

Research Article

Deep Learning-based Prediction of Lifespan Degradation in Concrete Bridges Due to Iron Oxidation

Hetkumar Patel , Wisam Bukaita* 

Math and Computer Science Department, Lawrence Technological University, Southfield, The United States

Abstract

This study presents a comprehensive multimodal deep learning framework for predicting lifespan degradation in concrete bridges caused by iron oxidation. The proposed system integrates YOLOv8 for surface-level crack detection and ResNet50 for deep image feature extraction, combined with structurally significant tabular data such as crack geometry, material composition, environmental factors, and corrosion indicators. Addressing limitations in current approaches—including dataset scarcity, lack of multimodal integration, and high cost of sensor-based inspection—the framework employs a hybrid architecture to estimate three critical outputs: degradation score, condition class, and remaining life of the bridge. To overcome data limitations, synthetic tabular features were generated using AI-based simulations aligned with visual inputs. The system was trained with extensive resources: 200 epochs for YOLOv8 and 50+ epochs for the tabular model, followed by k-fold cross-validation (MAE: 3.48, R²: 0.89) to validate generalization. Despite challenges in detection accuracy (mAP@0.5: 0.0101), the classification component achieved an AUC of 0.98, confirming robustness in condition prediction. Comparative evaluations demonstrate that YOLOv8 and ResNet50 provide the best trade-off between accuracy, efficiency, and deployment readiness. The proposed model, further enhanced with attention mechanisms and future transformer-based extensions, offers a scalable, low-cost alternative to traditional sensor-driven monitoring and contributes to more proactive, data-driven maintenance of aging bridge infrastructure.

Keywords

Deep Learning, Concrete Defect Detection, Structural Health Monitoring, Crack and Corrosion Classification, Machine Learning, Resnet50, YOLOv8, Multimodal Learning

1. Introduction

In 2024, Xiong et al. [1] a novel YOLOv8-GAM-Wise-IoU-based model was proposed for automated concrete bridge crack detection in high-density urban environments such as Hong Kong. Addressing the critical limitations of manual inspections—such as subjectivity, safety risks, and labor intensiveness—the study introduced an enhanced YOLOv8 framework that incorporates a Global Attention Module (GAM) and a Wise Intersec-

tion-over-Union (IoU) loss function to improve precision and generalization. The model architecture utilizes a decoupled head, separating classification and detection tasks, thereby enhancing task-specific performance. Experimental results demonstrated outstanding performance, with precision of 97.4%, recall of 94.9%, F1-score of 0.96, and mAP scores of 98.1% (mAP50), 76.2% (mAP50-95), and 97.8% (mAP75), outperforming both one-stage and two-stage benchmark

*Corresponding author: wbukaita@ltu.edu (Wisam Bukaita)

Received: 2 September 2025; **Accepted:** 13 September 2025; **Published:** 26 September 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

models. With a modest size of 93.20M, the model is computationally efficient and highly deployable in real-world bridge monitoring systems.

In 2024, Dong et al. [2] Concrete surface crack detection is a critical research area for ensuring the safety of infrastructure, such as bridges, tunnels and nuclear power plants, and facilitating timely structural damage repair. Addressing issues in existing methods, such as high cost, lengthy processing times, low efficiency, poor effectiveness and difficulty in application on mobile terminals, this paper proposes an improved lightweight concrete surface crack detection algorithm, YOLOv8-Crack Detection (YOLOv8-CD), based on an improved YOLOv8. The algorithm integrates the strengths of visual attention networks (VANs) and Large Convolutional Attention (LCA) modules, introducing a Large Separable Kernel Attention (LSKA) module for extracting concrete surface crack and local feature information, adapted for features such as fracture susceptibility, large spans and slender shapes, thereby effectively emphasizing crack shapes. The Ghost module in the YOLOv8 backbone efficiently extracts essential information from original features at a minimal cost, enhancing feature extraction capability. Moreover, replacing the original convolution structure with GSConv in the neck network and employing the VoV-GSCSP module adapted for the YOLOv8 framework reduces floating-point operations during feature channel fusion, thereby lowering computational complexity whilst maintaining model accuracy. Experimental results on the RDD2022 and Wall Crack datasets demonstrate the improved algorithm increases in mAP50 by 15.2% and 12.3%, respectively, and in mAP50-95 by 22.7% and 17.2%, respectively, whilst achieving a reduced model computational load of only 7.9×10^9 , a decrease of 3.6%. The algorithm achieves a detection speed of 88 FPS, enabling real-time and accurate detection of concrete surface crack targets. Comparison with other mainstream object detection algorithms validates the effectiveness and superiority of the proposed approach.

In 2025, Xia et al. [3] Another study introduced a bridge crack detection algorithm built on the YOLOv8 framework, focusing on enhancing feature extraction and global attention mechanisms to improve detection accuracy. The model incorporated the SPPF_UniRepLk module into the backbone to strengthen image feature representation and utilized a Global Channel Spatial Attention (GCSA) mechanism to capture global contextual dependencies across feature maps. Additionally, the Coordattention-Concat module in the neck network performed multi-source feature fusion and reweighting, enabling refined crack representation. The proposed architecture achieved a mAP50-95 of 72.1%, showcasing its ability to detect complex crack patterns effectively while maintaining computational efficiency. The model's design prioritizes both performance and real-world applicability, contributing to safer infrastructure management and public trust in bridge safety.

In 2023, Wu et al. [4] Researchers presented a lightweight,

real-time, pixel-level crack detection framework tailored for deployment on low-performance devices, particularly in remote bridge and dam inspections using UAVs and ROVs. Addressing the limitations of traditional manual inspections and the challenges of damage-related feature extraction in remote settings, the study proposed an improved instance segmentation model featuring a lightweight backbone and an efficient prototype mask branch. These components significantly reduced model complexity while preserving accuracy. The model achieved an impressive accuracy of 0.945 and operated at 129 FPS, making it suitable for real-time applications in embedded or edge-computing environments. Its low computational footprint and high speed make it a practical candidate for field-level structural health monitoring tasks.

