



**OPEN**  
**DATA DESCRIPTOR**

## FaultSeg: A Dataset for Train Wheel Defect Detection

Muhammad Zakir Shaikh<sup>1,2,3</sup>✉, Sahil Jatoi<sup>1,3</sup>, Enrique Nava Baro<sup>4</sup>✉, Bhagwan Das<sup>5</sup>, Samreen Hussain<sup>6</sup> & Bhawani Shankar Chowdhry<sup>1,3</sup>

Wheels are a critical component of railway infrastructure and work as the load carrier of the train. However, defective wheels pose a serious risk to safety that can gravely jeopardize people's safety. There is a significant risk of injury or death from defective wheels, endangering the lives of individuals. In this research, FaultSeg dataset is presented for automatic train wheel defect detection for railway transportation around the world. The FaultSeg consists of 829 manually annotated images of faulty wheels acquired by an indigenously developed wayside data acquisition system. Expert Annotators have manually annotated three classes of potential defects: Cracks/Scratches, Shelling, and Discoloration. To assess the practicality of the FaultSeg dataset for training and testing advanced deep learning (DL) models, the dataset was used to train and evaluate the YOLOv9 instance segmentation algorithm. The model achieves an approximate score of 87% accuracy. These results showcase the usability of the FaultSeg dataset in automatic inspection systems and data driven predictive maintenance strategies to safeguard and ensure the safety of railway transportation.

### Background & Summary

Railway is considered an integral part of the transport network due to its speed and ease of travel. It is also regarded as the most economically viable option compared to other modes of transport for instance, road, air, or sea<sup>1</sup>. Railway transport also ensures public safety and operates with high efficiency depending on the state of the infrastructure components in use, for instance rolling stock. The wheels are considered as the most important rolling component of the infrastructure, primarily responsible for the movement of the carriage as well as load balancing. The wheels also suffer due to extreme wear-out leading to small or big faults. These defects cause a disturbance in the conical structure of the wheel that is responsible for creating a centripetal force that keeps the wheels aligned with the track, even on curve<sup>2</sup>.

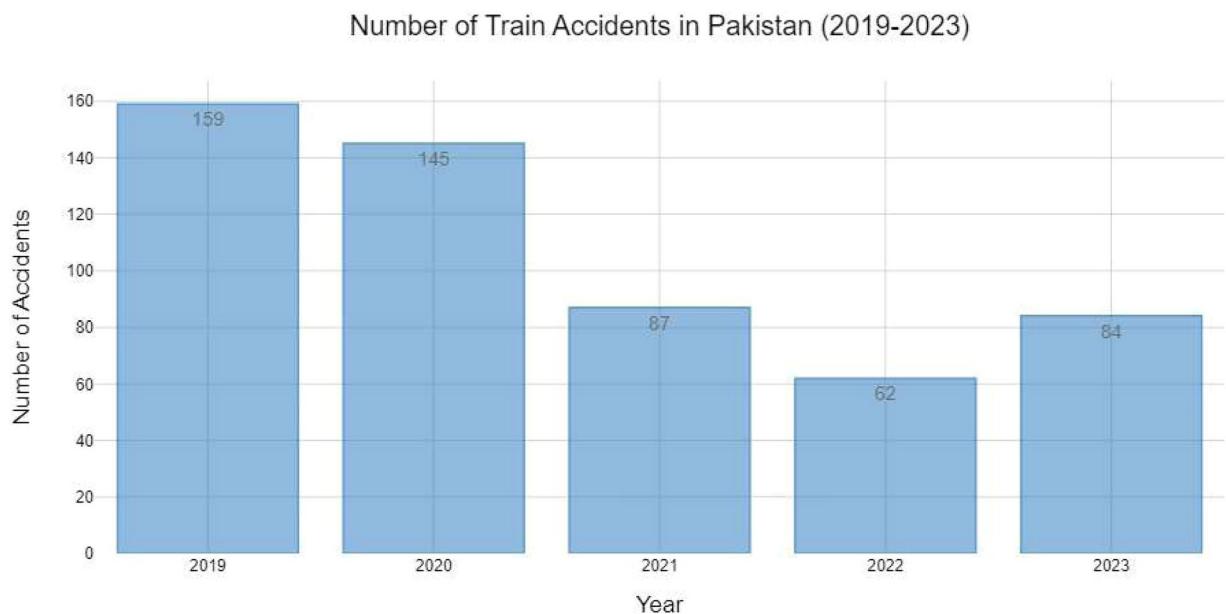
China's high-speed railway network has exceeded 40,000 kilometers in running mileage, enabling it to become the nation with the largest network in this regard by the end of 2023. With the significant rise in operational mileage, speed, and amount of passenger and freight traffic, the wheel-rail interaction has deteriorated over time, especially in relation to wheel defects. Friction and rolling contact between the train wheelset and railway track surfaces significantly result in wear and distortion.

