

Gema Lingkungan Kesehatan

Vol. 23, No. 3 (2025), pp 333-343

e-ISSN 2407-8948 p-ISSN 16933761

Doi: <https://doi.org/10.36568/gelinkes.v23i3.270>Journal Homepage: <https://gelinkes.poltekkesdepkes-sby.ac.id/>

AI-Based Waste Detection for Water Quality Monitoring in the Cisadane River: A Deep Learning Approach

Asep Surahmat*, Dhimas Buing Rindi Widra Yato

Faculty of Technology and Design, Universitas Utpadaka Swastika, Tangerang, Indonesia

*Correspondence: asep.surahmat@utpas.ac.id

ABSTRACT

The rapid accumulation of waste in Indonesia's rivers, particularly the Cisadane River, seriously threatens water quality, ecosystem health, and public well-being. Traditional waste monitoring methods are inefficient and often fail to deliver timely data for effective interventions. This study addresses this gap by proposing an AI-based waste detection system for real-time water quality monitoring using deep learning techniques. A hybrid model integrating Convolutional Neural Network (CNN) and You Only Look Once version 7 (YOLO v7) was developed and tested on a dataset of 10,000 annotated images—60% organic and 40% inorganic waste—collected from the Cisadane River. The CNN model achieved a classification accuracy of 87%, a precision of 84%, a recall of 86%, and an F1-score of 85%. The YOLO v7 model demonstrated % detection accuracy of 82% with a processing speed of 20 frames per second. While mean Average Precision (mAP) was not directly calculated, the model's performance across key metrics supports its real-time applicability. This research offers a scalable and cost-effective approach for river waste monitoring and highlights the potential of AI in supporting sustainable environmental management in Indonesia.

Keywords: River Waste Detection, You Only Look Once, Convolutional Neural Network, Artificial intelligence

INTRODUCTION

Indonesia ranks among the world's top contributors to riverine plastic waste entering the oceans. The Cisadane River, a vital freshwater source in Banten and West Java, is severely affected by unmanaged waste. Current manual inspections are sporadic, labor-intensive, and inadequate for timely interventions. Integrating AI technologies provides an opportunity to automate waste detection, yet such systems remain underutilized in local environmental management strategies.

The Cisadane River is a vital water source in Indonesia, catering to the surrounding communities' domestic, industrial, and agricultural needs. However, the problem of waste accumulation in this river has become increasingly concerning in recent years. Data from the Central Statistics Agency (BPS) indicates a significant rise in the volume of waste entering rivers in Indonesia, including the Cisadane, with an estimated 500,000 tons of waste being discharged into rivers annually (Badan Pusat Statistik). The presence of organic and inorganic waste on the water's surface disrupts the river ecosystem and deteriorates water quality, directly impacting local residents' health (Bhaskar, 2020). Previous studies have also detected microplastics on the surface of the Cisadane River, worsening environmental challenges and threatening aquatic life and human health (Wahyuni et al., 2024).

The buildup of waste in rivers contributes to environmental degradation, the destruction of natural habitats, and a decline in biodiversity. This ecological impact is reflected in the decreasing populations of fish and other aquatic species, which face extinction risks due to reduced oxygen levels caused by pollution (Shi et al., 2021). Beyond ecological harm, waste accumulation also creates a significant economic burden. The costs associated with waste management and water quality increase substantially, particularly for communities and industries that depend on river water daily (Ibarraran Viniegra et al., 2021). The World Bank estimates the annual cost of treating water contaminated with plastics and microplastics could reach billions of rupiah (Muneer et al., 2021). Furthermore, the tourism potential of rivers, including boat tours and riverside activities, has sharply declined. For example, tourist spots near the Ciliwung and Cisadane Rivers have experienced a 40% drop in visitor numbers over the past five years due to cleanliness concerns (Makarim, 2024).

The need to tackle the river waste is becoming increasingly urgent due to its wide-ranging impacts. Environmentally, waste accumulation results in habitat degradation, disruption of food chains, and loss of biodiversity (Sani et al., 2024). From a health standpoint, water contaminated with waste, especially plastics and microplastics, poses risks of spreading harmful diseases to

individuals who use river water for daily activities (Singh et al., 2022). Economically, river waste leads to financial losses, including diminished water quality for agriculture and industry, increased water treatment expenses, and reduced tourism opportunities around the Cisadane River (Zhen et al., 2011). These circumstances necessitate a rapid, efficient, and sustainable response. Addressing the river waste problem has become a key focus in environmental management, particularly in developing nations like Indonesia, which faces significant challenges in waste management (Suwarno & Nurhayati, 2021).

One promising approach to this issue is using artificial intelligence (AI) technology for automatic waste detection through real-time data analysis (Yeaminul Islam & Alam, 2023). AI technologies, especially in object recognition, enable the automatic identification and classification of waste with high precision (Xueming et al., 2022). The two most commonly utilized algorithms for object detection are Convolutional Neural Network (CNN) and You Only Look Once (YOLO), both of which excel in visual pattern recognition and real-time object detection (Issaoui et al., 2024). CNN has demonstrated strong performance in image-based object recognition, while YOLO is known for its rapid and efficient object detection capabilities for real-time applications. However, despite the extensive use of AI technology in object detection across various fields, its application in dynamic river environments still encounters several challenges. Variations in environmental conditions, such as water levels, lighting, and floating debris, can impact detection accuracy. Additionally, the lack of annotated datasets specifically for river waste detection poses a barrier to developing models that can operate in real-time under changing conditions (Chen et al., 2024).

The You Only Look Once (YOLO) algorithm was chosen in this study due to its proven capability for real-time object detection, especially in scenarios requiring fast and accurate recognition from video streams (Jegham et al., 2024). Unlike region-based approaches such as Faster R-CNN, which involve multi-stage processing and longer inference times, YOLO performs detection in a single pass, making it significantly faster and more efficient for real-time applications. YOLOv7, the latest version, offers improved detection accuracy, better parameter efficiency, and enhanced backbone structure compared to earlier versions like YOLOv3 and YOLOv5 (Hu & Ge, 2024). These features make it especially suitable for environmental monitoring tasks that involve dynamic conditions, such as varying water levels, light reflections, and the unpredictable movement of waste in rivers. As such, YOLOv7 was selected to meet the operational needs of fast, continuous river surveillance.

