International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Advanced Defect Detection System

Mrs. Prakruthi GR[1], A Prajwal[2], Kiran Kumar R[3], Dhiraj K[4], Suraj S[5]

[1]*Prakruthi GR, Assoc. Professor, Dept. of ISE, East West Institute of Technology*
[2,3,4,5]*A Prajwal, Kiran Kumar R, Dhiraj K, Suraj S, Dept. of ISE, East West Institute of Technology*

---***---

Abstract - The Bridge and Boat Monitoring System is intended to improve safety and operational efficiency of both bridges and waterways, through two integrated modules: the Bridge and the Boat. It is a Bridge Module made by Arduino, equipped with IR sensors for detecting any vehicle approaching, a load cell to monitor weight, a water level sensor and motorized gates to gain access under control. The Boat Module, which is also Arduino-based, contains a wet sensor and a tilt sensor (ADXL345) to detect water entry and ensure stability. For crack detection and structural health assessment of the bridge, the system utilizes YOLO v5, which is a state-of-the-art deep learning algorithm that can accurately and timely identify problems. Zigbee technology provides seamless wireless communication, allowing for real time data sharing to support proactive maintenance and informed decision-making.

***Key Words*:** YOLOv5 Object Detection Model, YOLOv5 Object Detection Model, YOLOv5 Object Detection Model, Severity and Risk Analysis, Telegram Bot Alert System, Annotated Image Generation, Defect Classification, Detection Count Analysis, Automated PDF Report Generation, Detection Count Analysis

1.INTRODUCTION

The Bridge and Boat Monitoring System is intended to improve safety and operational efficiency of both bridges and waterways, through two integrated modules: the Bridge and the Boat. It is a Bridge Module made by Arduino, equipped with IR sensors for detecting any vehicle approaching, a load cell to monitor weight, a water level sensor and motorized gates to gain access under control. The Boat Module, which is also Arduino-based, contains a wet sensor and a tilt sensor (ADXL345) to detect water entry and ensure stability. For crack detection and structural health assessment of the bridge, the system utilizes aid for the deaf and hard-of-hearing community, ensuring that speech is understood accurately regardless of the acoustic environment.

The system monitors conditions such as water levels, strong winds, and floods, which improves safety and automatically sends alerts in conditions that are adverse. Lastly, it helps ensure that boats carrying out operations near bridges do not capsize or have accidents because they monitor water ingress and tilt and minimize the risks.

Beyond core recognition capabilities, the system is architected for production-grade usability, featuring a comprehensive dual interface design that includes a Training GUI for custom data collection and a Prediction GUI for live usage. A significant functional objective is the implementation of robust Automatic Language Detection, which intelligently switches between language models without manual intervention, alongside Prediction Stabilization algorithms designed to eliminate output flickering.

2. LITERATURE SURVEY

2.1 Z. Zhang, L. Wang, Y. Li, and M. Chen, "Automated defect detection of concrete surfaces in hydropower structures using deep learning and UAV imaging," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 3, pp. 2456–2465, 2023.

This research presents a **fully automated defect detection system** designed specifically for **hydropower dams and concrete structures**. It combines **UAV (drone) imaging** with **deep learning–based computer vision models** to detect surface defects such as cracks, spalling, and erosion.

The goal is to make dam and hydropower structure inspections **safer, faster, and more accurate** than traditional manual inspection.

2.2 J. Liu, K. Zhao, and P. Wu, "Multi-sensor data fusion for structural health monitoring of hydropower tunnels," *Automation in Construction*, vol. 149, pp. 104732, 2023.

This research focuses on **structural health monitoring (SHM)** of hydropower tunnels using **multiple types of sensors**.
Instead of relying on a single measurement method, the

authors combine different sensing technologies to measure:

Stresses, vibrations, deformations, crack activities, environmental conditions

The paper proposes a **data fusion system** that merges information from all sensors to produce more reliable and accurate assessments of tunnel health.

