HPC-enabled Deep Learning Multi-model for Road Damage Detection

Prakhar Verma[1*], Shreya Chaturvedi[2], Sivakumar V[1], Sajeevan G[1]

[1]Centre for Development of Advanced Computing (C-DAC), Pune, Maharashtra, India

[2]Student of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology, Pune, Maharashtra, India

*prakharv@cdac.in

Keywords: HPC, Deep Learning, AI/ML, Object Detection, Road Damage, Predictive Maintenance

Abstract

Predictive maintenance for road infrastructure is a way-forward approach that replaces traditional reactive and routine maintenance with insights derived from AI. Accurate prediction of surface defects facilitates timely intervention, prioritizes repair effort, and optimizes cost. However, it is challenging due to lack of appropriate labelled dataset, larger data volume, photographic angle, poor resolution and surface types. In this paper, we propose a deep-learning pipeline that seamlessly integrates task-specific models for detection of pothole and waterlogging along with a crack detection model trained on a subset of publicly available RDD2022 dataset. GhostConv was adopted to enhance the YOLOv11 for crack detection. The model was trained on 640 x 640 resolution image with standard augmentations like flipping, scaling, hue and saturation using C-DAC PARAM Siddhi-AI HPC system. The augmentation techniques improved the model robustness across varied light and surface conditions. This multi-model approach allows new damage categories to be incorporated easily by retaining individual components without changing the core system. The results indicate that the multi-model pipeline is capable of detecting road damages such as pothole, waterlogging, longitudinal, transverse and alligator cracks from images. Our output model can be seamlessly integrated with online geospatial applications such as GeoSevak and GeoSadak (PMGSY National GIS), that can result in incredible increase in the implementation efficiency and transparency.

1. Introduction

The road networks serve as the backbone of economic growth and social mobility. Traditional maintenance strategies, which rely on visible damage or scheduled inspections at fixed intervals, are both inefficient and costly. By the time repairs are scheduled, damage often progresses to a severe stage, resulting in higher restoration expenses and increased safety risks. Modern predictive maintenance offers a transformative alternative by applying artificial intelligence to road photographs and video. This approach can detect early-stage distress patterns, such as micro cracks, emerging potholes and waterlogging, which helps in scheduling repairs before problems worsen. However, automating surface damage detection presents significant challenges. Real-world images differ widely in lighting, weather, camera angle and resolution. Additionally, Indian roads comprise a wide range of surface types. These factors make it difficult to develop a single model that performs reliably under all scenarios. Moreover, the scarcity of large, diverse labeled datasets forces training on limited examples, which limits the model's ability to generalize. This paper adapts a multi-model deep learning pipeline designed to detect cracks, potholes, and waterlogging on Indian roads.

2. Related Work

Automated road-damage detection methods relied on manually designed descriptors such as Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT) key points, which were classified via support vector machines (Dalal and Triggs, 2005; Lowe, 2004; Chang and Lin, 2011). Achieving invariance to scale, illumination and occlusion required exhaustive hyperparameter tuning, and model efficacy deteriorated when inputs deviated from the narrow conditions represented in the training data. The inherent rigidity of these deterministic feature extractors led to a shift toward deep convolutional neural networks, which perform end-to-end representation learning directly from raw pixel intensities. Stacked convolutional layers learn to abstract spatial edges, textures and patterns across multiple scales, which makes convolutional neural networks inherently robust to changes in illumination, viewpoint and surface material (Krizhevsky et al., 2012; He et al., 2016). Modern detection architectures leverage these learned feature hierarchies to deliver superior generalization across diverse operational environments, eliminating the need for manual feature design.

Eisenbach et al. (2017) focused on making pavement datasets for deep-learning training, emphasizing label consistency and class balance. Seichter et al. (2018) proposed pre-annotation pipelines and active learning strategies to reduce labeling effort that effectively minimized human intervention during training data preparation. Maeda et al. (2018) released one of the earliest large-scale road damage datasets captured via smartphones and evaluated SSD-based real-time object detection. Their Road Damage Detection (RDD) dataset laid the groundwork for benchmarking future architectures. Chun and Ryu (2019) applied fully convolutional networks (FCNs) in a semi-supervised setting to improve segmentation when ground truth masks are limited, showing improvements in generalization.