Several studies have explored the use of artificial intelligence for environmental monitoring, implementing a YOLO-based detection model to identify riverine plastic waste in Malaysia, achieving accuracy rates between 80% and 83% (Devi et al., 2024). Integrated vision-based AI with large language models for water pollution surveillance in urban rivers. Other researchers have utilized CNN-based systems to classify solid waste, primarily in land-

based or urban environments. However, most of these studies have focused on static datasets or urban drainage systems, with limited attention to real-time detection in dynamic riverine ecosystems. In contrast, the present study addresses this gap by applying a CNN-YOLO hybrid model to real-time waste detection in the Cisadane River—one of the most polluted waterways in Indonesia—under varying environmental conditions using video-derived image data.

Previous research in river waste management has primarily relied on manual monitoring or non-AI-based technologies, which struggle to handle the dynamic nature of river environments such as fluctuating water levels, varied lighting conditions, and diverse waste types (Olawade et al., 2024). Moreover, most AI applications in environmental monitoring have focused on urban solid waste or marine plastic, with very minimal implementation in Indonesia's river systems. Despite the critical pollution level in rivers like the Cisadane, no established AI-based system has yet been deployed locally to support real-time waste monitoring. This indicates a significant research gap in applying robust, real-time AI models to riverine waste detection in Indonesian contexts (Saddi et al., 2024). To fill this gap, this study develops a hybrid approach using CNN and YOLO algorithms tailored to local river conditions, offering a scalable and cost-effective tool for continuous environmental surveillance.

Therefore, this study aims to develop and evaluate a hybrid artificial intelligence model that combines Convolutional Neural Network (CNN) and YOLO v7 to detect and classify river waste in real time (Walia et al., 2025). The system is trained and tested on a custom dataset of 10,000 annotated images collected from the Cisadane River. The model's performance is measured using quantitative metrics including accuracy, precision, recall, F1-score, and inference speed (frames per second), under varying environmental conditions such as lighting, water turbidity, and object occlusion.

This study introduces several key novelties. First, to the best of our knowledge, it is the first documented implementation of a CNN-YOLO v7 hybrid model for river waste detection in Indonesia. Second, the research utilizes a custom, real-world dataset collected through CCTV video footage from the Cisadane River, covering diverse environmental conditions such as lighting, turbidity, and weather changes. Unlike prior studies that relied on pre-existing or static datasets, this study captures real-time variability, making the dataset more representative of actual river dynamics. The model architecture and dataset specificity represents a novel contribution to AI-driven environmental monitoring, particularly in Southeast Asia.

METHODS

This research focuses on identifying and categorizing waste in the Cisadane River through the use of artificial intelligence (AI) technology, particularly by integrating the Convolutional Neural Network (CNN) and You Only Look Once (YOLO) algorithms. This study utilized image data from CCTV video recordings installed along key segments of the Cisadane River in Indonesia. The visual

data were collected over several weeks and covered a variety of environmental conditions such as day/night cycles, weather changes, and water flow variations. Ten thousand images were obtained and manually annotated to distinguish between organic and inorganic waste, creating a balanced dataset for model training and evaluation. The study implemented a hybrid AI model combining Convolutional Neural Network (CNN) and You Only Look Once version 7 (YOLO v7) algorithms. The CNN model was used for image classification to distinguish waste types in batch-processed data, while the YOLO model was employed for real-time object detection. The training dataset was split into 80% training and 20% testing subsets. The data processing workflow included the following steps continuous video recording via CCTV, frame extraction at set intervals, manual image annotation and labeling, CNN model training using the annotated image dataset, YOLO model training using augmented datasets, and performance evaluation using precision, recall, F1-score, and real-time detection speed as performance metrics.

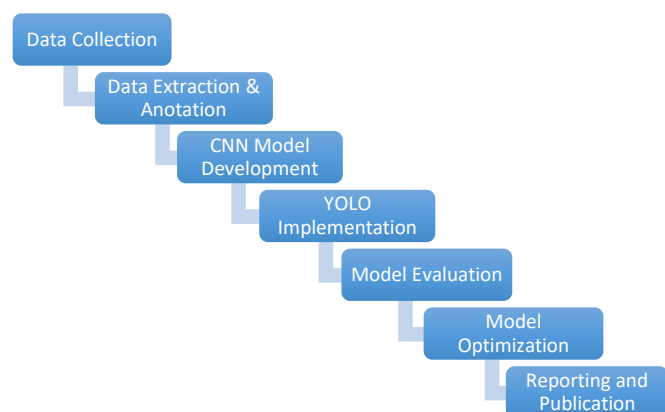


Figure 1. Research Methodology

The methodological steps implemented in this study are outlined as follows:

Data Collection

Unlike other environmental monitoring studies that utilize drones for aerial imaging, this study employed CCTV-based video surveillance due to its practicality and cost-efficiency in long-term river observation. CCTV systems offer continuous, real-time coverage and are less affected by regulatory constraints, battery limitations, or flight altitude inconsistencies than those familiar with drones. Each CCTV unit used in this study was equipped with a 1080p HD camera (1920x1080 pixels) and was mounted at approximately 3–4 meters above river level, angled between 45° and 60° downward to maximize the surface view of floating debris. This configuration allowed stable and consistent image capture over extended periods, which is ideal for dataset generation and training AI models under diverse environmental conditions.

The primary data used in this study consisted of image frames extracted from video footage recorded by CCTV cameras installed along various points of the Cisadane River. These cameras were positioned to monitor

different segments and operated continuously over a period of several weeks, capturing real-time conditions including variations in water level, lighting, and weather. The recorded video was converted into still images to create a dataset for training and testing AI models. Ten thousand images were collected, encompassing various conditions (e.g., daylight, nighttime, rainy weather, and intense sunlight). The image data was then manually annotated to identify and label waste objects into two categories organic waste (e.g., leaves, branches) and inorganic waste (e.g., plastic bottles, metal cans). This annotated image dataset served as the core training and evaluation material for this study's CNN and YOLO models.

