Deep Learning Approach for Automatic Detection and Characterization of Structural Cracks

Raphael EHIGIATOR-IRUGHE, Nigeria; Joseph ODUMOSU, Namibia; Hilda MARCUS, Nigeria.

Key words: Structural health monitoring; Crack detection; Deep learning; Mask R-CNN; Geospatial AI; Image segmentation.

SUMMARY

The condition and safety of concrete infrastructure are of growing concern worldwide, as cracks often represent early indicators of structural deterioration, overloading, or material fatigue. Conventional crack inspection methods are largely manual, subjective, time-consuming, and potentially hazardous, limiting their effectiveness for large-scale and continuous structural assessment. This study presents a deep learning–based framework for the automated detection and characterization of cracks in concrete structures using image data, with the aim of supporting safer, faster, and more objective infrastructure inspection practices. A custom dataset consisting of 1,077 high-resolution images of concrete and masonry surfaces was developed and annotated under varying lighting, texture, and environmental conditions. The proposed approach employs a Mask R-CNN architecture integrated with a ResNet-101 backbone and a Feature Pyramid Network (FPN) to enable robust multi-scale feature extraction and pixel-level segmentation. Transfer learning and optimized hyperparameter tuning were applied to improve model generalization and training efficiency on the domain-specific dataset. The results demonstrate strong performance, achieving a detection accuracy of 89.8%, an F1-score of 84.4%, and a mean Intersection over Union (mIoU) of 75.8%. The framework reliably detects and delineates cracks of varying widths and orientations, enabling quantitative measurements of crack geometry such as length, width, and surface area. These capabilities represent a significant improvement over traditional inspection methods, which typically provide only qualitative assessments. The proposed framework offers a practical solution for automated infrastructure inspection and has direct relevance to surveying, geomatics, and structural monitoring applications. It supports objective condition assessment, reduces field risk, and enhances decision-making for maintenance and asset management.

Deep Learning Approach for Automatic Detection and Characterization of Structural Cracks

Raphael EHIGIATOR-IRUGHE, Nigeria; Joseph ODUMOSU, Namibia; Hilda MARCUS, Nigeria.

1. Introduction

The structural integrity of concrete infrastructure is a critical concern in modern civil and geotechnical engineering, as cracks often serve as early indicators of deterioration, material fatigue, or overloading and potential failure in concrete infrastructure (Cha et al., 2017; Sony et al., 2019). Reliable detection and characterization of such cracks are therefore essential for ensuring the safety, durability, and serviceability of built structures. Traditional inspection techniques, such as visual observation, ultrasonic pulse velocity (UPV), or total-station monitoring, while effective in localized contexts, are often time-consuming, subjective, and limited in spatial coverage, making them risky and unsuitable for continuous structural health monitoring (SHM) in large-scale infrastructure (Macchiarulo et al, 2022; Crosetto et al., 2016; Kosh et al, 2015). These limitations have accelerated the need for automated, and non-contact approaches that can provide high-resolution, repeatable measurements of surface deformation and crack progression. With the rise of Artificial Intelligence (AI) and Deep Learning (DL), computer vision techniques now enable objective and automated assessment of civil structures.

Recent advances in remote sensing and computer vision have enabled transformative innovations in SHM through the integration of photogrammetry, LiDAR, and artificial intelligence (AI). Among these, Structure-from-Motion (SfM) photogrammetry has emerged as a particularly effective tool for reconstructing three-dimensional (3D) models of structural surfaces using overlapping two-dimensional (2D) imagery (Eltner & Sofia, 2020). SfM enables the derivation of high-density point clouds comparable to LiDAR data but at significantly lower cost and equipment complexity (Haralick et al, 1979; Zhang & Zhu, 2023). When merged with AI-driven feature extraction techniques, the duo can provide tremendous solution for automated crack detection, localization, and quantification (Mohammadzadeh et al, 2024).