Meng (2021) proposed crack segmentation with a ResNet101-based end-to-end model, demonstrating improved refinement of edges and generalization to various concrete environments. The network deep residual design helps capture multiscale features crucial for semantic segmentation in variable lighting and texture scenarios. A UAV-based approach by Silva et al. (2023) validated the effectiveness of combining aerial imagery and deep detectors (YOLOv5/YOLOv7) for large-scale road surveys, emphasizing the importance of overhead context in infrastructure analysis.

Han et al. (2020) proposed GhostNet, which uses GhostConv models to reduce redundant feature maps. This innovation enabled efficient inference without sacrificing semantic richness. Lightweight and mobile-efficient models are becoming central in real-world deployment. Wu et al. (2023) introduced YOLO-LWNet, a compact variation of YOLOv5 that integrates attention models for improved localization on mobile hardware. They demonstrated that the architecture balanced model complexity and detection precision on RDD-2020 datasets. Ding et al. (2023) improved YOLOv8 for road damage by reconfiguring the neck and detection layers for clearer boundary prediction and fewer false positives. Zeng and Zhong (2024) proposed YOLOv8-PD, integrating BoTNet and GhostNet-derived C2fGhost models to improve detection of fine-scale pavement distress, especially longitudinal cracks. The use of lightweight attention blocks and an LSCD-Head enhances both detection precision and inference efficiency, showing a 4.2% gain in mAP on RoadDamage dataset.

Zhang et al. (2024) introduced CPCDNet, which leverages dynamic snake convolutions and a crack-align model to preserve the continuity of elongated cracks. Their work emphasizes geometric preservation in segmentation, offering insights into crack edge fidelity. Ma et al. (2025) introduced the UDTIRI-Crack benchmark and reviewed SoTA deep learning models, comparing semantic segmentation and object detection approaches across multiple datasets. Their results underscore the gap in generalization and the need for lightweight yet accurate models for deployment in autonomous road inspection vehicles. Li et al. (2025) extended YOLO with attention and fuzzy information theory to refine crack segmentation masks. Their method achieved both reduced model size and improved edge sharpness which is critical for outlining narrow road cracks.

Collectively most of these studies are region-specific, limiting their applicability to broader geographic contexts and not on generic dataset. Furthermore, the dataset used generally have limited variation in road conditions, particularly in terms of different illuminations and surface types. Our work extends these studies by integrating multi-model detectors (pothole, waterlogging, crack), enhancing the backbone, and decoupling damage types to improve cross-condition robustness in Indian terrains.

3. Data and Methods

3.1 Data

The proposed pipeline consists of damage specific models for cracks, potholes and waterlogging. The pothole and water logging are in-house trained models developed on custom datasets captured under diverse urban and rural conditions with seasonal surface variations. The crack detection model is an enhanced YOLOv11 model trained on the Indian subset of the RDD2022 dataset. The crack detection model leverages the RDD2022 India subset, downloaded from the GitHub repository (https://github.com/sekilab/RoadDamageDetector). This subset contains 2047 training images and 512 validation images captured across varied Indian road surfaces. Crack instances are categorized into longitudinal, transverse and alligator cracks.
The pretrained pothole and waterlogging models integrated in our framework was trained on the PMGSY rural roads citizen dataset, captured using mobile application. The data was accessed from GRRIS (https://www.pmgsy-grris.nic.in/). From this dataset, 2000 and 1000 images were utilized for training pothole and waterlogging models, respectively.

3.2 Preprocessing

To mitigate class imbalance, all crack images were resized to 640 × 640 px and normalized to [0, 1]. We applied a class-weighted augmentation scheme, each image received two base augmentations plus extra copies based on crack type i.e. three additional for longitudinal and transverse cracks, one additional for alligator cracks. The pipeline included random rotation, illumination adjustment, horizontal flip, vertical flip, perspective distortion, ISO noise injection, and grid dropout.

