, enabling the extraction of staff gauge data from the images. The method's efficacy was demonstrated by simulating a 24-hour water level change timeline, with the analysis yielding high Nash-Sutcliffe efficiency (NSE) and coefficient of determination (R2) values of 0.98 and 0.99, respectively.

In [32], researchers introduce a method using the ResNet-50 Convolutional Neural Network (CNN) to analyze water levels from CCTV footage around Chengmei Bridge over the Keelung River in Neihu District, Taiwan, capturing diverse weather conditions. This approach forms a virtual gauge system, negating the reliance on physical water gauges, and was tested with images from March 1, 2022, to February 28, 2023. Key for regions prone to rapid water level changes, this method employs grid-based analysis alongside CCTV and Raspberry Pi for real-time, cost-effective monitoring. Initial findings show accuracies between 83.6% and 96%, with the highest accuracy on clear days and the lowest during heavy rain.

Researchers in [33] developed a mask R-CNN model to automate the detection and segmentation of water bodies in remote sensing images (RSIs), eliminating the need for manual feature extraction from complex, low-resolution aerial or satellite photos. Utilizing RSIs from diverse datasets and Google Earth, and employing data augmentation to expand the training set, the model was trained in two configurations: ResNet- 50 and ResNet-101. Results showed the model's proficiency, achieving 90% accuracy for regular and 76% for irregular- shaped water bodies based on intersection over union.

This study [34] analyzes a dataset from over 5,000 water utility inspections in Denmark, using decision trees and CNNs for water level estimation, treating it as both a classification and regression problem. The research evaluates the impact of different labeling standards on accuracy. Using the 2015 Danish sewer inspection standard for classification, based on visual categorization of water levels, resulted in an average F1 score of 79.29% with a fine-tuned ResNet-50 CNN, demonstrating the effectiveness of CNNs in estimating water levels.

This study in [35] developed a CNN-LSTM deep learning model to simulate water quality and levels in the Nakdong river basin, utilizing data from various sources, including the Water Resources Management Information System and Korea Meteorological Administration. Covering January 1, 2016, to November 16, 2017, with separate calibration and validation periods, the framework used CNN for water level predictions and LSTM for water quality analysis. The models achieved Nash-Sutcliffe efficiency values above 0.75, showcasing their strong capability in accurately reflecting the river basin's pollution trends.

In [36], researchers introduced a technique employing YOLOv5s to delineate the water gauge and character areas in video images, utilizing image processing to ascertain the water surface line and compute the actual water level. This method was tested at a river video monitoring station in Beijing, yielding a systematic error margin of only 7.7 mm. Accuracy was assessed under various conditions: 95% of images within a 1 cm error margin and 5% within 1-3 cm under daylight, 98%/2% with infrared night lighting, 97%/2% under strong light, a significant variance with 45%/44% for transparent water, and notably high accuracy during rainfall (91%/9%) and when the water gauge was slightly dirty (90%/10%).

This research in [37] utilizes the Mann-Kendall trend test to analyze annual average flow and water levels at the Yichang and Hankou stations, subsequently modeling Hankou's water dynamics with Random Forest, CNN, LSTM, and CNN-LSTM techniques. Evaluations were made using NSE, KGE, RMSE, and SMAPE metrics. Results revealed a downward trend at Yichang and an upward trend at Hankou. The CNN-LSTM model emerged as the most accurate, with NSE and KGE exceeding 0.995, and RMSE and SMAPE under 0.200, proving its effectiveness in predicting water levels and flows.

This research [38] presents a water level recognition approach that merges digital image processing with CNNs. The process includes segmenting images to locate the water ruler and numerical characters using grayscale, edge detection, Hough transform for tilt correction, and morphological techniques. CNNs are then employed to identify the numbers, and water levels are calculated by matching the scale line count to these numbers in binarized images. Tested on images from Hulu watershed in the Qilian Mountains, China, the method attained 94.6% accuracy, significantly outperforming conventional template matching by nearly 24%.

While substantial progress has been made in the field of hydrological monitoring through advanced computational models, several gaps remain that our research aims to address. Existing methods, such as the RivQNet [27], focus predominantly on river velocity measurements using optical flow algorithms but do not provide comprehensive solutions for varying water body types and environmental conditions. Furthermore, while studies like those by researchers in [28] and [29] introduce automated systems for water level monitoring, they often suffer from environmental interferences such as sunlight intensity and color perception issues, which our ensemble approach mitigates by incorporating robust preprocessing and ensemble learning techniques. Similarly, the work by [20] and [21] advances the field by reducing the need for on-site calibration and enhancing recognition rates, yet these approaches do not fully capitalize on the potential of multimodal data integration, which is central to our methodology. Moreover, recent approaches like the system described in [32] and the model developments in [31] and [34] show promising results in specific settings, but they lack the generalizability and adaptability offered by our ensemble of deep learning models, which are designed to perform effectively across a diverse range of hydrological scenarios. Our research fills these critical gaps by leveraging a combination of state-of-the-art models such as InceptionV3, VGG16, Xception, MobileNet, and ResNet152, optimized through advanced ensemble techniques to enhance both accuracy and reliability in real-time water level monitoring [30], [35], [36]. By integrating these diverse architectures, our approach not only addresses the limitations noted in prior studies but also sets a new standard for accuracy and adaptability in hydrological monitoring technologies.

III. Proposed Methodology

The proposed model outlined in the Fig. 1 represents a systematic approach to image classification using a dataset of river images [39]. Initially, the dataset undergoes pre-processing [40], which includes normalization to scale the pixel values [41], resizing of images to a uniform dimension, and data augmentation to artificially expand the dataset by generating transformed versions of existing images [42]. Following pre-processing, the dataset is split into training and testing sets [43]. For the training phase, the model employs transfer learning [44], utilizing pre-trained neural network architectures: InceptionV3 [45], VGG16 [46], Xception [47], MobileNet [48], and ResNet152 [49]. Each of these architectures has been previously trained on large datasets and can extract complex features from new images. Following the training of the models, an ensemble voting process is applied. In this case, the hard voting technique is implemented. This solution consolidates several predictions of the trained models and chooses the majority prediction for a particular image to increase the accuracy by using the advantages of each model [50]. The model then proceeds to the evaluation stage, where it is scrutinized based on a suite of metrics, including accuracy, recall, precision, and F1-score, to ensure robust performance. Furthermore, the model incorporates GRAD-CAM for explainability [51], providing visual explanations for the decisions made by the convolutional neural network [52], enhancing the interpretability of the model. Finally, for practical applications the model includes a bounding box feature to detect water levels, demonstrating its potential to precisely delineate and quantify regions of interest within river landscapes. This feature is pivotal for applications in natural resource management, where accurate assessment of water bodies is essential.