LiDAR-based approaches, though precise, are often constrained by instrument cost, line-of-sight limitations, and lower textural information on concrete surfaces (Abogast et al., 2021). Conversely, SfM photogrammetry offers a texture-rich and geometrically detailed representation that facilitates both geometric and semantic crack interpretation (Das et al., 2024). AI and deep learning models have further advanced the extraction of meaningful damage features from such datasets, and this allows for the automatic detection and classification of micro- and macro-cracks in complex structural geometries (Ghisi et al., 2024; Chen et al., 2025).

Several studies have successfully demonstrated the potential of hybrid approaches that integrate image-based 3D reconstruction with intelligent crack analysis. Early efforts used image processing methods such as edge detection (Canny, 1986), thresholding (Otsu, 1979), and morphological filtering. However, these methods were sensitive to lighting, noise, and shadows (Dorafshan et al.,

2018). Subsequently, Machine Learning (ML) methods such as Support Vector Machine (SVM) and Random Forest (RF) were deployed and these led to improved accuracy (Vapnik, 1999; Breiman, 2001) yet required handcrafted features. The advent of Convolutional Neural Networks (CNNs) revolutionized this field by allowing automatic feature extraction from raw imagery (LeCun et al., 2015; Feng & Feng, 2018).

For structural crack analysis, two-stage detectors like Faster R-CNN and Mask R-CNN have demonstrated superior precision over single-stage models (Ren et al., 2015; He et al., 2017). While Faster R-CNN localizes cracks via bounding boxes, Mask R-CNN extends this by producing pixel-level segmentation masks that enable quantitative measurements such as crack width and area (Yang et al., 2019). Integrating Feature Pyramid Networks (FPN) and ResNet backbone further enhance multi-scale feature detection, improving robustness under diverse lighting and surface conditions.

Building upon recent advancements in deep learning for structural inspection, this study develops a comprehensive automated workflow for detecting and characterizing cracks in concrete structures through the integration of Faster R-CNN and Mask R-CNN architectures. The proposed framework leverages high-resolution image data to enable accurate detection, segmentation, and quantitative assessment of crack geometry for enhanced structural health monitoring.

2. Methodology

2.1 Data Acquisition and Preprocessing

High-resolution images of concrete and masonry surfaces were captured from buildings exhibiting various crack patterns under different lighting and environmental conditions. Each image was manually annotated with bounding boxes for object detection and pixel-level masks for segmentation, forming a comprehensive labelled dataset for supervised learning. The final dataset comprises 1,077 high-resolution images of building surfaces, including concrete, masonry, captured under varying illumination, contrast, and weather conditions. All crack instances were annotated, ensuring no data leakage. The dataset used comprises of various types of cracks as illustrated in Figure 1.