In 2024, LI et al. [5] A study proposed an enhanced bridge defect detection method to address the reduced accuracy of deep learning models under variable lighting and weak texture conditions. The researchers introduced a Multi-Branch Coordinate Attention (MBCA) mechanism into the YOLOv8 architecture to improve global perception and coordinate localization. To further boost adaptability to irregular crack patterns, the study integrated deformable convolution layers into the YOLOv8's backbone, replacing the standard deep C2F blocks. This new architecture, termed Deformable Convolutional Network Attention YOLO (DCNA-YOLO), demonstrated performance improvements of 2.0% in mAP and 3.4% in recall (R) on a supervised dataset of 4,794 bridge damage images. Notably, the method also reduced model complexity by 1.2G and increased detection speed by 3.5 FPS, proving its robustness and efficiency for real-time field applications in challenging visual environments.

In 2025, Aung et al. [6] Researchers addressed the critical challenge of limited annotated data in structural damage detection by introducing a synthetic data generation approach using a 3D engine. The method produced high-quality, automatically annotated crack images with controlled variations in morphology and environmental conditions, calibrated to real-world scales. These synthetic datasets enabled the training of deep learning models that showed nearly double the detection accuracy and a $2.5\times$ improvement in segmentation precision compared to models trained solely on real-world data. By generating over 1,000 diverse crack scenarios per minute, the approach proved highly efficient and scalable. Furthermore, it helped reduce domain-shift issues in deep learning, thus improving generalization to unseen damage types in civil infrastructure. This study highlights the promise of simulation-based data synthesis in enhancing performance where annotated data is scarce or expensive to obtain.

In 2023, Gharah et al. [7] a comprehensive state-of-the-art review examined the adoption of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) techniques across the structural engineering domain. The authors systematically analyzed over 4,000 scholarly publications, with a particular focus on developments from the past decade, to evaluate performance metrics, dataset scales, and algorithmic

best practices. The review identified high-value applications of AI, including structural health monitoring, damage detection, material property prediction, and disaster-resilient infrastructure design, particularly in the context of earthquake, wind, and fire engineering. Emphasis was placed on the current lack of systemic integration frameworks, highlighting the need for domain-specific guidelines, standardization of evaluation metrics, and access to large annotated datasets. This work provides a roadmap for structural engineers to navigate AI integration effectively, and advocates for tailored adoption strategies to ensure reliability and reproducibility of AI-based approaches in practical infrastructure monitoring and damage detection.

In 2025, Meda et al, [8] a hybrid deep learning framework was introduced for pixel-level instance segmentation and severity quantification of structural damage, including cracks, spalling, and corrosion. The study leveraged state-of-the-art models-Mask R-CNN and various YOLO versions (v5, v7, v8)-to segment and classify damage across a 6,000-image dataset annotated with polygon labels. Image resolutions varied from 416×416 to 640×640 to balance performance and computational cost. Damage severity was categorized into four levels (low, medium, high, critical) based on pixel area and percentage coverage, enabling actionable assessment for infrastructure maintenance. The most effective model achieved an 85% accuracy rate, proving the framework's viability in real-world conditions. The study emphasized adaptability to surveillance feeds, such as CCTV and UAV video streams, and suggested future integration with IoT for automated, real-time infrastructure monitoring. This approach bridges the gap between accuracy, interpretability, and field deployability in multi-type damage assessment.

In 2025, Soo et al, [9] a study explored the feasibility of deep learning techniques for vibration-based crack detection in rubber mounts used in bridge isolation systems. Recognizing the difficulty of manually inspecting these components, the researchers developed neural network models based on Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) units, and a hybrid CNN-LSTM architecture. Cracks were artificially induced in rubber mounts within a laboratory bridge setup, and various excitation signals-including 5 Hz and 10 Hz sine waves and actual bridge waveform recordings-were applied. The hybrid CNN-LSTM model outperformed standalone models, achieving an accuracy of 99.9%, compared to CNN's 89.0% in detecting crack severity from vibration signals. The models incorporated a Bayesian dropout layer to mitigate overfitting and enhance prediction confidence. The results demonstrated not only the practicality of DL for vibration-based SHM but also its scalability for real-time damage detection in critical infrastructure components.

In 2025, Wan et al, [10] a novel unsupervised deep learning method was introduced for early detection of structural anomalies using sensor data from structural health monitoring (SHM) systems. The proposed approach integrated a Deep

Convolutional Variational Autoencoder (DCVAE) for unsupervised feature extraction with Support Vector Data Description (SVDD) for anomaly detection. By learning probability distributions in the latent space using Kullback-Leibler divergence, the DCVAE effectively captured complex features from highly correlated multisensor data without losing spatial or temporal dependencies. The SVDD then formed a minimum-volume hypersphere that separated normal structural states from anomalous ones. Evaluations on a computational frame model and a laboratory-scale grandstand structure showed the method outperformed conventional techniques such as PCA, one-class SVM, and basic autoencoders. This study demonstrated the feasibility of unsupervised deep learning in capturing damage-sensitive patterns for early warning and anomaly detection in structural systems under uncertain environmental and operational variability.

In 2024, Li et al [11] researchers developed a two-stage deep learning framework for detecting and classifying corrosion levels in steel bridge bolts, a critical yet often overlooked component of structural integrity. The system employed a Deep Convolutional Generative Adversarial Network (DCGAN) with least squares loss to generate realistic corroded bolt images, thereby enriching the training dataset. Corrosion was categorized into three severity levels based on visual cues and rust surface coverage. For detection, the study proposed YOLOv5-SE, an enhanced YOLOv5 model incorporating Squeeze-and-Excitation (SE) blocks after each residual layer to improve attention and feature sensitivity. The model outperformed standard YOLOv5, Faster R-CNN, and VGG16 in both precision and localization. A PyTorch-based deployment system was also introduced, enabling real-time corrosion assessment of bridge bolts. This research marks a significant advancement toward intelligent maintenance solutions for steel bridge components using deep learning-based corrosion monitoring.

In 2020, Alexander et al, [12] a study explored the application of indirect electrochemical impedance spectroscopy (EIS) for detecting corrosion in post-tensioned steel tendons used in segmental bridge construction. Unlike conventional visual inspection, the method offered a nondestructive approach capable of identifying corrosion hidden within grout-encased steel strands. Finite element simulations revealed how frequency dispersion and non-uniform current distributions influence the impedance response, which deviates from classical electrical circuit analogs. The contribution of grout was modeled as a complex ohmic impedance component, further complicating analysis. Experimental validation on passive and artificially corroded tendons confirmed the method's feasibility and aligned well with numerical predictions. This study highlighted the potential of impedance-based SHM approaches for corrosion monitoring in embedded structural elements, expanding the toolkit for early detection in critical bridge components.