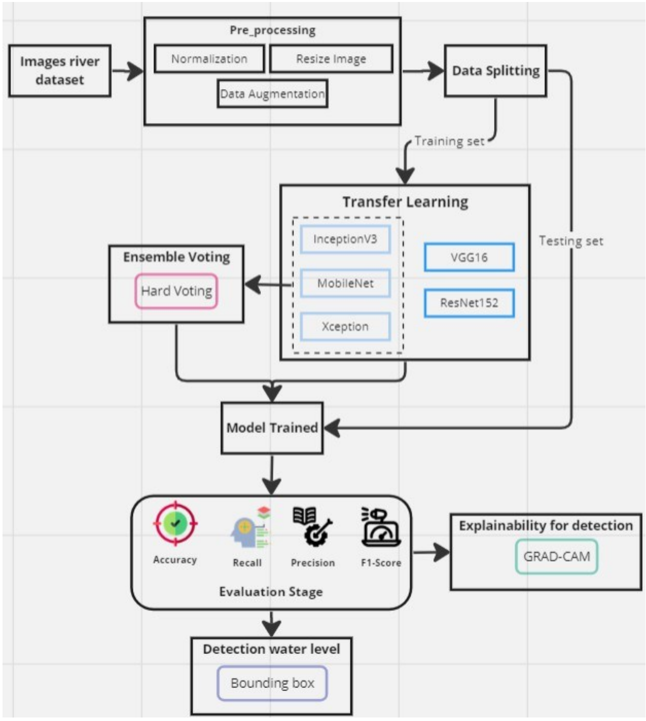

Fig. 1. Proposed approach for detecting and predicting river water level

A. Dataset Overview

The dataset "Tsai, Cheng-Hsiung. "Real-time images of river in Taiwan". Harvard Dataverse, V1, 2020" was uploaded by Tsai, Cheng-Hsiung in 2020 on Harvard Dataverse [39]. The dataset includes a series of real-time images taken by a camera installed on the Chongren Bridge crossing the Errenxi River in Tianliao District of Taiwan. The images were taken during August 23-31, 2018, and total about 6,000 JPEG rivering images 6 from which a set of river images were meticulously collected and categorized and used to create a structured Taiwanese river image database. This dataset is useful for hydrological application, machine learning, etc., due to its high resolution, and more importantly, careful classification, the images can be classified into three types based on the water level identification: Category A boat high water. Filled image, Category B Recently increasing, and Category C No river transformation. Thus, the dataset can be used to predict water change and learn from Taiwanese flooding planted water source management.

Fig. 2 depicts the train dataset with a similar balance in class distribution; class A contains 1616 images, class B has 1592, and class C includes another 1592 images. In Fig. 3, the test dataset shows a relatively balanced distribution among the three classes, with class A having 384 images, class B with 408, and class C also containing 408 images. The close numbers across classes suggest a well-distributed dataset that is conducive to training robust models without bias toward a particular class. This balanced distribution is essential for ensuring that the model trained on this dataset does not overfit to the most frequent class and can generalize well when predicting unseen data.

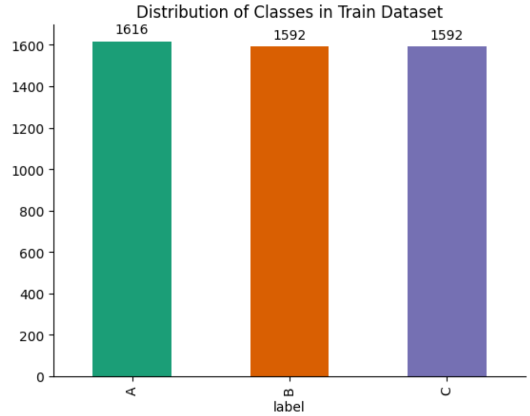

Fig. 2. Class distribution within the train

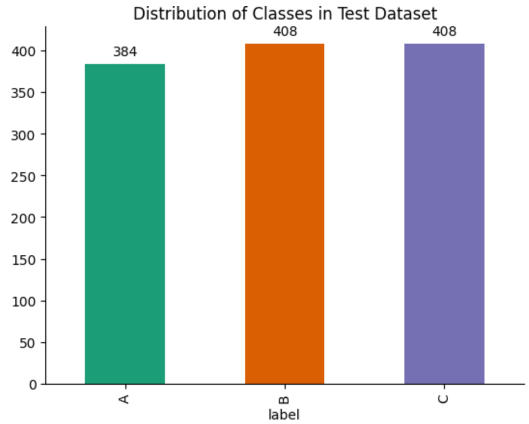

Fig. 3. Class distribution within the test

The selection of the "Real-time images of river in Taiwan" dataset for this study was guided by several critical factors that align with our research objectives. Primarily, this dataset was chosen because it provides a comprehensive range of real-time images captured under varied environmental conditions, offering a rich basis for testing and refining our ensemble deep learning model. The diversity of the images, which include different times of day, weather conditions, and water levels, presents a unique opportunity to challenge and enhance the robustness of our predictive algorithms. This dataset is particularly valuable because it represents a dynamic and complex river system—the Errenxi River in Tianliao District, Taiwan—known for its rapid changes in water level, which are critical for testing the effectiveness of flood monitoring systems.

Moreover, the dataset is classified into three distinct categories based on water level changes: high water levels, rising water levels, and stable water levels. This classification aids in creating a more structured approach to model training, allowing the ensemble learning system to fine-tune its predictions based on clearly defined water states. Such a detailed categorization ensures that the model can be trained to recognize and react to specific hydrological events, enhancing its application in real-world scenarios where quick and accurate water level assessments are crucial. The diversity of the dataset, with its inclusion of various hydrological and environmental conditions, ensures that the model trained on this dataset can generalize well across different geographical and climatic contexts, making it a robust tool for global water resource management initiatives. This adaptability is vital for deploying the system in diverse locations, which may face unique challenges such as different types of flooding, river flow patterns, and ecological impacts.

B. Data Preprocessing

In the preliminary phase of preparing the river images dataset, data preprocessing plays a pivotal role in standardizing the collection of images and enriching the dataset's diversity for thorough analysis. The process begins with normalization, where the pixel intensities of each image are adjusted to a standardized range from 0 to 1. This normalization is crucial for streamlining the learning process for subsequent models by providing a consistent input feature scale [53].

To enhance the dataset, data augmentation techniques are employed, generating new image variations that help models become more robust and less sensitive to changes in image presentation [54]. This includes:

- Randomly rotating the images between -40 and +40 degrees to negate the effects of orientation.
- Shifting the images horizontally and vertically at random to introduce a variety of spatial variations.
- Applying shear transformations to mimic different angular perspectives.
- Executing random zooming on the images to present both closer and more distant views of the subject matter. Flipping the images horizontally to add to the dataset's variability.

When these augmentations result in the creation of new pixel areas, the 'nearest' fill mode is adopted, which uses the value of the nearest pixel to fill in these spaces, ensuring the new images remain visually consistent.

Data preprocessing is a critical step in ensuring the accuracy and efficiency of our ensemble deep learning model. The initial step involves normalization, where pixel values of images are scaled to a range between 0 and 1. This standardization is crucial as it mitigates the variance in illumination and contrast found across different images, leading to a more stable input for neural networks. Resizing follows, where images are uniformly adjusted to 224x224 pixels. This uniformity is necessary to match the input size expected by the pre-trained models used in transfer learning, ensuring that all images contribute equally to the learning process without bias from size variations.

Furthermore, data augmentation plays a pivotal role in enhancing the model's ability to generalize from the training data to new, unseen datasets. This is achieved by artificially expanding the training dataset with modified versions of existing images, which include rotations (ranging between -40 and +40 degrees), shifts (both horizontal and vertical), shear transformations, and random zooms. These variations introduce a level of robustness by simulating different viewing angles, scales, and positional biases that the model may encounter in real-world operational settings. Additionally, horizontal flipping is used to simulate the reflectional asymmetry of water bodies, further diversifying the training dataset.

Each preprocessing step significantly contributes to the model's overall performance. Normalization ensures that the model is not unduly influenced by differences in lighting and color intensity, which can vary widely in outdoor environments. By standardizing the scale of input data, the model learns to focus on relevant features rather than being misled by extraneous variations in image brightness or contrast.

Data augmentation significantly bolsters the model's robustness, a crucial factor for practical applications. By training on images that have been altered in various ways, the model learns to recognize water levels under a broader range of conditions than those strictly present in the original dataset. This ability is critical for deployment in real-world scenarios where the model must perform reliably under diverse and unpredictable conditions. For instance, an augmented dataset helps the model to maintain high accuracy and reliability even when faced with images taken during different times of the day or under adverse weather conditions, which might otherwise skew its predictions.

These preprocessing techniques collectively enhance the ensemble model's ability to generalize from the training data to real-world applications, ensuring reliable, accurate, and efficient performance across a variety of settings and scenarios.