3.3 Modified YOLOv11

The crack detection backbone is an improved YOLOv11 with scale “x” architecture optimized for enhancing the prediction accuracy of small road damages. We integrated GhostConv layers in the early and mid-level stages to cut parameter redundancy. C3k2 blocks enrich multi-scale feature aggregation with minimal additional FLOPs. A C2PSA model provides lightweight channel-spatial attention to boost sensitivity to global context. Features from different scales are merged through an FPN-style neck before separate classification and localization branches in the decoupled detection heads. Our model accepts 640×640 inputs and comprises 51.7 million parameters.

3.4 Training Setup

Model training was performed on HPC PARAM Siddhi-AI, equipped with four GPU NVIDIA A100-SXM4 (40 GB) under CUDA 12.4. The software environment comprised Ultralytics detection framework v8.3.158, Python 3.12.3, and PyTorch 2.5.0+cu124. Training parameters were configured as follows:

- **Batch size**: 4 images per iteration on a single GPU
- **Epochs**: 200 with early stopping (patience = 15)
- **Optimizer**: AdamW with initial learning rate lr0 = 0.001, final factor lrf = 0.01, weight decay = 0.0005
- **Precision**: Automatic mixed precision (AMP = True)
- **Input size**: multi-scale resizing to 640×640 px

Training ran for 117 epochs (8 hours 46 minutes), with the best weights reached at 102 epoch.

3.5 Multi Model Integration

Crack, pothole and waterlogging each present unique visual characteristics and structural complexity. We integrated our crack model with C-DAC pothole and waterlogging detection models. We trained and tested the model on RDD dataset for identifying crack and pothole but prediction was not satisfactory, indicates that relying on a single model/dataset can limit accuracy. To overcome this, we introduce a multi-model integration pipeline. Each model is independently trained and optimized for a specific category of road damage, and together they provide a comprehensive assessment. The pipeline includes the following:

- A crack detection model that identifies longitudinal, transverse, and alligator cracks. This model uses object detection to draw bounding boxes around damaged regions
- A pothole segmentation model that outputs pixel-level masks to capture the irregular geometry and extent of potholes
- A waterlogging segmentation model that highlights waterlogging, which can vary greatly in shape depending on road conditions and lighting

Each model processes the input image separately, and the results are combined through a coordinated visualization pipeline.

4. Results and Discussion

The YOLO11x-GhostConv crack detection model was evaluated on a validation set of 512 images comprising 797 annotated crack instances.

4.1 Quantitative Evaluation

F1–Confidence Curve shows that score peaks at 0.51 for a confidence threshold of 0.434, indicating an optimal trade-off between precision and recall when tuning the detection threshold.

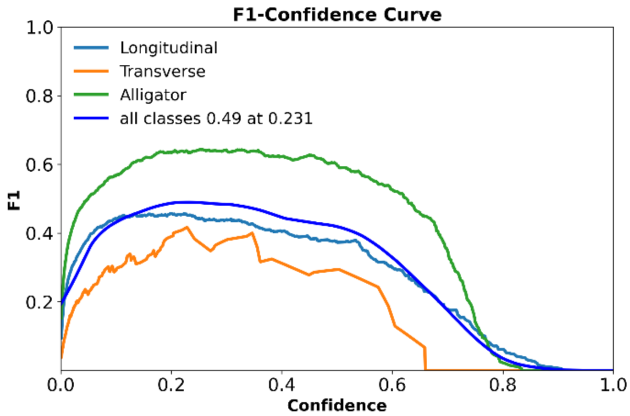

Figure 1: F1–confidence curve

4.2 Confusion Matrix Analysis

Analysis of the normalized confusion matrix in Figure 2 shows the model demonstrated strong class-specific learning, particularly for the alligator crack category, achieving a high true positive rate (0.71), indicating its potential to effectively localize complex crack patterns under favourable conditions. However, the model misclassified background pixels as crack instances, generating 246 false positives for the longitudinal class and 108 for the alligator class. This bias contributed significantly to overall error rates and reflected the inherent difficulty of distinguishing crack patterns from uniform pavement textures. Conversely, the network failed to detect a substantial number of true crack pixels, with 170 longitudinal cracks and 222 alligator cracks erroneously assigned to the background class.