Numerous train accidents have occurred worldwide as a result of the poor condition of the wheels. The wheel suffers from different types of faults having different level of severity. The most prevalent forms of faults found on the surfaces of the wheels are cracks, shelling, and discoloration, shown in Fig. 1. Cracks appear when there is a reduction in the structure of metallic materials, typically results from thermal stress or fatigue and are commonly classified as surface cracks and structural cracks. Shelling occurs due to the fatigue caused by the rolling of the wheel against the track. This leads to severe damage on a small portion of wheel that affects the geometry of the component. Severe shelling may also cause the derailment. Discoloration is caused due to the overheating of the metal surface. Scratches occur due to the debris present on the track.

<sup>1</sup>National Center for Robotics, Automation and Artificial Intelligence, Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan. <sup>2</sup>Mechanical Engineering and Energy Efficiency, School of Industrial Engineering, University of Malaga, Malaga, Spain. <sup>3</sup>NCRA-CMS Lab, Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan. <sup>4</sup>Departamento de Ingeniería de Comunicaciones, Campus de Teatinos, Universidad de Málaga, Málaga, Spain. <sup>5</sup>Centre for Artificial Intelligence Research and Optimization (AIRO), and Design and Creative Technology Vertical, Torrens University, Adelaide, Australia. <sup>6</sup>Dawood University of Engineering and Technology, Karachi, Pakistan. ✉e-mail: muhammadzakirshaikh90@gmail.com; zakir.shaikh@faculty.muet.edu.pk; zakir.shaikh@uma.es; en@uma.es



**Fig. 1** Various Forms of Defect.



**Fig. 2** Yearly Railway Accidents in Pakistan (2019–2023).

According to<sup>3,4</sup> 28% of the operational wheels have small cracks and 15% have visible shelling issues. Whereas 20% have discoloration in them due to excessive heating. Similarly, 10% and 12% of operational wheels either have a flat spot or a scratch. Continuous condition monitoring of wheels is mandatory for ensuring the safety and sustainability of railway operations. Traditionally, the task is performed using a manual, periodic method, where a group of skilled personnel painstakingly inspect each wheel for any possible problem. A more detailed evaluation is also carried out after every three to eight years period where the entire train is taken out of operation for a holistic evaluation at a maintenance facility. These approaches have their limitations in terms of cost, time consumption, as well as the high probability of error due to human negligence or weakness. The crux of the matter is that the most of train mishaps and derailments occur in developing nations where the quality of train infrastructure and transportation is inadequate. It is undeniable that a significant number of individuals worldwide, especially in impoverished nations, prefer travelling via train. Similarly, the majority of individuals in Pakistan opt for train travel due to financial constraints that make air travel unaffordable. Furthermore, Pakistan has reported a total of 537 major accidents in the last five years leading to a high number of casualties<sup>5</sup>. Yearly statistics of railway accidents in Pakistan is shown in Fig. 2.