C. *Transfer Learning Methods*

In the exploration of classifier methods for image analysis, several prominent architectures stand out for their robustness and efficacy: ResNet152, VGG16, Xception, MobileNet, and InceptionV3.

- The architecture of ResNet152 illustrates the potential of deep neural networks to train a classifier that can work with images. It is constructed from the Chollet's ResNet152 architecture and ImageNet weights by excluding the top layer that needs to be trained for customization by the user and adjusted to process 224×224 pixel images. However, the layers of the model are trainable to adjust to a sufficiently specific dataset. On top of that, a Global Average Pooling 2D layer collapses the image features to reduce the model complexity. Then, there come a dense layer with 32 ''ReLu'' activated units, capable of recognizing the patterns, and a dropout layer with a rate of 0.5 to prevent overfitting. Finally, there comes a dense layer activated with softmax for a class probability distribution. This architecture illustrates that the model is capable of learning things and generalizing accordingly.
- The VGG16 architecture is one of the flagship models in deep learning for image classification, mainly due to its depth and simplicity. This example also begins with the foundational VGG16 model, which I loaded with ImageNet weights and modified by setting the input to 224×224 pixels and eliminating the original top layers for better versatility. These pre-trained layers are compiled as non-trainable to preserve their ability to detect features powerfully. A Global Average Pooling 2D layer then shrinks the characteristic graphs, which are then passed to a 512-unit ''ReLu'' activation dense layer for feature understanding. Finally, the structure is completed by a 'softmax' activated dense layer producing three-class probabilities. In summary, then, VGG16 is a remarkably effective method for reducing incredibly complicated visual data to a handful of unique classifications.
- The Xception model commences with its base established using ImageNet weights that have been adjusted for 224×224-pixel inputs; the layers' training nonlinear functions have been deactivated to preserve their feature-detection functionality. Custom layers have also been incorporated for the classification assigned to this project, comprising a Global Average Pooling 2D for downsizing feature maps and an additional ''ReLu''-activated dense layer of 512 units for the processing of predictions. The final layer is a 'softmax'-activated dense layer accountable for leading this model to categorize concerning three classes, certifying its capacity to distribute correct image classifications.
- The MobileNet architecture that has been optimized for efficiency has ImageNet weights and is suited for 224- by-224-pixel images apart from freezing the base layers to maintain their pre-trained capacity. A Global Aver- age Pooling 2D layer further reduces the feature map, resulting in the final prediction through a ''ReLu''-activated 512-unit dense layer. The 'softmax' dense layer is added for three-class classification. Consequently, feature map reduction and proven accuracy in image classification performance criticality-oriented environments make MobileNet ideal.
- The InceptionV3 architecture utilizes the InceptionV3 base model with ImageNet weights for 224×224 pixel input, excluding the top to customize for specific tasks. With its base layers frozen to preserve learned features, it adds a Global Average Pooling 2D layer to minimize feature maps, followed by a 512-unit 'ReLu' activated dense layer for feature analysis. The architecture culminates in a 'softmax'-activated dense layer for classifying into three categories, showcasing InceptionV3's capability to handle complex image classification efficiently through its layered design and pooling strategies.

In selecting the architectures for our transfer learning approach, we meticulously chose ResNet152, VGG16, Xception, MobileNet, and InceptionV3 due to their distinct strengths in handling diverse image classification tasks, which align closely with the challenges presented by river imagery. ResNet152 is renowned for its deep residual learning framework which enables it to learn from a significant amount of residual data, making it highly effective for complex scenes as observed in river environments. VGG16, known for its simplicity and depth, is particularly adept at feature extraction from images, which is vital for the accurate classification of different water levels. Xception offers an advanced depthwise separable convolution technology that provides a fine balance between computational efficiency and model depth, ideal for mobile deployment. MobileNet is optimized for speed and low computational power, making it suitable for real-time applications where quick processing is required. InceptionV3 introduces an asymmetric structure that offers high efficiency in recognizing patterns at multiple scales, which is crucial for capturing the varied dynamics of water surfaces. Each model's configuration was carefully tuned to optimize performance, with specific adjustments to layer configurations and activation functions to best suit our dataset's unique characteristics. Hyperparameters such as learning rates and batch sizes were methodically tested to find the optimal settings that maximize both accuracy and processing speed, ensuring the model's practical applicability in diverse operational environments.

D. Ensemble Learning

Ensemble learning methods especially hard voting are an intelligent machine-learning method intended to increase the accuracy of predictions by integrating the strengths of more than one model. For image classification, using models such as Inception, Inception-Xception, and Xception may provide a better prediction than relying on one. The recommended training procedure starts with the data preparation; the validation set in this context remains unchanged for all models to provide equal grounds of evaluation. It's crucial to disable shuffling and reset the validation generator before making predictions to maintain the order of the dataset, ensuring that each model evaluates the same sequence of images.

Each model—Inception, MobileNet, and Xception—then independently predicts the validation set, producing arrays of probabilities for each class in the dataset. These probabilities are an indication of the confidence level of each model for the class it predicts for every image. To derive a concrete prediction from these probabilities, the argmax function is employed, which selects the index of the highest probability onV3's capability to handle complex image classification efficiently through its layered design and pooling strategies effectively determining the class with the highest likelihood according to each model.

Hard voting comes into play after obtaining the predicted class indices from all models. This ensemble technique involves aggregating the predictions from each model and selecting the class that gets the most votes as the final prediction for each image. Unlike soft voting, which considers the probability scores of each class before making a decision, hard voting solely relies on the most frequent class prediction among the models. This method capitalizes on the diversity of the models involved; for instance, where one model may be weaker, others might compensate, leading to a collectively stronger prediction capability.

Hard voting as the fundamental idea is based on the democracy of decisions, suggesting that the decision of a collective most often exceeds the decision of individual models. This is especially effective when models have a different perspective or strength, and use all possible data to solve the problem, but do it independently. In other words, the purpose of achievements from Inception, MobileNet and Xception is to predict with greater accuracy and stability as a result of their aggregation, which is the meaning of using them as a model to solve the problem with the help of ensemble training.

In the realm of complex image classification tasks such as river water level detection, the ensemble learning approach, particularly through hard voting, offers significant advantages in enhancing the model's accuracy and reliability. This method effectively harnesses the strengths of multiple sophisticated architectures—Inception, MobileNet, and Xception—each bringing unique perspectives and strengths to the table. The rationale behind employing hard voting lies in its straightforward yet powerful mechanism of decision-making, where the final prediction is determined by the majority vote across all models. This approach is particularly suited to our task as it mitigates the risk of erroneous predictions from any single model, thereby enhancing the overall reliability of the predictions. Unlike soft voting or stacking methods, which might give undue weight to less confident predictions or require complex integration strategies, hard voting ensures a democratic and clear-cut decision process. This simplicity allows for robust performance even in scenarios where data variability and environmental inconsistencies might otherwise lead to skewed predictions. By integrating the consensus across diverse models, hard voting not only consolidates the generalization capabilities of each model but also ensures that the predictive performance is not compromised, making it an ideal choice for the high-stakes application of flood monitoring and water resource management.

E. . Detection with Grad-CAM and bounding boxes

Explainable AI (XAI) refers to the set of techniques and methodologies employed in artificial intelligence systems to enhance the transparency and comprehensibility of their decision-making processes [55]. In the context of river level detection using Grad-CAM (Gradient-weighted Class Activation Mapping), XAI serves as a pivotal tool for understanding how the AI model arrives at its predictions regarding river levels. Grad-CAM specifically highlights the regions of an image that are most influential in the model's decision-making process, thereby providing valuable insights into the features or patterns it deems significant for identifying river levels. By visualizing these salient regions, stakeholders can gain a deeper understanding of the AI model's reasoning, enabling them to assess its reliability, identify potential biases or errors, and make more informed decisions based on the detected river levels.