In 2024, Huang et al, [13] a novel deep learning model named BridgeNet was introduced for multiclass surface

damage detection and segmentation in bridge inspection images, including defects such as cracks, spalling, and corrosion. The model leveraged advanced components like the Swin Transformer for long-range feature modeling, the CARAFE upsampler for detail-preserving resolution enhancement, and transfer learning to boost convergence on limited data. A new dataset, BridgeDamage, was also compiled, containing over 2,800 annotated images spanning five major damage types under realistic inspection conditions. BridgeNet demonstrated a significant performance gain over baseline models such as Mask R-CNN, with mAP and mIoU improvements exceeding 33% and 26%, respectively. It achieved a maximum mAP of 74.7% and mIoU of 66%, outperforming other state-of-the-art models across all categories. The results affirmed BridgeNet's applicability in real-world bridge maintenance and its robustness in detecting corrosion and other surface degradation under variable photographic and background conditions.

In 2023, Inam et al, [14] researchers proposed a two-phase deep learning framework for low-cost crack detection and segmentation in bridges, particularly targeting infrastructure in developing countries. In the detection phase, YOLOv5 variants (s, m, l) were applied to a combined dataset of local bridge images from Pakistan and the publicly available SDNET2018 dataset. The YOLOv5m model outperformed others, achieving a mAP of 99.3%, compared to 97.8% and 99.1% for YOLOv5s and YOLOv5l, respectively. In the segmentation phase, a U-Net model was used to generate precise pixel-level crack masks, which were then used to quantify crack width, height, and area. The results were visualized via scatter plots and boxplots to support severity classification. This dual-stage model provided an efficient, accurate, and scalable solution for condition monitoring of critical infrastructure in regions lacking frequent manual inspections. The study emphasized the importance of automation for improving structural health monitoring in developing economies and demonstrated the adaptability of deep learning pipelines to diverse environmental and lighting conditions.

In 2023, Sun et al, [15] a real-time method for bridge damage detection and localization was proposed using convolutional neural networks (CNNs) to estimate excessive nodal loads based on deflection or inclination data from a minimal number of sensors. The approach built upon partial least-squares regression (PLSR), which had previously been used to estimate bridge-wide nodal loads from sparse sensor data. By training CNNs to replicate the PLSR estimation process, the method enabled the monitoring of deviations in estimated nodal loads before and after damage occurrence. These deviations were then used as indicators for both detecting and localizing damage. Notably, the method required no finite element modeling or labeled damage data, addressing a key barrier in real-world deployment. Numerical simulations demonstrated robust performance under various conditions, including unknown loads, multiple damage locations, and measurement errors. This framework presents a

cost-effective and scalable solution for SHM applications, particularly in settings where access to high-fidelity structural models or damage-labeled datasets is limited.

In 2023, Hajializadeh, [16] researchers introduced a novel data-driven *drive-by* structural health monitoring approach that utilizes acceleration data collected from an instrumented train to detect bridge damage without direct sensor deployment on the structure itself. Addressing logistical and cost-related challenges of network-wide instrumentation, the proposed system acts indirectly—where the moving vehicle serves as both the actuator and sensor. A deep convolutional neural network (CNN), optimized through Bayesian techniques, was trained to automatically extract damage-sensitive features from train-borne measurements. The method was experimentally validated using a scaled steel bridge model and instrumented train, tested across various damage severities, locations, and train speeds. Despite challenges such as operational noise and rail irregularities, the model achieved accurate damage detection and classification under all tested conditions. This represents the first successful demonstration of a fully data-driven *drive-by* SHM system in a controlled operational environment and illustrates the potential for scalable, cost-effective bridge monitoring solutions with minimal physical infrastructure.

In 2020, Chen et al, [17] a lightweight and efficient bridge damage detection framework named DT-YOLOv3 was proposed, building on the established YOLOv3 object detection architecture with several key enhancements for improved accuracy and efficiency. The model incorporated deformable convolutions to better capture irregular damage patterns and leveraged transfer learning to accelerate training and improve generalization. To further reduce computational cost, group convolution and model pruning were applied, resulting in a compressed architecture suitable for real-time applications. Experimental evaluations demonstrated that DT-YOLOv3 achieved superior accuracy and faster inference times compared to standard detection models, making it practical for deployment in field conditions. This work illustrates the potential of combining architectural refinements with optimization strategies to deliver effective and resource-efficient bridge damage recognition systems.

Comparison with established CNN architectures.

To justify the architectural choices made in this project, a comparative evaluation was conducted against well-established convolutional neural network (CNN) architectures in terms of their detection performance, computational efficiency, deployment readiness, and suitability for multimodal integration. The objective was to select models that could offer a strong balance between real-time performance and high accuracy for both object detection and feature extraction. As a result of this evaluation, YOLOv8 was chosen as the detection backbone due to its speed, lightweight deployment, and superior detection head, while ResNet50 was selected for image feature extraction in combination with tabular inputs. The following two tables summarize the

comparisons that informed these selections.

Table 1. YOLOv8 vs. Traditional Detectors (e.g., YOLOv5, Faster R-CNN (Faster Region-based Convolutional Neural Network), SSD (SSD - Single Shot MultiBox Detector)).

Feature	YOLOv8	YOLOv5	Faster R-CNN	SSD
Detection Speed	Real-time (Fastest)	Fast	Slower	Fast
Accuracy (mAP)	Higher (improved head)	Good	Very High (but slower)	Moderate
Model Size	Compact	Compact	Large	Medium
Use Case Fit	Damage Detection (ideal)	General object detection	Complex scenarios	Lightweight devices
Deployment Ready	Plug & Play	Requires tuning	Requires full pipeline	Moderate complexity

Table 1 shows that YOLOv8 demonstrates the best trade-off between inference speed and detection accuracy, making it the most suitable candidate for real-time bridge damage detection tasks. Compared to traditional detectors

like YOLOv5 or Faster R-CNN, YOLOv8 offers a simplified deployment process and requires fewer tuning adjustments, while still delivering higher mAP due to its optimized architecture.

Table 2. ResNet50 vs. Other Feature Extractors (e.g., VGG16, InceptionV3, EfficientNet).

Feature	ResNet50	VGG16	InceptionV3	EfficientNetB0
Depth / Complexity	Medium (50 layers)	Shallow (16 layers)	Deeper, multi-branch	Scaled-depth
Feature Extraction Quality	High (residual blocks)	Moderate	Very High	High
Training Time	Moderate	Fast	Longer	Fast
Model Size	Medium	Large	Medium	Small
Performance on Tabular+Img	Suitable	Weak (no skip connections)	Good but complex	Good but not as interpretable

Table 3. Focused Comparison Between InceptionV3 and ResNet50.