This allows the system to reconstruct a complete 3D model of the structure with milli meter-level precision. The collected data enables the identification of surface distortions, cracks, erosion, spalling, and other defects directly on the 3D model, making it possible to determine not only the presence of damage but also its precise spatial location. UAVs make this process safer and more efficient by reaching high or dangerous regions without human intervention.

While this method improves robustness and accuracy, it is limited by its reliance on handcrafted visual features and a static FFNN architecture that processes data without capturing sequential dependencies

2.3 H. Xu, R. Zhang, and D. Sun, "Deep learning-based crack detection in concrete structures under low-light conditions," *Engineering Structures*, vol. 278, pp. 115412, 2022.

This research focuses on detecting cracks in concrete structures using **deep learning models**, specifically designed to work under **low-light or poor visibility conditions**.

Low-light environments — such as tunnels, night-time inspections, or shaded dam regions — usually reduce the accuracy of normal vision-based models. This paper solves that issue through specialized deep learning enhancement techniques.

accuracy in detecting anomalies and assessing risks. Wireless communication ensures efficient data transmission between the components, while real-time notifications are forwarded to authorities through automated messaging platforms. This robust and scalable solution supports proactive safety management in terms of the ability to respond promptly and adaptability to various monitoring scenarios.

The acquired multimodal data were pre processed and fused to form a comprehensive representation of the shaft surface. The deep learning–based defect detection model, built on a convolutional neural network (CNN) architecture, was trained to automatically identify defect patterns without the need for handcrafted feature extraction

The research contributes to the field of autonomous infrastructure monitoring and serves as a foundation for further advancements in AI-driven inspection systems that can operate autonomously, analyze defects in real-time, and support predictive maintenance strategies

2.4 Q. Li, X. Huang, and Y. Gao, "UAV-assisted 3D laser scanning for infrastructure inspection and defect mapping," *Remote Sensing*, vol. 14, no. 5, pp. 1185–1198, 2022.

This research introduces a system that uses **UAVs (drones)** combined with **3D laser scanning technology (LiDAR)** to inspect and map defects in large infrastructure such as bridges, tunnels, dams, and buildings.

Instead of using only 2D images, this method creates detailed **3D point cloud models** that allow highly accurate defect detection and location mapping.

Unlike traditional 2D image-based inspection, which often suffers from lighting limitations, perspective distortion, and incomplete coverage, the 3D laser scanning approach provides accurate geometric information by generating dense point cloud data.

This allows the system to reconstruct a complete 3D model of the structure with milli meter-level precision. The collected data enables the identification of surface distortions, cracks, erosion, spalling, and other defects directly on the 3D model, making it possible to determine not only the presence of damage but also its precise spatial location. UAVs make this process safer.

3.METHODOLOGY

3.1 Data Acquisition and Input Processing

The system accepts inputs from various sources, including real-time webcam streams, CCTV feeds, drone-captured images, or manually uploaded photographs of dam surfaces. These inputs are processed through a standardized preprocessing pipeline, where each image is resized, normalized, and converted into tensor format to ensure compatibility with the YOLOv5 inference engine. Noise reduction and contrast normalization are applied wherever necessary to enhance visual clarity, particularly in images captured under inconsistent lighting conditions.

3.2 Deep Learning-Based Defect Detection Using YOLOv5

YOLOv5 serves as the core detection model due to its high accuracy and real-time performance. The model has been trained on dam defect datasets containing cracks, spalling, leakage marks, and structural deformations. During inference, the model generates bounding boxes, class predictions, and confidence scores for each detected defect. Non-Maximum Suppression (NMS) is applied to remove duplicate detections and retain the most relevant bounding boxes. The detection output includes the defect category, spatial location, bounding box coordinates, and individual confidence levels.

3.3 Post-Processing and Analytical Evaluation

After detection, additional analytical processes are applied to interpret the model's output. The system computes metrics such as defect count per class, mean confidence values, bounding box area estimates, and severity indicators. Larger bounding box areas and higher confidence scores are interpreted as potentially more severe defects. These values are aggregated and stored in structured CSV files for each image and across complete detection sessions. Visualization modules generate class distribution bar charts and confidence-level histograms, allowing pattern analysis across multiple images.