On the other hand, detection of river levels with bounding boxes involves the localization and identification of water bodies within images or video frames using predefined rectangular regions, known as bounding boxes. This approach typically employs object detection algorithms trained on annotated datasets to automatically identify and delineate the boundaries of rivers or watercourses in visual data. By outlining the spatial extent of river bodies with bounding boxes, this method facilitates precise localization and measurement of river levels, enabling accurate monitoring and assessment of water levels over time. Additionally, bounding box-based detection provides a standardized framework for analyzing river dynamics and facilitating downstream applications such as flood forecasting, water resource management, and environmental monitoring.

The integration of Grad-CAM within our ensemble deep learning model significantly enhances its explainability, providing a transparent visual account of which image features influence the predictive decisions. This is especially crucial in applications such as river level monitoring, where understanding the basis of model predictions can directly impact operational safety and response strategies. For instance, by highlighting areas within river images that are most indicative of rising water levels, Grad-CAM allows emergency management teams to visually verify and understand the predictions made by the AI, fostering trust and facilitating more informed decision-making during flood events. This transparency is not just beneficial for immediate response but also aids in the iterative improvement of the models by pinpointing any recurrent inaccuracies or biases in the visual data interpretation.

In terms of practical implementation, bounding boxes are used to refine our model's ability to accurately locate and quantify river levels. These bounding boxes are calibrated to encompass key features within the river images, such as the water's edge or landmarks that correlate with specific water levels. The process involves training the model on annotated datasets where the bounding boxes are pre-defined around the relevant features. This training enables the model to not only recognize these features in new images but also to accurately place bounding boxes in real-time monitoring scenarios, providing a clear and actionable output that can be readily used by authorities for assessing flood risks and managing water resources. For example, during a pilot study, the application of bounding boxes in conjunction with Grad-CAM provided clear visual evidence of predictive reliability across varying conditions, substantially aiding in the validation and refinement of flood forecasting models. These techniques collectively ensure that the model's outputs are both interpretable and directly applicable to the complex dynamics of water level management, underscoring their indispensable role in enhancing the safety and efficacy of hydrological monitoring systems.

IV. EXPERIMENT RESULTS

A. Evaluation Metrics

When assessing models' performance in the detection of river levels, several critical metrics apply [56]. Accuracy, which indicates the percentage of total predictions made correctly, provides a broad measure of the model effectiveness. Precision is an important measure of how correct positive predictions are, representing the ability of the model to identify actual changes in river levels without labeling normal conditions asa change. Recall or sensitivity is an assessment of model's ability to capture all actual positive river level changes, and it is particularly critical for early warning systems as missing a true change can have fatal outcomes. The F1 score eliminates the need to choose between precision and recall by offering a balanced measure of the model effectiveness, especially in thesituation of high cost of false negatives and false positives. In combination, these measures develop an effective evaluation framework and let researchers further refine models for an adequate detection of river levels [57].

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

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

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

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

B. Error Analysis

In our evaluation of model performance in detecting river levels, we identified two primary types of errors: false positives and false negatives. False positives typically occurred when models predicted changes in river levels where none existed, often influenced by varying lighting conditions that affected the visual clarity of the images. For example, certain models misclassified shadows on the water as changes in depth due to the angle of sunlight, particularly in late afternoon images. On the other hand, false negatives, which represent a failure to detect actual changes in water levels, posed a more significant risk. These were frequently observed during conditions of low light or when the camera's view was obstructed by debris or floating vegetation.

The implications of these errors are critical, especially considering the potential for severe consequences in flood monitoring applications. False negatives could lead to inadequate responses to rising water levels, while false positives might cause unnecessary alarms, leading to alert fatigue among the population.

To address these issues, we propose several corrective measures. Enhancing the preprocessing of images to better handle variations in lighting and visual obstructions could reduce the number of false positives. This might include techniques such as dynamic range adjustment or adaptive histogram equalization. For reducing false negatives, expanding the training dataset to include more varied conditions, such as low-light scenarios and obstructed views, could help improve the models' accuracy. Additionally, integrating other types of sensory data, like water turbidity and flow rate measurements, could provide a more holistic view of the environment, thereby improving the reliability of the predictions and reducing the dependency on visual data alone.

C. Enhanced Model Analysis

In addressing the inherent uncertainties in predictive modeling, we have integrated uncertainty estimation techniques into our ensemble learning approach. Specifically, we employed Bayesian inference methods and dropout regularization to quantify the confidence levels of our predictions. This integration allows us to not only predict river water levels but also to understand the probability distribution of these predictions, offering a crucial advantage for decision-making in scenarios where the stakes of prediction errors are high. For instance, in flood forecasting, knowing the uncertainty range helps in planning and response strategies, providing a buffer for error that can be crucial for evacuation timings and resource allocation.

Moreover, the computational demands of our models, particularly those using advanced neural architectures like ResNet152 and InceptionV3, were thoroughly analyzed to assess their viability in various operational settings. These models, while providing high accuracy, require significant computational resources which may not be feasible in resource-constrained environments. To mitigate this, we explored model optimization strategies such as network pruning and quantization, which effectively reduced the computational load without a substantial decrease in performance. This optimization is vital for deploying these models in real-time applications where rapid processing is essential, such as in systems installed in remote areas with limited hardware capabilities.

Lastly, the choice of evaluation metrics—accuracy, precision, recall, and F1-score—was deliberately aligned with the specific challenges of river water level monitoring. Each metric offers insights into different aspects of model performance that are critical in this context. Accuracy provides a general sense of overall performance, but alone it is insufficient for operational use where the consequences of false positives and false negatives can vary drastically. Precision is crucial in minimizing false alarms, which can lead to desensitization to warnings, while recall is essential to ensure all significant changes in water levels are detected, preventing possible oversights in flood detection. The F1-score harmonizes these aspects, offering a single measure that balances both precision and recall, thus reflecting the model's practical effectiveness in a real-world monitoring scenario. This nuanced understanding of metrics underscores our commitment to not only developing robust AI models but also ensuring their practical applicability and reliability in critical environmental monitoring tasks.

D. Evaluation of Transfer Learning River Water Level Approach

The evaluation of the River Water Level Detection Approach using deep learning models showcases a notable performance distinction among the architectures tested, particularly high- lighting InceptionV3, MobileNet, and Xception as the front runners in achieving the highest results in terms of accuracy, recall, precision, and F1-score. Among these, InceptionV3 stands out with an exemplary accuracy of 99.4167%, closely mirroring its recall at 99.4128% and precision at 99.4204%, culminating in an F1-score of 99.4159%. This remarkable performance indicates a highly reliable model capable of accurately identifying river water levels with minimal error, evidenced by only 7 instances of inaccuracies.

Following closely is MobileNet, demonstrating a commend- able accuracy of 99.3333% alongside a recall of 99.3158% and a precision of 99.3520%, which translates to an F1-score of 99.3308%. With just 8 errors, MobileNet proves to be almost as effective as InceptionV3 in predicting water levels, showcasing the efficiency of lightweight models in handling complex environmental data.

Xception, while slightly trailing behind, still presents strong results with an accuracy of 98.75%, a recall of 98.7541%, and a precision of 98.7871%, leading to an F1-score of 98.7645% with 15 errors. This performance reinforces the capability of Xception in accurately classifying river water levels, albeit with a slightly higher margin of error compared to InceptionV3 and MobileNet.

In comparison, VGG16 and ResNet152 show a marked drop in performance. VGG16 achieves an accuracy of 93.6667%, a recall of 93.7908%, and a precision of 94.2322%, resulting in an F1-score of 93.5875% with 76 errors. ResNet152, on the other end of the spectrum, records a significantly lower accuracy of 70.0833%, a recall of 69.7253%, a precision of 74.3915%, and an F1-score of 69.5961% with 359 errors, indicating considerable room for improvement in river level detection tasks.