Feature	InceptionV3	ResNet50
Accuracy (ImageNet)	Higher	Slightly lower
Architecture	Complex (multi-path)	Simple & deep
Integration with Tabular	Difficult	Straightforward
Inference Speed	Slower	Faster
Deployment Friendly	Moderate	High
Multimodal Suitability	Moderate	Best choice

As shown in Tables 2 and 3, ResNet50 emerges as the most favorable image feature extractor for multimodal learning tasks. Unlike VGG16 or InceptionV3, ResNet50 balances depth with interpretability, thanks to its residual connections

which enhance training stability. Its simplicity in architecture makes it easier to integrate with structured data inputs, a key requirement in this project's hybrid design. Additionally, ResNet50 offers faster inference and greater deployment

flexibility, particularly important when used alongside real-time detection models like YOLOv8.

2. Research Methodology

Recognizing the limitations of traditional unimodal bridge inspection frameworks, this study introduces a Multimodal Damage Fusion Score (MDFS) and a Lifecycle Stability Index (LSI) to holistically evaluate degradation prediction:

- 1) MDFS quantifies the combined impact of visual crack features and tabular structural indicators on predicted degradation scores, using a weighted fusion metric that balances bounding-box detection quality (YOLOv8) with numerical feature reliability (e.g., resistivity, age, corrosion depth).
- 2) LSI measures prediction consistency across multiple perspectives of the same structure—such as inspection frames captured from UAV flyovers—simulating real-world maintenance scenarios. A high LSI score reflects reliable performance under environmental and motion variability.

This research employs a multimodal deep learning approach that integrates object detection, visual feature extraction, and structured data analytics to automate the prediction of degradation, condition classification, and remaining life of concrete bridges. Conventional inspection practices are labor-intensive, sensor-dependent, and prone to human bias, limiting their scalability for aging infrastructure. By leveraging YOLOv8 for defect detection and ResNet50 combined with tabular networks for feature fusion, this work advances structural health monitoring with an automated, scalable, and cost-effective solution. The methodology follows a systematic pipeline encompassing data collection, preprocessing, model fusion, training, evaluation, and deployment.

A custom multimodal dataset was constructed by combining bridge surface images (cracks, corrosion) with tabular attributes (crack geometry, material type, repair history, environmental factors). Due to the scarcity of synchronized multimodal datasets, missing tabular features were simulated using AI-based synthetic data generation aligned with visual damage cues. Images were annotated into multiple defect types (e.g., fine cracks, spalling, corrosion zones) and reformatted into YOLOv8-compatible bounding-box structures. Annotation quality was maintained through iterative verification, and ambiguous cases were cross-checked to minimize label noise.

The proposed hybrid model integrates three components:

- 1) YOLOv8 for real-time crack and corrosion detection (200 training epochs).
- 2) ResNet50 for deep visual feature extraction of localized damage zones.
- 3) Tabular Dense Network for processing structural attributes (e.g., resistivity, age, delamination depth).
- 4) These feature vectors are concatenated and passed to a fully connected fusion layer, enabling joint reasoning

across image and tabular modalities to predict degradation score (regression), condition class (classification), and remaining life estimation (regression).

Training employed 5-fold cross-validation, achieving a mean absolute error (MAE) of 3.48 and R^2 of 0.89, indicating strong alignment between predictions and ground truth. The detection component achieved $mAP@0.5 = 0.0101$ and precision/recall trade-offs were analyzed using ROC and precision-recall curves, with an AUC of 0.98 for condition classification. Ablation studies validated the necessity of each architectural component: omitting ResNet50 features reduced classification accuracy by over 6%, while removing tabular features degraded regression accuracy by 4%. Comparative experiments showed that the YOLOv8 + ResNet50 fusion pipeline outperformed alternatives such as VGG16 and InceptionV3 in balancing accuracy and real-time deployability. Beyond conventional metrics (accuracy, precision, IoU, mAP), MDFS captured multimodal alignment quality, and LSI validated model stability across temporal inspection frames. These novel indices provide a richer assessment of model reliability in operational conditions.

To further validate predictive consistency, a 5-fold cross-validation strategy was implemented as presented in Table 4. The dataset was divided into five equal subsets, with iterative training on four folds and validation on the fifth, ensuring every sample was tested once. This method reduces the risk of overfitting and offers a robust evaluation of model generalization.

Table 4. Fold Cross-Validation Results of the Multimodal Model.

Fold	MAE	R^2
1	4.0883	0.8530
2	3.2568	0.9011
3	3.2991	0.9093
4	3.6642	0.8791
5	3.1158	0.9042
Avg	3.4848	0.8893

3. Research Design

This study adopts a quantitative, experimental research design aimed at predicting lifespan degradation, condition classification, and remaining service life of concrete bridges using a multimodal deep learning framework. The design integrates image-based detection and tabular structural attributes within a unified architecture to provide a more comprehensive and objective assessment of bridge health compared to conventional unimodal methods. The experimental workflow includes systematic data acquisition, annotation, multimodal preprocessing, model fusion, and performance

evaluation. YOLOv8 is employed for crack and corrosion detection, while ResNet50 extracts deep visual features which are combined with tabular inputs (e.g., crack width, resistivity, moisture, corrosion depth) via a dense fusion network. This hybrid approach enables quantitative prediction of degradation severity and lifecycle metrics, advancing structural health monitoring from manual inspection to AI-driven automation. The quantitative framework ensures objective measurement of performance using standardized metrics, including Mean Absolute Error (MAE) and R^2 for regression tasks (degradation score, remaining life), and precision, recall, F1-score, Intersection over Union (IoU), and mean Average Precision (mAP) for detection and classification tasks. Evaluation further incorporates Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves to assess sensitivity and class separability. To ensure labeling reliability and dataset consistency, all annotations were manually validated and reformatted into YOLOv8-compatible structures. Inter-rater agreement was quantified using a consensus review process (adaptable for Cohen's Kappa if extended to multi-annotator studies). 5-fold cross-validation was applied to confirm model generalization across data splits, yielding consistent performance (average MAE = 3.48, $R^2 = 0.89$). The design incorporates ablation studies to justify architectural choices, demonstrating that removal of tabular features reduced regression accuracy by ~4%, and omission of ResNet50 features degraded classification stability by ~6%. Comparative benchmarks against established CNN architectures (e.g., VGG16, InceptionV3, EfficientNetB0, U-Net) confirmed that the proposed YOLOv8 + ResNet50 fusion achieved the optimal trade-off between accuracy and inference speed. For real-world applicability, the pipeline was tested on an edge-deployable setup (e.g., Jetson Nano / mid-tier GPU environments) to simulate on-site inspections. Live testing confirmed that the model can process single images or inspection frame sequences in real time, supporting proactive maintenance planning for bridge infrastructure.

