



OPEN A contribution to reliable rail transport: AI-powered real-time wheel defect detection

Muhammad Zakir Shaikh^{1,3,4}✉, Enrique Nava Baro² & Sahil Jatoi^{3,4}

The railway system is a green mode of transport that is essential for contributing significantly to economic growth, enhancing accessibility, and easing regional integration. The safety and efficiency of railway transport systems depend on the condition of wheels, as deterioration of wheels is a major cause of both human life and financial loss. Therefore, real-time monitoring is essential for early detection and preventing failures. This study presents an AI-based framework for real-time railway wheel defect detection, leveraging advanced You Only Look Once (YOLO) models (v5–v12) and a Real-Time Detection (RTD) Transformer model. A custom wayside imaging system was developed, capturing high-resolution images to construct the FaultSeg dataset, addressing class imbalances and annotation challenges. Eight YOLO models and the RTD Transformer were evaluated, with extensive hyperparameter tuning, to identify defects such as wheel flats, shelling, discoloration, and cracks/scratches. The YOLOv5-seg model demonstrated superior performance with 91% precision, 90% recall, and 92% mAP@0.5, achieving real-time processing at 30 FPS with latency under 30 ms. The optimized model was deployed on an edge device for operational railway environments, showcasing its feasibility for real-time defect detection, enhancing predictive maintenance, and improving railway safety. This work contributes to the advancement of AI in condition monitoring by providing a publicly available dataset and demonstrating the practical deployment of a real-time defect detection system.

Keywords Defect detection, Railway transportation, Deep learning, Fault diagnosis, YOLO models, Transformer

Railway transport is an essential part of the transportation infrastructure in many countries across the world. Beyond its contribution to the national economy, railway transportation plays a significant role in reducing greenhouse gas emissions, thereby supporting the environmental goals of nations¹. According to the International Union of Railways, approximately eight billion passengers use train service annually, and rail transport helps to reduce traffic jams, saving countries close to \$100 billion each year². Research indicates that 45.6% of the total annual accidents are caused due to the derailment of wheels³. While systems rely heavily on regular inspection, these approaches are not capable of promptly detecting defects under extreme weather conditions. Hence, for effective and safe railway operations, real-time condition monitoring of wheelsets is highly critical for predictive maintenance, reducing downtime, improving safety, and lowering lifecycle maintenance costs⁴. Railway transportation is not ideal for existing condition monitoring systems, as they are limited to periodic inspections, which are expensive and unreliable in the detection of defects in real-time⁵. These systems are not used optimally under many operating conditions because they struggle with unbalanced datasets and sparse annotations.

$$P(x, y) = P_0 \left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2} \right)^{1/2} \quad (1)$$

¹School of Industrial Engineering, University of Malaga, Malaga, Spain. ²Departamento de Ingeniería de Comunicaciones, Universidad de Malaga, Campus de Teatinos, Málaga, Spain. ³National Center for Robotics, Automation and Artificial Intelligence, Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan. ⁴NCRA-Condition Monitoring Systems Lab, Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan. ✉email: zakir.shaikh@faculty.muet.edu.pk; zakir.shaikh@uma.es; muhammadzakirshaikh90@gmail.com; en@uma.es

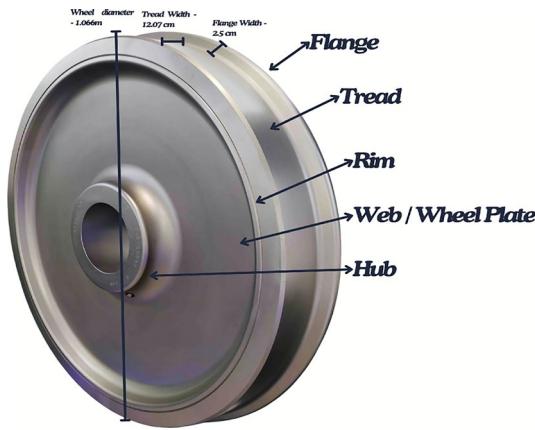


Fig. 1. External structure of a single freight train wheel, including flange, tread, rim, wheel plate, and hub.

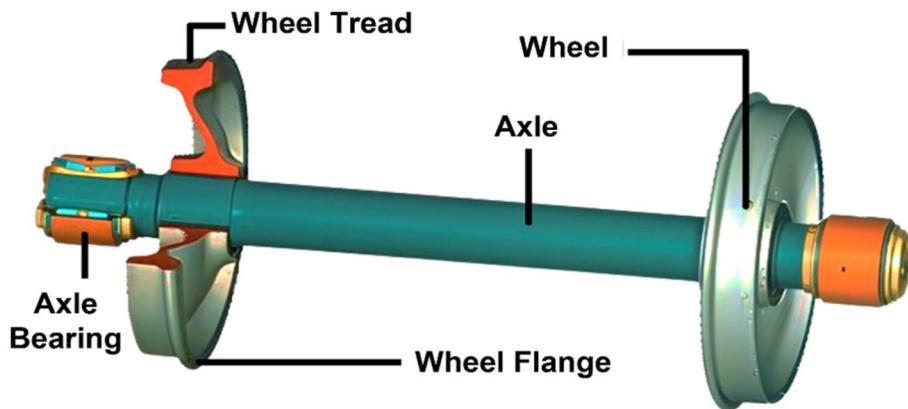


Fig. 2. Internal structure of freight train wheelsets including tread, flange, axle, and axle bearing.

Equation 1 represents the Hertzian contact stress equation for wheel rail interaction, where $P(x, y)$ shows the pressure at a given coordinate on wheel-rail contact. P_0 , a, b, and x, y shows the maximum contact pressure, semi-axes of the elliptical contact and coordinates on the wheel respectively.

$$P_0 = \frac{3F}{2\pi ab} \quad (2)$$

Equation 2 represents the maximum contact pressure where F shows the vertical force acting on the wheel and a, b shows the contact patch semi-axes. The structure of a train wheel consists of several components, including tread, flange, wheel plate, rim, and hub. Figures 1 and 2 presents that the tread is 12.07 cm wide, that ensures smooth contact with the track surface, while the flange is 2.5 cm wide, that prevents derailment by guiding the wheel along the rails. The wheel plate connects the rim to the hub, which is mounted on the axle, enabling the wheel's rotation. The wheelset consists of two wheels connected by a 1.75-meter long axle, matching the standard gauge track. Each wheel has a diameter of 1.066 meters and a circumference of 3.35 meters. Axle bearings enable smooth, frictionless rotation of the wheels, ensuring stability during operation.

Wheel defects often cause substantial accidents and substantial economic losses.

$$F_{\text{impact}} = ma_{\text{impact}} \quad (3)$$

Equation 3 represents the defect impact on train dynamics where F_{impact} is the impact force, m is the mass of the train while a_{impact} is the acceleration spike due to defect.

$$f = \frac{v}{\lambda} \quad (4)$$

Equation 4 represents the excitation frequency due to defect where f is the frequency impact, v is the train velocity and λ is the defect length.

The most prevalent defects include flat spots, which typically result from wheel sliding⁶⁷. The major defects after flat spots are shelling, cracks, scratches, and discoloration⁸. Shelling is a result of surface fatigue that develops

when the wheel gets soft from excessive heat under braking, which creates fissures and spalls. Cracks grow from excessive pressure. Scratches and discoloration are caused by friction and wear. This can cause vibrations and rails damage, which raises maintenance costs. The early diagnosis of wheel defects is crucial for the maintenance of safe railways and requires cutting-edge diagnostic techniques.

Defect detection has become a critical technique in industrial applications, particularly in image processing, where it is used to identify defects and errors. To address manual defect detection problems, deep learning (DL) techniques have successfully been employed. Convolutional neural network (CNN) are most commonly used in image classification tasks to achieve better accuracy⁹. For example, Mask R-CNN is a real-time object detection algorithm that is capable of pixel-level defect detection, which makes it a more reliable technique. Another prominent DL technique is YOLO, which is one of the most popular models to detect objects in real time¹⁰.

$$\mathcal{L} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{\text{conf}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \quad (5)$$

Equation 5 represents the DL model representation where S is the grid size, B is the number of bounding boxes per grid cell. (x_i, y_i) are the center coordinates of the detected defect, C_i is the confidence score of the predicted bounding box, and λ_{coord} , λ_{conf} are the hyperparameters controlling the localization and the weight of the loss of confidence.

For instance segmentation

$$\mathcal{L}_{\text{seg}} = \sum_{k=1}^K \mathbb{1}_k^{\text{obj}} \sum_{(i,j)} (M_{ij} - \hat{M}_{ij})^2 \quad (6)$$

Equation 6 represents instance segmentation mask prediction loss where K, M_{ij} , and \hat{M}_{ij} shows the number of detected objects, ground truth segmentation mask for the defect and predicted segmentation mask.

It operates in a single stage instead of a two-stage object detection process, which enables it to achieve a faster processing speed than other DL algorithms. In railway systems, the defect detection process have evolved over time and has significantly improved due to recent advancements in DL techniques.

This research proposes an online defect detection system for operational train wheels, which effectively achieves outstanding performance over different configurations of YOLO models and Real time detection (RTD) transformer model. The contribution of this proposed research are:

- The design and development of a custom wayside imaging system for operational railway wheel defect detection.
- The creation of the FaultSeg dataset, a publicly available dataset for training deep learning models on railway wheel defects.
- A comprehensive evaluation of eight YOLO models and an RTD Transformer, highlighting their effectiveness in detecting various defects in real-time.
- The deployment of the YOLOv5-seg model on an edge device, demonstrating its real-time defect detection capabilities with 30 FPS and latency under 30 ms.

The rest of the paper is organized as follows: Section “Related work” provides an overview of traditional and automated methods for defect detection using manual and DL based techniques. Section “Methods” details the overall methodology of the research, including data collection, preprocessing, experimental setup, instance segmentation and object detection implementation. Section “Results” discusses the analysis of model performance, including computing resources, hyperparameters, model evaluation, cross validation, comparison, and model deployment. Section “Conclusion” concludes the paper and summarizes the key findings of this study.

Related work

Introduction to traditional railway wheel inspection methods

Traditional manual inspection methods, such as visual checks and wheel hammering, are limited in detecting subsurface defects and are prone to human error. These methods also struggle under varying environmental conditions. According to Sundaram and Zeid et al.¹¹, manual inspection methods yield an average accuracy rate of 80%, and it suffers from inconsistency. Researchers have investigated new sensor-based techniques to enhance the reliability of detection methods. A reflective optical position sensor which uses laser-based displacement analysis to detect wheel flats was proposed by Gao et al.¹². Similarly, Gazafrodi et al.¹³ created an advanced non-contact laser triangulation system that precisely measures wheel and rail corrugation and is more efficient than manual and vibration-only assessments. But with the progress of technology more effective forms of condition monitoring were researched, such as railway rolling noise analysis, time-frequency analysis, adaptive noise-cancelling¹⁴ etc. During the time-frequency analysis, both time and frequency information are maintained for fault diagnosis.

$$X(t, f) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j2\pi f \tau} d\tau \quad (7)$$

Equation 7 represents the mathematical model for traditional railway wheel inspection, where $X(t, f)$, $x(\tau)$, $w(t)$, f , and $e^{j2\pi ft}$ represent the time-frequency representation, input signal from wheel vibration, window function centered at t , frequency, and Fourier transform kernel, respectively.

Emergence of advanced techniques for condition monitoring

In a study Asplund et al.¹⁵ used wheel profile parameters including flange thickness, tread width, wheel diameter, flange angle, profile wear (contact area), track gauge compatibility, cross-section profile, wheelset load, and fatigue strength. However, in another study, Bevan et al.¹⁶ used magnetic flux techniques to detect wheel tread damage in railway wheels. However, Zhang et al.¹⁷, utilized eddy current pulsed thermography for rolling contact fatigue crack detection in wheel tread. Additionally, Chen et al.¹⁸ introduced a two-level adaptive chirp mode decomposition method to detect wheel flats under variable speed and noise. Mosleh et al.¹⁹ developed a spectral kurtosis method for extracting wheel flat features from intricate vibration signals, which showed effective in operational environments. Xu et al.²⁰ and Zhou et al.²¹ examined the mechanical effects of flats and polygonization, highlighting the pressing need of precise detection methods to reduce structural damages. All these methods yielded more accurate results than traditional manual inspection and hammering. However, they required some manual intervention, which was susceptible to errors influenced by environmental conditions and required a high initial implementation cost.

$$\frac{da}{dN} = C(\Delta K)^m \quad (8)$$

Equation 8 represents the modeling of defect growth over time using fracture mechanics, where $\frac{da}{dN}$, C , and K are the crack growth rate per cycle, material-dependent constants, and stress intensity factor range, respectively.

$$\Delta K = Y\sigma\sqrt{\pi a} \quad (9)$$

Equation 9 represents ΔK , where Y , σ , and a are the correction factor based on crack geometry, applied stress, and crack length, respectively. This equation demonstrates why manual inspection sometimes fails to detect cracks, as they can propagate undetected until reaching critical failure.

Non-destructive testing methods in wheel inspection

Modern railway wheel inspections heavily rely on non-destructive testing methods including Ultrasonic Testing (UT), Eddy Current Testing (ECT), and Magnetic Flux Leakage (MFL). The Ultrasonic Testing technique is widely used for identifying internal cracks within wheel rims and hubs, while phased array UT enhances both coverage and resolution introduced by Fry et al.²². Motion-Induced Eddy Current Array (MIECA), eddy current methods enable high-speed detection of both surface and near-surface defects. Magnetic Flux Leakage (MFL) systems combine with magnetization methods and leakage field analysis to detect surface cracks and material losses in wheel treads conducted by Feng et al.²³. These techniques provide real-time feedback, high sensitivity, and applicable in automated systems.