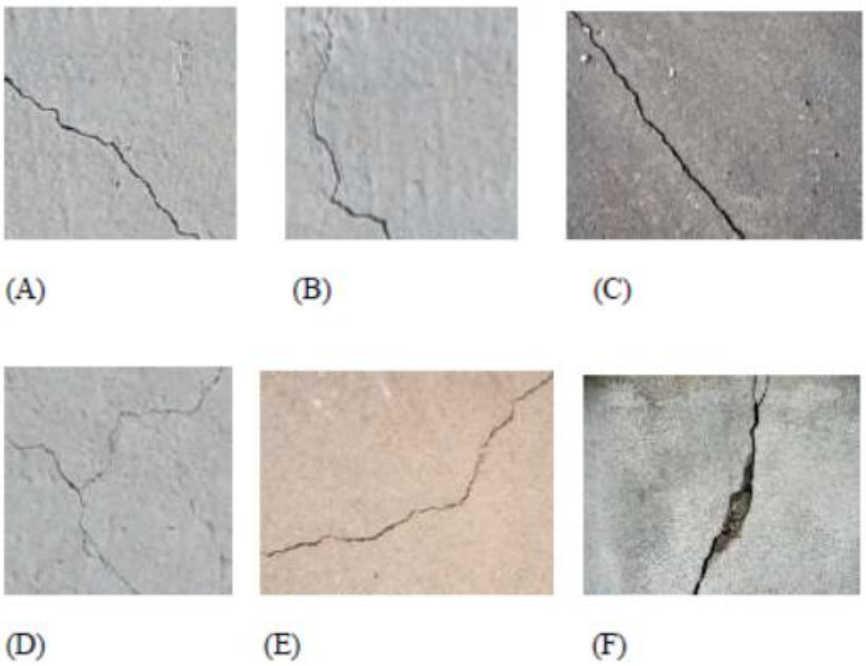


Figure 1: Diversity of the dataset showing (a) grayscale concrete with high contrast crack, (b) grayscale concrete with fine hairline crack having lower contrast, (c) low contrast crack on rough concrete surface, (d) grayscale concrete with high contrast crack, (e) lightening condition on concrete surface, (f) deeper crack on a concrete surface.

The dataset was divided into training (70%), validation (20%), and testing (10%) subsets. To enhance the model's generalization and reduce overfitting, data augmentation techniques were employed, including geometric transformations (rotation, flipping, and scaling) and radiometric adjustments (contrast and brightness normalization). All images were standardized to 1024 × 1024 pixels and normalized to a [0,1] range for consistent feature scaling across input batches.

2.2 Model Architecture

The proposed deep learning framework integrates Faster R-CNN for region proposal generation and Mask R-CNN for precise instance segmentation, allowing both detection and pixel-wise delineation of cracks. The ResNet-101 backbone coupled with a Feature Pyramid Network (FPN) enables robust multi-scale feature extraction for detecting both fine and coarse crack features.

The Region Proposal Network (RPN) was customized with elongated anchors (aspect ratios of 1:3 and 1:5) to improve the detection of thin, elongated crack structures. Pre-trained COCO dataset weights were used to initialize the model, leveraging transfer learning to accelerate convergence and improve training stability on the relatively small, domain-specific dataset.

2.3 Training and Validation Procedure

Model training followed a three-stage approach to ensure stable convergence and optimal performance:

Stage 1: Training of classification and mask heads while freezing the ResNet backbone.

Stage 2: Fine-tuning of upper ResNet layers to enhance high-level feature discrimination.

Stage 3: End-to-end fine-tuning of all network layers using a lower learning rate (1e-4) for optimal weight adaptation.

A multi-task loss function combining region proposal, classification, bounding-box regression, and segmentation losses was minimized during training. Model performance was quantitatively assessed using Precision, Recall, F1-score, mean Average Precision (mAP), and Mean Intersection over Union (mIoU) on the validation and test sets.

2.4 Training Regimen and Hyperparameter Tuning

2.4.1 Transfer Learning Strategy

To mitigate the limitations of a modest custom dataset, transfer learning was employed. The ResNet-101 backbone was initialized using weights pre-trained on large-scale datasets such as ImageNet and COCO, allowing the model to inherit general visual representations. This approach substantially reduced training time and enabled the network to focus on learning domain-specific crack features with limited labelled examples, thereby improving convergence efficiency and accuracy.

2.4.2 Hyperparameters and Loss Function

Model training was configured using optimized hyperparameters: an initial learning rate of 1×10^{-3}, momentum of 0.9, and weight decay (L2 regularization) of 5×10^{-4}. The optimization process minimized a composite multi-task loss function, defined as:

$$L = L_{RPN} + L_{RCNN} + L_{Mask} \quad (1)$$

Where L_{RPN} corresponds to the region proposal loss, L_{RCNN} represents classification and bounding-box regression losses, and L_{Mask} denotes the segmentation noise.

Because cracks occupy only a small fraction of the image area, standard loss functions such as Cross-Entropy or Dice Loss tend to bias learning toward the dominant background class. To address this class imbalance, a Focal Loss function and a weighted Cross-Entropy loss (background-to-crack ratio of approximately 100:1) were implemented. This ensured stable convergence and accurate segmentation of sparse crack pixels.

Empirical observations showed that with appropriate hyperparameter tuning and attention-based optimization, the model achieved significant performance stabilization within approximately 20 training epochs.