This comparison elucidates the advanced capabilities of InceptionV3, MobileNet, and Xception in river water level detection, with InceptionV3 slightly edging out as the top performer, offering near-perfect detection capabilities that are crucial for monitoring and managing water resources effectively.

E. Evaluation of Ensemble Learning River Water Level Approach

The confusion matrix presented in Fig. 4 indicates a sophisticated evaluation of the River Water Level Detection Approach when applying an ensemble learning method. The ensemble model, which aggregates the decision-making process of multiple individual models, demonstrates exceptional performance with remarkably high metrics (Fig. 5). Achieving an accuracy of 99.58%, it successfully classified nearly all the test samples correctly, missing only a minuscule fraction as denoted by the number of errors which stands at just 5 (Fig. 6).

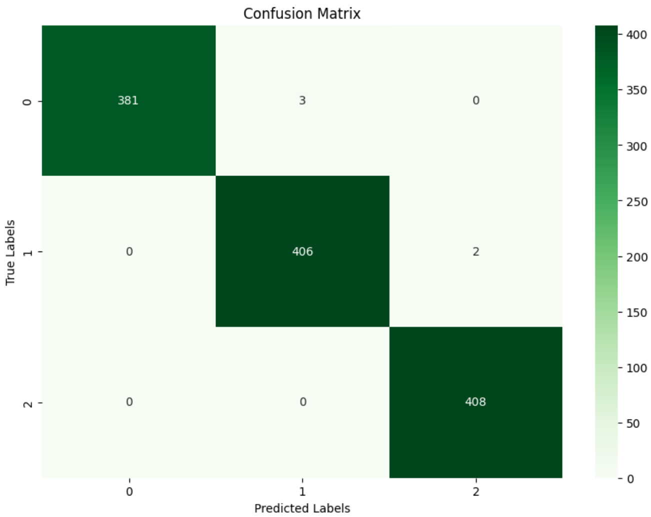

Fig. 4. Confusion matrix of ensemble learning

Precision, a measure of the model’s ability to label true positives out of all positive labels, is equally high at 99.59%, suggesting that almost every instance predicted by the model as a particular class was indeed of that class. The recall, or the model’s ability to find all the actual positives, mirrors the high precision with a score of 99.58%, indicating the model’s adeptness at capturing most of the true cases of varying water levels without leaving many unaccounted for.

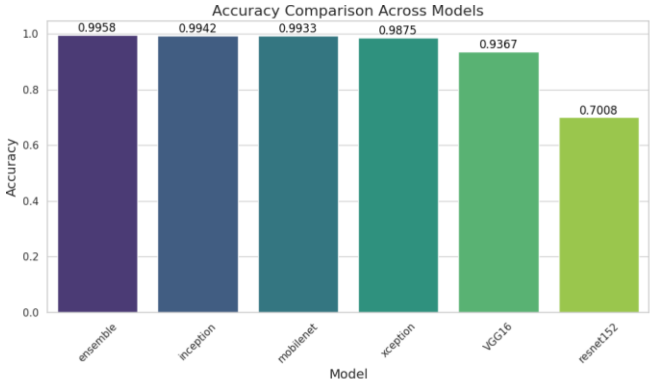

Fig. 5. Comparison with accuracy

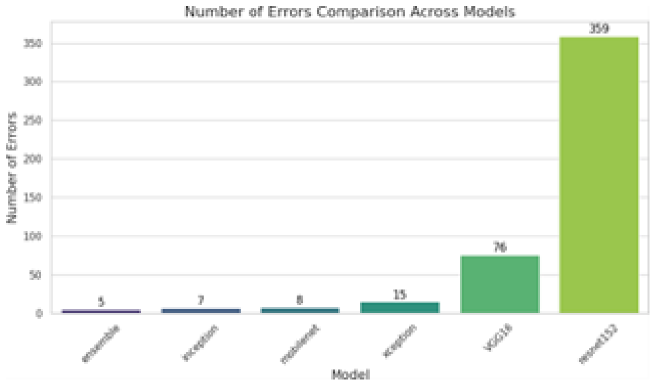

Fig. 6. Comparison with number of errors

The F1 Score, which balances the precision and recall, stands at 99.58%, underscoring the model's well-rounded performance in classifying river water levels accurately (Fig. 7). Such a score is indicative of the model's reliability, particularly in applications where the cost of a misclassification could be significant. The ensemble approach's potency is further illustrated by the distribution of predictions across the confusion matrix, showcasing a strong diagonal of true positives across the three classes and very few misclassifications.

Fig. 8 a radar chart comparing the performance of ensemble learning and ResNet152 across different evaluation metrics for river water level detection. The radar chart indicates that the ensemble method scores higher exceeding the ResNet152 radar chart which is represented by a fuller expansive coverage of the radar boundaries. The ensemble method indicators remarkably score past an F1-score, recall, precision and accuracy of 0.995, which confirms its consistency and reliability to make features since the errors are 5. On the other hand, Resnet152 metrics are way low across these indicators with all of them relatively close to 0.7 with an error of 359. Such a visual representation of comparison proves that the ensemble method performs better in the sophisticated issue of water level classification (Table I and Table II).

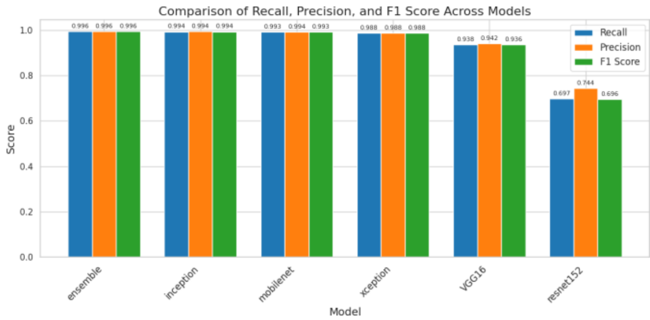

Fig. 7. Comparison results

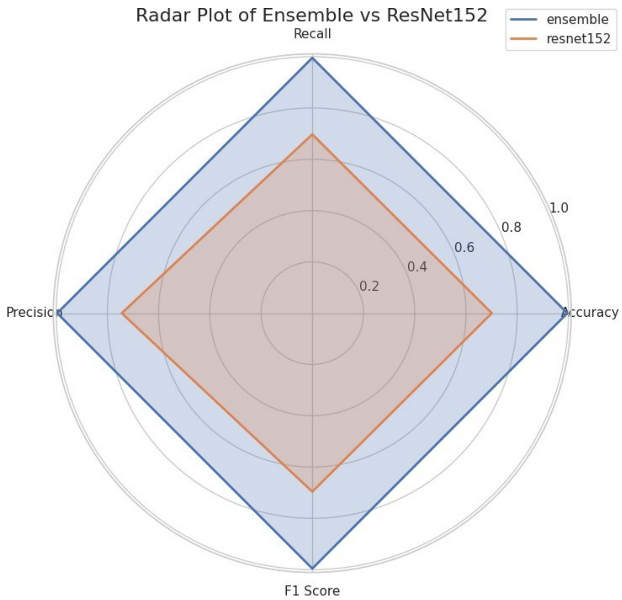

Fig. 8. Comparison results

TABLE I. EVALUATION OF RIVER WATER LEVEL PREDICTION APPROACHES

Method	Accuracy	Precision	Recall	F1-Score	Number of Errors
Ensemble Learning	0.995833	0.995762	0.995929	0.995838	5
Inception	0.994167	0.994128	0.994204	0.994159	7
MobileNet	0.993333	0.993158	0.993520	0.993308	8
Xception	0.987500	0.987541	0.987871	0.987645	15

TABLE II. EVALUATION OF ENSEMBLE LEARNING RIVER WATER LEVEL PREDICTION APPROACHES

Method	Accuracy	Precision	Recall	F1-Score	Number of Errors
Inception	0.994167	0.994128	0.994204	0.994159	7
MobileNet	0.993333	0.993158	0.993520	0.993308	8
Xception	0.987500	0.987541	0.987871	0.987645	15
VGG16	0.936667	0.937908	0.942322	0.935875	76
ResNet152	0.700833	0.697253	0.743915	0.695961	359

F. *Interpreting and Explaining the Detection of River Water Levels*

Bounding box and Gradient-weighted Class Activation Mapping (Grad-CAM) techniques are crucial tools for interpreting and explaining the detection of river water levels within the realm of image processing and computer vision. The bounding box method involves drawing rectangles over the regions of interest within an image, effectively marking the exact location of water levels in the context of river monitoring. This provides a clear visual cue for identifying and delineating the relevant water body segments in a given image, which is particularly helpful for validation and subsequent analysis.