DL has found numerous applications in the domain of condition monitoring and predictive maintenance in multiple sector<sup>6–9</sup>. A considerable amount of research has been done for semi/fully automating the process of condition monitoring of train wheels. The explored techniques mostly focus on vibration analysis, acoustic signal identification, as well as rule-based image analysis. For instance, in study<sup>10</sup> the Fuzzy-logic method is proposed for diagnostics of railway wheels with defects detection by measuring wheel vibration. Compared to the healthy wheels, the faulty wheels showed spikes in their feature values. Similarly, researchers proposed<sup>11</sup> an algorithm of frequency-domain Gramian angular field (FDGAF) for the first time to represent the vibration

signatures of wheel flat by featured image so that the concept of wheel flat can be intelligently diagnosed by machine learning based on vibration image classification, which is of great significance to railway vehicles performance maintenance.

The use of sensory devices coupled with DL is also a very popular research area in the condition monitoring of railway infrastructure. For example, the use of an ultra-low power sensor node for railway wagon onboard monitoring with an analog defect detector<sup>12</sup>. A deep belief network and cuckoo search algorithm were also used for fault identification using multi-sensor signals<sup>13</sup>. A comprehensive review study also studied the advancement of recent wayside railway wheel flat detection intelligent methods and concluded that the direction of the development of the wayside railway wheel flat detection techniques is to simplify devices, give preference to the use of multi-sensor fusion, enhance the accuracy of the algorithm and improve the sensing equipment operation intelligence<sup>14</sup>. Another study provides a comprehensive description of all the faults of railway track surface (cracks, holes, deformations of surface) which is useful for training machine learning models for automated fault detection<sup>15</sup>. Moreover, another comprehensive review evaluates various data acquisition systems and analysis methods to enhance the efficiency and reliability of condition monitoring systems for railway wheels<sup>16</sup>. Moreover, The predictive maintenance framework was proposed using various machine learning models such as boosting as well as bagging algorithms to predict failure of air production unit (APU) in electric trains<sup>17</sup>. However, a study leverages YOLOv8, an anchor free object detection model, to detect real-time railway wheels faults<sup>18</sup>.

Image processing is also widely used in the railways for fault diagnosis, condition monitoring of rolling stock, and detecting various railway faults<sup>19–26</sup>. It consists of innovative approaches such as identifying geometric characteristics of rolling stocks, implementing of condition monitoring system for switches, using a vision inspection system for rail components, and deploying an AI-based surveillance system for railway crossing. Additionally, multiple studies have used techniques like adaptive multiscale morphological filtering for defect identification by measuring the wheel diameter of railroad vehicles by using a digital camera<sup>27,28</sup>. Furthermore, particle swarm optimization along with morphological filtering method was used for fault diagnosis of railway vehicle bearings<sup>29</sup>. Cepstrum analysis was also found to be suitable for wheel flat detection<sup>30</sup>. Another study emphasizes on visual inspection systems specifically providing its significant advantages over traditional sensor based methods<sup>31</sup>. Moreover, A study introduces an Anchor-Free Yolov8 (AFYv8) model to detect and track bogie parts in real-time without manual intervention<sup>32</sup>. Additionally, image processing is also used for wheel profile measurement<sup>33</sup> to measure abnormalities of the wheel profile by images taken from the cutting plane of railway wheels, and then made a comparison with the wheel profile drawing pro forma in the original plan.

However, to the best of our knowledge, there is currently no benchmark publicly available dataset of wheel defect detection using image segmentation. Various types of train wheel defects include cracks, scratches, shelling, discoloration, spalling, fatigue, and peeling. The most common defects are cracks, shelling, and discoloration from which DL models get confused between scratch, crack, and peeling so it is recommended to consider them in a single class. Recent progress in DL and computer vision provides a potential solution for automatic defect detection using high-resolution images and state-of-the-art DL models. By automated defect detection, it would possibly ensure the safety and reliability of railway operations.