2.5 Computational Environment

All experiments were conducted using Python 3.10, TensorFlow 2.10, and Keras, running on an NVIDIA RTX T4 GPU (16 GB) with 32 GB RAM. Training time averaged approximately 3.5 hours per 50 epochs, depending on the augmentation settings and batch size. Matplotlib was employed for visualization purposes.

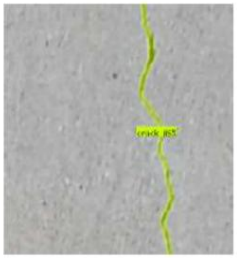
2.6 Evaluation Metrics

Model evaluation was performed using both object-level and pixel-level accuracy measures. Object-level metrics included Precision, Recall, F1-score, and mean Average Precision (mAP), while segmentation quality was assessed using Mean Intersection over Union (mIoU). These metrics ensured comprehensive validation of both detection reliability and segmentation accuracy.

3. Results and Discussion

3.1 Detection and Segmentation Performance

The optimized Mask R-CNN framework demonstrated strong performance across all evaluation metrics. The model achieved a detection accuracy of 89.8%, F1-score of 84.4%, mean Intersection over Union (mIoU) of 75.8%, and a mean Average Precision ($mAP_{50–95}$) of 79.2%. Figure 3 shows some crack detection and segmentation results obtained from the model across various challenging conditions.

(a) Thin crack on concrete surface (CI = 85%)


(b) deep crack on rough concrete (CI = 81%)


(c) reduced CI (66%) in deep continuous crack


(d) smooth, dry, dusty surface (CI = 61%)


(e) rough decaying wall (C.I = 74%)


(f) Low CI (56%) due to poor lightening

Figure 3: Some results of cracks identification by the model

These results confirm the robustness and reliability of the framework in detecting and delineating cracks across diverse illumination conditions, surface textures, and background noise. Visual inspection of the predicted segmentation masks revealed that the model effectively captured fine crack geometries with high boundary precision while minimizing false detections in shadowed or unevenly illuminated regions. The integration of the Feature Pyramid Network (FPN) proved particularly beneficial in preserving multi-scale feature consistency, enhancing the model's sensitivity to thin and elongated crack structures.

Despite the strong overall performance, a few instances of missed or incorrect detections were recorded (Figure 4). These limitations were observed under three main conditions:

(i) False negatives, occurring when fine or faint hairline cracks were not detected due to low contrast or poor illumination.

(ii) False positives, where dark stains or wet surface patches were erroneously classified as cracks; and

(iii) Under-segmentation, observed in regions containing dense or intersecting crack networks that the model merged into a single instance.

These error patterns highlight the challenges of distinguishing subtle surface discontinuities from background textures and emphasize the need for adaptive thresholding and improved contextual learning in future model iterations.

Figure 4: (a) False negative (b) False positive (c) Under segmentation

3.2 Model Performance Enhancement through FPN and Deep Backbone Integration

The integration of a Feature Pyramid Network (FPN) with a deeper ResNet-101 backbone produced a substantial improvement in detection and segmentation accuracy, reflected by an approximate 15% increase in the primary $mAP_{50–95}$ metric. The integration improved both detection accuracy and segmentation precision, especially for small or discontinuous crack regions (Tables 1 and 2).

Table 1: Detection and classification performance metric

Model Configuration	Precision (%)	Recall (%)	F1 – score (%)	Accuracy (%)

Faster R-CNN (ResNet-50)	82.4	79.6	78.8	85.2
Masked R-CNN (ResNet-101)	88.7	89.5	84.4	89.8

This outcome shows that multi-scale feature extraction enabled by FPN is essential for capturing cracks of varying dimensions. The enhanced architecture demonstrated superior capability in detecting fine, small-scale cracks using high-resolution feature maps from shallow layers, while simultaneously preserving robust detection of larger, more pronounced cracks through semantically enriched features from deeper layers (Figure 5). The high mAP_{50} further shows the model's excellent recall and precise localization performance across diverse crack patterns for both the Faster R-CNN and masked R-CNN.