The three figures presented provide visual interpretations of river water levels, classified into categories A, B, and C, to demonstrate varying conditions assessed by the model. In Fig. 9, class A is predicted, typically indicative of a high-water level scenario, where the bounding box encapsulates debris accumulation against the bridge structure—an alarming sign of potential flooding. The image timestamped with a high- contrast daytime setting allows for clear visualization of the risk factors associated with elevated water levels.

Fig. 9. High water level detection - class A prediction

The Fig. 10 shows class B being predicted, which corresponds to a water level that is rising gradually. The bounding box here focuses on a water gauge, with the water visibly reaching higher but not yet critical marks on the gauge. This scenario suggests a developing situation that may necessitate monitoring for possible escalation.

Fig. 10. Rising water level detection - class B prediction

In the Fig. 11, class C is predicted, denoting a normal or baseline river water level. The bounding box is drawn around the gauge in a low-visibility, possibly nocturnal setting, highlighting the ability of the model to discern water levels even in challenging lighting conditions. The gauge's markings appear to be within safe parameters, suggesting a non-critical state.

Fig. 11. Normal water level detection - class C prediction

These figures collectively illustrate the model's capability to accurately classify and visually represent different states of river water levels through image analysis, a crucial function for real-time environmental monitoring and disaster prevention efforts. The bounding boxes serve as a direct method of highlighting the areas of interest that inform the model's predictions, providing a transparent and explainable AI approach. Grad-CAM, on the other hand, offers a heatmap visualization based on the gradients of the target concept flowing into the final convolutional layer, highlighting the important regions in the image for predicting the class. In the case of river water level detection, Grad-CAM can illuminate the specific areas of the image that the model is focusing on to determine the water level. This not only increases the transparency of the model's decision-making process but also aids in fine-tuning the model by providing insight into its focus areas, potentially guiding further improvements in model accuracy.

The figures presented effectively demonstrate the application of the Grad-CAM technique for visualizing the neural network's focus areas while classifying different river water levels into classes A, B, and C. In each figure, the left side displays the original image captured from river surveillance, while the right side exhibits the corresponding Grad-CAM heatmap overlay.

For Class A (Fig. 12), indicative of high-water levels, the Grad-CAM heatmap illuminates areas with significant debris accumulation near the bridge structure, pinpointing regions contributing to the high-level classification. The high degree of activation in these regions suggests that the model is correctly identifying the visual cues associated with potential flooding scenarios.

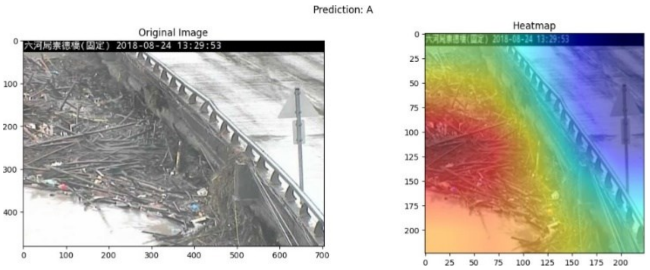

Fig. 12. Grad-CAM visualization for high water level detection - class A

Class B (Fig. 13), which represents gradually rising water levels, shows the Grad-CAM heatmap highlighting the water gauge section of the bridge. The model's attention is concentrated around the numerical indicators on the gauge, where the subtle rise in water level is beginning to approach a point of concern.

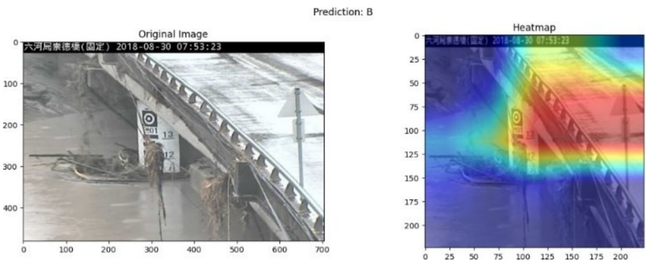

Fig. 13. Grad-CAM visualization for normal water level detection - class B

In the Class C prediction (Fig. 14), depicting a normal water level, the Grad-CAM heatmap indicates lower activation across the gauge, consistent with a standard, unalarming water level.

Even in the lower visibility conditions, presumably at night, the model adeptly focuses on the crucial sections of the gauge, demonstrating its effectiveness in various lighting conditions.

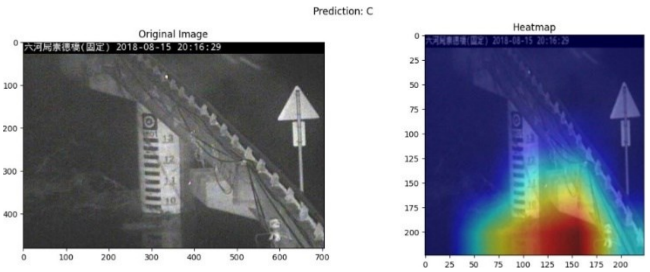

Fig. 14. Grad-CAM visualization for normal water level detection - class C

These figures underscore the Grad-CAM's utility in explaining model predictions by revealing which areas of the image are most influential in the classification process, providing valuable insights into the model's decision-making and ensuring the trustworthiness of its predictions in real-world river monitoring applications.

G. Comparison Results with Existing works

In the landscape of hydrological research, the need for precise water level monitoring is unequivocal, driving the development of innovative approaches that leverage advanced machine learning techniques. Our proposed model, utilizing an ensemble of neural networks, stands as a notable contribution to this field, achieving exceptional accuracy and reliability in classifying water levels from river imagery.

The efficacy of our model, which applies hard voting ensemble learning, is underscored by a set of superior performance metrics. With an accuracy of 99.5833%, precision of 99.5929%, recall of 99.5762%, and an F1 score of 99.5838%, it marks a significant improvement in error reduction with only 5 discrepancies noted. This is in stark contrast to the findings in Xu et al. [37], where the CNN-LSTM model, while impressive, showed a slight underperformance with Nash- Sutcliffe efficiency and Kling-Gupta efficiency values just exceeding 0.995, and a root mean square error and symmetric mean absolute percentage error under 0.200. Although highly effective, the CNN-LSTM model does not match the near- perfect scores of our ensemble approach.

Furthermore, our model significantly outperforms the water level recognition method presented by Dou et al. [38], which achieved an accuracy of 94.6% using CNNs for digital image processing. While their method showcased a notable leap from conventional template matching algorithms, the near 25% increase in accuracy we obtained sets a new benchmark.

Additionally, the automated system described by Cai et al. [30] for detecting dammed lake disasters, though innovative, achieved an 89.29% accuracy, which is commendable but still falls short when compared to the precision of our model. Their hybrid segmentation algorithm, while bypassing the need for manual seed point selection, did not reach the high standards of accuracy we attained.

Lastly, the real-time analysis method utilized by Chen et al. [32] to monitor water levels using CCTV footage and grid-based analysis obtained accuracy rates ranging from 83.6% to 96%. These results, while practical for live monitoring, highlight the challenge of achieving high accuracy across all weather conditions—a challenge that our model addresses with its robust ensemble learning method.