Automation in railway wheel defect detection

After advancements in hardware capabilities and computational power, researchers began exploring new avenues for automation in industrial condition monitoring. In a study, Zhang et al.²⁴ explored defect detection technology in railway wheel images, using DL to automate ultrasonic testing of train wheels. In another study, Bernal et al.²⁵ introduces a sensor node architecture for onboard railway wagon monitoring with an Analog Fault Detector that can save power and reduce hardware costs. However, Shruthi et al.²⁶ discusses the use of MFL detection with autonomously driven testing trains capable of identifying and inspecting railway tracks for cracks, as well as machine-learning techniques for defect detection using tachographs on railway components. However, Cavuto et al.²⁷ used Finite Element Modeling (FEM) for optimizing test setups and improving laser ultrasonics defect detection capabilities in train wheelsets. The research presented an advanced FEM model that replicated ultrasonic wave generation and propagation inside axle materials through air. The authors used modeling to identify the best experimental conditions and improved detection sensitivity for defects in complex regions like wheel-axle fittings. The integration of edge computing and artificial intelligence (AI) models enables new avenues for real-time defect detection. Wang et al.²⁸ created a YOLOv5s-VF model for edge computing that achieves 93.5% rail defect detection accuracy and significantly reducing the computational power being required. The STrans-YOLOX model integrates Swin Transformer with YOLOX to improve detection accuracy of small and complex defects across different tracks and lighting conditions²⁹. These models prioritize long-range spatial dependencies, which leads to better performance than previous CNN-only models in complex environments.

Digital image processing for defect detection

Another area of research is the use of digital image processing for automatic or semi-automatic defect detection of wheels. However, these methods rely on classical image processing algorithms, which are data-intensive and prone to errors.

$$\varphi(X) = W_3\sigma(W_2\sigma(W_1X + b_1) + b_2) + b_3 \quad (10)$$

Equation 10 represents the machine learning (ML) based feature extraction for defect detection, where W_1 , W_2 , and W_3 are the weight matrices of the convolutional layers, b_1 , b_2 , and b_3 are the biases, and $\sigma(\cdot)$ is the non-linear activation function. It differentiates DL-based defect detection from traditional techniques.

Traditional image processing techniques are limited by variations in image quality and environmental conditions. To overcome these challenges, Zhang et al.³⁰ developed a Deep Extractor system that integrates bilateral convolutional networks with conditional random fields to improve small-defect detection on rail surfaces. Soleimani et al.³¹ introduced an image-based profiling system for railway wheels that enables precise wear evaluation against CAD reference standards. However, CycleGAN represents a GAN-based technique that helps generate synthetic defect images for deep learning model training. GAN-based methods including CycleGAN developed by Guo et al.³² address data limitations through creation of synthetic defect images, which enhances deep learning model training. Edge computing-based train obstacle detection algorithms powered by YOLOv3 have been proposed by Li et al.³³ to save onboard equipment's computation resources. Diagnostic tools for detecting and quantifying wheel-flat in trains using wavelet transform were validated through experiments conducted by Belotti et al.³⁴. Additionally, ML models are proposed by Krummenacher et al.³⁵ for automatic wheel defects detection using vertical force measurements from a sensor system on the railway network. Moreover, the predictive maintenance framework was proposed by Shaikh et al.³⁶ using various ML models such as boosting and bagging algorithms to predict the failure of the air production unit (APU) in electric trains. However, another study proposed by Shaikh et al.³⁷ anchor-free YOLOv8 object detection model was used to detect real-time railway wheel defects.

Infrared thermography in railway wheel inspection

Infrared Thermography (IRT) has become a widely used technique for non-contact defect detection in railway wheels. Qu et al.³⁸ showed that IRT could detect wheel surface cracks and wear through temperature differential mapping. Active thermography identifies subsurface cracks through controlled heating and thermal response observation, that other optical techniques fail to detect conducted by³⁹. The integration of IRT with acoustic and vibration sensors leads to better defect localization while simultaneously increasing fault classification accuracy⁴⁰.

The role of visual inspection in modern condition monitoring

Despite advancements in automation, visual inspection is still required. There is ongoing discussion about the importance of large annotated datasets for the analysis of historical documents, and the value of annotating such data to emphasize the role of visual inspection in automated systems. The hybrid approach of integrating a visual inspection with YOLO models provides a balanced methodology, emphasizing the human visual aspect while integrating DL to enhance the computational power of the human expert for defect detection and predictive analysis. Latest techniques for detecting defects in wheel images have been proposed by Redmon et al.⁴¹, based on the improved YOLOX model. An approach named UT-YOLO has been developed by Zhang et al.⁴² to detect defects in the high-speed railway train wheel using the computer vision algorithm YOLO. Transformer models have been applied for Vision-based problems including object detection, defect detection, and multiple-object tracking. It enhanced the detection accuracy in different applications⁴³⁴⁴⁴⁵. Despite advances in automation technologies, visual inspections utilize advanced tools to remain relevant. Remote inspection systems using thermal cameras, GoPro high resolution video, or drone-based recording can be used in ML pipelines for hybrid inspection frameworks proposed by⁴⁶. By integrating these technologies, higher efficiency was achieved while enabling human expert evaluation of AI results that ensures safe operations in critical environments.

Artificial intelligence and machine learning for railway safety

Recent technological developments in AI and ML have significantly improved railway safety. In a study, Shaikh et al.⁵ discusses the importance of condition monitoring systems for railway wheels, focusing on problems of wheel wear and dynamic stress such as skidding while analyzing various data acquisition and analysis approaches to enhance monitoring precision and reliability. Braga et al.⁴⁷ and Caldeira et al.⁴⁸ used survival modeling alongside multivariate statistical methods to predict trends in wheel wear. The research by Baasch et al.⁴⁹ used cepstral analysis on axle box acceleration data to continuously monitor equipment conditions.

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \int_0^1 P(r) dr \quad (11)$$

Equation 11 represents the performance evaluation metrics for railway wheel defect detection where N is the number of object classes and P(r) is the precision-recall curve for each class. It shows that defect detection models optimize both precision and recall in critical application like railway safety.

Another study proposed by Shaikh et al.⁵⁰ considers the essential contribution of wheelset failure to train accidents. They created a cheap, AI driven, visual inspection system for real-time wheel defect detection with GoPro Hero 9 cameras to capture ultra high-resolution images. The system exemplifies the integration of AI and readily available hardware for self-trained, preemptive defect detection for operational safety are required for defect detection and condition monitoring of rolling stock components like bogie and wheels. The AFYv8 DL bogie component real-time identification and tracking feature, with a 99.1% mean Average Precision (mAP), can optimize TRSE and prevent derailment hazards related to wheel and bogie component failure conducted by Zakir et al.⁵¹. The paper⁵² presents a novel machine learning-based system to detect and predict deterioration in railway turnouts caused by flooding. Using convolutional neural networks (CNN) trained on nonlinear finite element models validated with field data, the system analyzes time-series data representing turnout support

behavior under flood conditions. The model achieves over 98% accuracy in recognizing and classifying flood-induced deterioration, which can help schedule timely maintenance and enhance railway infrastructure resilience. This intelligent automation offers a promising tool for managing railway infrastructure under extreme weather impacts. The integration of these advanced techniques with edge devices supports real-time adaptive decision-making, which minimizes unexpected downtimes and strengthens operational resilience. Collectively, these experiments demonstrate the impact of AI, DL and real-time monitoring on creating safer and more effective railway operations by early detection.

Recent advances in YOLO models for defect detection

DC-YOLOv8 was introduced by Lou et al.⁵³ for detecting small-size defects, that demonstrated superior performance compared to YOLOv5 and YOLOX in terms of accuracy and inference time. Zhang et al.⁵⁴ deployed an enhanced YOLOX model with image processing techniques to identify micro-defects on rail surfaces and achieved a 2.42% improvement in mean average precision (mAP). STrans-YOLOX transformer-based models utilize self-attention mechanisms to enhance spatial understanding that leads to promising results on railway defect detection across various environmental conditions²⁹. Major developments in defect detection have been discussed in the literature review, from traditional approaches to modern automated systems. Improved YOLO models have played a key role in developing systems that can detect more defects efficiently. By combining visual inspection with YOLO, it can reach the highest detection accuracy in short time.

$$\text{Operations} = S^2 B(5 + C) \quad (12)$$

Equation 12 represents the object detection performance modeling of YOLO model where S, B, C shows the grid size, number of bounding boxes per grid cell and number of classes.

These studies illustrate how powerful YOLO and transformer based object detection systems are for various industrial applications of defect detection.

Hybrid learning strategies: integrating supervised and reinforcement learning for railway wheel wear management

A hybrid learning strategy integrating supervised and reinforcement learning techniques has been proposed in this study⁵⁵ to effectively manage railway wheel wear under limited measurement data constraints. The supervised learning component relies on a predictive model developed from validated simulations, focusing on tread wear, flange height, and flange thickness, which enhances prediction accuracy with an R² improvement from 0.94 to 0.95. Reinforcement learning complements this by optimizing maintenance scheduling decisions based on real-time operational parameters such as train weight, speed, and track characteristics, thereby minimizing the need for constant inspections. The integration forms an agent that learns from wear states predicted by the supervised model and refines decision-making through feedback rewards, resulting in reduced maintenance costs, defect rates, and inspection times. Additionally⁵⁶ applied artificial neural networks to predict wheel wear using historical wear measurements, achieving effective performance in wear forecasting through supervised training techniques. These studies collectively demonstrate the potential of combining machine learning techniques to improve predictive maintenance strategies, reduce operational costs, and enhance safety and ride quality in railway systems.

Machine learning applications in defect detection, severity classification, and localization

The evolution of machine learning techniques has significantly advanced railway defect detection capabilities beyond traditional inspection methods, enabling sophisticated defect severity classification and precise localization systems. Sresakoolchai and Kaewunruen⁵⁷ pioneered the application of deep learning techniques including Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) for wheelflat detection and severity classification, achieving remarkable performance with DNN demonstrating 96% overall accuracy, 100% wheelflat detection accuracy, and superior capability in classifying wheelflat severity into light, intermediate, and severe categories using axle box acceleration data from 1,608 simulation samples. Advanced defect localization frameworks have been developed to provide precise spatial information about detected defects, with Vishwakarma and Mohapatra⁵⁸ introducing clustering-based algorithms utilizing axle box acceleration data processed through filter banks and Principal Component Analysis (PCA) to accurately detect and locate multiple track defects using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithms. They successfully classified track defects by applying the DBSCAN algorithm on selected decomposed signals and achieved accurate defect detection and localization without being affected by external parameters such as vehicle speed and axle load. Shaikh et al.⁵⁹ developed a hybrid deep learning approach combining Multi-Layer Perceptron (MLP) with various machine learning models including Random Forest, Support Vector Machine, and k-Nearest Neighbors for wheel defect detection, achieving exceptional performance with MLP-RF demonstrating 99% accuracy, 99% precision, and balanced recall in detecting wheel defects using exclusively accelerometer vibration data, while MLP-DT achieved 98.9% accuracy in classifying conicity values for different severity levels. Krummenacher et al.³⁵ introduced novel time-series features combined with Support Vector Machine classifiers and neural network approaches for automated wheel defect detection, successfully outperforming classical detection methods for flat spots while demonstrating prediction capabilities for shelling and non-roundness defects for the first time using vertical force measurements from permanently installed sensor systems, achieving accuracy rates between 81–92% with their 2D-convolutional neural networks applied to time-domain data. Furthermore, comprehensive machine learning frameworks have been developed for predictive maintenance applications, where Yang et al.⁶⁰ presented YOLOv5s-based detection and classification systems for railway track surface defects including erosion-caused

holes and scratches, demonstrating superior performance compared to YOLOv3-Tiny in terms of precision, recall, and mean average precision (mAP) while utilizing convolutional neural networks and computer vision techniques for robust real-time defect monitoring and spatial localization across various environmental conditions. Kaewunruen et al.⁶¹ investigates the use of machine learning as a non-destructive approach to assess the condition and dynamic properties of railway track components. The study develops and tests several machine learning models including Linear Regression, K-Nearest Neighbors, Gradient Boosting, and Convolutional Neural Networks (CNN), finding that the tuned CNN model offers the highest accuracy with minimal error in predicting dynamic properties from field measurements. This approach provides a fast, cost-effective alternative to traditional inspection methods, enabling efficient in-situ condition assessment without expensive equipment or lengthy analysis. The work highlights the potential of machine learning to improve railway maintenance practices by enabling accurate and timely track condition monitoring.

Gaps and future directions in railway wheel defect detection

However, there is a significant gap that still needs to be addressed. Advanced models like Vision Transformers, DETR, DINO, ViT need to be integrated. The optimization of real-time detection efficiency is essential, necessitating the development of more adaptable hybrid systems that integrate manual and automated processes. The absence of larger public datasets for railway wheel defects remains a critical problem. Researchers need to create such datasets with appropriate annotations and transformer-based models are needed to enhance specific applications for multi-object detection and tracking. The railway industry demands comprehensive solutions that integrate image acquisition with real-time AI analysis and automated report generation. Domain-adaptive pretraining along with federated learning and lightweight deployment on edge devices represent the new areas of focus. Future scalability and adoption of railway systems depend on the creation of models that perform reliably across different track and weather conditions. All these deficiencies need to be considered so that the defect detection systems can be improved, bringing railway safety and operational efficiency to a new level.

Methods

Comprehensive data acquisition and curation processes – faultseg dataset

A custom wayside data acquisition system was developed to capture high-resolution images of moving train wheels. Three cameras were positioned on one side of the track with a 0.86 radian field of view, ensuring coverage of the wheelset region of interest.

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T \quad (13)$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} x'/z' \\ y'/z' \\ 1 \end{bmatrix} \quad (14)$$

Equation 13 and 14 represent the camera projection model where (X, Y, Z) , (x', y', z') , (u, v) , and K show the real world coordinates, camera coordinates after applying the rotation matrix R and translation vector T , 2D pixel coordinates on the camera image, and the camera intrinsic matrix, respectively.

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (15)$$

Equation 3.1 represents the field of view where f_x , f_y are focal lengths in pixels, and c_x , c_y are the principal points of the image.