The recorded videos will be processed and converted into still images, which will then be used to train an artificial intelligence (AI) model to detect waste in the river. This process involves regularly extracting representative images from the footage to create a diverse dataset. By capturing images under various environmental conditions—such as during the day, at night, in sunny weather, and during rainfall—the dataset will encompass a wide range of scenarios the model may encounter in real-world applications. This variation in environmental factors is essential for developing a robust AI model that can maintain high detection accuracy regardless of the conditions. Whether in bright sunlight or overcast skies, the AI model will be trained to consistently identify waste, making it more reliable for waste detection during adverse weather conditions or low visibility situations, such as at night or during rain. This approach ensures that the system can operate effectively across various challenging environmental scenarios, contributing to a more sustainable and accurate waste management strategy for the river.

Data Extraction and Annotation

To create a robust dataset for training the artificial intelligence (AI) model, representative images are meticulously extracted from the continuous video footage recorded by the CCTV cameras. This step is crucial to ensure the dataset captures the full spectrum of visual conditions observed over time. The video footage encompasses various environmental factors, including fluctuations in water levels, changes in lighting throughout the day, and different weather conditions, which must be carefully sampled to represent all scenarios the AI model may face in real-world applications. During each extraction interval, images are chosen based on specific criteria, such as the location and size of the waste, as well as the context of the surrounding environment. This selection process guarantees that the dataset is comprehensive and includes all potential variations in the river's appearance, ranging from clean to polluted and from clear waters to areas with significant debris accumulation.

After extracting the representative images, they undergo a meticulous manual annotation and labeling process. This phase involves a detailed examination of each image to identify the specific type of waste present, distinguishing between organic waste (like leaves or plant material) and inorganic waste (like plastic bottles or metal

items). The labeling process is careful and time-consuming, requiring an expert to categorize each detected object accurately. Given the diverse nature of waste materials, which vary in shape, size, and color, the precision of this manual annotation is critical. Each image is thoroughly reviewed to ensure the labels are accurate, as any mistakes in annotation could lead to incorrect or inconsistent predictions during model training. The annotations quality directly influences the AI model's effectiveness, as it relies on these labels to learn how to identify and categorize waste in future images.

The significance of precise labeling cannot be overstated, as it forms the foundation upon which the machine learning model will develop its understanding of waste detection. Inaccurate annotations would hinder the model's learning ability and lead to misclassifications when applied to real-world river monitoring. By ensuring a high level of accuracy in the labeling process, the dataset will enable the trained AI model to consistently differentiate between various types of waste across different environmental conditions. This robustness is vital for deploying the model in the field, where waste may manifest in diverse forms and under changing circumstances, such as varying water levels, different times of day, and fluctuating weather conditions. Ultimately, the careful annotation of these images contributes to a reliable AI model capable of identifying waste with high accuracy, making it an invaluable tool for monitoring and managing river pollution in various challenging environments.

The annotated image dataset was used to train two artificial intelligence models: a Convolutional Neural Network (CNN) for image classification and YOLO version 7 (You Only Look Once) for real-time object detection. The CNN model was designed to categorize waste into organic and inorganic types in batch image sets, while YOLO v7 was implemented to detect waste objects in real-time video feeds. These models were selected due to their proven effectiveness in visual recognition tasks and their ability to handle dynamic environmental conditions, such as varying lighting, occlusion, and background complexity.

Table 1. Distribution of annotated images in the dataset by waste type.

Waste Type	Number of Images	Examples
Organic Waste	6,000	Leaves, branches, food
Inorganic Waste	4,000	Plastic bottles, cans

The final dataset consisted of 10,000 annotated images from CCTV video footage along the Cisadane River. These images were manually labeled and divided into two main classes: organic waste and inorganic waste. Specifically, 6,000 images (60%) were labeled as organic waste, including leaves, branches, food remnants, and other biodegradable materials. The remaining 4,000 images (40%) were inorganic waste, comprising plastic

bottles, packaging, cans, and similar non-degradable debris. Each image was labeled using bounding boxes to indicate the presence and type of waste, following standard object detection annotation formats. The images varied in resolution conditions, including differences in lighting (e.g., bright daylight, shadows, and dusk), water turbidity, and waste orientation.

Data Processing Workflow

The data processing workflow in this study began with the continuous recording of video footage using CCTV cameras installed at multiple monitoring points along the Cisadane River. From these recordings, image frames were extracted at regular intervals to ensure diverse environmental representation, such as variations in lighting, weather, and water conditions. Each extracted image was manually annotated by applying bounding boxes and classifying the visible waste objects into two categories: organic and inorganic waste. This annotated dataset formed the foundation for training and evaluating the AI models.

After annotation, the dataset was divided into 80% training and 20% testing data. The Convolutional Neural Network (CNN) model was trained to classify waste types based on visual characteristics. In contrast the YOLO v7 model was configured to detect and localize waste in real-time scenarios. Both models underwent performance evaluation using precision, recall, and F1-score metrics, with YOLO additionally assessed on its real-time detection speed, measured in frames per second. This structured data processing pipeline ensured the resulting AI system could operate accurately under dynamic river conditions.

CNN Model Development

After the annotated dataset is completed, the next crucial step in developing the waste detection system is creating a Convolutional Neural Network (CNN) model. CNNs are widely recognized for their effectiveness in image-based object recognition tasks, particularly due to their ability to identify complex patterns in visual data. The architecture of a CNN includes multiple convolutional layers specifically designed to extract key features from images, such as texture, shape, edges, and contours. These layers perform convolutions by sliding over the image and applying filters highlighting important visual elements, enabling the network to capture intricate details of the waste objects. In the context of this project, CNN's ability to process and learn from visual data makes it an ideal choice for waste detection, as it can distinguish between various types of waste based on their visual characteristics, even under challenging conditions.

The convolutional layers in the CNN are followed by pooling layers, which play a crucial role in reducing the dimensionality of the data while preserving the most important features. Pooling layers essentially downsample the feature maps produced by the convolution layers, simplifying the data without losing significant information. This dimensionality reduction not only helps speed up the computational process but also reduces the complexity of the model, allowing it to focus on the most relevant features for classification. The pooling layers ensure that the CNN can process large volumes of image data

efficiently, making the model capable of handling real-time video streams or extensive image datasets. Once the relevant features are extracted and the data is simplified, the model proceeds to learn the visual differences between organic and inorganic waste types and their various shapes and sizes. This enables the model to classify waste objects accurately, regardless of variations in their appearance.