When compared to the related works as in Table III, our model not only sets a new high standard for accuracy but also demonstrates the potential of ensemble learning methods in significantly reducing the number of errors in water level detection. This indicates a promising future for employing such sophisticated machine learning strategies in critical environmental monitoring and resource management applications.

TABLE III. COMPARISON OF WATER LEVEL DETECTION METHODS

Research	Model Type	Accuracy
Xu et al. [37]	CNN-LSTM	> 99.5
Dou et al. [38]	CNN	94.6%
Cai et al. [30]	Hybrid Algorithm	89.29%
Chen et al. [32]	Real-Time CCTV	83.6-96%
Our Proposed Model	**Ensemble Learning**	**99.5833%**

V. CONCLUSION

Water level monitoring remains an indispensable aspect of environmental management, essential for sustaining bio-diversity, agriculture, and human habitation [58][59]. The proposed model represents a significant advancement in the automated classification of river images, providing an efficient and ac- curate means of assessing water levels. It hinges on a pre- processing regimen that ensures image uniformity and enhances the dataset, followed by leveraging the sophisticated feature extraction capabilities of pre-trained networks such as InceptionV3, VGG16, Xception, MobileNet, and ResNet152 through transfer learning. The introduction of an ensemble voting system, particularly hard voting, further refines the model's accuracy by amalgamating the predictive prowess of individual models into a cohesive decision-making process. The ensemble model achieves a stellar accuracy of 99.5833%, precision of 99.5929%, recall of 99.5762%, and an F1 score of 99.5838%, with only 5 errors noted across the evaluations. These figures represent a benchmark in water level detection, demonstrating the model's superior capability to identify and classify water levels with high reliability. Such a system can be instrumental in early flood warning mechanisms, contributing to disaster preparedness and mitigation efforts.

Future work will aim to boost the model's accuracy by including datasets with a wider variety of hydrological conditions and implementing the model for real-time data analysis. Upcoming research will also test cutting-edge neural network designs and learning methods to refine the model further. Additionally, the model's applicability to water quality as- assessment and ecosystem monitoring will be explored [60-63]. Efforts to enhance the model's computational efficiency for use in resource-constrained environments are also planned, extending the benefits of this technology to more isolated regions.

In addressing the limitations inherent in our approach, it is crucial to recognize that while the model showcases exemplary performance metrics, these results are derived under controlled conditions that may not fully replicate the dynamic complexities of natural environments. The generalizability of the model across different geographical regions and varied river conditions remains an area for further validation. Moreover, the interpretability of the model's decisions, crucial for ensuring trust and transparency, especially in critical applications, needs more thorough exploration. The Grad-CAM visualization, although useful, provides only a surface-level insight into model reasoning, and deeper analysis is necessary to fully understand the underlying decision processes. Such enhancements in model transparency and understanding will not only foster greater confidence in the predictions but also enable stakeholders to make more informed decisions in managing water resources and responding to environmental challenges. This discussion points towards a continuous cycle of evaluation and refinement, underscoring the need for ongoing research and development to maintain and improve the efficacy and reliability of water level detection systems.

REFERENCES

[1] F. Lin, Z. Yu, Q. Jin, and A. You, "Semantic segmentation and scale recognition–based water-level monitoring algorithm," *Journal of CoastalResearch*, vol. 105, pp. 185–189, 2020.

[2] S.-W. Lo, J.-H. Wu, F.-P. Lin, and C.-H. Hsu, "Visual sensing for urbanflood monitoring," *Sensors*, vol. 15, no. 8, pp. 20 006–20 029, 2015.

[3] U. Iqbal, P. Perez, W. Li, and J. Barthelemy, "How computer vision canfacilitate flood management: A systematic review," *International Journalof Disaster Risk Reduction*, vol. 53, p. 102030, 2021.

[4] C. X. Hui, G. Dan, S. Alamri, and D. Toghraie, "Greening smart cities: An investigation of the integration of urban natural resources and smart city technologies for promoting environmental sustainability,"*Sustainable Cities and Society*, vol. 99, p. 104985, 2023.

[5] M. I. H. Z. Farouk, Z. Jamil, and M. F. A. Latip, "Towards online surface water quality monitoring technology: A review," *EnvironmentalResearch*, p. 117147, 2023.

[6] S. S. Al-Hasnawi, "Water quality index of tanjero river basin near sulaymania city," *Al-Mustansiriyah J. Sci. Res*, vol. 23, pp. 193–200, 2012.

[7] S. H. Halos and S. Mahdi, "Effect of climate change on spring massive sand/dust storms in iraq," *Al-Mustansiriyah Journal of Science*, vol. 32, no. 4, pp. 13–20, 2021.

[8] S. H. Ewaid, T. D. Hussein, and F. K. Emran, "Fuzzy logic inference index to assess the water quality of tigris river within baghdad city," *Al-Mustansiriyah Journal of Science*, vol. 29, no. 3, pp. 16–20, 2018.

[9] C. Sweetapple, J. Webber, A. Hastings, and P. Melville-Shreeve, "Realising smarter stormwater management: A review of the barriers and a roadmap for real world application," *Water Research*, p. 120505, 2023.

[10] G. Zheng and H. Zong, "High accuracy surface perceiving water level gauge with self calibration," in *2009 International Conference on Mechatronics and Automation*, pp. 3680–3686, 2009.

[11] D. A. Segovia-Cardozo, L. Rodr´ıguez-Sinobas, F. Canales-Ide, and S. Zubelzu, "Design and field implementation of a low-cost, open-hardware platform for hydrological monitoring," *Water*, vol. 13, no. 21, p. 3099, 2021.

[12] T. S. R. Pereira, T. P. de Carvalho, T. A. Mendes, and K. T. M. Formiga, "Evaluation of water level in flowing channels using ultrasonic sensors,"*Sustainability*, vol. 14, no. 9, p. 5512, 2022.

[13] D. E. Alsdorf, E. Rodr´ıguez, and D. P. Lettenmaier, "Measuring surfacewater from space," *Reviews of Geophysics*, vol. 45, no. 2, 2007.

[14] G. R. Brakenridge, S. V. Nghiem, E. Anderson, and S. Chien, "Space-based measurement of river runoff," *Eos, Transactions American Geophysical Union*, vol. 86, no. 19, pp. 185–188, 2005.

[15] Z. Zhang, Y. Zhou, H. Liu, L. Zhang, and H. Wang, "Visual measurement of water level under complex illumination conditions," *Sensors*, vol. 19, no. 19, p. 4141, 2019.

[16] D. Scherer, C. Schwatke, D. Dettmering, and F. Seitz, "Icesat-2 based river surface slope and its impact on water level time series from satellite altimetry," *Water Resources Research*, vol. 58, no. 11, p. e2022WR032842, 2022.

[17] L.-C. Kuo and C.-C. Tai, "Automatic water-level measurement system for confined-space applications," *Review of Scientific Instruments*, vol. 92, no. 8, 2021.

[18] A. Royem, C. Mui, D. Fuka, and M. Walter, "Proposing a low-tech, affordable, accurate stream stage monitoring system," *Transactions of the ASABE*, vol. 55, no. 6, pp. 2237–2242, 2012.

[19] S. Etter, B. Strobl, I. van Meerveld, and J. Seibert, "Quality and timing of crowd-based water level class observations," *Hydrological processes*,vol. 34, no. 22, pp. 4365–4378, 2020.

[20] Z. Zhang, Y. Zhou, H. Liu, and H. Gao, "In-situ water level measurement using nir-imaging video camera," *Flow Measurement and Instrumenta- tion*, vol. 67, pp. 95–106, 2019.