The setup enables high-resolution imaging and precise mapping of the region of interest (ROI) within a single working distance. Figure 3 and Table 1 illustrates the camera setup parameters on the indigenously designed wayside system. Videos were captured using GoPro Hero 9 cameras at 2704×1520 resolution and

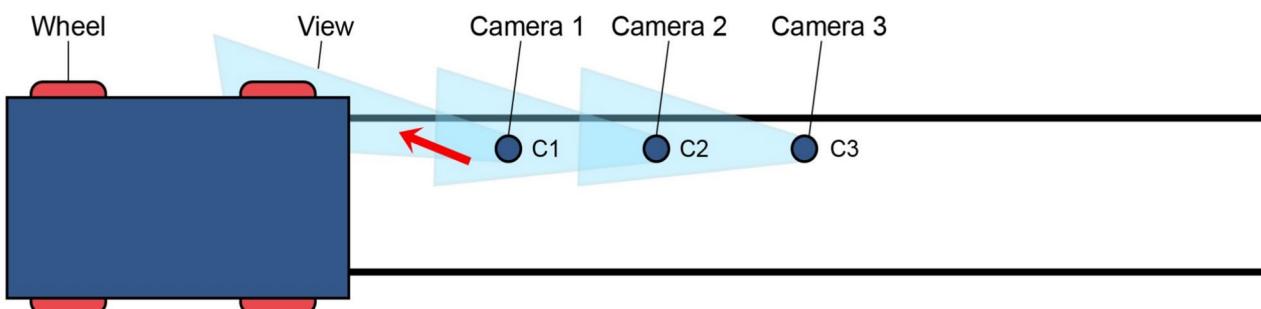


Fig. 3. Camera configuration of proposed case study illustrating indigenously designed wayside system.

Parameter	Details
Camera Model	GoPro Hero 9
Height from Base (hb)	3 inches
Distance from Track (dt)	10 inches
Weight (with batteries)	158 g (0.35 lb/5.57 oz)
Dimensions	71 × 55 × 34 mm (2.8 × 2.17 × 1.34)
Resolution	2704 × 1520

Table 1. Camera Setup Parameters.

	Assembly	Base	Camera Setup	Enclosure	Lead Panel	Led
CAD Models						
3D Components						

Fig. 4. Indigenously designed wayside system breakdown: CAD to 3D Integration.

24 FPS. Cameras were positioned 10.16 cm above the track, 73.66 cm from the rails, with a 0.26 radian tilt and zero roll/yaw to ensure consistent imaging. Experiments were performed on our automated wheelset test rig that is used for wheel defect detection, which was developed in the condition monitoring systems laboratory of the university.

Figure 4 represents a comprehensive overview of CAD models and 3D components of essential design elements. It includes assembled wayside system, base, camera setup, camera enclosure, LED panel, and LED strip both technical drawings and 3D models. For the purpose of this research, a wayside system was designed from scratch, as shown in Fig. 5. The system consists of rotatable wooden camera holders mounted on a mechanical bed-like structure that is 138.25 inches long and 68.45 inches wide.

$$B = \frac{v \cdot t_{\text{exp}}}{s} \cos(\phi) \quad (16)$$

Equation 16 represents the motion blur estimation for moving train wheels where v , t_{exp} , s , and ϕ represent the train speed in meter per second, exposure time of the camera in seconds, pixel size of the camera sensor in meter per pixel, and angle between the train motion and the camera video direction, respectively.

To remove motion blur, a wiener deblurring filter is applied:

$$\hat{I}(u, v) = \frac{I_{\text{blurred}}(u, v) \cdot H(u, v)}{|H(u, v)|^2 + \frac{1}{\alpha}} \quad (17)$$

Equation 17 represents the Wiener deblurring filter where $I_{\text{blurred}}(u, v)$, $H(u, v)$, and α are the Fourier transform of the blurred image, motion blur transfer function, and regularization term to prevent division by zero, respectively.

The system also has an attached illumination setup consisting of two LED boards attached to the wayside bed to cater to various lighting conditions throughout the day. A key feature of the illumination setup is its ability to emit unidirectional light radiation, which minimizes glare caused by excessive brightness, especially in daytime conditions.

$$I_T = \sum_{i=1}^N (I_i \cdot \rho \cdot \cos(\theta_i)) \left(\frac{1}{d_i^2} \right) \quad (18)$$

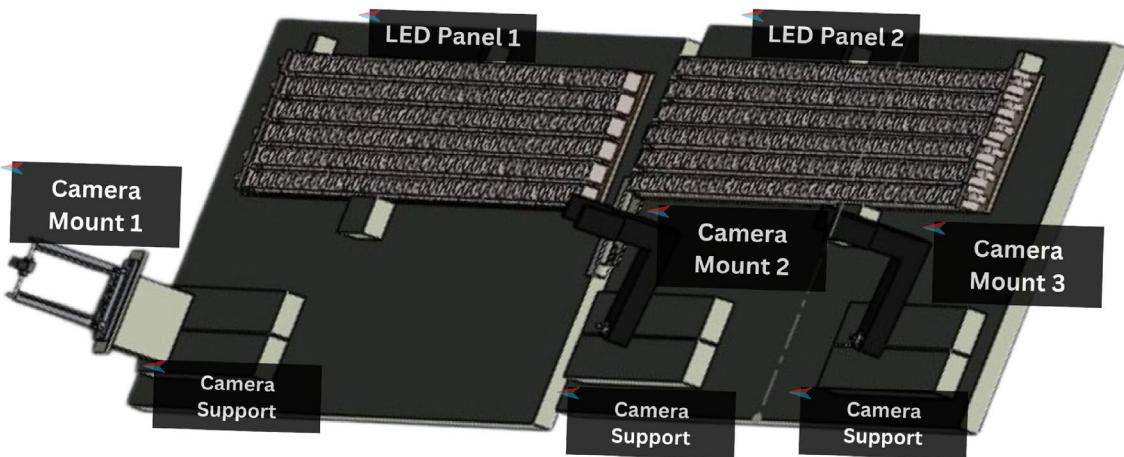


Fig. 5. CAD model of indigenously designed wayside system for data acquisition.

Aspect	Details
Setup Dimensions	Length: 138.25 inches, Width: 68.45 inches
LED Panel Dimensions	Length: 39.37 inches, Width: 34.45 inches
Number of LED Panels	2
Power Source	UPS (Stabimatic, 3000 VA)
Laptops	2 for live inspection
AC Adapter	Included
Extension Boards	2

Table 2. Specifications of Data Acquisition Setup.

Equation 18 represents the extended illumination model for the wayside setup and adaptive brightness where I_T , I_i , ρ , θ_i , d_i , and N are the total light intensity received by the train wheel, luminous intensity of each LED source, surface reflectance of the train wheel, angle between the LED and wheel surface, distance from the LED to the wheel, and number of LEDs in the setup, respectively.

To minimize lighting, the minimum brightness required for clear defect visibility is:

$$I_{\min} = \frac{255}{1 + e^{-\alpha(I_{\text{threshold}} - I_{\max})}} \quad (19)$$

Equation 19 represents the minimum brightness required for the setup where I_{\min} is the adaptive brightness level, while $I_{\text{threshold}}$ is the minimum threshold brightness for defect detection and α is an adjustable scaling factor. It helps to understand the dynamics of adjustment of LED power to ensure consistent lighting across different train speeds and ambient conditions.

Additionally, the indigenously designed setup gives a bright white-light source that is both controllable and energy efficient. During the installation of the system, the safety precaution guidelines established by the relevant authorities were strictly followed. Precautions were taken to ensure that the power supply equipment was located no closer than 6 meters from the track, 3 inches below the train, and installed at least 10 minutes before the train arrived.

The specifications of the wayside setup are given in Table 2 including the setup dimensions, power source, and additional components.

The FaultSeg dataset was collected by our team of researchers from the School of Industrial Engineering from the University of Malaga, NCRAAI, and NCRA-CMS Lab at MUET through the deployment of the wayside data acquisition system at Kotri Railway Station for multiple days at different timings. It is pertinent to mention that the data gathered at night, particularly after 1800 hours, was acquired using LED panels integrated into our indigenously designed wayside system to reduce shadow effects. Certain trains depart from Karachi to Kotri Railway Station and continue to northern cities, including Mirpurkhas, Sialkot, Lahore, Peshawar, and Multan. The trains departing from Karachi and traveling northward are designated as up trains due to their direction towards the North. Nonetheless, certain trains from northern regions arrive at Kotri Railway Station via another track, ultimately destined for Karachi. These trains are generally categorized as down trains due to their downward trajectory from the north. Moreover, not all trains halt at the Kotri Railway station. Certain trains halt at the Kotri Railway Station, whilst others transit without stopping. Our team gathered data on the several trains proceeding northward^{7,8}. Table 3 presents a concise summary of the different trains our team encountered during the data acquisition procedure, along with their schedules.

Train route and bogie numbers	Train name	Arrival time at Kotri station	Train stop information	Departure (Origin)	Destination
09 UP	Allama Iqbal Express	18:16	Yes (Departure at 18:18)	Karachi Cantt	Sialkot
10 DOWN	Allama Iqbal Express	06:33	Yes (Departure at 06:35)	Sialkot	Karachi Cantt
33 UP	Pak Business	18:38	No	Karachi Cantt	Lahore
34 DOWN	Pak Business	07:35	No	Lahore	Karachi Cantt
15 UP	Karachi Express	20:38	No	Karachi Cantt	Lahore
16 DOWN	Karachi Express	10:30	No	Lahore	Karachi Cantt
151 UP	Shah Latif Express	10:15	Yes (Departure at 10:17)	Karachi City	Mirpurkhas
152 DOWN	Shah Latif Express	16:43	Yes (Departure at 16:45)	Mirpurkhas	Karachi City
47 UP	Rehman Baba	14:49	No	Karachi Cantt	Peshawar
48 DOWN	Rehman Baba	13:31	No	Peshawar	Karachi Cantt
25 UP	Bahauddin Zakaria Express	21:06	Yes (Departure at 21:08)	Karachi City	Multan

Table 3. A Brief Overview of the Various Trains Encountered During the Data Acquisition Process, Together With Their Schedule.

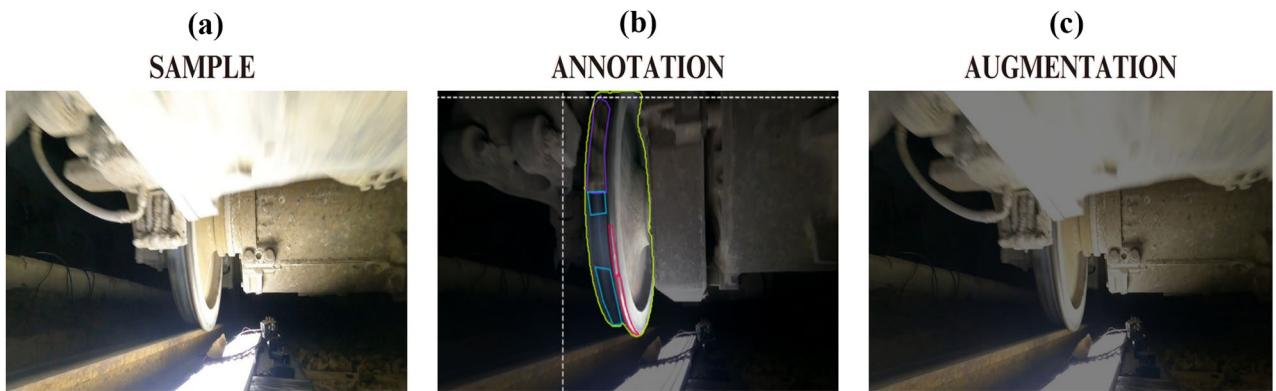


Fig. 6. (a) Sample Image, (b) Annotated Image, and (c) Augmented image.

The dataset consists of four target classes, i.e., Wheel, Shelling, Discoloration, and Cracks/Scratches. Cracks and scratches were considered as a single class instead of two because it is very hard to distinguish between their features. The dataset was annotated in several formats; however, the YOLO instance segmentation and detection annotation format was utilized for this research. Figure 6 below shows the original image while Fig. 6 below illustrates an example of a ground truth image with all 4 classes used in it. The data acquired from the wayside data acquisition system suffered from class imbalance problems. In order to deal with it, data augmentation was performed on the dataset to balance out the classes and improve overall better performance of models. Augmentation also enhanced the diversity of the FaultSeg dataset, and the generalization of models tested over it. Figure 6 depicts an augmented image to enhance diversity in the dataset.

Data augmentation included rotation, scaling, brightness adjustment, and Gaussian noise addition to simulate real-world variability and improve model generalization.

$$X' = T_{\text{brightness}}(T_{\text{scaling}}(T_{\text{rotation}}(T_{\text{translation}}(X)))) \quad (20)$$

Equation 20 represents the data augmentation transformation, where Rotation by θ degrees:

$$T_{\text{rotation}}(X) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} X \quad (21)$$

Scaling by factor s :

$$T_{\text{scaling}}(X) = \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} X \quad (22)$$

Brightness adjustment:

$$T_{\text{brightness}}(X) = \alpha X + \beta \quad (23)$$

where α controls contrast and β adjusts brightness.

Gaussian Noise Addition:

$$T_{\text{noise}}(X) = X + N(0, \sigma^2) \quad (24)$$

where $N(0, \sigma^2)$ is zero-mean Gaussian noise.

Equations 20, 21, 22, 23 and 24 helps to simulate real-world variations in the dataset and increases robustness of the DL model.

The dataset, acquired and developed by the authors of this research study using an indigenously designed system for Railway Wheel Defect Detection, is publicly accessible at the following link: <https://doi.org/10.5281/zendodo.13162335> and <https://doi.org/10.6084/m9.figshare.27996866.v1>

The initial data annotation in the FaultSeg dataset exhibited an imbalance in the distribution of defect class. The class-wise ratios were as follows: Cracks/Scratches: 24.93%, Discoloration: 10.50%, Shelling: 32.28% and wheel: 32.28%. To address this problem, the dataset was rebalanced by adding more samples of Cracks/Scratches and Discoloration. Additionally, data augmentation was used to increase the number of wheel samples. Ultimately, the data was redistributed, resulting in wheel representing 30% of the total, while Cracks/Scratches, Discoloration and Shelling were each balanced at 23.33%.

Table 4 represents the number of samples at each processing step. The dataset consists of 1,872 annotated images. After the implementation of augmentation techniques, there are a total of 22,464 images.