In this research work, the FaultSeg dataset is introduced for the development of DL model to create an automated system for wheel defect detection. This dataset includes high-resolution images for various defects and annotated for instance segmentation. It reduces inspection time and enhances detection accuracy. It is crucial for trend analysis, pattern recognition, and in planning maintenance for railway operations, assuring safety and efficiency. The key contributions of this work are:

- FaultSeg dataset is introduced first time that reports the high-resolution images precisely marked for instance segmentation of wheel defects. This dataset is created to enhance the performance of deep learning models in detecting and classifying wheel defects.
- FaultSeg dataset helps in the proposed DL model that significantly reduces inspection time while improving detection accuracy. These efforts facilitates real-time monitoring and maintenance planning, ensuring operational safety and efficiency in railway systems.

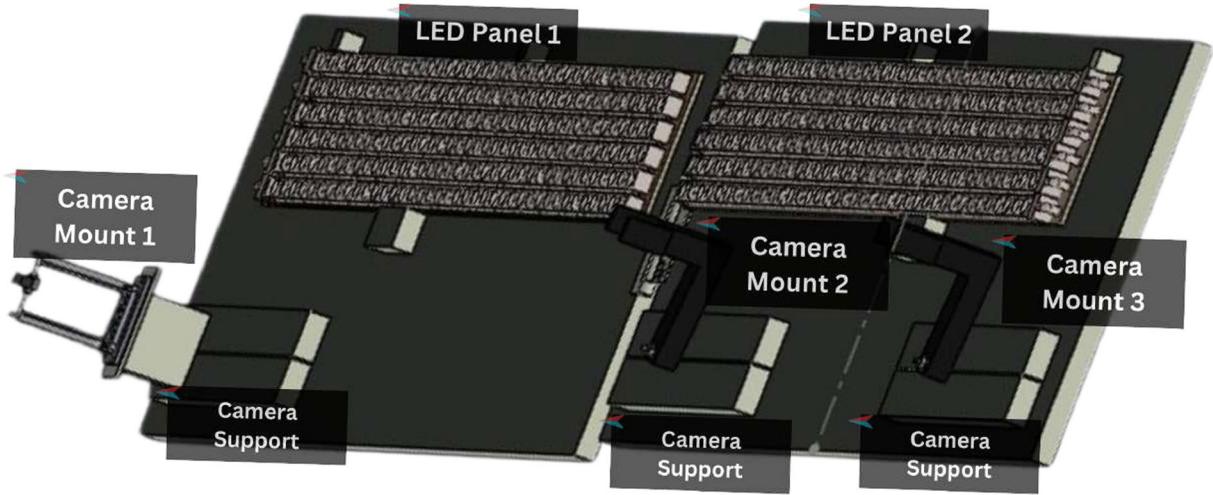
Our study offers a comprehensive analysis of the data collecting procedure, experimental framework, technological requirements, and validation techniques. This comprehensive methodology not only showcases the dependability and relevance of the FaultSeg dataset but also provides useful perspectives for future investigations and practical applications in the domain of automated defect identification and maintenance planning

## Methods

**Data collection.** The data was collected using multiple cameras attached to an indigenously designed wayside inspection system. The videos were recorded with GoPro Hero 9 cameras. The camera configuration for data collection was fixed at a specific height and distance from the track, and that can create a viewpoint limitation. In order to mitigate this limitation, a variety of data augmentation techniques including random rotations, translations, flipping were used to extent in generalizing model to different viewpoints. Frames were extracted and selected based on visual quality using a Python script. The positioning parameters used for ideal camera setup are shown in Table 1. The optimal values for  $h_b$  and  $d_t$  were determined based on experimental camera calibration performed using our automated wheelset test rig system for wheel set fault detection that was developed domestically in the Condition Monitoring Systems laboratory of the University.