[21] G. Chen, K. Bai, Z. Lin, X. Liao, S. Liu, Z. Lin, Q. Zhang, and X. Jia, "Method on water level ruler reading recognition based on imageprocessing," *Signal, Image and Video Processing*, vol. 15, pp. 33–41, 2021.

[22] J. Jamaludin, K. N. Z. Ariffin, and W. M. Wan Ahmad Kamil, "Advancements in monitoring water quality based on various sensing methods: a systematic review," *International Journal of Environmental Research and Public Health*, vol. 19, no. 21, p. 14080, 2022.

[23] Y. Kim, H. Park, C. Lee, D. Kim, and M. Seo, "Development of a cloud-based image water level gauge," *IT Converg. Pract. (INPRA)*, vol. 2 , no. 1, pp. 22–29, 2014.

[24] K. Hiroi and N. Kawaguchi, "Floodeye: Real-time flash flood prediction system for urban complex water flow," in *2016 IEEE sensors*, pp. 1–3, 2016.

[25] L. Sabbatini, L. Palma, A. Belli, F. Sini, and P. Pierleoni, “A computer vision system for staff gauge in river flood monitoring,” *Inventions*, vol. 6, no. 4, p. 79, 2021.

[26] R. Narayanan, V. Lekshmy, S. Rao, and K. Sasidhar, “A novel approachto urban flood monitoring using computer vision,” in *Fifth International Conference on Computing, Communications and Networking Technolo- gies (ICCCNT)*, pp. 1–7, 2014.

[27] S. Ansari, C. Rennie, E. Jamieson, O. Seidou, and S. Clark, “Rivqnet: Deep learning based river discharge estimation using close-range water surface imagery,” *Water Resources Research*, vol. 59, no. 2, p. e2021WR031841, 2023.

[28] L. Kuswidiyanto *et al.*, “Automatic water level monitoring system based on computer vision technology for supporting the irrigation modernization,” in *IOP Conference Series: Earth and Environmental Science*, vol. 686, no. 1, p. 012055, 2021.

[29] J. Pan, Y. Yin, J. Xiong, W. Luo, G. Gui, and H. Sari, “Deep learning-based unmanned surveillance systems for observing water levels,” *IEEE Access*, vol. 6, pp. 73 561–73 571, 2018.

[30] Z. Cai *et al*., “Automatic monitoring alarm method of dammed lake based on hybrid segmentation algorithm,” *Sensors*, vol. 23, no. 10, p. 4714, 2023.

[31] G. Bai, J. Hou, Y. Zhang, B. Li, H. Han, T. Wang, R. Hinkelmann, D. Zhang, and L. Guo, “An intelligent water level monitoring method based on ssd algorithm,” *Measurement*, vol. 185, p. 110047, 2021.

[32] J.-F. Chen, Y.-T. Liao, and P.-C. Wang, “Development and deployment of a virtual water gauge system utilizing the resnet-50 convolutional neural network for real-time river water level monitoring: A case study of the keelung river in taiwan,” *Water*, vol. 16, no. 1, p. 158, 2024.

[33] F. Yang, T. Feng, G. Xu, and Y. Chen, “Applied method for water-bodysegmentation based on mask r-cnn,” *Journal of Applied Remote Sensing*, vol. 14, no. 1, 2020.

[34] J. B. Haurum, C. H. Bahnsen, M. Pedersen, and T. B. Moeslund, “Water level estimation in sewer pipes using deep convolutional neural networks,” *Water*, vol. 12, no. 12, p. 3412, 2020.

[35] S.-S. Baek, J. Pyo, and J. A. Chun, “Prediction of water level andwater quality using a cnn-lstm combined deep learning approach,” *Water*, vol. 12, no. 12, p. 3399, 2020.

[36] G. Qiao, M. Yang, and H. Wang, “A water level measurement approach based on yolov5s,” *Sensors*, vol. 22, no. 10, p. 3714, 2022.

[37] Y. Xu, C. He, Z. Guo, Y. Chen, Y. Sun, and Y. Dong, “Simulation of water level and flow of catastrophic flood based on the cnn-lstm couplingnetwork,” *Water*, vol. 15, no. 13, p. 2329, 2023.

[38] G. Dou, R. Chen, C. Han, Z. Liu, and J. Liu, “Research on water-level recognition method based on image processing and convolutional neuralnetworks,” *Water*, vol. 14, no. 12, p. 1890, 2022.

[39] C.-H. Tsai. *Real-time images of river in Taiwan*. Harvard Dataverse, V1, 2020, https://doi.org/10.7910/DVN/8FDC7P.

[40] A. P. Joshi and B. V. Patel, “Data preprocessing: the techniques for preparing clean and quality data for data analytics process,” *Orient. J. Comput. Sci. Technol*, vol. 13, pp. 78-81, 2021.

[41] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” *Applied Soft Computing*, vol. 97, p. 105524, 2020.

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

[43] B. Vrigazova, “The proportion for splitting data into training and test set for the bootstrap in classification problems,” *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, vol. 12, no. 1, pp. 228-242, 2021.

[44] F. Zhuang *et al*., “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43-76, 2020.

[45] C. Szegedy *et al*., “Going Deeper with Convolutions,” *arXiv preprint arXiv: 1409.4842*, 2014.

[46] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv preprint arXiv: 1409.1556*, 2014.

[47] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” *arXiv preprint arXiv: 1610.02357*, 2016.

[48] A. G. Howard *et al*., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *arXiv preprint arXiv: 1704.04861*, 2017.

[49] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *arXiv preprint arXiv: 1512.03385*, 2015.

[50] M. Atif, F. Anwer, and F. Talib, “An ensemble learning approach for effective prediction of diabetes mellitus using hard voting classifier,” *Indian Journal Of Science And Technology*, vol. 15, no. 39, pp. 1978-1986, 2022.

[51] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” *arXiv preprint arXiv: 1610.02391*, 2016.

[52] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: analysis, applications, and prospects,” *IEEE transactions on neural networks and learning systems*, vo. 33, no. 12, pp. 6999-7019, 2021.

[53] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization Techniques in Training DNNs: Methodology, Analysis and Application," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 10173-10196, Aug. 2023.

[54] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, 2022.

[55] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu, “Explainable AI: A brief survey on history, research areas, approaches and challenges,” in *Natural language processing and Chinese computing: 8th cCF international conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II 8*, pp. 563-574, 2019.

[56] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A Survey on Performance Metrics for Object-Detection Algorithms," *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237-242, 2020.

[57] Ž. Vujović, “Classification model evaluation metrics,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599-606, 2021.

[58] C. Z. Zulkifli *et al*., “IoT-based water monitoring systems: a systematic review,” *Water*, vol. 14, no. 22, p. 3621, 2022.

[59] S. Kossieris, V. Tsiakos, G. Tsimiklis, and A. Amditis, “Inland Water Level Monitoring from Satellite Observations: A Scoping Review of Current Advances and Future Opportunities,” *Remote Sensing*, vol. 16, no. 7, p. 1181, 2024.

[60] J. O. Ighalo and A. G. Adeniyi, “A comprehensive review of water quality monitoring and assessment in Nigeria,” *Chemosphere*, vol. 260, p. 127569, 2020.

[61] M. H. Al-Tai, B. M. Nema, and A. Al-Sherbaz, "Deep learning for fake news detection: Literature review," *Al-Mustansiriyah Journal of Science*, vol. 34, no. 2, pp. 70-81, 2023.

[62] S. Ali, J. H. Al’Ameri, and T. Abbas, "Real Time E-learning Students Monitoring for Optimization Facial Landmark Recognition Based on Hybrid Deep Learning Techniques," *Full Length Article*, vol. 10, no. 2, 2023.

[63] N. M. Khassaf and S. H. Shaker, "Image Retrieval based Convolutional Neural Network," *Al-Mustansiriyah Journal of Science*, vol. 31, no. 4, pp. 43-54, 2020.