YOLO and RTD transformer based instance segmentation and detection models

Object detection and Instance segmentation play a crucial role in classification. Instance segmentation performs pixel-level classification, enabling the model to detect objects that are not easily visible. Wheel defect detection is not as viable as standard object detection, as observed in experiments through object detection models, so it is preferred to use instance segmentation. Instance segmentation helps to pinpoint the position of every defective instance on an image with their assigned label. Incorporating expert annotated data in model training enabled the system to detect defects with exceptional precision. Eight architectures of YOLO models, YOLOv5, YOLOv7, YOLOv8, YOLOv9, YOLOv11 instance segmentation, YOLOv6, YOLOv10, YOLOv12 detection and RTD Transformer models were used to perform instance segmentation and object detection respectively. Figure 7 illustrates the step-by-step process from obtaining government permission and implementing prerequisites to data collection, preprocessing, and model training.

Table 5 provides a comprehensive comparison of various instance segmentation and object detection models including YOLOv5-seg, YOLOv6, YOLOv7-seg, YOLOv8-seg, YOLOv9-seg, YOLOv10, YOLOv11-seg, YOLOv12 and RTD Transformer. It outlines the backbone, neck, head, loss function, key features, and notable enhancements of each model. It highlights differences in model architecture, advancements like attention mechanism, enhanced feature extraction and real-time processing capabilities.

Performance evaluation metrics and experimental configuration

The experiments were performed on a system that has an Intel i7 9th generation CPU having 16 GB of RAM, and an NVIDIA GeForce RTX 2060 Super GPU with 8 GB of VRAM. This led to efficient training as well as testing of the YOLO instance segmentation and object detection models. The training for each model used to differ depending on the complexity of the architecture. For instance, YOLOv9-seg architecture required a training time of 24 hours on 100 epochs. The dataset was divided into three sets, training set, validation set, and testing set into a ratio of 70 by 20 by 10 respectively. This helped prevent overfitting during the training phase, thereby ensuring that the models remain reliable.

Model performance was evaluated using standard metrics: precision, recall, mean average precision (mAP@0.5), and normalized confusion matrices. These metrics provide a comprehensive assessment of both classification and localization accuracy. Each metric reflects different aspects of the model's effectiveness in identifying defects, as discussed below:

Confusion Matrix: A confusion matrix is used to visualize the result of classification, represented as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). It helps identify the type of errors made by the model.

$$\begin{bmatrix} \text{TP} & \text{FN} \\ \text{FP} & \text{TN} \end{bmatrix} \quad (25)$$

Precision: Precision is the percentage of true positives among all predicted positives. It evaluates the accuracy of positive predictions. The equation is given below:

Label	Number of samples
Original number of images	1872
Annotated number of images	1872
Augmented number of images	20592
Total number of images	22464

Table 4. Composition of the dataset, including original, annotated, and augmented image counts.

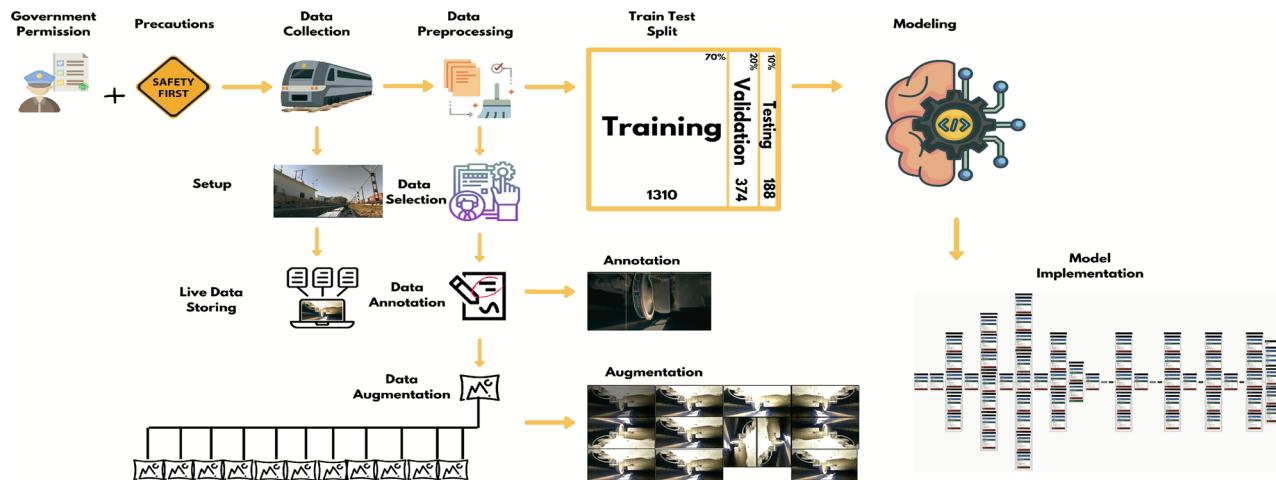


Fig. 7. Comprehensive workflow diagram of entire wheel defect detection system.

Model	Backbone	Neck	Head	Loss function	Key features	Notable enhancements
YOLOv5-seg	CSPDarknet53	FPN, PANet	YOLO layers for detection and classification	Localization, objectness, classification loss, binary cross-entropy for classification	Lightweight architecture, fast inference speed, cross-stage partial network for feature extraction	Residual connections, CSPNet for gradient flow
YOLOv6	EfficientRep Backbone	Rep-PAN	Decoupled head with SIoU loss	Classification loss, box regression loss (GIoU + DFL), knowledge distillation loss	Hardware-aware design, efficient decoupled head, state-of-the-art accuracy on COCO dataset	Bi-directional Concatenation (BiC) module, Rep-PAN neck, anchor-aided training (AAT) strategy
YOLOv7-seg	CSPDarknet53 (enhanced)	FPN, PANet	YOLO layers, enhanced E-ELAN	Same as YOLOv5-seg	Improved detection with reduced false positives, enhanced feature aggregation	Focus layer, RepConv, ADown for better feature extraction
YOLOv8-seg	CSPDarknet53 (with enhancements)	FPN, PANet	YOLO layers, upgraded detection head	Same as YOLOv5-seg	Improved mAP and mAP50, optimized for small and overlapping objects	Refined backbone and head, extra layers in backbone and neck
YOLOv9-seg	GELAN (Generalized ELAN)	FPN, PANet	Attention-enhanced detection head	Localization, objectness, classification loss, PGI for gradient updates	Advanced dynamic convolution, attention mechanism for geometric feature capture	Auxiliary reversible branches, programmable gradient information (PGI)
YOLOv10	Enhanced CSPNet	PANet	One-to-many and one-to-one heads	Classification loss, localization loss (CIoU), objectness loss, confidence loss	Dual assignment strategy for NMS-free inference, efficient parameter utilization	One-to-many and one-to-one heads, consistent dual assignments for NMS-free training
YOLOv11-seg	CSPNet with advanced feature extraction	FPN, PANet, multi-scale fusion	Segmentation head, decoupled detection head	CIoU, soft-NMS, multi-part loss function	Optimized for small and occluded objects, high-quality defect masks	Enhanced feature extraction (C3X2 Block), advanced loss functions (CIoU, Soft-NMS)
YOLOv12	R-ELAN (Residual Efficient Layer Aggregation Network)	Area Attention (A2C2f)	Detection and segmentation heads	Classification loss, localization loss (CIoU), objectness loss, attention loss	Attention-centric design, FlashAttention acceleration, real-time processing at high resolutions	R-ELAN backbone, A2C2f modules, FlashAttention-driven area-based attention
RTD Transformer	ResNet or Swin	Transformer encoder-decoder	Cross-attention mechanism for refined predictions	Feature uncertainty loss, query uncertainty in loss function	Context-driven approach for high variability defect detection, multi-head attention	Self-attention function, temporal resistance, multi-scale feature extraction

Table 5. Comparison of model architectures, loss functions, key features, and notable enhancements for all models.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (26)$$

Recall: Sensitivity or Recall is the true positive predictions divided by the total actual positives, indicating how efficiently the model identifies relevant instances. The equation to calculate recall is given below:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (27)$$

The metrics were calculated to provide a detailed evaluation of model performance by analyzing the confusion matrix, precision and recall to determine the best model for distinguishing the wheel defects.

Results

Key hyperparameters for YOLO and RTD transformer models

Hyperparameters, including learning rate, batch size, weight decay, and box loss gain, were carefully tuned for each model to optimize detection accuracy, reduce overfitting, and ensure stable training.

$$\begin{aligned} \mathcal{L}_{\text{total}} = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{H}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \lambda_{\text{conf}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{H}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\ & + \lambda_{\text{cls}} \sum_{i=0}^{S^2} \mathbb{H}_{ij}^{\text{obj}} \sum_{c \in \text{classes}} (p(c) - \hat{p}(c))^2 + \lambda_{\text{seg}} \sum_{k=1}^K \sum_{(i,j)} (M_{ij} - \hat{M}_{ij})^2 \end{aligned} \quad (28)$$

Equation 4.1 represent hyperparameter optimization and loss function formulation where L_{total} , λ_{coord} , λ_{cls} , λ_{seg} , S , B , K , (x_i, y_i) and (\hat{x}_i, \hat{y}_i) , C_i , \hat{C}_i , $p(c)$ and $\hat{p}(c)$, M_{ij} and \hat{M}_{ij} show the scaling factor for localization, confidence, classification, and segmentation, grid size, bounding box count per grid cell, and the number of detected objects, ground truth bounding box coordinates, predicted coordinates, ground truth and predicted confidence scores, class probability and predicted class probability for each class, true and predicted segmentation masks, respectively.

Table 6 illustrates hyperparameters chosen for different YOLO instance segmentation, object detection models and RTD transformer model. The initial learning rate (lr0), momentum, and weight decay were fixed to stabilize the training of models. YOLO instance segmentation, object detection and RTD transformer models have multiple loss components: objectness loss, classification loss, and bounding box regression loss. The box loss gains in YOLOv6, YOLOv8-seg, YOLOv9-seg, YOLOv11-seg, YOLOv12 and RTD transformer are set to a higher value (7.5) compared to earlier versions, emphasizing precise localization of bounding boxes. This modification, along with deeper architecture, improves object detection and instance segmentation accuracy that reduces errors by increasing the model's sensitivity to bounding-box predictions. Decoupled Focal Loss (DFL) was set to 1.53 to help with the class imbalance problem. For the auto-augmentation method, randaugment technique was employed to apply a wide range of augmentations during training, with the aim to improve the generalizability of the models. These specific parameters were fine-tuned through extensive training to achieve the best possible results for each model. The hyperparameter chosen optimized the training time, minimized redundancy and overfitting, maximized the model accuracy. This ensured an unbiased assessment of the model training, which generalized the unseen data.

Comparative performance evaluation

Table 7 shows the results of different class-wise and overall metrics. It consists of precision, recall, and mean average precision (mAP) at an IoU threshold of 0.5 for all models on four classes. It can be observed that YOLOv5-seg yields high precision and recall for all classes except cracks-scratches. It has high recall and precision for discoloration, and wheel, whereas recall for shelling is below 80%. Similarly, YOLOv7-seg, YOLOv8-seg, YOLOv9-seg, and YOLOv11-seg also have optimal precision and recall for discoloration and wheel, whereas recall and precision are unstable for cracks-scratches and shelling. In comparison, YOLOv6, YOLOv10, and YOLOv12 shows balanced performance across all classes, with high mAP@0.5 values and improved performance of the Cracks-Scratches and Shelling classes. RTD transformer has achieved the consistent precision and recall among all classes, and it has the highest mAP@0.5.

The Figure 8 shows the performance of all models on these metrics: Precision (blue), recall (green), mAP50 (red) and mAP50-95 (orange). These illustrate the performance of each model; average performance increases over training. YOLOv8-seg is the most stable in both precision and recall, while mAP50 and mAP50-95 are better in the RTD transformer, YOLOv7-seg and YOLOv9-seg. RTD transformer has the highest achieved mAP50 and mAP50-95, indicating superior performance. YOLOv6 demonstrates prompt convergence with stable precision and recall, also balanced mAP score throughout training. Additionally, YOLOv10 and YOLOv12 shows strong

Hyperparameter	YOLOv5-seg	YOLOv6	YOLOv7-seg	YOLOv8-seg	YOLOv9-seg	YOLOv10	YOLOv11-seg	YOLOv12	RTD Transformer
Initial Learning Rate (lr0)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Momentum	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937
Weight Decay	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
Box Loss Gain	0.05	7.5	0.05	7.5	7.5	7.5	7.5	7.5	7.5
Mosaic Augmentation	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Epochs	100	100	100	100	100	100	100	100	100
Batch Size	16	16	16	16	16	16	16	16	16
Image Size	640	640	640	640	640	640	640	640	640

Table 6. Hyperparameter configuration for all models.

Model	Class	Precision	Recall	mAP@0.5
YOLOv5-seg	Cracks-Scratches	0.93	0.73	
	Discoloration	1.00	1.00	0.80
	Shelling	1.00	0.75	
	Wheel	1.00	1.00	
YOLOv6	Cracks-Scratches	0.90	0.73	
	Discoloration	0.94	1.00	0.96
	Shelling	0.88	0.95	
	Wheel	0.98	1.00	
YOLOv7-seg	Cracks-Scratches	0.78	0.92	
	Discoloration	1.00	1.00	0.82
	Shelling	0.55	1.00	
	Wheel	1.00	1.00	
YOLOv8-seg	Cracks-Scratches	0.93	0.65	
	Discoloration	1.00	1.00	0.81
	Shelling	0.96	0.66	
	Wheel	1.00	1.00	
YOLOv9-seg	Cracks-Scratches	0.80	0.57	
	Discoloration	1.00	1.00	0.75
	Shelling	0.83	0.68	
	Wheel	1.00	1.00	
YOLOv10	Cracks-Scratches	0.93	0.92	
	Discoloration	0.86	1.00	0.97
	Shelling	1.00	0.96	
	Wheel	1.00	0.98	
YOLOv11-seg	Cracks-Scratches	0.70	0.53	
	Discoloration	0.98	1.00	0.89
	Shelling	0.89	0.91	
	Wheel	0.99	1.00	
YOLOv12	Cracks-Scratches	1.00	0.70	
	Discoloration	1.00	0.95	0.97
	Shelling	0.80	0.83	
	Wheel	0.97	1.00	
RTDETR	Cracks-Scratches	0.89	0.86	
	Discoloration	0.95	1.00	0.98
	Shelling	0.92	0.97	
	Wheel	0.97	1.00	

Table 7. Class-wise precision and recall, and mAP@0.5 for all models across different defect classes.

convergence with minimal fluctuations, showing their reliable performance in training. These plots provide an understanding of the convergence and overall generalization of these models.