Figure 3 shows the CAD model of the data acquisition system used for gathering the data. It consists of a wooden hardware setup measuring 138.25 inches in length and 68.45 inches in width upon which the entire system was built. The base consists of moveable wooden camera holders. These holders are designed to capture maximum tread area of a running wheel. Along with that, an illumination setup was also integrated into the

Parameter	Details
Camera Model	GoPro Hero 9
Height from Base ( $h_b$ )	3 inches
Distance from Track ( $d_t$ )	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.**Fig. 3** CAD Model of the Data Acquisition Setup.

system using LED panels to mitigate the shadow effect and adjust varying light conditions throughout the day for effective data capture. Table 2 shows the specifications of data acquisition setup that includes setup configurations, power source, and additional components.

Kotri Railway Station is one of the important and busiest railway junctions of Pakistan. The system was deployed on an operational track at the Kotri Railway Station. The temperature during data acquisition ranged from 15 to 25 degrees Celsius during the winter season to collect data in real-time as well as to test the reliability of the setup. This narrow range may not reflect full spectrum of real-world scenarios. To overcome this restriction, preprocessing of data used brightness adjustment, contrast scaling, and noise addition. These techniques reproduce environmental variations such as lighting and weather, etc and help to strengthen the model.

Figure 4(a) shows the system placed on an actual track to collect the data and Fig. 4(b) shows the image captured by the system. The data was collected for multiple consecutive days when numerous trains passed over the system. Furthermore, the data collected during nighttime, particularly after 1800 hours, was obtained utilizing LED panels, which are already part and parcel of our developed wayside system, in order to mitigate the impact of shadows. Some trains arrive at Kotri Railway Station from Karachi and head towards cities in the north, such as Mirpurkhas, Sialkot, Lahore, Peshawar, and Multan. The trains that are leaving Karachi and heading north are referred to as Up Trains as they are going towards the North. However, some trains from North areas are arriving at Kotri Railway Station on the other track, with Karachi as their eventual destination. These trains are typically classified as down trains because they are travelling southward from the north. Furthermore, not every train stops at the Kotri Railway station. Some trains stop at the Kotri Railway Station, while others pass through without stopping. Our Team collected data on the six trains which were heading towards the north. Table 3 includes a brief overview of the various trains our team encountered during the data acquisition process together with their schedule.

Acquiring real-time data on the operational train station was itself one of the main obstacles keeping the safety of the team and the passengers nearby in mind. Prior to moving forward for the acquisition of the data, official approval from the government in general and the Railway department in particular was required. Installing the setup in accordance with the train timetable was also a challenge. Installation of the indigenously designed wayside data acquisition system had to be completed at least ten minutes before the train arrives. To guarantee the safety procedure, the data acquisition team had to maintain at least a 6-meter buffer distance from the track.

Multiple videos of varying lengths were recorded on several consecutive days. The videos were recorded at 2704 × 1520 resolution, at 24 frames per second (FPS). The total length of the shortlisted videos were 28 minutes and 52 seconds. Frame extraction resulted in 10,392 frames at a rate of 6 FPS. These frames were further trimmed to 1872 frames to only include the instances where the wheel was within the field of view. A cohort

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.