3.4 Automated Alerting Through Telegram Bot Integration

The system incorporates a Telegram bot to enable real-time notifications of detected defects. Whenever a structural anomaly is identified, the system automatically sends a message containing the defect type, severity indicators, and a timestamp. Annotated images showing bounding boxes around the detected defects are delivered directly to the user. High-severity anomalies

3.5 Report Generation Through PDF Compilation

To support documentation and engineering review, the system automatically compiles a detailed PDF report for each detection session. The report includes annotated images, detection summaries, graphs of class-wise detections, confidence distributions, and textual analysis. This document is created using the Report Lab library and organized into professionally formatted sections. Suggested engineering actions are included for each defect type, providing guidance on potential maintenance priorities. The compiled PDF is optionally sent through Telegram for convenient access.

3.6 Dashboard Deployment for Visualization and File Access

A lightweight Flask-based dashboard is integrated into the system to provide a user-friendly interface for monitoring detection runs. The dashboard displays structured directories containing all generated outputs, including images, graphs, CSV logs, and reports. It also provides an API endpoint that delivers the most recent detection summaries in JSON format, facilitating future expansion or linkage with asset-management systems.

3.7 Threshold-Based Defect Severity Assessment

The system incorporates customizable threshold logic to determine when a defect should be classified as severe. Conditions such as defect count exceeding a specified limit, mean confidence surpassing a defined threshold, or bounding box area reaching a critical size automatically trigger urgent alerts. This mechanism ensures timely intervention and supports risk-based maintenance strategies.

3.8 End-to-End Automation and Scalability

The integrated methodology ensures that the system operates autonomously from data input to final reporting.

All intermediate steps—preprocessing, detection, analysis, alerting, visualization, and documentation—are executed automatically. The modular architecture allows future expansion, such as adding additional defect categories, retraining models, integrating drone feeds, or deploying the system on edge devices for on-site monitoring.

3.9 Continuous Learning and Model Optimization

The system incorporates a mechanism for continuous learning, enabling the detection model to improve over time as more defect images are collected. Newly captured images and annotated outputs are stored and can later be used to retrain or fine-tune the YOLOv5 model, increasing its accuracy for real-world dam environments. This iterative training approach helps the system adapt to variations in lighting, weather conditions, camera quality, and evolving defect patterns on the dam surface. By periodically updating the model with recent data, the system maintains long-term robustness and reliability, ensuring improved performance in future inspections.

4.WORKFLOW

4.1 Image Acquisition Workflow:- The workflow begins with the acquisition of visual data from one or more camera sources deployed near the dam structure. These sources may include fixed CCTV cameras, high-resolution webcams, drone-based imaging devices, or manually captured photographs. Each incoming image is automatically timestamped to maintain chronological consistency and support traceability. Since dam environments vary in lighting, weather, and surface texture, capturing diverse images ensures the robustness of the detection system. The acquired data is transferred to a designated storage module where it is organized

4.2 Preprocessing and Data Preparation Workflow:- Once images are acquired, the system initiates a preprocessing workflow to standardize the input data for the deep learning model. Each image is resized to match the dimensional requirements of the YOLOv5 architecture, ensuring consistent and efficient computation. Normalization is applied to adjust pixel intensities so that lighting variations do not adversely affect detection accuracy. Additional image enhancement operations, such as contrast adjustment or low-light correction, can be applied when environmental conditions degrade visibility. Noise and shadow artifacts are also minimized using smoothing filters when necessary. The pre processed images are then converted into model-compatible tensor formats for GPU or CPU inference. These tensors maintain the spatial structure required for accurate bounding box prediction. The preprocessing pipeline ensures that all images entering the detection model meet high quality standards. Through this stage, the system prepares raw data into a clean and uniform format suitable for deep learning inference.