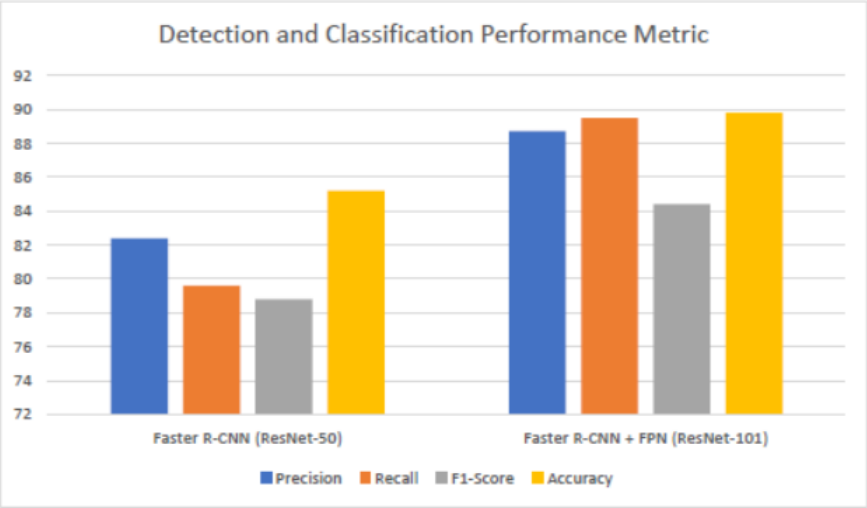


Figure 5: Detection and Classification Performance Metric

Table 2 presents result of the instance segmentation performance. The segmentation results closely reflect the detection performance, with the optimized Mask R-CNN achieving a Mask $mAP_{50–95}$ of 0.489 and a mean Intersection over Union (mIoU) of 0.758.

Table 2: Instance segmentation Performance

Model Configuration	mAP_{50-95}	mAP_{50}	m_{IoU}
Faster R-CNN (ResNet-50)	0.421	0.745	0.712
Masked R-CNN (ResNet-101)	0.489	0.792	0.758

The high mIoU value indicates a strong correspondence between the predicted segmentation masks and the ground truth annotations, enabling precise quantification of crack geometry, including parameters such as length, width, and surface area. This pixel-level accuracy (Figure 6) represents a significant advancement over conventional detection-only approaches, as it provides detailed geometric information essential for predictive maintenance and long-term structural health monitoring (SHM) applications.

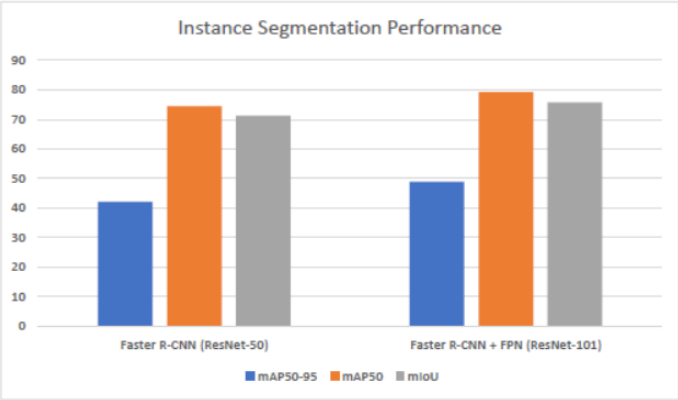


Figure 6: Instance segmentation performance

While the Mask R-CNN incurred a modest reduction in inference speed (averaging approximately 5 frames per second) due to its two-stage architecture, this trade-off was compensated by a significant gain in spatial precision and semantic segmentation quality. For structural health monitoring applications where accurate delineation and quantification of crack geometry are more critical than real-time speed, the increased computational cost is acceptable and justified.