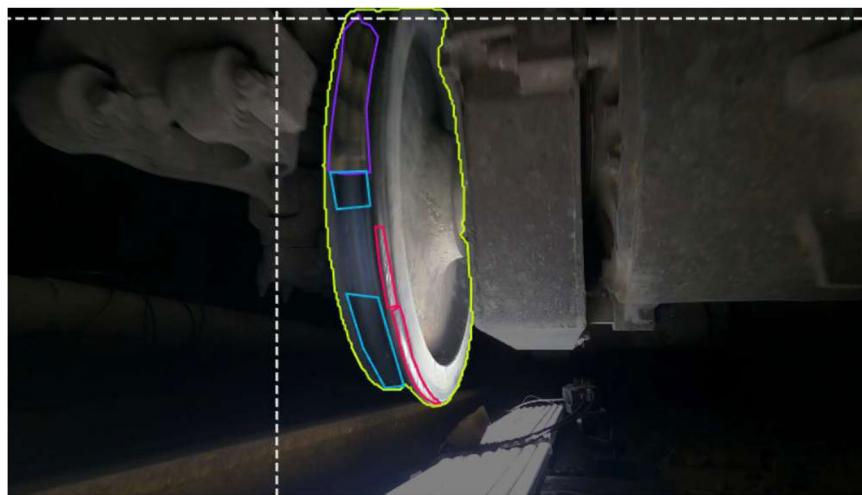
Train Route And Bogie Numbers	Train Name	Tentative Timing of arrival at Kotri Station	Train Stop Information	Departure (Origin)	Destination
09 UP	Allama Iqbal Express	18:16	Yes (Departure at 18:18)	Karachi Cantt	Sialkot
10 DOWN	Allama Iqbal Express	06:33	Yes (Departure at 06:35)	Sialkot	Karachi Cantt
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
26 DOWN	Bahauddin Zakaria Express	06:12	Yes (Departure at 06:14)	Multan	Karachi City

**Table 3.** A Brief Overview of the Various Trains our team encountered during the Data Acquisition Process together with their Schedule.**Fig. 4** (a) Deployed Data Acquisition System and (b) Sample Image Captured.

of 829 high-resolution images were selected for further processing including resize and augmentation. During the annotation process, defect labels were identified, and each image was annotated for their respective ground truth values such as, shelling, discoloration, and cracks/scratches with the collaboration and assistance of the Railway department to strengthen the reliability of the FaultSeg dataset. Manual annotations were performed using Roboflow. During the annotation process, it was verified whether the FaultSeg adequately represented the flaws on the train wheels or not. The annotated ground truths in the FaultSeg dataset are provided in multiple formats including TXT, XML, JSON, CSV, and tfrecord for each image.

Figure 5 shows the screenshot of an image annotated for the segmentation task using Roboflow. Each class label has been assigned to a separate consistent color for ease of visualization. Red indicates the places with cracks or scratches on the surface of the wheel, Cyan shows the signs of discoloration. Green outlines the edge of the wheel. Moreover, Purple marks the locations with shelling. These ground truth values help to find out the different types of defect classes of the wheel.

**Data preprocessing.** The annotated images were also subjected to different types of data preprocessing techniques including auto-orient, resize, and auto-adjust contrast using contrast stretching. Furthermore, numerous data augmentation techniques were also employed to increase the generalization capability of the DL models. The purpose of using preprocessing techniques was to eliminate class imbalance in the dataset like data augmentation



**Fig. 5** Annotated Sample Image using Roboflow.

Technique	Description
Cropping	Images were dynamically cropped by 2% on both the horizontal and vertical axes.
Translation or Flip	Images were horizontally and vertically flipped to introduce variations in the dataset.
Brightness Adjustment	The brightness was adjusted between $\pm 1\%$ to replicate various lighting scenarios.
Shear	Shear was applied to images with a horizontal and vertical distortion of $\pm 1^\circ$ pixels.
Rotation	Images were rotated 90° clockwise, counter-wise, and upside down.
Hue	Hue-intensity was used $\pm 1^\circ$ to shift the dominant colors in the images.
Saturation	Saturation was varied by $\pm 10\%$ units to alter the intensity of colors.
Exposure	Exposure was adjusted by $\pm 1\%$ units to modify the overall brightness of images.
Blur	A minimal blur of 0.1% pixel was applied to enhance the dataset.
Noise	A small amount of noise was used equivalent to 0.02%.