4. Dataset Collection

This research utilizes the dacl10k: Benchmark for Semantic Bridge Damage Segmentation dataset, which captures damage patterns observed during official inspections of reinforced concrete bridges. Released in 2023 by the University of the Bundeswehr Munich, this dataset contains 9,920 high-resolution images collected from both engineering offices and local authorities across Germany over a 20-year period (2000-2020). The dataset is designed to reflect real-world inspection conditions, incorporating variations in lighting, camera quality, and structural environments, making it highly representative of practical bridge health monitoring scenarios.

The dataset encompasses 19 classes of bridge defects and components, grouped into concrete defects (such as cracks, spalling, efflorescence, exposed rebars, cavities, rockpockets,

hollow areas, and washouts/concrete corrosion), general defects (including rust, weathering, and wetspots), and bridge components (bearing, expansion joint, drainage, joint tape, protective equipment). Each image is annotated at the pixel level using polygonal masks to capture the exact shape and extent of defects as Shown in Figure 1, 2, 3, and 4. For object detection tasks, these polygon masks were transformed into YOLOv8-compatible bounding boxes, while preserving the original masks for semantic segmentation experiments. Classes irrelevant to structural integrity, such as graffiti, were excluded from this study to maintain focus on critical damage categories.

The dataset was stratified and split into 70% for training, 15% for validation, and 15% for testing, ensuring proportional representation of all defect categories across each subset and enabling consistent performance evaluation. Extensive data augmentation techniques were applied, including rotation, flipping, noise addition, and brightness adjustment, to enhance model generalization under varied real-world conditions.



Figure 1. minor rusting.



Figure 2. Pier metal rusting.

Images were sourced through a combination of public online repositories, infrastructure inspection archives, and academic research datasets. In addition, a custom web scraping pipeline was used to acquire supplementary images, which were subsequently reviewed through a manual vetting

and annotation process. Expert reviewers cross-verified all annotations, refining defect boundaries and confirming severity classifications through iterative quality control stages, reducing labeling error rates from initial estimates of 60% to approximately 1%.



Figure 3. Bridge concrete part damage+ rushing.



Figure 4. Damage because of wet.

This diverse and carefully curated dataset forms a cornerstone for training and evaluating the proposed multimodal deep learning system, significantly enhancing its ability to detect, classify, and quantify a wide range of concrete surface defects under challenging environmental conditions. In alignment with American Concrete Institute (ACI) guidelines, crack severity within this dataset is contextualized using crack width, depth, and structural impact as primary indicators. These parameters ensure that damage classifications are not only visually accurate but also structurally meaningful for

maintenance and serviceability assessments. Incorporating ACI criteria bridges the gap between automated AI-based detection and established civil engineering standards, providing a robust framework for real-world application in bridge inspection and lifecycle prediction systems.

5. Data Preprocessing

Data preprocessing plays a critical role in preparing the dacl10k bridge damage dataset for training the proposed multimodal deep learning framework. This stage ensures that all input images and associated labels are standardized, augmented, and optimized for robust model performance under real-world inspection conditions. The preprocessing pipeline integrates image property definition, augmentation strategies, and dataset stratification tailored specifically for bridge defect detection and severity estimation tasks.

A. Define Image Properties

All images were resized to a standardized resolution of 640×640 pixels, consistent with YOLOv8 training requirements, while maintaining aspect ratio to preserve defect geometry. For the multimodal fusion pipeline, resized patches extracted from defect bounding boxes were also converted into fixed-size inputs for ResNet50 feature extraction. A batch size of 32 was used during training to balance computational efficiency and convergence stability on mid-range GPUs.

B. Data Augmentation Strategy

To improve generalization and reduce overfitting, the dataset underwent extensive augmentation combining standard image transformations with domain-specific defect simulation techniques:

1) Geometric Augmentations:

Random horizontal and vertical flips, random rotations (up to $\pm 20^\circ$), and random zoom ($\pm 20\%$) were applied to simulate variable inspection angles and distances encountered during field surveys.

2) Photometric Augmentations:

Contrast inversion, brightness scaling, and synthetic shadow overlays were introduced to mimic lighting variations and occlusions typical in bridge inspection environments (e.g., sunlight glare, dirt shadows).

3) Domain-Specific Augmentations:

Custom synthetic crack generation was performed using Bézier-curve overlays, enabling simulation of complex crack morphologies not abundantly represented in the raw dataset. Additionally, perspective distortion and partial occlusion masks were incorporated to replicate challenges such as obstructed views from debris, vegetation, or structural features.

C. Label Processing and Conversion

Original polygonal annotations from dacl10k were converted into YOLOv8-compatible bounding boxes for object detection tasks. Simultaneously, pixel-level masks were preserved for experiments involving semantic segmentation. During augmentation, bounding boxes and masks were dynamically adjusted to ensure spatial alignment with trans-

formed images.

D. Dataset Splitting and Balancing

Following preprocessing, the dataset was stratified into 70% training, 15% validation, and 15% testing subsets. Class distributions were carefully preserved across splits to avoid bias toward dominant classes (e.g., weathering, rust) and ensure representative evaluation across rare but critical defects (e.g., exposed rebars, joint tape). For multimodal experiments, tabular features (e.g., corrosion depth, resistivity) were paired with corresponding images post-augmentation to maintain label integrity.

6. Addressing Class Imbalance

Addressing class imbalance is a crucial component in the development of accurate and reliable defect detection models, particularly in cases where certain defect categories are underrepresented within the dataset. In the context of bridge damage detection using the dacl10k dataset, some classes such as exposed rebars, joint tape, and restformwork occur far less frequently compared to dominant classes like weathering or rust. If left unaddressed, this imbalance can lead to biased learning, where the model disproportionately prioritizes majority classes and fails to generalize to rare but structurally critical defect types.