Figure 9 illustrates the differences in precision, recall, and mean average precision (mAP) at IoU thresholds of 0.5 among all models. YOLOv12 achieves highest precision of 100%, while YOLOv10 achieves highest recall of 98%, and RTD achieves highest mAP@0.5 of 98%, and YOLOv8-seg balances near-perfect recall 98% and precision of 97%. Overall, the bar graph highlights uniformly high detection capability, with RTD transformer and YOLOv8-seg as the most consistent and accurate contenders among all models.

The results in Fig. 9 provide valuable insights into the performance of all models:

Real-time detection transformer

RTD Transformer performs the best overall with both precision and recall of 0.97, and mAP@0.5 of 0.98. It indicates that this model is highly effective in identifying all objects correctly. These improvements may be due to changes in model architecture and training techniques in RTD transformers compared to YOLO models.

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (29)$$

Equation 29 represents the scaled dot-product attention where Q , K , V , d_k are the query matrix, key matrix, value matrix and dimension of keys and queries. To enhance more attention, multihead self-attention (MHSA) is used:

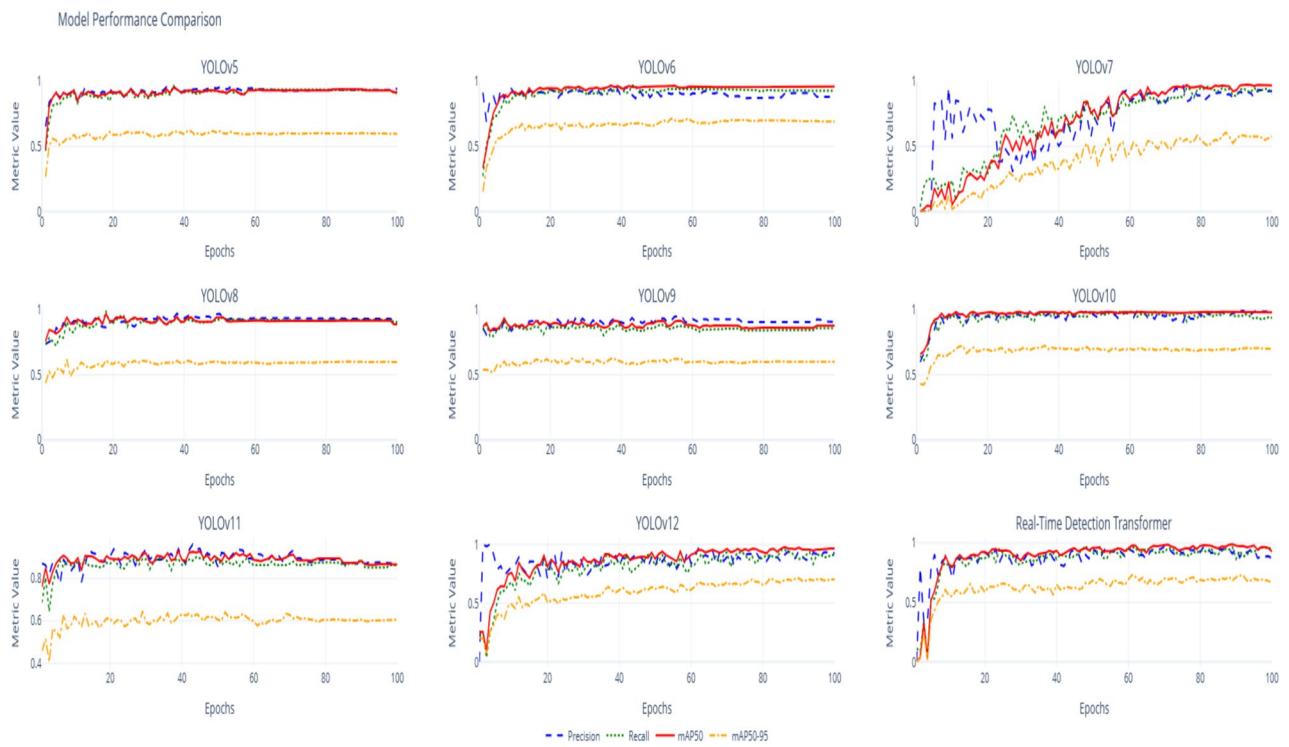


Fig. 8. Performance metrics (Precision, Recall, mAP@0.5, and mAP@0.5:0.95) across training epochs of all nine models.

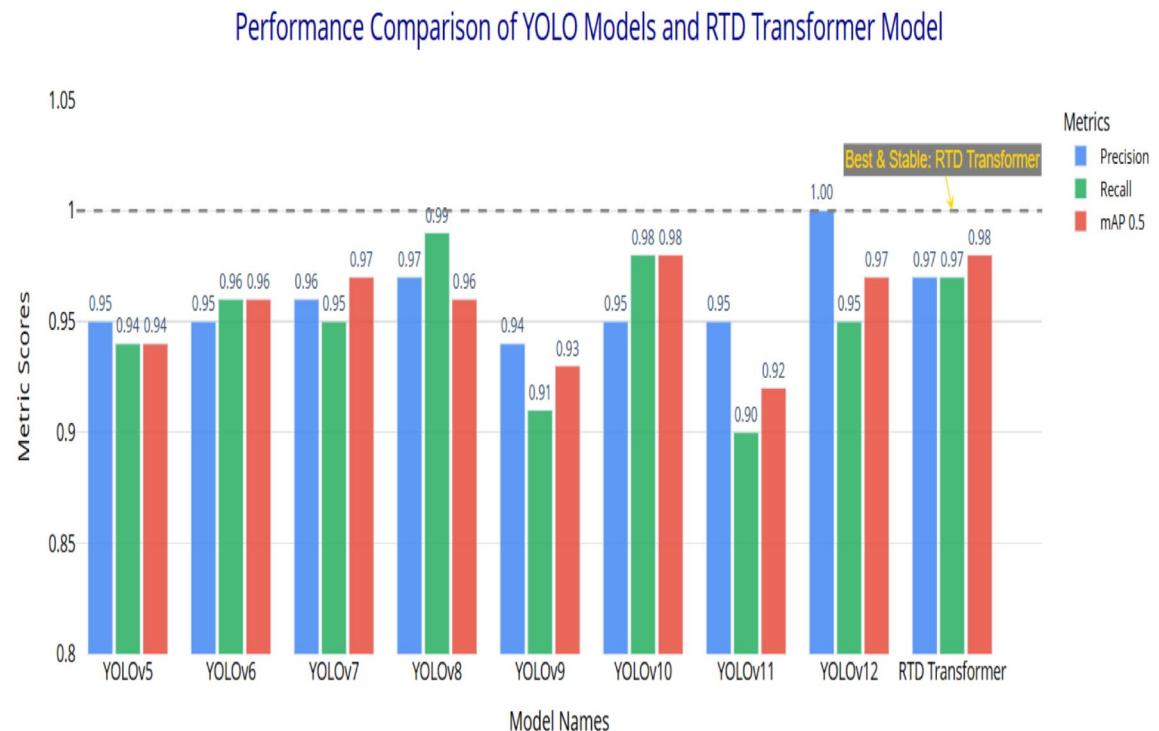


Fig. 9. Comparison of Precision, Recall, and mAP@0.5 scores across all models.

$$\text{MHSA}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (30)$$

Equation 30 represents MHSA where $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$ is the projective matrix and W^O is the output weight matrix.

YOLOv8-seg

YOLOv8-seg performs better but not the best, with a recall of 0.99 and precision of 0.97 which indicates that this model is highly effective in not only detecting the objects (recall) but also identifying them correctly (precision).

YOLOv12

YOLOv12 has achieved recall of 0.95 but attains a perfect precision of 1.00 and strong mAP@0.5 of 0.97. It indicates that v12 does not give any false positives and maintains localization accuracy.

YOLOv10

YOLOv10 shows the recall and mAP@0.5 of 0.98. It detects all objects keeping localization precise, while its precision is 0.95 that indicates over detection in some cases.

YOLOv6 and YOLOv7-seg

YOLOv6 and YOLOv7-seg are balanced with 0.95–0.97, 0.95, 0.97 metrics around all evaluation metrics. YOLOv7-seg achieves mAP@0.5 of 0.97 that reflects consistent localization, on the other hand YOLOv6 has high recall and it has more object coverage at modest computational power.

YOLOv5-seg

The precision, recall and mAP@0.5 of the YOLOv5-seg model are 0.95, 0.94, 0.94 respectively. It is a well-balanced model with much better detection performance. It has good optimization that keep inference fast and resource-friendly.

YOLOv9-seg and YOLOv11-seg

Both YOLOv9-seg and YOLOv11-seg achieves lower precision and recall score that is in between 0.91–0.95, 0.91–0.95. On the other hand, mAP@0.5 is between 0.92–0.93, 0.92–0.93 that gives a hint of over-parametrization. Both have architectural depth that increases training time and difficulty in training as well. It leads to under-generalization and slower real-time performance models.

Inference speed is another important thing that influences preference for a given model.

$$T_{\text{YOLO}} = S^2 B(5 + C) \cdot \frac{1}{\text{FPS}} \quad (31)$$

Equation 31 represents the computational complexity and latency where S is the grid size, B is the number of bounding boxes per grid cell, C is the number of detected classes, and FPS is the frame per second.

$$T_{\text{RTD}} = O(N^2 d) \quad (32)$$

Equation 32 represents the transformer complexity where N is the number of image patches and d is the embedding dimension.

$$\text{Latency} = \frac{1}{\text{Throughput (Images/sec)}} \quad (33)$$

Equation 33 represents real time inference latency.

Since YOLOv8-seg has higher precision and recall while being slightly more time-consuming to inference in terms of the number of weights, it is likely that its inference speed may increase incrementally compared to RTD transformer and YOLOv7-seg, with much better performance metrics. YOLOv10 and YOLOv12 showed better detection performance, as YOLOv10 achieved the highest recall and YOLOv12 achieved perfect precision, but they have moderate inference speed due to computational complexity. On the other hand, YOLOv6 is stable and maintain balanced performance with relatively faster inference. While YOLOv11-seg, YOLOv9-seg are the latest model, but both have been introduced with more layers or more computationally complex operations. It has even slower inference speeds compared to RTD transformers and YOLOv7-seg, thus contributing to its lower performance metrics. On the other hand, RTD transformer have the best performance in terms of mAP@0.5 because they have achieved a more favorable trade-off between better detection accuracy during inference and the associated time. It may make it more versatile in many more real-time applications. RTD transformer is also likely to have a much better inference speed than YOLOv8-seg, as it has undergone good optimization.

Real-time evaluation metrics

The primary objective of this research was to develop an online wheel inspection system for live condition monitoring; therefore, real-time evaluation of each model was extremely important. Each model was evaluated on real-time videos of moving trains captured at 30 FPS with an average latency of 30ms to make sure timely detection of defects. Normalized confusion matrix was used for visualizing the accurate predictive ability of each model for individual classes, i.e., Cracks-Scratches, Discoloration, Shelling, Wheels, and Background, as shown in Fig. 10. During real-time evaluation, it is noted that YOLOv8-seg struggled to accurately identify the shelling

class, whereas YOLOv9-seg and YOLOv12 faced challenges in identifying Cracks/Scratches due to the similarity of low-level features and their inherent complexity. These variations in model performance, particularly for TP and FP, can intuitively be attributed to several factors. YOLOv5-seg demonstrated robust feature extraction capabilities, enabling it to detect the most shelling instances. YOLOv6 and YOLOv10 exhibited moderate performance but misclassified between background and shelling. However, RTD transformer and YOLOv7-seg demonstrated the most consistent performance. YOLOv11-seg had good results in identifying wheel defects but also faced issues with shelling. Despite YOLOv12 and YOLOv9-seg architectural advancements, they exhibited generalization problems, whereas RTD achieved the best balance across all classes.

Figure 11 represents the results of a 3-fold Cross Validation analysis on unseen data of different YOLO series models and a transformer model. The performance of each model is evaluated based on Precision, Recall, mAP50, and mAP50-95. Each metric is represented with a distinct color: green for Precision, orange for Recall, blue for mAP50, and purple for mAP50-95. The RTD transformer achieved the overall best performance, with high Precision, Recall, mAP50, and mAP50-95 scores. YOLOv8 showed balanced performance across all metrics, particularly exceeding in mAP50, while its mAP50-95 is slightly lower. On the other hand, YOLOv6 performed the lowest across all metrics, especially in Recall and mAP50-95. Overall, RTD has the highest score, while YOLOv12, YOLOv6 and YOLOv9-seg achieved the lowest Recall and mAP50-95.

Figure 12 shows the training and validation loss curves for box and segmentation/DFL loss model training. The RTD transformers and YOLOv7-seg had the lowest training and validation loss across 100 epochs, indicating superior stability and model convergence, generalization for both detection and segmentation applications. YOLOv5-seg and YOLOv6 also showed stable training, with slight higher but smooth loss curves. In contrast, YOLOv8-seg, YOLOv9-seg, YOLOv11-seg, YOLOv12 exhibited higher training and validation losses, while YOLOv9 and YOLOv12 showed the slowest convergence and the most persistent loss values. Overall, the RTD transformer showed the best learning dynamics, supporting its superior performance observed in detection metrics and based on Table 7 and Fig. 9, RTD transformer emerges as the best model, with balanced accuracy, highest mAP@0.5, and the lowest loss values.