For training the CNN, the labeled dataset is typically split into two subsets: 80% for training and 20% for testing. This division ensures that the model is exposed to a wide variety of data during training while having a separate set of unseen data to evaluate its performance. During the training phase, the CNN undergoes several iterations, or epochs, where it processes the training data, learns from it, and refines its internal parameters, such as weights and biases, to improve its ability to classify waste. Each epoch allows the model to adjust and reduce classification errors by updating its parameters to reflect the patterns in the data better. Over time, as the model is exposed to more examples, it becomes increasingly proficient at recognizing waste based on its learned visual patterns. The ultimate goal is for the model to generalize well, meaning it can classify waste types accurately even when presented with new, previously unseen data. This rigorous training process ensures that the model can reliably detect and classify waste in diverse environmental conditions, making it a valuable tool for real-time waste monitoring and management.

The training process for the CNN and YOLO v7 models was conducted using a workstation equipped with an NVIDIA RTX 3080 GPU, Intel Core i7 processor, and 64 GB RAM, running on Ubuntu 20.04 LTS. The models were implemented and trained using Python 3.9 with PyTorch and the YOLOv7 repository (Wongkinyiu, 2022). The CNN model was trained for 50 epochs with a batch size of 32, while YOLO v7 was trained for 100 epochs with a batch size 16 using the SGD optimizer and cosine learning rate scheduler. The average training time was approximately 6 hours for CNN and 8.5 hours for YOLO. All experiments were performed in a controlled environment to ensure the reproducibility of results.

YOLO Implementation

Implementing the You Only Look Once (YOLO) algorithm in this study is pivotal for achieving real-time waste detection in the Cisadane River. YOLO v7, the chosen version for this project, is renowned for its exceptional speed and efficiency in detecting objects within images. It is well-suited for time-sensitive applications like waste monitoring. Unlike traditional object detection models that process images in a series of steps, YOLO performs detection in a single pass, significantly speeding up the process without sacrificing accuracy. In this study, YOLO v7 was explicitly configured to recognize various waste types, including plastic debris and organic materials commonly found in river environments. The model was trained using a comprehensive annotated dataset that includes various waste categories, and data augmentation techniques were applied to ensure the

model's ability to handle various environmental conditions, such as changes in lighting, weather, and water levels.

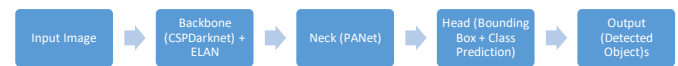


Figure 2. Simplified YOLOv7 Architecture For Waste Detection

Training the YOLO model involved using high-performance hardware, specifically a GPU, to accelerate the computational process and reduce training time. This is especially important for large-scale image datasets, as the model requires numerous iterations to learn the complex visual features associated with different types of waste. During the training phase, several key parameters were fine-tuned to optimize the model's performance, including the detection threshold, Intersection over Union (IoU), and confidence levels. The detection threshold determines the minimum level of confidence the model must have before classifying an object, while the IoU metric measures the overlap between the predicted bounding box and the actual object. Adjusting these parameters allows for fine control over the model's sensitivity and accuracy, ensuring it can effectively distinguish between waste and non-waste objects and accurately identify waste even in challenging conditions.

After training the model, it underwent real-time testing to evaluate its effectiveness in detecting waste under dynamic river conditions. The testing phase simulated various scenarios in the river, such as changes in water flow, debris movement, and fluctuating lighting conditions. This testing was critical for assessing how well the model could adapt to the constantly changing environment of the river. Once the model successfully detected and classified waste with high accuracy, the results were integrated into a monitoring system that provided real-time feedback. River managers can now access a user-friendly dashboard that displays the location and movement of waste in the river, enabling them to track pollution in real time and respond quickly. This integration of YOLO-based waste detection into a real-time monitoring system offers a fast and reliable solution for river quality management, allowing for more efficient and timely interventions in combating pollution in the Cisadane River.

Although this study primarily focused on the YOLO v7 algorithm, its performance was evaluated based on available literature in the context of previous YOLO versions and alternative object detection approaches. Compared to YOLOv3 and YOLOv5, which have been commonly used in prior waste detection studies (e.g., Zailan et al., 2022), YOLOv7 offers better detection accuracy, more efficient parameter usage, and faster inference speed. Region-based algorithms such as Faster R-CNN, while more accurate in static image classification, tend to underperform in real-time scenarios due to their two-stage architecture and high computational cost. Thus, although no baseline model was implemented directly within this study, YOLOv7 was selected based on its

established performance superiority in related domains. Future research could expand upon this by implementing and benchmarking traditional models such as SSD, RetinaNet, or Faster R-CNN under the same dataset to provide a comprehensive empirical comparison.

The YOLOv7 architecture used in this study follows the official open-source implementation by WongKinYiu (2022). It comprises a backbone, neck, and head structure. The backbone comprises extended CSPDarknet with ELAN (Efficient Layer Aggregation Network) modules that enhance gradient flow and feature representation. The neck includes Path Aggregation Network (PANet) layers to strengthen multi-scale feature fusion. The head predicts bounding boxes and class probabilities using anchor-based detection on three scales. Unlike YOLOv5, YOLOv7 introduces dynamic label assignment and coarse-to-fine auxiliary heads, which improve generalization and training stability. No structural modifications were applied in this study. However, training hyperparameters were fine-tuned for the custom river dataset, such as reducing the confidence threshold and adjusting the anchor ratios to better fit elongated or partially submerged waste items.