**Table 4.** Augmentation Techniques.

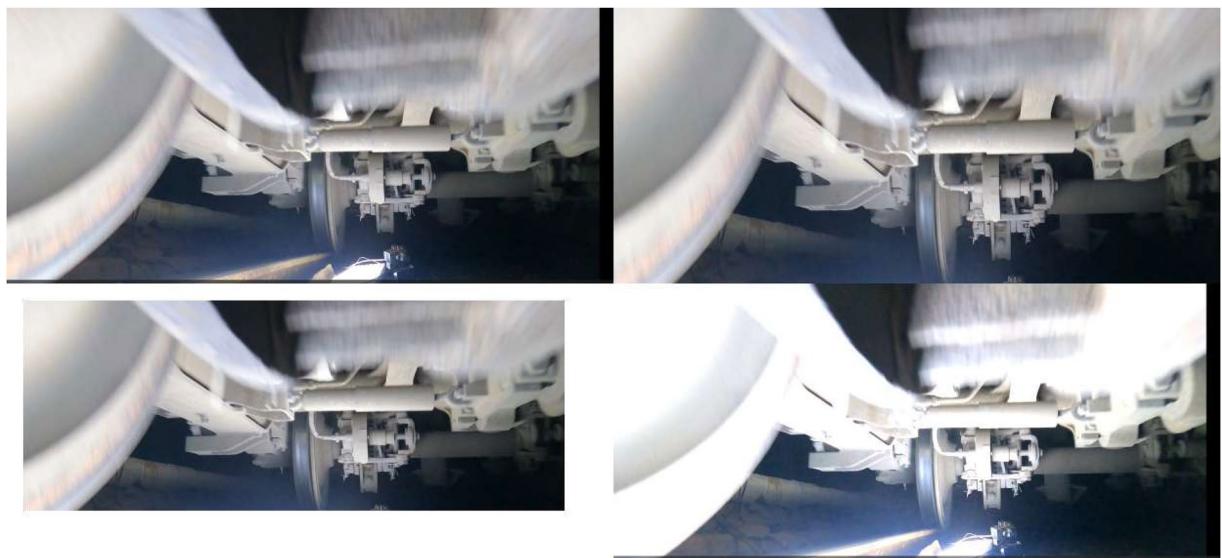
of images for minority defect classes. Particularly, random rotations, brightness adjustments and flipping were strategically employed on defects under-represented such as shelling and discolouration. Moreover, class-weights adjustments were also incorporated to the loss function, enhancing model sensitivity to less frequent defects. The combination of these methods create a more balanced model and reduces the model's inclination towards dominant classes like 'Wheel'. The augmentation techniques are detailed in Table 4. A visual representation of these techniques is also shown in Fig. 6.