The heatmap shown in Fig. 13 compares the performance (Precision, Recall, mAP50, mAP50-95) of YOLOv5–YOLOv12 instance segmentation, object detection and RTD transformer over the last 50 epochs. YOLOv9-seg consistently outperforms the other models in every metric with better precision, recall, and mAP. While YOLOv8-seg shows low precision, particularly in the late epochs that exhibit higher false positive rates. YOLOv6, YOLOv10, and YOLOv12 exhibit steady performance gains, contributing to the overall trend of improvement across models. YOLOv5-seg and YOLOv7-seg show very similar performance; each model is

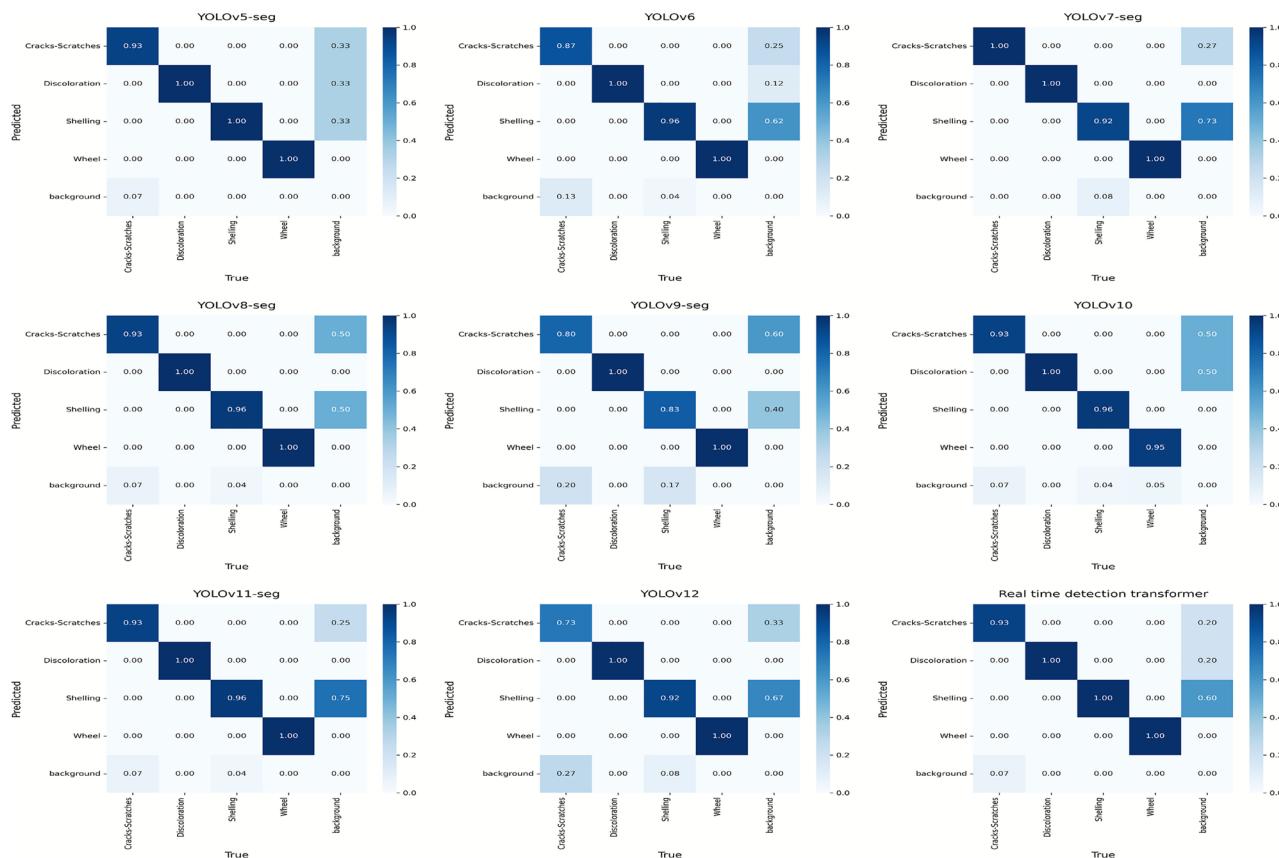


Fig. 10. Normalized confusion matrix of all models, showing classification performance across defect detection classes.

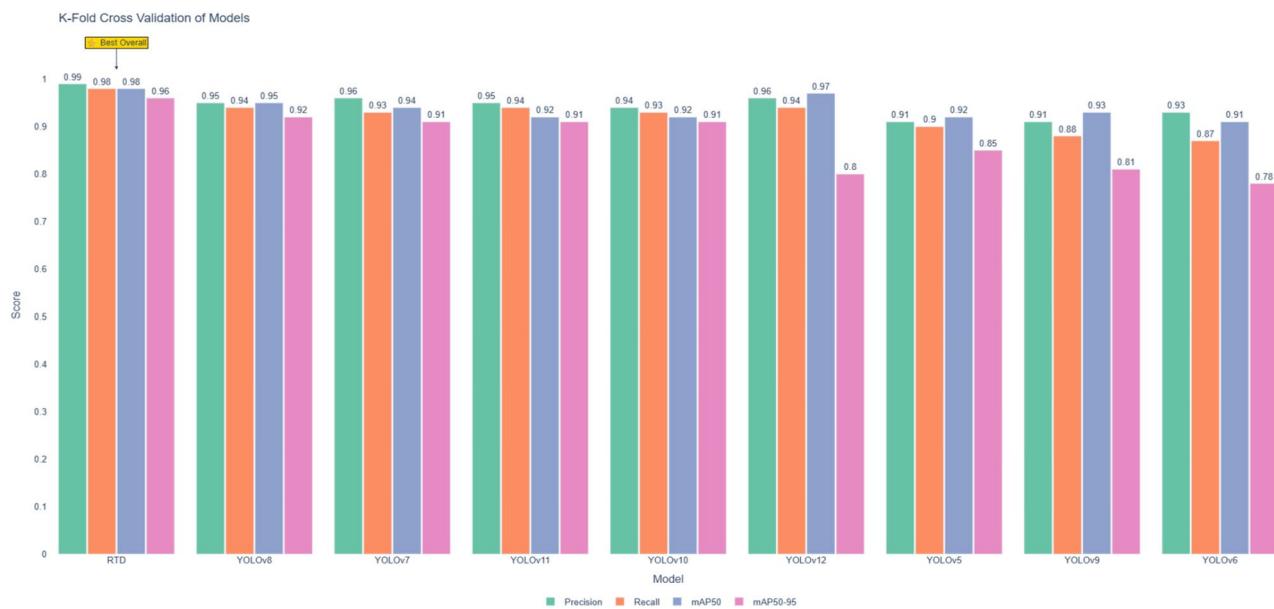


Fig. 11. K-Fold Cross Validation Performance of Different Models based on Precision, Recall, mAP50 and mAP50:95.

slightly better than the other. The heatmap also shows overall performance improvement through the epochs for all models, although at different rates of improvement varies. The heatmap also highlights consistent performance improvement across all models, with YOLOv11-seg and RTD transformers showing steady advancements. These findings provide an indication that additional hyperparameter tuning can be made, especially for YOLOv8-seg, in order to bring it close to YOLOv9-seg.

The proposed defect detection technique using RTD Transformer for railway wheels outperforms previous approaches by identifying a broader range of defects, including Shelling, Discoloration, and Cracks/Scratches. This technique is highly accurate. On the other hand, the selection of the final model is considered on the basis of Flops (Floating point operations), number of parameters and accuracy in edge device. YOLOv5-seg has the lowest parameters (1.9 million) and Flops (4.5 billion), while moderate precision (91%) recall (90%) and mAP50 (92%) in cross validation, representing the most efficient choice for deployment on an edge device. The RTD-transformer has the highest number of parameters, flops and accuracy, but it is not the best choice for edge device, so YOLOv5-seg is chosen as it has superior performance in accurately detecting railway wheel defects with the least computational resources.

Table 8 presents a comprehensive comparative analysis of previous studies and the proposed deep learning-based approach for railway wheel defect detection. It systematically outlines the evolution of machine learning and computer vision techniques from 2018 to 2024, emphasizing variations in model architectures, dataset characteristics, hyperparameters, and performance metrics. The comparison highlights that earlier studies primarily focused on limited defect classes such as normal, flat, and spalling, achieving moderate accuracy levels, whereas the proposed model incorporates four defect categories—wheel, shelling, crack/scratch, and discoloration—using high-resolution imagery and advanced architectures like YOLO and RTD Transformer. Demonstrating a significant improvement in precision (97%), recall (97%), and mAP@0.5 (98%), the proposed system surpasses prior works in both detection accuracy and real-time efficiency, thereby establishing a robust framework for reliable, real-time condition monitoring in railway operations.

The best-trained model was deployed on an edge device to evaluate the system's performance in real-world. Figure 14 shows the interface of the TrainScan: AI mobile application, where users can either upload images from storage or tap on the camera icon to capture live video streams. In both scenarios, the defects on the surface of a railway wheel were detected in real time without the aid of additional hardware. This is important as it enables the automation of highly predictive features at the point of inspection, which mitigates downtime and accelerates the inspection process.

Conclusion

This study demonstrates the importance of real-time monitoring for railway wheelsets, which are critical to the safety and reliability of railway systems. The proposed framework, which combines YOLO instance segmentation, object detection models, and the RTD Transformer, effectively identifies railway wheel defects. YOLOv5-seg achieved 91% precision, 90% recall, and 92% mAP@0.5 in cross-validation, showcasing its strong performance. Additionally, it excels in real-time defect detection with a processing speed of 30 FPS and 30 ms latency, making it the most reliable solution for operational railway environments. While other models, such as YOLOv6, YOLOv7-seg, YOLOv8-seg, YOLOv9-seg, YOLOv11-seg, YOLOv12, and RTD Transformers, also demonstrated strong performance, YOLOv5-seg remains the best choice for defect detection due to its balanced

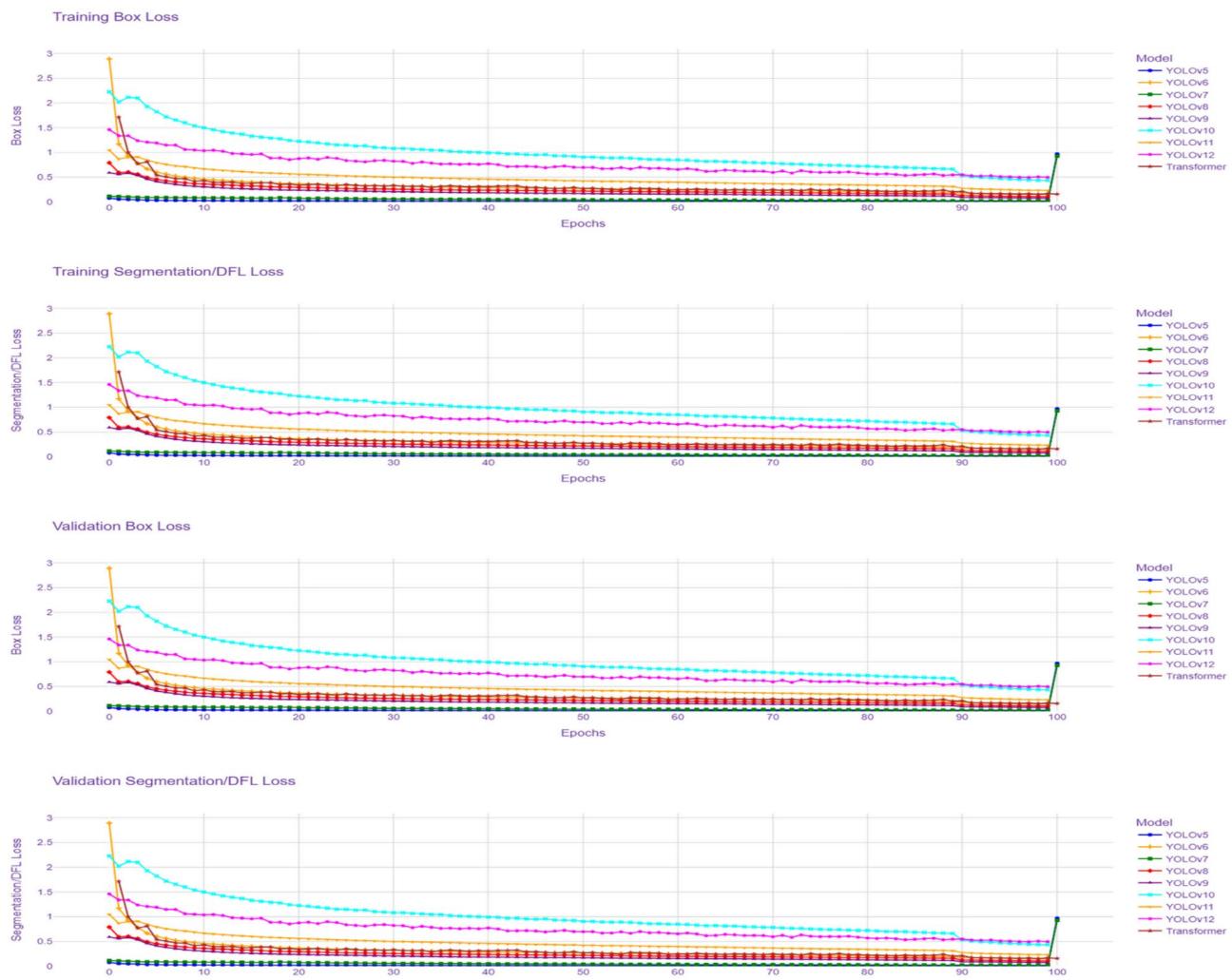


Fig. 12. Training and validation box and segmentation loss trends for all models over 100 epochs.

precision, recall, and real-time efficiency. Despite these successes, challenges remain, particularly in detecting defects in heavily occluded or small regions. Future work will focus on incorporating sensor data for enhanced defect detection and scaling the system to support extensive real-time monitoring of train systems, ultimately improving safety and operational efficiency.