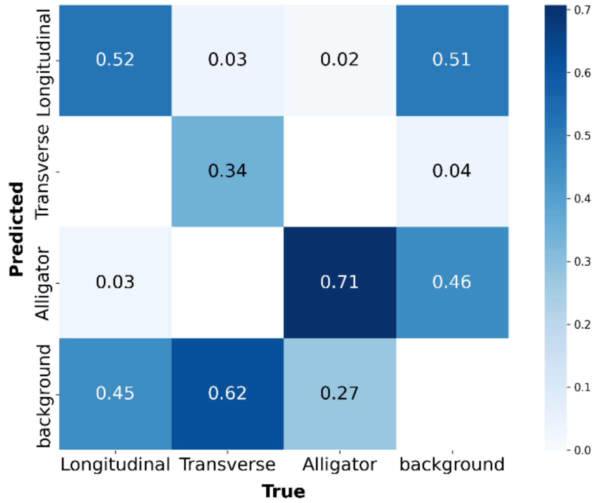

Figure 2: Normalized confusion matrix

4.3 Model Performance

Figure 3 illustrates the progression of training and validation loss components along with key performance metrics across 120 epochs. The training losses, shows a consistent and smooth downward trend that indicates stable optimization and effective convergence of the model. An important early-phase behavior is the rapid reduction in training DFL and classification loss, suggesting that the model quickly learns the structural patterns associated with different crack types.

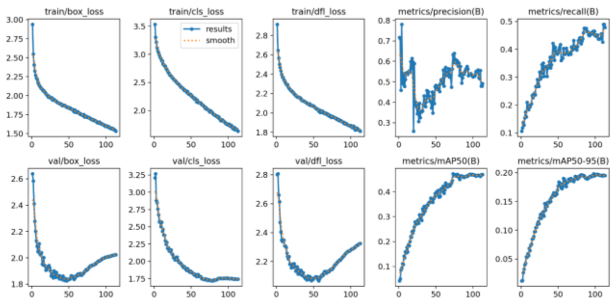

Figure 3: Model performance

The performance metrics are as follows:

- **Precision (B)** fluctuated initially but stabilized beyond 40 epoch, which shows moderate confidence in predictions
- **Recall (B)** steadily improved throughout training, reached **0.48**, shows enhanced ability to detect true positives as the model matures
- **mAP50 (B)** and **mAP50-95 (B)** both show a consistent upward trend, saturating after 90 epoch at around **0.56** and **0.20** respectively. This indicates improved overall object localization and classification performance over time

4.4 Qualitative Assessment

YOLO11x-GhostConv architecture delivered strong performance on alligator cracks (mAP50 = 0.673), benefiting from GhostConv efficient feature generation and the lightweight channel-spatial attention (C2PSA) that enhanced context sensitivity. The moderate performance on longitudinal cracks (mAP50 = 0.417) shows that elongated patterns are well captured by the hierarchical feature fusion. The remarkably low recall for transverse cracks (R = 0.310) highlighted the adverse effects of class imbalance and the scarcity of representative training samples. Table 1 shows performance metrics.

Metric	Crack			
	All	Longitudinal	Transverse	Alligator
Instances	797	375	29	393
Precision	0.63	0.52	0.66	0.7
Recall	0.44	0.39	0.31	0.62
mAP50	0.49	0.41	0.40	0.67
mAP50–95	0.21	0.18	0.12	0.32

Table 1: Detection performance metrics

Figure 4 shows the pothole prediction. The red outlined polygon is the exact shape where the pothole is present.

Figure 4: Pothole detection

Figure 5 shows Waterlogging detection. The blue polygon outline's the waterlogged area on the road surface.