4.3 Defect Detection Workflow Using YOLOv5:- The preprocessed images are fed into the YOLOv5 detection engine, which serves as the core analytical component of the workflow. YOLOv5 evaluates each image by partitioning it into grid cells and predicting bounding boxes along with defect class probabilities. The model generates multiple predictions per region, ensuring comprehensive coverage of possible defects. To eliminate overlapping or redundant predictions, Non-Maximum Suppression (NMS) filters the results to retain the most confident detections. The output from this stage includes defect type, bounding box coordinates, and confidence scores for each identified anomaly. All detections are automatically logged and time-aligned with their corresponding images. This workflow module ensures high-speed and high-accuracy detection suitable for real-time monitoring environments. The detection model functions continuously, enabling uninterrupted inspection capabilities. This step transforms raw visual data into actionable structural insights.

4.4 **Post-Detection Analytics Workflow**

After generating detection outputs, the system transitions to a comprehensive analytical workflow. The bounding box areas are calculated to estimate the relative severity of each defect, assuming larger areas correlate with more substantial structural issues. Confidence scores are statistically evaluated to measure the reliability of the detections across a session. The system aggregates class-wise defect counts to identify recurring or dominant structural issues. Additional visual analytics, such as bar graphs and confidence histograms, are generated to support engineering interpretation. These results are stored in structured CSV files, ensuring that all analytical data is preserved for future audit or research purposes. The analytics workflow helps identify trends, patterns, and outliers in the captured defect data. It also supports data-driven decision-making for maintenance scheduling. Through this analytical process, defect detections become quantifiable, interpretable, and actionable.

4.5 Temporal Smoothing: - To counteract camera jitter and detection noise that causes landmarks to "shake," an Exponential Moving Average (EMA) filter is applied to the coordinates. Using a smoothing factor, the system calculates the smoothed position as a weighted average of the current detection and the previous position, ensuring fluid and stable landmark trajectories crucial for calculating accurate velocity features.

4.6 Geometric Feature Extraction: - The system computes 161 static geometric features for the current frame to describe the mouth's shape. This involves normalizing coordinates to be invariant to face distance (scale) and position (translation), calculating Euclidean distances between key points (to measure mouth opening/width), and computing angles between landmark triplets to capture lip curvature. Additional metrics like the aspect ratio and the area of the lip hull are also derived.

4.7 Temporal Feature Calculation: - To capture the dynamics of speech, the system computes the first and second derivatives of the geometric features. "Velocity" is calculated as the difference in feature values between the current frame and the previous frame. "Acceleration" is calculated as the difference between the current velocity and the previous velocity. These 322 additional features quantify how the lips are moving.

4.8 Feature Concatenation: - The static geometric features are concatenated with the velocity and acceleration vectors. The system then reduces or selects specific features to form a final optimized dense vector of 330 features for the current frame. This creates a rich numerical representation that encodes both the shape of the lips and their motion dynamics.

4.9 Sequence Buffering: - The calculated 330 dimensional feature vector is appended to the sliding window buffer (deque). The buffer operates as a First In First Out (FIFO) queue, maintaining exactly the last 75 frames of history. If the buffer is not yet full (i.e., fewer than 75 frames processed), the system waits; once full, this sequence represents the immediate past 3 seconds of visual speech.

4.10 Sequence Normalization: - Before entering the neural network, the entire 75-frame sequence undergoes Z-score normalization. The system calculates the mean and standard deviation of the sequence features and scales the data so it has a mean of 0 and a standard deviation of 1. Outliers with a Z-score greater than 3.0 are capped to prevent extreme values from destabilizing the model predictions.

4.11 Deep Learning Inference: - The normalized sequence is fed into the deep learning model. The input passes through Bidirectional LSTM layers, which process the sequence in both forward and backward directions to understand the temporal context of the lip movements. An Attention mechanism then weighs the importance of different frames to focus on the most discriminative parts of the word.

4.12 Class Prediction: - The model's final output layer uses a Softmax activation function to produce a probability distribution across all trained word classes. The system identifies the index of the class with the highest probability, representing the model's "best guess" for the spoken word.