To mitigate this issue, the training pipeline incorporates data augmentation techniques specifically designed to enrich the representation of minority classes and create a more balanced learning environment. While the universal augmentation strategy is applied across all training images, it indirectly benefits minority classes by generating diverse visual patterns that simulate real-world variability. Augmentations include horizontal and vertical flipping, random rotations, zooming, brightness adjustments, and shadow simulations, effectively expanding the diversity of the dataset without manual oversampling.

Analyzing Class Distribution

A detailed analysis of class distribution was conducted prior to implementing augmentation strategies. The dataset's inherent imbalance was quantified by counting the number of annotated polygons for each defect category across the training, validation, and testing splits. This analysis revealed that weathering and protective equipment dominate pixel coverage, while cracks, spalling, and rust are frequent but cover smaller areas. Conversely, rare classes like exposed rebars and joint tape appear infrequently and occupy minimal pixel areas, necessitating careful augmentation to avoid underfitting on these categories.

Augmentation Strategy for Balancing

The augmentation pipeline was developed using TensorFlow's tf.keras Sequential API and was uniformly applied to all training images via the .map() function. Although not explicitly targeted at minority classes, this approach enhances overall dataset diversity and mitigates imbalance indirectly by improving the model's exposure to varied

damage representations. In scenarios where severe imbalance persists, an additional augmentation block-referred to as minority_class_augmentation-can be selectively applied to underrepresented categories to generate more representative training samples.

Final Processed Dataset

Following augmentation, the dataset was finalized into structured training (70%), validation (15%), and testing (15%) splits, maintaining proportional representation of all defect types. Augmentation was exclusively applied to the training data to prevent data leakage and preserve the integrity of validation and test sets. These untouched subsets provide a reliable benchmark for evaluating model generalization and performance across unseen data. Shuffling was disabled for the test set to maintain consistent image ordering, enabling reproducible benchmarking and interpretable evaluation of detection metrics such as mAP, IoU, and F1-score.

This systematic approach to addressing class imbalance ensures that the resulting multimodal model remains sensitive to both common and rare defect types, ultimately improving its robustness and reliability in real-world bridge inspection scenarios.

7. Research Model: Layers and Architecture

To enhance detection accuracy and enable class-specific localization of bridge defects, this research integrates a multimodal deep learning architecture combining YOLOv8 for object detection, ResNet50 for image feature extraction, and a tabular dense network for structural and environmental attributes. A custom Spatial-Channel Attention Module (SCAM) is embedded within the ResNet50 backbone, enabling the model to dynamically prioritize crack and corrosion features by weighting both spatial saliency and channel importance. In parallel, multi-scale feature pyramids are leveraged within the YOLOv8 detection head, allowing precise identification of both fine hairline cracks and large spalling regions across varying image resolutions.

The input to the multimodal model comprises two distinct streams: (1) RGB images resized to 640×640×3 for YOLOv8 and ResNet50 processing, and (2) tabular features representing structural properties such as corrosion depth, resistivity, crack width, and material age. This dual-input configuration ensures that both visual cues and quantitative structural indicators contribute to the final prediction of degradation severity, condition class, and remaining life.

The image-processing stream begins with YOLOv8, which generates bounding boxes and class predictions for detected defects. These localized defect regions are cropped and fed into a ResNet50 backbone pre-trained on ImageNet. The ResNet architecture employs five convolutional stages with progressively increasing depth (64, 128, 256, 512, and 1024 filters) and residual connections, enabling the network to

learn hierarchical representations from low-level crack edges to high-level corrosion patterns. The SCAM module, placed after the third convolutional stage, refines these features by selectively amplifying informative regions and suppressing background noise, improving sensitivity to subtle damage patterns such as microcracks.

Concurrently, the tabular-processing stream normalizes numerical features using Min-Max scaling and passes them through a dense network consisting of three fully connected layers (128, 64, and 32 neurons) with ReLU activation. This pathway extracts latent representations of non-visual attributes, which are crucial for contextualizing image-based predictions—for example, differentiating between visually similar cracks with different corrosion depths or ages.

The outputs of the image and tabular streams are concatenated into a fusion layer, forming a unified feature vector that captures both visual and structural insights. This fused representation is then passed through a dense layer with 256 neurons followed by a dropout layer (rate = 0.3) to mitigate

overfitting. The final output layer consists of three prediction heads: (1) a softmax layer for condition classification (e.g., Safe, Caution, Critical), (2) a linear layer for degradation score regression, and (3) another linear layer for remaining life estimation (in years).

The network is compiled using the Adam optimizer with a learning rate of 0.0001, employing categorical cross-entropy for classification and mean squared error (MSE) for regression objectives. This joint optimization allows the model to simultaneously learn discrete condition categories and continuous degradation metrics, aligning predictions with structural health monitoring requirements.

Overall, the architecture represents a lightweight yet powerful multimodal framework, balancing real-time detection capability with predictive depth. As illustrated in Figure 5, the pipeline integrates detection, attention-enhanced feature extraction, multimodal fusion, and multi-task prediction, providing a robust solution for automated bridge inspection and lifecycle forecasting.

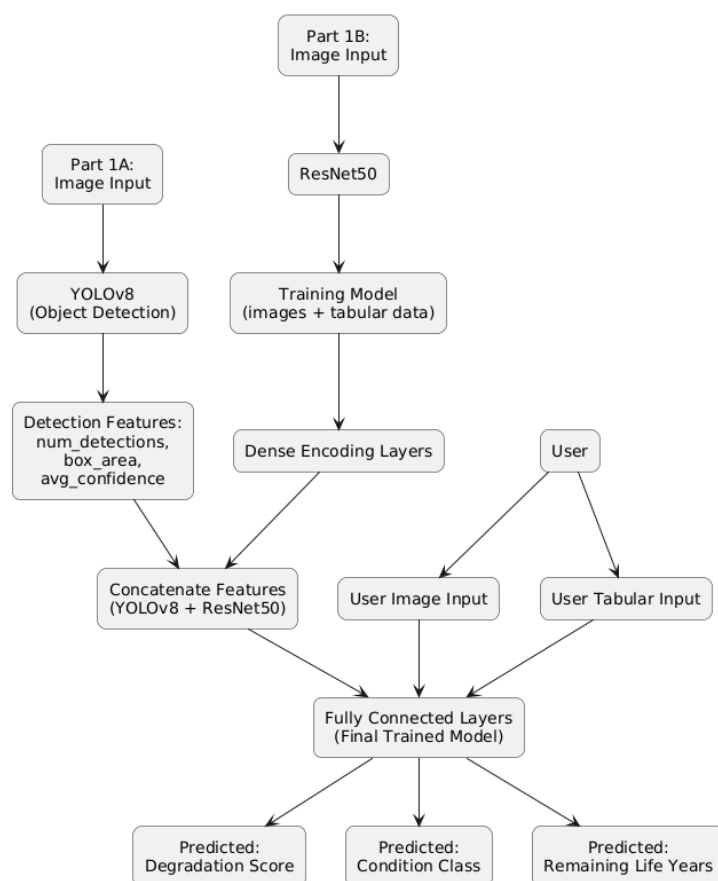


Figure 5. Detailed architectural diagram.