3.3 Computational Feasibility and Practical Application

Model training and evaluation were conducted on an NVIDIA T4 GPU (16 GB VRAM) using TensorFlow 2.10 and Keras frameworks. The average training time per 50 epochs was approximately 3.5 hours, depending on augmentation configuration and batch size. Inference testing achieved near real-time performance (approximately 5 fps), indicating the method's potential for on-site deployment when integrated with UAV-based data acquisition systems.

The framework's generalization capability across multiple datasets and lighting conditions suggests its suitability for automated infrastructure inspection, bridge deck monitoring, and building façade assessment. Furthermore, its quantitative outputs, such as crack width, length, and spatial distribution, provide a reliable basis for predictive maintenance and decision support within structural health monitoring workflows.

4.0 Conclusion, Recommendations, and Future Work

4.1 Conclusion

This study successfully developed and validated a deep learning–based automated framework for the detection and characterization of structural cracks in concrete surfaces. By integrating the Mask R-CNN architecture with a ResNet-101 backbone and Feature Pyramid Network (FPN), the system achieved high detection accuracy and precise segmentation performance ($mAP_{50–95}$ = 0.489; mIoU = 0.758), showing robustness across diverse illumination and texture conditions.

The developed workflow enables pixel-level delineation and quantitative geometric analysis of cracks, including width, length, and propagation. This represents a significant advancement over traditional inspection methods and also eliminates the subjectivity of manual assessment, minimizes field risk, and enhances predictive maintenance and Structural Health Monitoring (SHM) capabilities.

The integration of FPN and a deeper ResNet-101 backbone proved effective in enhancing multi-scale feature extraction, and enables reliable detection of both fine hairline cracks and major structural fractures. While minor misclassifications occurred in shadowed or stained regions and under-segmentation was noted in dense crack networks, overall system performance reveals its applicability for large-scale SHM applications.

4.2 Recommendations

Based on the research findings, the following recommendations are made:

Further investigation and refinement of this framework so that it can be used by regulatory bodies and engineering institutions for large-scale inspection of bridges, pavements, and retaining walls.

The framework should be integrated into existing Building Information Modeling (BIM) and asset management systems to support automated defect mapping and predictive maintenance scheduling.

Collaborative efforts should be made by researchers and industry towards developing standardized annotated datasets for cracks of varying scales, orientations, and materials to support continuous model improvement.

4.3 Future Work

Future studies will aim to enhance the developed framework through the following directions:

(i) Temporal Monitoring by integrating time-series analysis to track crack propagation and deformation trends over successive inspection epochs.

(ii) Incorporation of integrated multispectral and thermal sensors to improve discrimination between cracks, stains, and surface discolorations.

(iii) Combining UAV-based photogrammetry and Structure-from-Motion (SfM) techniques for real-time, 3D cracks mapping of large infrastructures via 3D reconstruction

Acknowledgement:

The authors gratefully acknowledge Mr. Chris Leonard Ehis, an undergraduate student at the University of Benin, for his contributions to the foundational aspects of this work, which formed part of his undergraduate dissertation.

References

Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. *Computer-Aided Civil and Infrastructure Engineering, 32* (5), 361–378

Sony, S., Laventure, S., & Sadhu, A. (2019). A Literature Review of Next-Generation Smart Sensing Technology in Structural Health Monitoring. *Structural Control and Health Monitoring, 26* (3), e2321.

Macchiarulo, V., Milillo, P., Blenkinsopp, C., & Giardina, G. (2022). Monitoring deformations of infrastructure networks: A fully automated GIS integration and analysis of InSAR time-series. *Structural Health Monitoring, 21* (4), 1849-1878. DOI:10.1177/14759217211045912

Crosetto, M., Monserrat, O., Cuevas-González, M. Devanthéry, N., Crippa, B., 2016. Persistent Scatterer Interferometry: a review. ISPRS Journal of Photogrammetry and Remote Sensing, 115, 78-89

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. (2015). A Review on Computer Vision Based Defect Detection and Condition Assessment of Concrete and Asphalt Civil Infrastructure. *Advanced Engineering Informatics, 29*(2), 196–210.