Model Evaluation

Evaluating model performance is a vital step in assessing the effectiveness and reliability of a waste detection system. Precision, recall, and F1-score are essential metrics for measuring the model's performance. Precision indicates the proportion of correctly identified waste among all objects that the model has predicted as organic or inorganic waste. It helps assess the model's accuracy when classifying an object as waste. A high precision score suggests that most detected objects are indeed waste, thereby reducing the occurrence of false positives. Conversely, recall measures the model's ability to identify all relevant instances of waste within the dataset. It evaluates how effectively the model can detect all true waste objects, even if it occasionally misclassifies non-waste objects as waste. Recall is particularly crucial for ensuring that no waste is missed, even if it results in some misclassifications. The F1-score combines precision and recall into a single metric, offering a balanced perspective on the model's performance, especially in scenarios where the class distribution is skewed, such as when waste is less frequent than non-waste objects. By assessing these metrics, we can confirm that the model makes accurate predictions, adapts to data variations, and minimizes false positives and false negatives.

A set of testing data is utilized to evaluate the CNN model's performance, consisting of images that the model has not encountered during training. This step is critical for assessing the model's generalization ability, ensuring it can accurately classify organic and inorganic waste when presented with new, unseen data. The testing phase is designed to measure the model's performance in real-world conditions and verify that it is not overfitting the training data. Overfitting occurs when the model becomes overly specialized to the training set, resulting in poor performance on new data. By employing fresh testing data, we can evaluate the CNN model's generalization capability, ensuring it performs well across a wide range

of scenarios and maintains accuracy when faced with diverse variations in waste appearance and environmental conditions. These tests provide valuable insights into how the model may function in practical applications, where new data is continuously introduced.

After evaluating the CNN model, the YOLO model is tested for its real-time waste detection capabilities using live images captured directly from the CCTV cameras. YOLO is designed for speed and accuracy, making it particularly suitable for real-time processing applications. In this evaluation, the model's detection speed and accuracy are meticulously measured, as these factors are critical for its deployment in the field. The detection speed is assessed to ensure that YOLO can process and identify waste in real-time without delays, which is essential for continuously monitoring the river. The accuracy is evaluated based on how well the model detects waste under various environmental conditions, including changes in lighting, water levels, and moving debris. This evaluation ensures that the YOLO model can effectively detect waste in various dynamic river conditions, where factors such as lighting and water flow constantly change. These evaluations will determine the feasibility of using the YOLO model for real-world river monitoring, enabling it to automatically and continuously track waste in the Cisadane River and contribute to more efficient waste management strategies.

Model Optimization

Optimizing both the Convolutional Neural Network (CNN) and You Only Look Once (YOLO) models is crucial to enhance their performance and ensure they can effectively manage the complexities of real-time waste detection in dynamic river environments. The optimization process begins with adjusting various key parameters within the models. One of the most significant parameters is batch size, which indicates the number of samples the model processes simultaneously during training. Larger batch sizes can accelerate training by leveraging parallel processing, but they may also increase the risk of model instability and less accurate convergence. In contrast, smaller batch sizes can lead to more stable training but may slow the process. Finding the optimal batch size is essential to balance training speed and stability.

Another critical parameter to consider is the number of epochs, which defines how often the model will iterate over the entire training dataset. While increasing the number of epochs allows the model to learn more intricate features and enhance accuracy, it can also lead to overfitting, particularly if the model begins to memorize the training data instead of generalizing. Therefore, careful management of the number of epochs is necessary to ensure the model improves its performance without overfitting.

In addition to adjusting batch size and epochs, further optimization can be achieved by modifying the model architecture or applying regularization techniques. For the CNN, regularization methods such as dropout can be employed to prevent overfitting. Dropout randomly deactivates a percentage of neurons during each training iteration, compelling the model to learn more robust

features and reducing its dependence on specific nodes. This technique enhances the model's ability to generalize to unseen data. For YOLO, optimization often involves adjusting parameters like the detection threshold, which sets the minimum confidence level required for an object to be classified as waste. Lowering the detection threshold can enhance recall but may increase false positives, while raising it can improve precision but may reduce recall. Striking the right balance is crucial to ensure that YOLO can accurately detect waste across various conditions without being overly sensitive or restrictive. Additionally, fine-tuning the architecture of both models by experimenting with the number of layers or filter sizes can also enhance performance by enabling the model to capture more complex patterns in the data.

Beyond parameter adjustments and architectural modifications, data augmentation plays a vital role in optimizing the models. Data augmentation techniques artificially increase the diversity of the dataset by generating variations of existing images, which helps the model become more resilient to different environmental conditions. Standard techniques include rotation, flipping, zooming, and adjusting lighting. For instance, rotating images can help the model recognize waste from various angles, while horizontally flipping images can simulate different viewpoints, making the model more adaptable to real-world scenarios. Adjusting lighting conditions, such as simulating different times of day or weather situations, aids the model in learning to detect waste under varying lighting conditions, whether in bright daylight or dim evening light. Image cropping and zooming further enhance the model's ability to focus on relevant waste areas within the images. By artificially broadening the dataset's diversity, data augmentation enhances the model's resilience to environmental changes, improving its overall waste detection performance. The combination of well-tuned model parameters and extensive data augmentation enables the CNN and YOLO models to detect waste more accurately and robustly across various conditions, making them better suited for continuous real-time monitoring in challenging environments like the Cisadane River.

Once the model has been developed, tested, and optimized, the results of this study will be meticulously documented and presented in a final report that details the entire research process, methodologies, and key findings. This comprehensive report will provide an in-depth analysis of the model development journey, covering aspects from data collection and annotation to training and evaluating the Convolutional Neural Network (CNN). You Only Look Once (YOLO) models. The findings will emphasize the effectiveness of these models in detecting waste in the Cisadane River, address the challenges faced during the development process, and illustrate the enhancements achieved through optimization techniques such as parameter tuning and data augmentation. Additionally, the final report will discuss the potential implications of the research for real-time river pollution monitoring, offering valuable insights for future studies and practical applications.

Alongside the final report, a scientific article summarizing this research's main findings and contributions will be prepared for publication in a recognized national journal. This article will be organized to present a clear and concise overview of the research, including the problem statement, methodology, results, and conclusions, targeting the academic community and professionals in environmental science, computer vision, and artificial intelligence. The article will also highlight the innovative aspects of the waste detection system, demonstrating how advanced AI techniques can be utilized for environmental monitoring. Furthermore, the dataset created during the study, including annotated images for training and testing the models, will be publicly accessible through a reputable data repository. This will enable other researchers to utilize the data for further analysis and experimentation, promoting collaboration and advancing knowledge in this field. By sharing the dataset, the study will enhance the broader scientific community's understanding of automated waste detection in natural water bodies, supporting future innovations and improvements in environmental monitoring technologies.