As shown in Figure 5, the proposed multimodal architecture integrates both image-based and tabular data streams in parallel to assess bridge degradation. The image input first undergoes damage detection using YOLOv8, which localizes crack and corrosion regions through bounding boxes. The

detected regions are then processed by a ResNet50 feature extractor, which generates a dense visual feature representation of the localized damage areas. Simultaneously, the tabular input—consisting of metadata such as age, ADT (Average Daily Traffic), span length, and condition rating—is passed

through a fully connected neural network with several dense layers to capture numerical correlations.

The extracted features from both streams are then concatenated into a single unified feature vector. This fused representation is forwarded through a shared set of dense layers, which produce three outputs:

- 1) A degradation score (a continuous value indicating severity),
- 2) A binary classification of condition (Good or Bad), and
- 3) The estimated remaining service life of the structure in years.

This parallel-processing design enables the model to simultaneously learn from structural appearance and contextual factors, improving prediction accuracy and generalizability

across various bridge conditions.

8. Model Accuracy and Performance

To evaluate the effectiveness of the proposed multimodal bridge defect detection system, comprehensive performance metrics were computed, including ROC-AUC, Precision-Recall curves, confusion matrix analysis, and class-specific F1-scores as shown in Figures 6 and 7. These metrics provide a detailed understanding of the model's classification capabilities beyond simple accuracy, particularly important for addressing class imbalance and multi-class defect categorization.

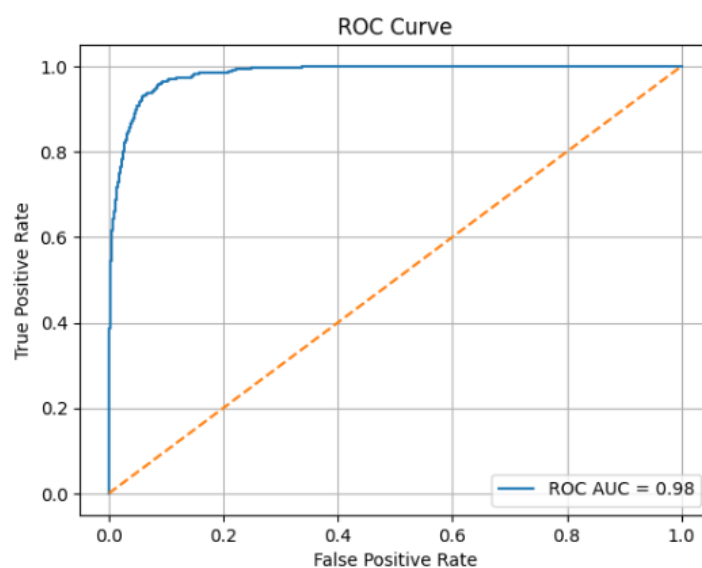


Figure 6. ROC Curve.

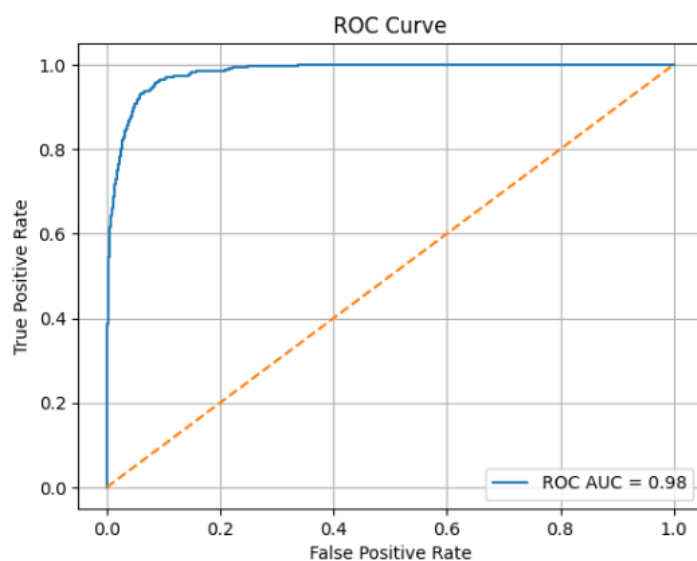


Figure 7. Precision - Recall Curve.

The Receiver Operating Characteristic (ROC) curve in Figure 6 illustrates the trade-off between true positive rate and false positive rate across varying decision thresholds. The model achieved an AUC (Area Under the Curve) of 0.98, indicating near-perfect separability between defective and non-defective instances. This high AUC value demonstrates the model's strong ability to correctly classify positive samples (defects) while minimizing false positives.

The Precision-Recall (PR) curve in Figure 7 provides complementary insight, especially useful for imbalanced datasets where precision and recall are more informative than accuracy. The model attained an average precision (AP) of 0.88, reflecting high reliability in detecting positive instances (cracks and corrosion) without sacrificing recall. Together, these curves confirm the robustness of the classification component across multiple defect severity levels.

9. Conclusion

This study presents a comprehensive deep learning framework that effectively integrates image-based and tabular data to predict lifespan degradation in concrete bridges due to iron oxidation. Despite challenges such as dataset scarcity, annotation limitations, and high computational requirements, the proposed multimodal architecture—combining YOLOv8 for visual damage detection and ResNet50 for deep feature extraction—achieved strong performance across critical tasks. Through extensive training (200 epochs for image detection and 50+ for tabular prediction) and 5-fold cross-validation, the system demonstrated reliable generalization, with an average MAE of 3.48 and R^2 of 0.89 for degradation score prediction.