Eltner, A., & Sofia, G. (2020). Structure from Motion Photogrammetric technique; In Developments in Earth Surface Processes by Tarolli, P., & Mudd, S. M. (Eds). Elsevier

Haralick, R.M, Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern. SMC-3*, 610–621.

Zhang, Z., & Zhu, L. (2023). A Review on Unmanned Aerial Vehicle Remote Sensing: Platforms, Sensors, Data Processing Methods, and Applications. *Drones*, *7*, 398.

Mohammadzadeh, M., Kremer, G.E.O., Olafsson, S., & Kremer, P.A. (2024). AI-Driven Crack Detection for Remanufacturing Cylinder Heads Using Deep Learning and Engineering-Informed Data Augmentation. *Automation*, 5, 578–596. https://doi.org/10.3390/ automation5040033

Abogast, K., Awadaljeed, M., Chatzi, E., Clinton, J., Rossi,Y., Rothacher, M., & Tatsis, K. (2021). Kalman Filter-Based Fusion of Collocated Acceleration, GNSS and Rotation Data for 6C Motion Tracking. Sensors, 21(4), 1543. https://doi.org/10.3390/s21041543

Das, A., Mohamed, S. M., Mukhopadhyay, S., Nagarajaiah, S., & Pal, A. 2024. Signal-based online acceleration and strain data fusion using B-splines and Kalman filter for full-field dynamic displacement estimation. https://doi.org/10.48550/arXiv.2411.19282

Ghisi, A.F., Mariani, S., Qiu, H., & Rosafalco, L. (2024). Structural health monitoring strategy based on adaptive Kalman filtering, in Proceedings of the 11th International Electronic Conference on Sensors and Applications, 26–28 November 2024, MDPI: Basel, Switzerland, doi:10.3390/ecsa-11-20493

Chen,C., Han, H., Li, D., Wang, J., Wang, L., & Xiao, X. (2025). Dynamic Deformation Analysis of Super High-Rise Buildings Based on GNSS and Accelerometer Fusion. Sensors, 25(9), 2659. https://doi.org/10.3390/s25092659.

Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8*(6), 679–698.

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics, 9*(1), 62–66.

Dorafshan, S., Thomas, R. J., & Maguire, M. (2018). Comparison of Deep Convolutional Neural Networks and Edge Detectors for Image-Based Crack Detection in Concrete. *Construction and Building Materials, 186*, 1031–1045.

Vapnik, V. N. (1999). The Nature of Statistical Learning Theory (2nd ed.). Springer.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

Feng, C., & Feng, M. Q. (2018). Vision-Based Concrete Crack Detection Using a Convolutional Neural Network. *Proceedings of the 2018 ASCE International Conference on Engineering, Construction, and Operations in Challenging Environments*

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature, 521*(7553), 436–444

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems (NeurIPS), 28*.

Yang, X., Li, H., Yu, Y., Luo, X., Huang, T., & Yang, X. (2019). Automatic Pixel-Level Crack Detection and Measurement Using Fully Convolutional Network. *Computer-Aided Civil and Infrastructure Engineering, 34*(8), 616–634.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *IEEE International Conference on Computer Vision (ICCV)*.

BIOGRAPHICAL NOTES

CONTACTS

Prof. Raphael EHIGIATOR-IRUGHE
University of Benin.
Department of Geomatics,
Faculty of Environmental Sciences,
University of Benin,
Benin City, Edo State
NIGERIA
+234 803 368 1019
raphael.ehigiator@uniben.edu

Dr. Joseph ODUMOSU
Namibia University of Science and Technology.
Department of Land and Spatial Sciences (Geomatics Division),
Faculty of Engineering and Built Environment,
Namibia University of Science and Technology
Windhoek,
NAMIBIA
+264 81 84 319 32
jodumosu@nust.na

Hilda MARCUS
University of Benin.
Department of Geomatics,
Faculty of Environmental Sciences,
University of Benin,
Benin City, Edo State
NIGERIA
+234 818 287 0709
hilda.marcus@uniben.edu