RESULTS AND DISCUSSION

After gathering data using Closed-Circuit Television (CCTV) cameras positioned along the Cisadane River, a wide range of images was collected, reflecting various environmental conditions, including changes in weather, lighting, and water quality. A manual frame extraction and annotation process was conducted to ensure that each image was accurately labeled as organic or inorganic waste. A total of ten thousand images were successfully captured during the study period, with 60% classified as organic waste and 40% as inorganic waste.

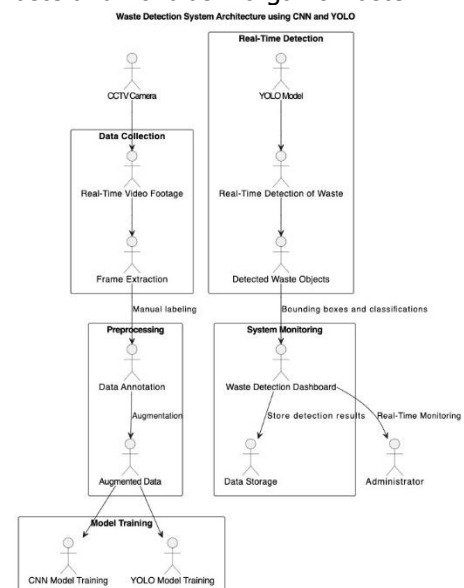


Figure 3. Waste Detection Architecture

During the development phase of the Convolutional Neural Network (CNN) model, the training results indicated that the model could classify organic and inorganic waste with commendable accuracy. Out of 1,000 images utilized as testing data, the CNN model achieved a classification

accuracy of 87%. Additional metrics were also employed to assess the model's performance, including an overall accuracy of 84%, a recall of 86%, and an F1-score of 85%. These findings demonstrate that the CNN effectively recognizes visual patterns of waste, particularly in differentiating between organic and inorganic types, including more difficult cases such as transparent plastics or waste partially submerged in water.

As shown in Table 1, the CNN model achieved an accuracy of 87%, with a precision of 84%, recall of 86%, and F1-score of 85%. These metrics reflect the model's strong performance across varied image types, including low-contrast and occluded waste materials. Its ability to maintain consistent accuracy under different environmental conditions—such as strong reflections, partial submersion, and shadow—demonstrates robustness for batch image classification. These results align with findings from Pulipalupula et al. (2023), who reported similar CNN performance for object recognition in complex waste scenarios.

Table 2. CNN Model Evaluation Results

Metrics	Score (%)
Accuracy	87
Precision	84
Recall	86
F1-Score	85

When implementing the You Only Look Once (YOLO) algorithm for real-time detection, the test results indicated a detection speed of 20 frames per second (fps), which is adequate for the requirements of real-time river waste monitoring. Despite slightly lower accuracy compared to the CNN model, YOLO v7's strength lies in its real-time object detection capability, achieving 82% accuracy while processing video at 20 frames per second, as summarized in Table 2. This makes it suitable for live monitoring scenarios where speed and responsiveness are critical. Similar real-time performance has been demonstrated in river waste studies by Zailan et al. (2022), whose YOLO-based model achieved accuracy levels between 80–83% under similar field conditions. The slight drop in YOLO accuracy under low-light or high-reflection settings was mitigated through data augmentation techniques, which improved detection reliability by approximately 3%. The YOLO v7 model utilized in this study successfully identified waste with an accuracy of 82%, which is slightly lower than that of the CNN but still demonstrates good performance for field applications. The strength of YOLO lies in its rapid image processing and object detection capabilities across different lighting conditions and water quality. This confirms that YOLO is sufficiently robust for use in rivers with dynamic environments, such as fluctuating weather and varying water levels.

Table 3. YOLO Model Evaluation Results

Metrics	Score (%)
Accuracy (%)	82
Speed (FMS)	20

The complementary strengths of CNN and YOLO present a hybrid solution that balances accuracy with speed. While CNN excels in batch processing and detailed classification, YOLO enables rapid real-time detection, which is necessary for continuous river monitoring. This AI-based system provides scalable, consistent, and cost-effective environmental surveillance compared to manual monitoring methods. These findings are consistent with studies, who highlighted the effectiveness of integrating vision-based AI systems in environmental quality management.

Table 4. Performance comparison by waste class

Waste Class	Precision (%)	Recall (%)	F1-Score (%)
Inorganic	85	87	86
Organic	82	84	83

A comparative performance analysis assessed how effectively the model distinguishes between different types of river waste, specifically, inorganic (e.g., plastic bottles, cans) versus organic (e.g., leaves, branches) materials. The evaluation showed that the model performed marginally better in detecting inorganic waste. This is likely because plastic and metal items tend to have consistent shapes, more precise edges, and higher contrast against the water background, making them easier for the model to detect. In contrast, organic waste such as leaves or food scraps presents a more irregular appearance. It can be visually similar to natural river textures, especially under low lighting or turbid water conditions. These factors reduce detection confidence and lead to occasional false negatives. Based on class-specific evaluation metrics, the model achieved an F1-score of 86% for inorganic waste, compared to 83% for organic waste. The recall for inorganic waste was also higher (87%) than that for organic waste (84%).

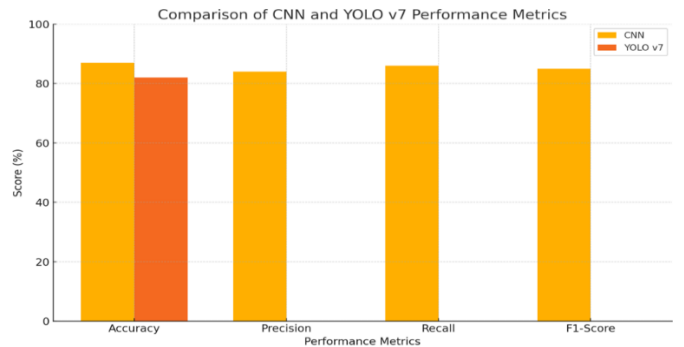


Figure 4. Comparison of CNN and YOLO v7 performance metrics

Waste detection in low-light conditions or when waste is partially submerged in water presents challenges.