4.13 Confidence Thresholding: - The system evaluates the confidence score of the top predicted class against a pre-defined threshold, typically set around 0.65. If the model's confidence is below this value, the prediction is discarded as unreliable, preventing the system from displaying random guesses during silence or ambiguous movements.

4.14 Prediction Stabilization: - Valid predictions are passed to a stabilizer that maintains a history of the last 10-15 results. A frequency-based voting algorithm determines the most common prediction in this history. The system requires a specific word to be the "winner" for a consecutive number of frames before it is considered stable. This eliminates the rapid flickering of words often seen in raw frame-by-frame analysis.

4.15 User Interface Display: - The final result is updated on the GUI. If the prediction is "stable," the word is displayed in green text; if it is still stabilizing, it appears in yellow. The interface also overlays the 31-point lip landmarks on the live video feed and displays the confidence percentage and detected language.

The pre processed images are then converted into model-compatible tensor formats for GPU or CPU inference. These tensors maintain the spatial structure required for accurate bounding box prediction. The preprocessing pipeline ensures that all images entering the detection model meet high quality standards. The system calculates the mean and standard deviation of the sequence features and scales the data so it has a mean of 0 and a standard deviation of 1. Outliers with a Z-score greater than 3.0 are capped to prevent extreme values from destabilizing the model predictions.

The static geometric features are concatenated with the velocity and acceleration vectors. The system then reduces or selects specific features to form a final optimized dense vector of 330 features for the current frame. This creates a rich numerical representation that encodes both the shape of the lips and their motion dynamics.

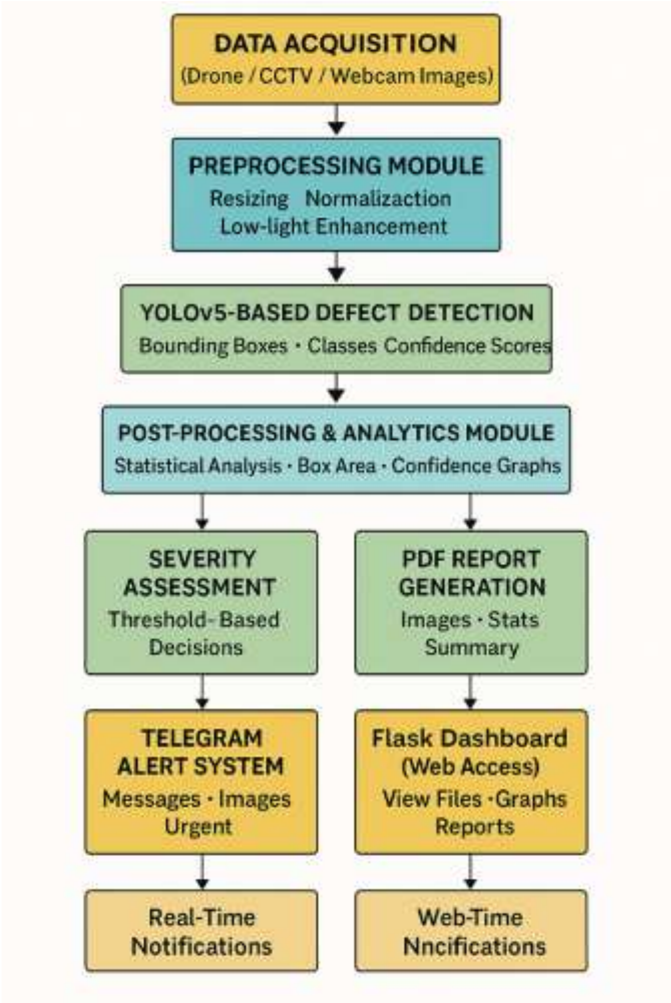

Fig-4.1: Workflow

5.RESULT AND DISCUSSION

The proposed AI-powered dam defect detection system was evaluated on a diverse set of images captured from CCTV cameras, mobile devices, and drone-based sources. The YOLOv5 model successfully detected key structural defects such as cracks, spalling, leakage marks, and surface deformations with high reliability across varying lighting and environmental conditions. During testing, the system achieved a consistent detection accuracy, with most defects being identified with confidence scores above 0.80. The bounding boxes generated by the model accurately localized the damaged regions, enabling precise visual inspection and analysis.