Performance Metrics Comparison: YOLOV5, YOLOv6, YOLOV7, YOLOV8, YOLOV9, YOLOv10, YOLOV11, YOLOv12 Transformer (Last 50 Epochs)

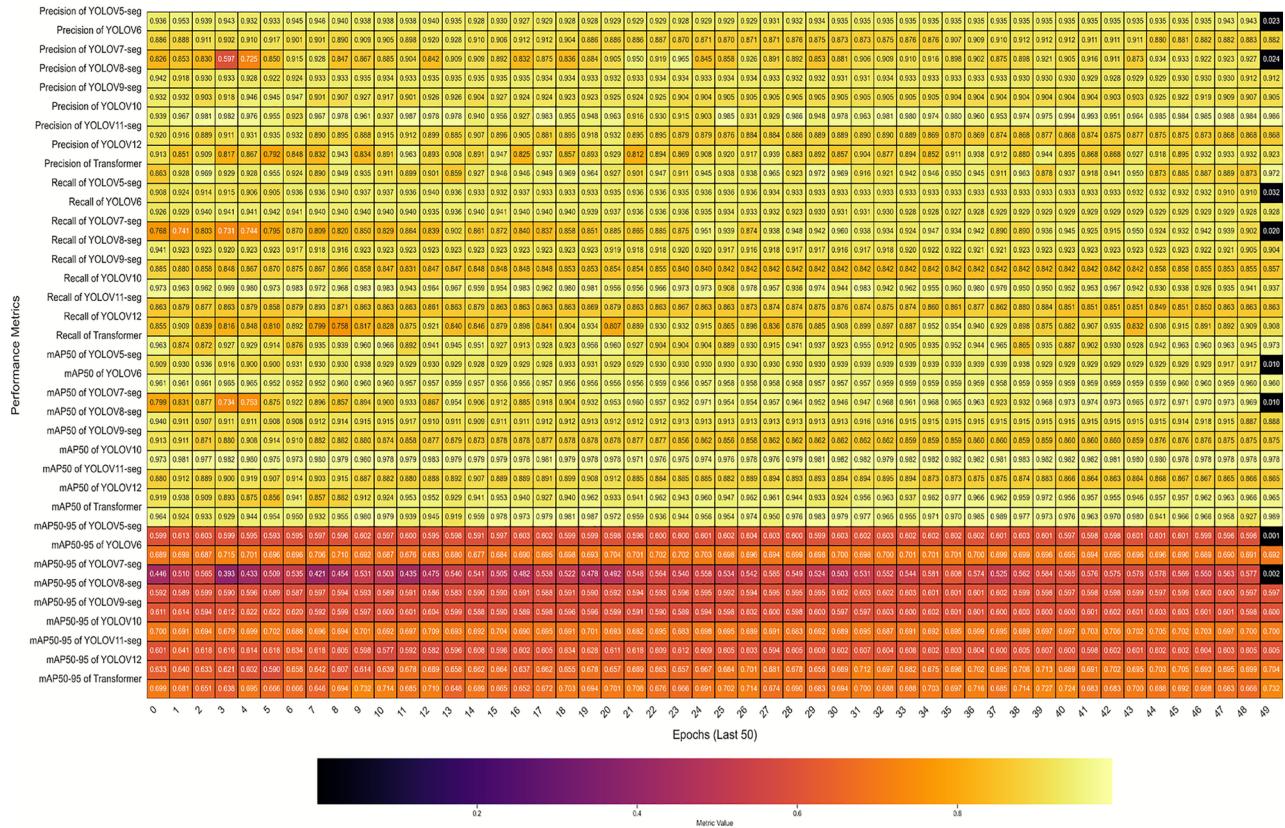
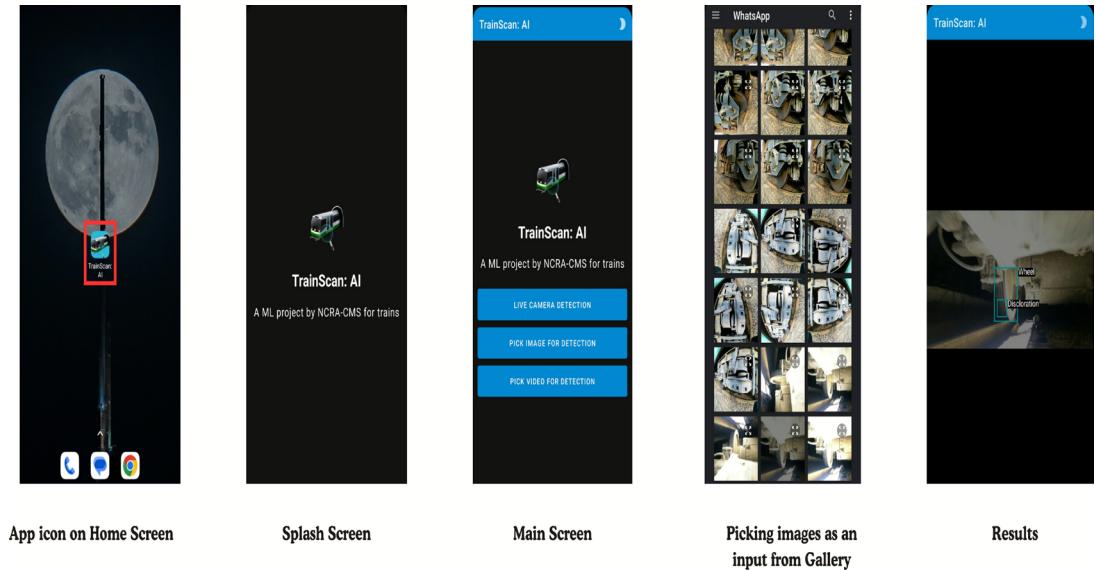


Fig. 13. Heatmap of performance metrics among all models (last 50 epochs).

Techniques Used (Reference)	Year	Classes	Hyperparameters	Resolution	FPS	Speed	Distance	Frames	Metrics
ML Techniques for Wheel Defect Detection ⁵⁵	2018	3 (normal, flat, spalling)	LR: 0.01, Batch: 64, Epochs: 100	1920×1080	30	20 km/h	2.0 m	1000	Accuracy: 92.0%, Precision: 90.5%, Recall: 89.0%
Improved YOLOv3 for Railway Tracks ⁶²	2020	5 (rock, tire, wood, plastic, metal)	Batch: 32, LR: 0.001, Anchor Boxes: Custom	1920×1080	25	25 km/h	2.0 m	4500	mAP: 92.4%, Precision: 88.7%, Recall: 86.2%
ML and DL for Wheel Defect Detection ⁶³	2022	3 (normal, flat, spalling)	LR: 0.001, Batch: 32, Epochs: 50	1920×1080	30	10 km/h	1.5 m	5000	Accuracy: 95.5%, Precision: 94.0%, Recall: 93.5%
Few-Shot Learning for Intrusion Detection ⁶⁴	2022	3 (person, vehicle, animal)	LR: 0.001, Batch: 16, Epochs: 100	1280×720	30	15 km/h	2.0 m	3000	Accuracy: 88.5%, F1-score: 0.85
Multi-Modal CNN for Object Detection ⁶⁵	2022	4 (plastic bag, paper, cloth, metal)	Batch: 32, LR: 0.001, Kernel: 3×3, Pooling: 2×2	1920×1080	30	20 km/h	1.8 m	4000	Precision: 90.2%, Recall: 87.4%, F1-score: 0.88
DL for Railway Wheel Defect Detection ⁶⁶	2023	3 (normal, flat, spalling)	LR: 0.001, Batch: 32, Epochs: 50	1920×1080	30	15 km/h	2.0 m	3000	Accuracy: 96.2%, Precision: 94.5%, Recall: 93.0%
Hybrid Deep Learning for wheel defect detection ⁵⁹	2024	3 (normal, flat, conical)	Batch: 100, Epochs: 100, Loss: Categorical Cross-entropy, Optimizer: Adam	Not mentioned	Simulated data	Not mentioned	Not mentioned	Not mentioned	Accuracy: MLP-88.6%, RF-99.0%, DT-98.9%, KNN-95.0%, SVM-83.1%
Deep learning for wheel detection ⁵⁰	2024	1 (Wheel)	LR: 0.001, Batch: 8, Epochs: 100	3840×2160	30	5 km/h	Center of tracks	7380	Precision: 92%, Recall: 91%, mAP@0.5: 94%, mAP@0.5–95.5: 43%
Lightweight CNN for Bolt Detection ⁶⁷	2024	2 (bolt present, missing)	Epochs: 1000, Batch: 256, Optimizer: SGD, LR: 0.02	4000×3000	30	Not mentioned	Not mentioned	2000	Accuracy: 95.5%, Precision: 94.0%, Recall: 93.5%
ML and DL for Wheel Defect Detection ⁶⁸	2024	3 (normal, flat, spalling)	LR: 0.001, Batch: 32, Epochs: 50	1920×1080	30	10 km/h	1.5 m	5000	Accuracy: 95.5%, Precision: 94.0%, Recall: 93.5%
Proposed DL for Railway Fault Detection ⁸	2024	4 (Wheel, shelling, crack/scratch, discolouration)	LR: 0.001, Batch: 8, Epochs: 100	2704×1520	24	0–50 km/h	10 inches	22464	Precision: 97%, Recall: 97%, mAP@0.5: 98%

Table 8. Comparison table for Summary of Contributions: Previous Studies vs. Proposed Work.**Fig. 14.** Interface of the prototype mobile application "TrainScan: AI".

Data availability

The dataset was collected with indigenously designed wayside system. The GoPro hero 9 cameras were mounted on setup. The optimal height and distance from the track were adjusted after several experiments on our automated wheelset test rig that is used for wheel defect detection which was developed in the condition monitoring systems laboratory of the university. The dataset was named as FaultSeg and it was collected by our team of researchers from the School of Industrial Engineering from the University of Malaga, NCRAAI, and NCRA-CMS Lab at MUET through the deployment of the wayside data acquisition system at Kotri Railway Station for

multiple days at different timings. The dataset is available online on the zenodo and figshare repository: <https://doi.org/10.5281/zenodo.13162335>, <https://doi.org/10.6084/m9.figshare.27996866.v1> respectively. The dataset paper is published and available on scientific data journal of nature, <https://doi.org/10.1038/s41597-025-04557-0>.