While the CNN model could recognize the visual features of waste, the precision of the YOLO model slightly decreased when confronted with images exhibiting strong light reflections on the water's surface. Data augmentation techniques were employed to address this issue, including variations in rotation, lighting, and cropping, which successfully improved detection accuracy by 3% in more difficult conditions. Qualitatively, the integration of CNN and YOLO in this system offers flexibility, with CNN being more suited for batch image processing, while YOLO is designed for real-time detection. This combination enables the system to operate in a continuous automatic monitoring mode, facilitating quicker decision-making in the field. Additionally, this research produced an annotated dataset containing over 5,000 accurately labeled images, which is anticipated to significantly contribute to advancing river waste detection technology in Indonesia.

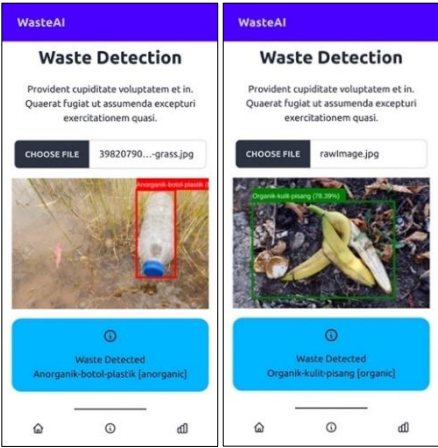


Figure 5. Examples of YOLO v7 prediction results on river waste images

The results and discussion of the research are presented comprehensively. The findings can also be illustrated through graphs, images, or tables. The layout, manuscript format, graphs, images, tables, and equations must adhere to specific guidelines. This research can enhance the accuracy of waste detection and offer a sustainable solution for monitoring river conditions. CCTV as a visual data source presents lower operational costs than more complex data collection methods, such as drones or satellite sensing. With these findings, this study makes a significant contribution to environmental management, particularly in monitoring river waste, which can assist the government in making data-driven decisions to improve water quality and mitigate the effects of waste on ecosystems and public health. The results underscore the potential for this system to be integrated into broader waste management strategies, aiding government efforts to monitor and reduce pollution in major rivers like the Cisadane. This approach could guide targeted clean-up efforts and inform policy-making by providing real-time data on waste accumulation.

Table 5. Comparison of this study's results with previous research on AI-based waste detection.

Study / Author	Model Used	Dataset Context	Accuracy	Real-Time (fps)	Notable Findings
This study	CNN + YOLO v7	Cisadane River, Indonesia	87% (CNN), 82% (YOLO)	20 fps	Strong performance under varied conditions
Zailan et al. (2022)	YOLO	Riverine waste, Malaysia	80–83%	Not stated	Focused on plastic detection
Pulipalupula et al. (2023)	CNN	Mixed solid waste	85–88%	Not applicable	High classification precision
Samuel et al. (2024)	AI + LLM	Urban river monitoring	~84%	Yes	Emphasis on AI integration and scalability

Table 5 summarizes a comparative analysis between this study and recent related works. The results demonstrate that the hybrid CNN-YOLO system developed in this study performs competitively, particularly in terms of classification accuracy and real-time capability under challenging river conditions.

While the proposed model performed well across various environmental conditions, several limitations should be acknowledged. First, the system showed reduced precision under low-light conditions and in the presence of intense surface reflections, which occasionally led to false positives or missed detections. Second, although diverse, the dataset was limited to the Cisadane River, raising questions about the generalizability of the model to other rivers with different waste compositions or visual characteristics. Lastly, although literature-based benchmarking was conducted, the study did not include an experimental baseline using alternative models such as SSD or Faster R-CNN under the same dataset, which may limit the strength of the performance comparison.

The results of this study are in agreement with several previous works. For example, Zailan et al. (2022) implemented YOLOv3 for river plastic detection and reported accuracy rates of 80%–83%, comparable to the 82% achieved by YOLOv7 in this study. Similarly, Pulipalupula et al. (2023) demonstrated CNN's capability to classify solid waste with 85–88% accuracy, aligning with our CNN results (87% accuracy and 85% F1-score). However, this study improves upon those approaches by

integrating both CNN and YOLO into a hybrid architecture, optimized for real-time detection in dynamically changing river environments—something few previous studies have addressed in depth.

Beyond technical contributions, this research has practical policy and river management implications. The proposed AI system can be integrated into national environmental monitoring programs, such as those led by the Ministry of Environment and Forestry (KLHK), to provide real-time waste data from major rivers. By automating surveillance and offering location-specific detection, the system supports data-driven interventions and policy enforcement under programs like Indonesia's National Plastic Action Partnership (NPAP). Furthermore, integrating this system with GIS platforms or waste cleanup scheduling could significantly improve the efficiency of river management operations at both national and municipal levels.

CONCLUSIONS

This study demonstrated the feasibility of integrating Convolutional Neural Network (CNN) and YOLO v7 algorithms to build a hybrid model for real-time river waste detection. The CNN model achieved a classification accuracy of 87% with an F1-score of 85%, while YOLO v7 reached 82% accuracy and processed video at 20 frames per second. These results confirm that the system performs reliably across varied river conditions, including differences in lighting, water turbidity, and waste types. The novelty of this research lies in its application of a CNN-YOLO hybrid model to real-world river data collected through continuous CCTV monitoring in Indonesia—a setting that has not been previously explored in published literature. The AI-based system developed in this study presents a scalable and operationally viable solution for river surveillance, offering real-time detection capabilities supported by empirical performance benchmarks. Rather than proposing a generalized solution, this work provides a tested, context-specific prototype that can be further integrated into environmental monitoring frameworks. Future research should focus on expanding the dataset to include multiple river systems, incorporating IoT-based sensors for cross-validation, and implementing comparative benchmarks with other object detection models such as SSD or Faster R-CNN. Additionally, long-term monitoring deployments can support policy development through empirical trend analysis in waste accumulation and seasonal pollution dynamics.