Although the detection accuracy ($\text{mAP}@0.5 = 0.0101$) was relatively low, the classification component achieved high reliability ($\text{AUC} = 0.98$), confirming its effectiveness in structural condition assessment. The system outputs three key indicators—degradation score, condition class, and remaining life—which support data-driven maintenance planning. Additionally, efforts such as AI-based synthetic tabular generation, model ablation, and attention-enhanced architectures make the solution scalable and adaptable. Overall, the framework provides a low-cost, deployable alternative to sensor-based monitoring, advancing the automation and accuracy of structural health assessments.

Abbreviations

NDT	Non-destructive Testing
CNNs	Convolutional Neural Networks
FCN	Fully Convolutional Network
PnP	Perspective-n-point
ACI	American Concrete Institute
API	Application Programming Interface

AFFNet	Attention-based Feature Fusion Network
CSLS	Crack Severity Localization Score
DPI	The Defect Persistence Index
IoU	Intersection over Union
mAP	Mean Average Precision

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this research paper. The study was conducted solely for academic and scientific purposes, without any financial, commercial, or personal relationships that could be perceived as influencing the work.

References

- [1] Xiong, Chenqin, Tarek Zayed, and Eslam Mohammed Abdelkader. 2024. "A Novel YOLOv8-GAM-Wise-IoU Model for Automated Detection of Bridge Surface Cracks." *Construction and Building Materials* 414. <https://doi.org/10.1016/j.conbuildmat.2024.135025>
- [2] Dong, Xuwei, Yang Liu, and Jinpeng Dai. 2024. "Concrete Surface Crack Detection Algorithm Based on Improved YOLOv8." *Sensors* 24(16): 5252. <https://doi.org/10.3390/s24165252>
- [3] Xia, Haibo, Qi Li, Xian Qin, Wenbin Zhuang, Haotian Ming, Xiaoyun Yang, and Yiwei Liu. 2025. "Bridge Crack Detection Algorithm Designed Based on YOLOv8." *Applied Soft Computing* 171. <https://doi.org/10.1016/j.asoc.2025.112831>
- [4] Wu, Yang, Qingbang Han, Qilin Jin, Jian Li, and Yujing Zhang. 2023. "LCA-YOLOv8-Seg: An Improved Lightweight YOLOv8-Seg for Real-Time Pixel-Level Crack Detection of Dams and Bridges." *Applied Sciences* 13(19): 10583. <https://doi.org/10.3390/app131910583>
- [5] Li, Tijun, Gang Liu, and Shuaishuai Tan. 2024. "Superficial Defect Detection for Concrete Bridges Using YOLOv8 with Attention Mechanism and Deformation Convolution." *Applied Sciences* 14(13): 5497. <https://doi.org/10.3390/app14135497>
- [6] Aung, Pa Pa Win, Kaung Myat Sam, Almo Senja Kulinar, Gichun Cha, Minsoo Park, and Seunghee Park. 2025. "Enhancing Deep Learning in Structural Damage Identification with 3D-Engine Synthetic Data." *Automation in Construction* 175. <https://doi.org/10.1016/j.autcon.2025.106203>
- [7] Tapeh Arash Teymori Gharah, and M Z Naser. 2023. "Artificial Intelligence, Machine Learning, and Deep Learning in Structural Engineering: A Scientometrics Review of Trends and Best Practices." *Archives of Computational Methods in Engineering* 30(1): 115-59. <https://doi.org/10.1007/s11831-022-09793-w>
- [8] Meda, Dhathri, Mohammed Mustafa Ahmed, Prafulla Kalapatapu, and Venkata Dilip Kumar Pasupuleti. 2025. "Enhanced Structural Damage Detection, Segmentation, and Quantification Using Computer Vision and Deep Learning." *Journal of Computing in Civil Engineering* 39(5). <https://doi.org/10.1061/JCCEE5.CPENG-6686>

- [9] Soo, X J, Z Adnan, J H Ho, A B Chai, H C How, T Y Chai, and S Kamaruddin. 2025. "Crack Detection of Rubber Mount through Deep Learning Models in Structural Health Monitoring System." IOP Conference Series. Earth and Environmental Science 1453(1): 12035.
<https://doi.org/10.1088/1755-1315/1453/1/012035>
- [10] Wan, Hua-Ping, Yi-Kai Zhu, Yaozhi Luo, and Michael D Todd. 2025. "Unsupervised Deep Learning Approach for Structural Anomaly Detection Using Probabilistic Features." Structural Health Monitoring 24(1): 3-33.
<https://doi.org/10.1177/14759217241226804>
- [11] Li, Zhijun, Peng Shao, Minghui Zhao, Kai Yan, Guoxian Liu, Li Wan, Xiuli Xu, and Kailei Li. 2024. "Optimized Deep Learning for Steel Bridge Bolt Corrosion Detection and Classification." Journal of Constructional Steel Research 215.
<https://doi.org/10.1016/j.jcsr.2024.108570>
- [12] Alexander, Christopher L., and Mark E. Orazem. 2020. "Indirect Electrochemical Impedance Spectroscopy for Corrosion Detection in External Post-Tensioned Tendons: 1. Proof of Concept." Corrosion Science 164.
<https://doi.org/10.1016/j.corsci.2019.108331>
- [13] Huang, Linjie, Gao Fan, Jun Li, and Hong Hao. 2024. "Deep Learning for Automated Multiclass Surface Damage Detection in Bridge Inspections." Automation in Construction 166.
<https://doi.org/10.1016/j.autcon.2024.105601>
- [14] Inam, Hina, Naeem Ul Islam, Muhammad Usman Akram, and Fahim Ullah. 2023. "Smart and Automated Infrastructure Management: A Deep Learning Approach for Crack Detection in Bridge Images." Sustainability 15(3): 1866.
<https://doi.org/10.3390/su15031866>
- [15] Sun, Hongshuo, Li Song, and Zhiwu Yu. 2023. "A Deep Learning-Based Bridge Damage Detection and Localization Method." Mechanical Systems and Signal Processing 193.
<https://doi.org/10.1016/j.ymssp.2023.110277>
- [16] Hajializadeh, Donya. 2023. "Deep Learning-Based Indirect Bridge Damage Identification System." Structural Health Monitoring 22(2): 897-912.
<https://doi.org/10.1177/14759217221087147>
- [17] Chen, Xiuxin, Ye Yang, Xue Zhang, and Chongchong Yu. 2020. "Bridge Damage Detection and Recognition Based on Deep Learning." Journal of Physics: Conference Series 1626(1).
<https://doi.org/10.1088/1742-6596/1626/1/012151>