Figure 5: Waterlogging detection

Figure 6 shows crack detection. The model detects alligator cracks (yellow box, 0.67 confidence) and longitudinal cracks (green box, 0.37 confidence), outputting class-specific bounding boxes and scores for damage mapping.

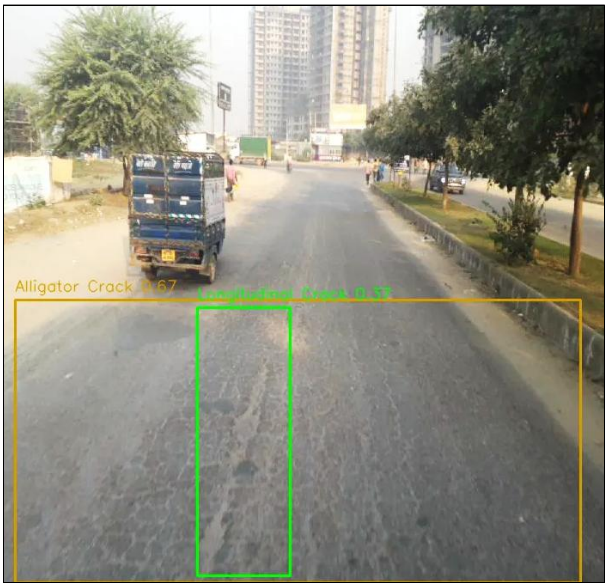

Figure 6: Crack detection

Figure 7 shows the integration of multi-model pipeline. Crack detections are overlaid as labeled boxes using distinct colors for each class. Pothole and waterlogging masks are converted into binary images and processed with contour extraction to preserve their spatial outlines. These contours are then rendered on the input image, ensuring that each damage type is clearly visible and correctly localized.

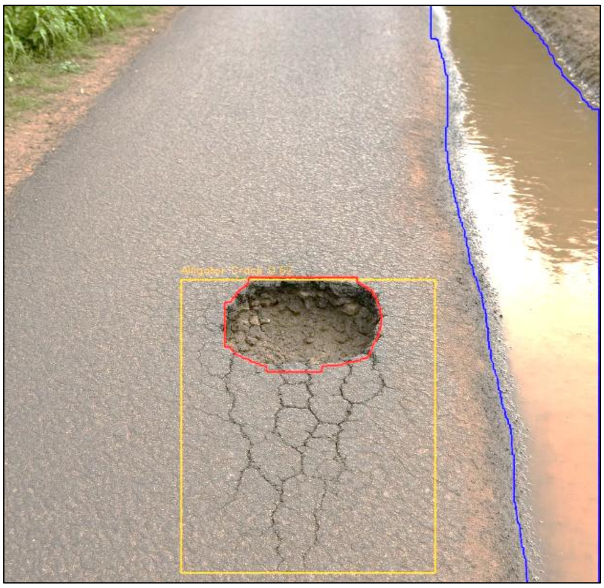

Figure 7: Multi-model detection. AI generated input image

The pipeline integrated three models to detect different types of road damage. Cracks categorized as longitudinal, transverse, and alligator which were identified using an object detection model that draws class-specific bounding boxes. Potholes and waterlogged regions were segmented at the pixel level to capture their irregular shapes. Each model processed the image independently, and the outputs are fused through a unified visualization pipeline.

5. Conclusions

This study presented an HPC-enabled, decoupled multi-model pipeline for road damage detection, combining a GhostConv-enhanced YOLOv11x crack detector with in-house pothole and waterlogging segmentation models. The integration of these damage-specific models enhanced the overall performance of the pipeline by enabling accurate and comprehensive detection of diverse road surface defects. Our results showed that accurate prediction of the irregular geometry and varying appearance of these damage types complemented the crack detection output and enabled a more comprehensive assessment of road conditions. Future scope of study would be addressing class imbalance and suppressing background noise. Migrating all detectors to segmentation-based architectures promises superior performance in real-world maintenance planning and deployment. The proposed pipeline is recommended for predictive maintenance of road infrastructure to enhance key performance indicators such as operational efficiency, road safety, transparency in maintenance planning, cost-effectiveness, timely intervention, and data-driven decision-making.