Received: 7 May 2025; Accepted: 19 November 2025

Published online: 27 November 2025

References

1. Fonseca-Soares, D.D., Eliziário, S. A., Galvacio, J. D. & Ramos-Ridao, A. Greenhouse gas emissions in railways: systematic review of research progress. *Buildings*, **14**(2), <https://doi.org/10.3390/buildings14020539> (2024).
2. UIC. Sustainable mobility reporting. <https://uic.org/sustainability/article/sustainable-mobility> (2022).
3. Srinath, M., Sreram, S., Jayanth, P. D. D., S. B. & Venkatachalapathi, S. S. Advanced monitoring of rail cracks using laser-emat system. *Ijarcce*, **5**(12), 413–416 <https://doi.org/10.17148/ijarcce.2016.51295> (2016).
4. Kostrzewski, M. & Melnik, R. Condition monitoring of rail transport systems: a bibliometric performance analysis and systematic literature review. *Sensors*, **21**(14), <https://doi.org/10.3390/s21144710> (2021).
5. Shaikh, M. Z., Mehran, S. & Nava, E. State-of-the-art wayside condition monitoring systems for railway wheels: a comprehensive review. *IEEE Access*, **XX**, <https://doi.org/10.1109/ACCESS.2023.3240167> (2023).
6. Arfiansyah, O. & Handoko, Y. A. Reliability improvement of railway vehicles: a review. *J. Nas. Apl. Mekatronika, Otomasi dan Robot Ind.*, **3**(1), 19, <https://doi.org/10.12962/j27213560.v3i1.18994> (2024).
7. Das, B., Shaikh, M. Z., Jatoi, S., Nava, E. & Chowdhry, B. S. FaultSeg: A Dataset for Train Wheel Defect Detection. Zenodo. <https://zenodo.org/records/12957455> (2024).
8. Shaikh, M. Z. et al. FaultSeg: A Dataset for Train Wheel Defect Detection. *Scientific Data* **12**, 1–12. <https://doi.org/10.1038/s41597-025-04557-0> (2025).
9. Magel, E. & Kalousek, J. Identifying and interpreting railway wheel defects. In *IHHA Conf. Int. Heavy Haul Assoc. Conf. Freight Car Truck*, 5.7–5.20 (1996).
10. Google. ML practicum: image classification. <https://developers.google.com/machine-learning/practica/image-classification/> (2022).
11. Sundaram, S. & Zeid, A. Artificial Intelligence-Based Smart Quality Inspection for Manufacturing. *Micromachines* **14**, 570 (2023).
12. Gao, R., He, Q., Feng, Q. & Cui, J. In-Service Detection and Quantification of Railway Wheel Flat by the Reflective Optical Position Sensor. *Sensors* **20**, 4969 (2020).
13. Gazafrudi, S. M. M., Younesian, D. & Torabi, M. A High Accuracy and High Speed Imaging and Measurement System for Rail Corrugation Inspection. *IEEE Transactions on Industrial Electronics* **68**, 8894–8903 (2021).
14. Liang, B., Iwnicki, S., Ball, A. & Young, A. E. Adaptive noise cancelling and time-frequency techniques for rail surface defect detection. *Mech. Syst. Signal Process.* **54**, 41–51. <https://doi.org/10.1016/j.ymssp.2014.06.012> (2015).
15. Asplund, M., Palo, M., Famurewa, S. & Rantatalo, M. A study of railway wheel profile parameters used as indicators of an increased risk of wheel defects. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit.*, **230**(2), 323–334. <https://doi.org/10.1177/0954409714541953> (2016).
16. Bevan, A. & Klecha, S. Use of magnetic flux techniques to detect wheel tread damage. *Proc. Inst. Civ. Eng. Transp.* **169**(5), 330–338. <https://doi.org/10.1680/jtran.16.00025> (2016).
17. Zhang, K. et al. Research on eddy current pulsed thermography for rolling contact fatigue crack detection and quantification in wheel tread. In *Proc. 2016 IEEE Int. Wheel. Congr. IWC*, 5–11, <https://doi.org/10.1109/IWC.2016.8068338> (2017).
18. Chen, S., Wang, K., Chang, C., Xie, B. & Zhai, W. A two-level adaptive chirp mode decomposition method for the railway wheel flat detection under variable-speed conditions. *Journal of Sound and Vibration* **498**, 115963 (2021).
19. Mosleh, A., Montenegro, P. A., Costa, P. A. & Calçada, R. Railway Vehicle Wheel Flat Detection with Multiple Records Using Spectral Kurtosis Analysis. *Applied Sciences* **11**, 4002 (2021).
20. Xu, J. et al. Effect of wheel flat on dynamic wheel-rail impact in railway turnouts. *Vehicle System Dynamics* **60**, 1829–1848 (2022).
21. Zhou, Z. et al. Dynamic response feature of electromechanical coupled drive subsystem in a locomotive excited by wheel flat. *Engineering Failure Analysis* **122**, 105248 (2021).
22. Fry, G. T. Ultrasonic Testing of Steel - Railway Age. <https://www.railwayage.com/mw/ultrasonic-testing-of-steel/>, (2022).
23. Feng, B., Wu, J., Tu, H., Tang, J. & Kang, Y. A Review of Magnetic Flux Leakage Nondestructive Testing. *Materials* **15**, 7362 (2022).
24. Zhang, Y., Gao, X., Peng, J. & Guo, J. Defect detection and recognition technology in wheel images based on deep learning. *Ciop* **12935**, 79. <https://doi.org/10.1117/12.3006110> (2023).
25. Bernal, E., Spiriyagin, M. & Cole, C. Ultra-low power sensor node for on-board railway wagon monitoring. *IEEE Sens. J.* **20**(24), 15185–15192. <https://doi.org/10.1109/JSEN.2020.3011132> (2020).
26. Shruthi, S. N., Iype, G. M., Kavana, K. C. P., Sharon, M. & Subhash, S. Rail track defect detection using enhanced method of magnetic flux leakage signal. In *Proc. - 2021 Int. Conf. Des. Innov. 3Cs Comput. Commun. Control. ICDI3C*, 277–280, <https://doi.org/10.1109/ICDI3C53598.2021.00062> (2021).
27. Cavuto, A., Martarelli, M., Pandarese, G., Revel, G. M. & Tomasini, E. P. FEM based design of experiment for train wheelset diagnostics by laser ultrasonics. *Ultrasonics* **113**, 106368. <https://doi.org/10.1016/j.ultras.2021.106368> (2021).
28. Wang, M., Li, K., Zhu, X. & Zhao, Y. Detection of Surface Defects on Railway Tracks Based on Deep Learning. *IEEE Access* **10**, 126451–126465 (2022).
29. Luo, H., Li, J., Cai, L. & Wu, M. STrans-YOLOX: Fusing Swin Transformer and YOLOX for Automatic Pavement Crack Detection. *Applied Sciences* **13**, 1999 (2023).
30. Zhang, Z., Liang, M. & Wang, Z. A Deep Extractor for Visual Rail Surface Inspection. *IEEE Access* **9**, 21798–21809 (2021).
31. Soleimani, H., Moavenian, M., Masoudi Nejad, R. & Liu, Z. An applied method for railway wheel profile measurements due to wear using image processing techniques. *SN Applied Sciences* **3**, 1–10 (2021).
32. Guo, Y. et al. Using ISU-GAN for unsupervised small sample defect detection. *Scientific Reports* **12**, 1–13 (2022).
33. Li, S., Zhao, H. & Ma, J. An edge computing-enabled train obstacle detection method based on YOLOv3. *Wirel. Commun. Mob. Comput.*, 2021, <https://doi.org/10.1155/2021/7670724> (2021).
34. Belotti, V., Crenna, F., Michelini, R. C. & Rossi, G. B. Wheel-flat diagnostic tool via wavelet transform. *Mech. Syst. Signal Process.* **20**(8), 1953–1966. <https://doi.org/10.1016/j.ymssp.2005.12.012> (2006).
35. Krummenacher, G., Öng, C. S., Koller, S., Kobayashi, S. & Buhmann, J. M. Wheel defect detection with machine learning. *IEEE Trans. Intell. Transp. Syst.* **19**(4), 1176–1187. <https://doi.org/10.1109/TITS.2017.2720721> (2018).
36. Shaikh, M. Z., et al. Predictive maintenance in urban railway systems using machine learning models. *Proc. 2024 Global Conf. on Wireless and Optical Technologies (GCWOT)*, 1–5. <https://doi.org/10.1109/GCWOT63882.2024.10805699> (2024).
37. Shaikh, M. Z., et al. AI-powered real-time detection of wheel defects in railways using YOLOv8. *Proc. 2024 Global Conf. on Wireless and Optical Technologies (GCWOT)*, 1–7. <https://doi.org/10.1109/GCWOT63882.2024.10805632> (2024).
38. Qu, Z., Jiang, P. & Zhang, W. Development and Application of Infrared Thermography Non-Destructive Testing Techniques. *Sensors* **20**, 3851 (2020).

39. Li, H. et al. Virtual point tracking method for online detection of relative wheel-rail displacement of railway vehicles. *Reliability Engineering and System Safety* **246**, 110087 (2024).
40. Benzeroual, H., Khamlichia, A. & Zakrifi, A. Inspection of transverse flaws for railways using phased array ultrasonic technique. *International Review of Applied Sciences and Engineering* **12**, 119–126 (2021).
41. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: unified, real-time object detection. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, (December), 779–788, <https://doi.org/10.1109/CVPR.2016.91> (2016).
42. Zhang, Q. et al. Advancing ultrasonic defect detection in high-speed wheels via UT-YOLO. *Sensors*, **24**(5), <https://doi.org/10.3390/s24051555> (2024).
43. Zhao, L., Guo, J., Xu, D. & Sheng, L. Transformer3D-Det: improving 3D object detection by vote refinement. *IEEE Trans. Circuits Syst. Video Technol.* **31**(12), 4735–4746. <https://doi.org/10.1109/TCSVT.2021.3102025> (2021).
44. Fang, Y. et al. You only look at one sequence: rethinking transformer in vision through object detection. In *Adv. Neural Inf. Process. Syst.* **31**, 26183–26197 (2021).
45. Wang, Q., Lu, C., Gao, L. & He, G. Transformer-based multiple-object tracking via anchor-based-query and template matching. *Sensors*, **24**(1), <https://doi.org/10.3390/s24010229> (2024).
46. Adortech. Visual Inspection vs. Laser Scanning for Wheel Profiling. <https://adortech.com/blog/a-comparison-of-visual-inspection-and-laser-scanning-for-wheel-profiling>.
47. Braga, J. A. P. & Andrade, A. R. Multivariate statistical aggregation and dimensionality reduction techniques to improve monitoring and maintenance in railways: The wheelset component. *Reliability Engineering and System Safety* **216**, 107932 (2021).
48. Caldeira, G. A. C., Braga, J. A. P. & Andrade, A. R. Assessing Wear Evolutions in Railway Wheelsets Using a Survival Modeling Approach. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, **7**, (2021).
49. Baasch, B., Heusel, J., Roth, M. & Neumann, T. Train Wheel Condition Monitoring via Cepstral Analysis of Axle Box Accelerations. *Applied Sciences* **11**, 1432 (2021).
50. Shaikh, M. Z., Mehran, S. & Nava, E. Design and development of a wayside AI-assisted vision system for online train wheel inspection. *Eng. Reports*, 1–13, <https://doi.org/10.1002/eng2.13027> (2024).
51. Zakir, M., Ahmed, Z., Nava, E., Hussain, S. & Milanova, M. Deep learning-based identification and tracking of railway bogie parts. *Alexandria Eng. J.* **107**(July), 533–546. <https://doi.org/10.1016/j.aej.2024.07.064> (2024).
52. Sresakoolchai, J., Hamarat, M. & Kaewunruen, S. Automated machine learning recognition to diagnose flood resilience of railway switches and crossings. *Sci. Rep.* **13**, 1–11 (2023).
53. Lou, H. et al. DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor. *Electronics* **12**, 2323 (2023).
54. Zhang, C., Xu, D., Zhang, L. & Deng, W. Rail Surface Defect Detection Based on Image Enhancement and Improved YOLOX. *Electronics* **12**, 2672 (2023).
55. Sresakoolchai, J., Ngamkhanong, C. & Kaewunruen, S. Hybrid learning strategies: integrating supervised and reinforcement techniques for railway wheel wear management with limited measurement data. *Front. Built Environ.* **11**, 1546957 (2025).
56. Shebani, A. & Iwnicki, S. Prediction of wheel and rail wear under different contact conditions using artificial neural networks. *Wear* **406–407**, 173–184 (2018).
57. Sresakoolchai, J. & Kaewunruen, S. Wheel flat detection and severity classification using deep learning techniques. *Insight: Non-Destructive Testing and Condition Monitoring*, **63**, 393–402, (2021).
58. Vishwakarma, A., DasGupta, A. & Racherla, V. Detecting and Locating Track Defects Using Clustering Algorithm. SSRN <https://doi.org/10.2139/ssrn.4711483> (2024).
59. Shaikh, K., Hussain, I. & Chowdhry, B. S. Wheel defect detection using a hybrid deep learning approach. *Sensors*, **23**, (14), <https://doi.org/10.3390/s23146248> (2023).
60. Yang, C., Dimyati, K. & Mohamad, M. Detection and Classification of Railway Track Surface Erosion-caused Holes and Scratches Defects Based on YOLOv5s. *Tehnicki Vjesnik* **31**, 1288–1296 (2024).
61. Kaewunruen, S., Sresakoolchai, J. & Stittle, H. Machine Learning to Identify Dynamic Properties of Railway Track Components. https://doi.org/10.1142/S0219455422501097_22 (2022).
62. Wei, X., Wei, D., Suo, D., Jia, L. & Li, Y. Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model. *IEEE Access* **8**, 61973–61988 <https://doi.org/10.1109/ACCESS.2020.2984264> (2020).
63. Xing, Z., Zhang, Z., Yao, X., Qin, Y. & Jia, L. Rail wheel tread defect detection using improved YOLOv3. *Meas. J. Int. Meas. Confed.* **203**, 111959. <https://doi.org/10.1016/j.measurement.2022.111959> (2022).
64. Gong, X., Chen, X., Zhong, Z. & Chen, W. Enhanced few-shot learning for intrusion detection in railway video surveillance. *IEEE Trans. Intell. Transp. Syst.* **23**(8), 11301–11313. <https://doi.org/10.1109/TITS.2021.3102613> (2022).
65. Ye, T., Zhang, J., Zhao, Z. & Zhou, F. Foreign body detection in rail transit based on a multi-mode feature-enhanced convolutional neural network. *IEEE Trans. Intell. Transp. Syst.* **23**(10), 18051–18063. <https://doi.org/10.1109/TITS.2022.3154751> (2022).
66. Yue, Y. et al. Automatic recognition of defects behind railway tunnel linings in GPR images using transfer learning. *Meas. J. Int. Meas. Confed.* **224**, 113903. <https://doi.org/10.1016/j.measurement.2023.113903> (2024).
67. Alif, M. A. R. & Hussain, M. Lightweight convolutional network with integrated attention mechanism for missing bolt detection in railways. *Metrology* **4**(2), 254–278. <https://doi.org/10.3390/metrology4020016> (2024).
68. Zhang, Z. et al. Railway obstacle intrusion warning mechanism integrating YOLO-based detection and risk assessment. *J. Ind. Inf. Integr.* **38**, 100571. <https://doi.org/10.1016/j.jii.2024.100571> (2024).

Acknowledgements

This research is supported by Departamento de Ingeniería de Comunicaciones, Universidad de Málaga, Spain, National Center for Robotics, Automation and Artificial Intelligence (NCRAAI MUET), Higher Education Commission Pakistan, Sindh Higher Education Commission Pakistan, NCRA-CMS Lab of Mehran University of Engineering and Technology, Jamshoro, and by the doctoral program of Mechanical Engineering and Energy Efficiency, School of Industrial Engineering, University of Málaga, Spain. We would also like to thank Pakistan Railways, Kotri Railway Station, and Carriage and wagon workshop Hyderabad near American quarters, Pakistan railways for their assistance and support. We would also like to express our sincere gratitude to Dawood University of Engineering and Technology for their continued support, with special thanks to Dr. Samreen Hussain, Vice Chancellor, for her leadership and invaluable contributions to this endeavor. Furthermore, we would like to acknowledge the International Electronic Machines Corporation of USA, EU Funded Erasmus Plus Capacity Building in Higher Education ACTIVE Climate Action Project and EU Funded Erasmus Plus Capacity Building in Higher Education Project CENTRAL (Capacity building and ExchaNge towards attaining Technological Research and modernizing Academic Learning). We are also thankful to INDUDECAM Project CPP2021-009117, financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea “NextGenerationEU”/PRTR, Universidad de Málaga for the support in our research work. We would also acknowledge Institute of Oceanic Engineering Research, Vicerrectorado de Movilidad y Proyectos Internacionales Vicerrectorado de internacionalización of the University of Málaga for their support. We would also like to acknowledge

Edibon, Spain and Paktech instruments Company, Pakistan for their guidance and collaboration. Special thanks to Professor Dr Mariofanna Milanova, University of Little Rock at Arkansas whose expertise and guidance in AI significantly contributed to the success of this research. We would also like to acknowledge and thank Engr. Muhammad Rizwan Khan, Senior Examiner of Patents and Design, The Patent office IPO-Pakistan for his guidance on Industrial Design and IPRs Awareness.

Author contributions

M.Z.S conceived the study and was responsible for data collection, formal analysis, investigation, software, validation, visualization, and writing the original draft as well as reviewing and editing. E.N.B was involved in conceptualization, funding acquisition, methodology, supervision, validation, collaboration, reviewing, management of the project, and project administration. S.J contributed to writing the original draft, validation, investigation, and software. All authors read and approved the manuscript.

Funding

This research is funded by University of Malaga, Spain, EU Funded Erasmus Plus Capacity Building in Higher Education ACTIVE Climate Action Project (ERASMUS-EDU-2023-CBHE Project:101082866 ACTIVE), EU Funded Erasmus Plus Capacity Building in Higher Education Project CENTRAL (Capacity building and Exchange towards attaining Technological Research and modernizing Academic Learning 598914-EPP-1-2018-1-DK-EPPKA2-CBHE-JP) and National Center of Robotics and Automation-Condition Monitoring Systems Laboratory, Mehran University of Engineering and Technology Grant/Award Number: 2(1076)/HEC/M/E/2018/704.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.Z.S. or E.N.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025