SUGGESTION

Here are some concise suggestions for further research and development in river waste detection using AI technology integrate IoT sensors for real-time environmental data, explore advanced machine learning models to enhance detection accuracy, test the system in other rivers to assess scalability, analyze historical data to identify waste accumulation patterns, collaborate with stakeholders to raise awareness and support, develop a mobile app for public waste reporting, improve data augmentation techniques for better recognition, evaluate

the environmental impact of detected waste, and consider using drones for aerial monitoring to complement the system.

REFERENCES

- Bhaskar, M. (2020). The impact of anthropogenic organic and inorganic pollutants on the Hasdeo River Water Quality in Korba Region, Chhattisgarh, India. *Bioinformation*, 16(4), 332–340. [[Crossref](#)], [[Publisher](#)]
- Chen, L., Huang, Y., Dong, J., Xu, Q., Kwong, S., Lu, H., Lu, H., & Li, C. (2024). *Underwater Object Detection in the Era of Artificial Intelligence: Current, Challenge, and Future*. [[Crossref](#)], [[Publisher](#)]
- Devi, B., Gupta, D., & Singh, R. (2024). *Yolo-based Deep Learning Techniques for Identifying Floating Bottles in Inland Water: A Comprehensive Analysis*. [[Crossref](#)], [[Publisher](#)]
- Hu, C., & Ge, Y. (2024). Improved YOLOv7 Algorithm for Drone Image Object Detection. *2024 International Conference on Artificial Intelligence and Power Systems (AIPS)*, 238–242. [[Crossref](#)], [[Publisher](#)]
- Ibarraran Viniegra, M. E., Saldaña-Vázquez, R., & Perez-Garcia, T. (2021). *The Cost of Pollution in the Upper Atoyac River Basin: A Systematic Review*. [[Crossref](#)], [[Publisher](#)]
- Issaoui, H., ElAdel, A., & Zaid, M. (2024). Object Detection Using Convolutional Neural Networks: A Comprehensive Review. *2024 IEEE 27th International Symposium on Real-Time Distributed Computing (ISORC)*, 1–6. [[Crossref](#)], [[Publisher](#)]
- Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). *YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions*. [[Crossref](#)], [[Publisher](#)]
- Makarim, M. N. (2024). Prevention of decreasing river water quality due to anthropogenic activities: A systematic review of water pollution on Cisadane River. *Journal of Marine Problems and Threats*, 1(1). [[Crossref](#)], [[Publisher](#)]
- Muneer, F., Azam, H. M. H., Zubair, M., Farooq, T., Ibrahim, M., Rasul, I., Afzal, M., Ahmad, A., & Nadeem, H. (2021). *Remediation of Water Pollution by Plastics* (pp. 89–117). [[Crossref](#)], [[Publisher](#)]
- Olawade, D. B., Fapohunda, O., Wada, O. Z., Usman, S. O., Ige, A. O., Ajisafe, O., & Oladapo, B. I. (2024). Smart waste management: A paradigm shift enabled by artificial intelligence. *Waste Management Bulletin*, 2(2), 244–263. [[Crossref](#)], [[Publisher](#)]
- Saddi, K. C., Miglino, D., Isgro, F., Emmerik, T. H. M. van, & Manfreda, S. (2024). Balancing Accuracy and Efficiency for River Plastic Monitoring. *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 3836–3840. [[Crossref](#)], [[Publisher](#)]
- Sani, M. T., Budihardjo, M. A., & Sarminingsih, A. (2024). Charting the currents of environmental concern: a decades-spanning scientometric analysis of riverine debris. *IOP Conference Series: Earth and*

- Environmental Science*, 1414(1). [[Crossref](#)], [[Publisher](#)]
- Shi, Z., Assis, J., & Costello, M. (2021). Vulnerability of Marine Species to Low Oxygen Under Climate Change. In *Reference Module in Earth Systems and Environmental Sciences*. [[Crossref](#)], [[Publisher](#)]
- Singh, S., Trushna, T., Kalyanasundaram, M., Tamhankar, A. J., & Diwan, V. (2022). Microplastics in drinking water: a macro issue. *Water Supply*, 22(5), 5650–5674. [[Crossref](#)], [[Publisher](#)]
- Suwarno, P., & Nurhayati, N. (2021). Traditional Views and Attitude Toward Waste and Rivers in Indonesia: Challenges of Cleaning Up the Indonesia Polluted Environment. *E3S Web of Conferences*, 317. [[Crossref](#)], [[Publisher](#)]
- Wahyuni, N. S., Efadeswarni, Syuhada, Liza, C., Sudaryanto, A., Sulistia, S., Wahyono, I. B., Witama, R. O., Shiddiq, M. A., Mardani, R., & Aditya, H. R. (2024). Abundance of microplastics in Cisadane river - Indonesia. *IOP Conference Series: Earth and Environmental Science*, 1388(1). [[Crossref](#)], [[Publisher](#)]
- Walia, J. S., Haridass, K., & Pavithra, L. K. (2025). Deep Learning Innovations for Underwater Waste Detection: An In-Depth Analysis. *IEEE Access*, 1. [[Crossref](#)], [[Publisher](#)]
- Xueming, Z., Yanyue, P., Fenghua, W., Hengheng, S., Rui, G., & Xinjie, L. (2022). Automatic Classification Bin Based on Artificial Intelligence Technology. *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 6, 408–413. [[Crossref](#)], [[Publisher](#)]
- Yeaminul Islam, S. M., & Alam, Md. G. R. (2023). Computer Vision-Based Waste Detection and Classification for Garbage Management and Recycling. In Md. S. Hossain, S. P. Majumder, N. Siddique, & Md. S. Hossain (Eds.), *The Fourth Industrial Revolution and Beyond* (pp. 389–411). Springer Nature Singapore. [[Crossref](#)], [[Publisher](#)]
- Zhen, S., Xiaohong, N., & Min, W. (2011). Empirical Analysis on Economic Losses of Urban River Pollution-A Case Study of Changzhou City of Jiangsu Province. *Energy Procedia*, 5, 2010–2014. [[Crossref](#)], [[Publisher](#)]