Acknowledgements

Authors gratefully acknowledge C-DAC for funding this R&D and providing access to PARAM Siddhi-AI HPC facilities.

References

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), Article 27:1–27:27. https://doi.org/10.1145/1961189.1961199

Chun, C., & Ryu, S.-K. (2019). Road surface damage detection using fully convolutional neural networks and semi-supervised learning. *Sensors, 19*(24), 5501. https://doi.org/10.3390/s19245501

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 886–893). IEEE. https://doi.org/10.1109/CVPR.2005.177

Eisenbach, M., Stricker, R., Seichter, D., Amende, K., Debes, K., Sesselmann, M., Ebersbach, D., Stoeckert, U., & Gross, H.-M. (2017). How to get pavement distress detection ready for deep learning? A systematic approach. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 2039–2047). IEEE. https://doi.org/10.1109/IJCNN.2017.7966101

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). GhostNet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1580–1589). https://doi.org/10.1109/CVPR42600.2020.00165

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778. https://doi.org/10.1109/CVPR.2016.90

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.

Li, Q., Wu, T., Xu, T., Lei, J., & Liu, J. (2025). A novel YOLO algorithm integrating attention mechanisms and fuzzy information for pavement crack detection. *International Journal of Computational Intelligence Systems, 18*, Article 158. https://doi.org/10.1007/s44196-025-00894-5

Li, Y., Chen, X., & Zhang, W. (2024). RDD-YOLO: Road damage detection algorithm based on improved YOLOv8. *Applied Sciences, 14*(8), 3360. https://doi.org/10.3390/app14083360

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Ma, N., Song, Z., Hu, Q., Liu, C. W., Han, Y., Zhang, Y., ... & Xie, L. (2025). Vehicular Road Crack Detection with Deep Learning: A New Online Benchmark for Comprehensive Evaluation of Existing Algorithms. https://doi.org/10.48550/arXiv.2503.18082

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., & Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. In *Proceedings of the IEEE International Conference on Big Data* (pp. 1–6). https://doi.org/10.1109/BigData.2018.8621824

Meng, X. (2021). Concrete crack detection algorithm based on deep residual neural networks. *Scientific Programming, 2021*, 3137083. https://doi.org/10.1155/2021/3137083

Seichter, D., Eisenbach, M., Stricker, R., & Gross, H.-M. (2018). How to improve deep learning-based pavement distress detection while minimizing human effort. In *2018 IEEE International Conference on Automation Science and Engineering (CASE)* (pp. 63–68). IEEE. https://doi.org/10.1109/COASE.2018.8560372

Silva, L. A., Leithardt, V. R. Q., Batista, V. F. L., Villarrubia González, G., & De Paz Santana, J. F. (2023). Automated road damage detection using UAV images and deep learning techniques. *IEEE Access, 11*, 62918–62931. https://doi.org/10.1109/ACCESS.2023.3287770

Wu, C., Ye, M., Zhang, J., & Ma, Y. (2023). YOLO-LWNet: A lightweight road damage object detection network for mobile terminal devices. *Sensors, 23*(6), 3268. https://doi.org/10.3390/s23063268

Wu, C., Zhang, L., & Fan, X. (2025). Object detection model design for tiny road surface damage. *Scientific Reports, 15*, 11032. https://doi.org/10.1038/s41598-025-95502-z

Zeng, J., & Zhong, H. (2024). YOLOv8-PD: An improved road damage detection algorithm based on YOLOv8n model. *Scientific Reports, 14*, 12052. https://doi.org/10.1038/s41598-024-62933-z

Zhang, J., Sun, S., Song, W., Li, Y., & Teng, Q. (2024). A novel convolutional neural network for enhancing the continuity of pavement crack detection. *Scientific Reports, 14*, 30376. https://doi.org/10.1038/s41598-024-81119-1