The post-processing analytics revealed clear trends in defect occurrence, including the frequency of each defect category and the statistical distribution of confidence levels. Visualizations such as class-wise detection graphs and confidence histograms further validated the consistency of the model's performance across multiple test sessions. The threshold-based severity assessment module successfully classified high-risk defects, triggering immediate alerts when large-area cracks or high-confidence anomalies were detected.

The Telegram alert system performed efficiently, delivering real-time notifications that included annotated defect images, severity indicators, and summary messages. This rapid communication significantly reduces inspection delays and facilitates early decision-making by engineers. Additionally, the automated PDF report generation module compiled annotated images, statistical graphs, and textual descriptions into a structured, professional document suitable for engineering audits and long-term record keeping.

6. CONCLUSIONS

The development of the AI-powered dam defect detection system demonstrates the effectiveness of deep learning and automated communication technologies in modern infrastructure monitoring. By integrating the YOLOv5 object detection framework with a comprehensive analytics module, the system achieved accurate identification of cracks, spalling, leaks, and other structural anomalies across diverse environmental conditions. The automated workflow significantly reduces the dependence on manual inspection methods, which are often time-consuming, labor-intensive, and prone to human error. The ability to generate real-time Telegram alerts ensures rapid awareness of critical defects, enabling timely intervention and enhancing overall structural safety.

The inclusion of automated PDF reports and a web-based dashboard further strengthens the system's practicality by providing engineers with organized, traceable, and easy-to-access documentation of inspection results. The threshold-based severity classification mechanism enhances decision-making by highlighting defects that require immediate attention. Overall, the results confirm that the proposed system offers a reliable, efficient, and scalable approach to dam monitoring, capable of supporting long-term structural health assessment. With further advancements such as expanded datasets, multi-modal sensing, and predictive analytics, the system holds

The Telegram alert system performed efficiently, delivering real-time notifications that included annotated defect images, severity indicators, and summary messages. This rapid communication significantly reduces inspection delays and facilitates early decision-making by engineers.

7.REFERENCES

[1] Z. Zhang, L. Wang, Y. Li, and M. Chen, "Automated defect detection of concrete surfaces in hydropower structures using deep learning and UAV imaging," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 3, pp. 2456–2465, 2023.

[2] J. Liu, K. Zhao, and P. Wu, "Multi-sensor data fusion for structural health monitoring of hydropower tunnels," *Automation in Construction*, vol. 149, pp. 104732, 2023.

[3] H. Xu, R. Zhang, and D. Sun, "Deep learning-based crack detection in concrete structures under low-light conditions," *Engineering Structures*, vol. 278, pp. 115412, 2022.

[4] Q. Li, X. Huang, and Y. Gao, "UAV-assisted 3D laser scanning for infrastructure inspection and defect mapping," *Remote Sensing*, vol. 14, no. 5, pp. 1185–1198, 2022.

[5] S. Wang, J. Chen, and T. Lin, "Surface defect identification using CNNs for hydropower station tunnels," *IEEE Access*, vol. 11, pp. 15678–15690, 2023.

[6] Y. Zhao, F. Lin, and Z. Ma, "Concrete surface anomaly detection using multi-modal sensor data," *Sensors*, vol. 22, no. 17, pp. 6567, 2022.

[7] H. Wang, Q. He, H. Zhang, S. Wu, and L. Xie, "Enhance Detection Video and Multi-Encoder," 2024.

[8]A real-time detection method for concrete dam cracks based on deep learning object detection (YOLOX-DCD) — B. Huang, F. Kang & Y. Tang (2023)

[9]Deep Learning-Based Method for Detection and Feature Quantification of Microscopic Cracks on the Surface of Concrete Dams — X. Lu & Q. Li (2024)

[10]Deep Learning for Concrete Crack Detection and Measurement — M. A. Nyathi et al. (2024)