#### Data Records

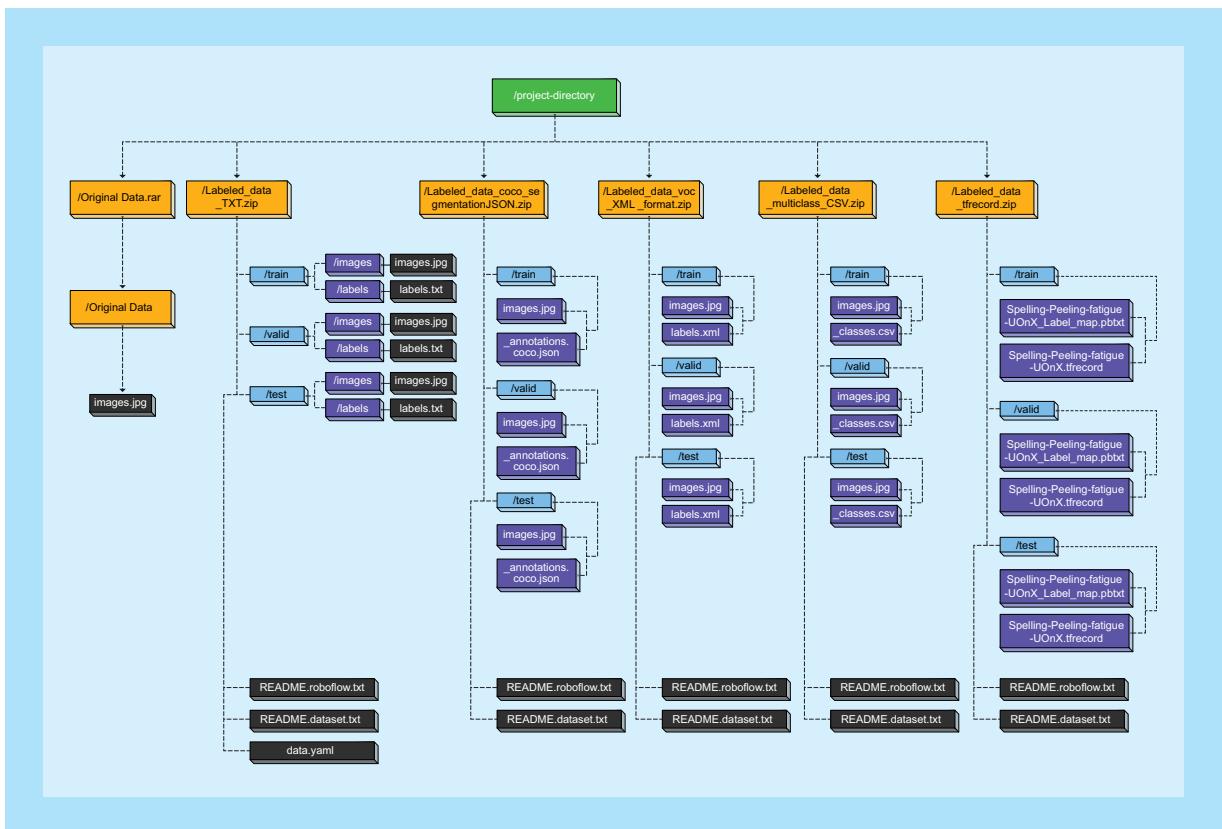
The FaultSeg dataset consists of high-resolution images of train wheels which could be used for developing and evaluating DL models for automatic defect detection through instance segmentation. The FaultSeg consists of 1872 frames where 829 frames were annotated for different defects that include cracks, scratches, shelling, and discoloration. Images were captured by three cameras mounted on the same side of the track to capture a full view. Setup was installed on track with prerequisites discussed in Table 1. The manual annotation process was thoroughly carried out and the results against the actual ground truth were revalidated by experts.

The images and annotation labels were organized into a repository tree to make access easy through the tree structure. Figure 7 illustrates how images and labels are organized into a repository named 'project-directory' based on tree structure. The directory tree includes zip files that consists of different repositories, images in JPG format, and labels in JSON, XML, TXT, CSV, and tfrecord format. Each labelled file consists of information of the defects associated with respective image – for example, the type of defect and its location (specifically coordinates) in the image – in a structured format that is ideal for training and evaluating DL models.

This FaultSeg is useful to researchers, engineers, software developers, and Railway Industries. Additionally, organizations and enterprises developing automated systems for detecting defects and maintaining railway transportation systems can utilize this dataset. The FaultSeg dataset is available under a free and open license for academic and research use on Zenodo at the following link "<https://zenodo.org/records/13162335>". Zenodo is an open repository supported by the European Organization for Nuclear Research (CERN). It is commonly utilized for depositing datasets, research papers, reports, and other research-related digital artifacts. The dataset also includes a detailed user guide to assist researchers in understanding and using the data. The release of the published dataset on Zenodo repository is 02 Aug 2024<sup>34</sup>. The data set is also available on the Figshare-an online open access repository: <https://doi.org/10.6084/m9.figshare.27996866><sup>35</sup>.



**Fig. 6** Augmented Image.



**Fig. 7** Project Directory Structure.

### Technical Validation

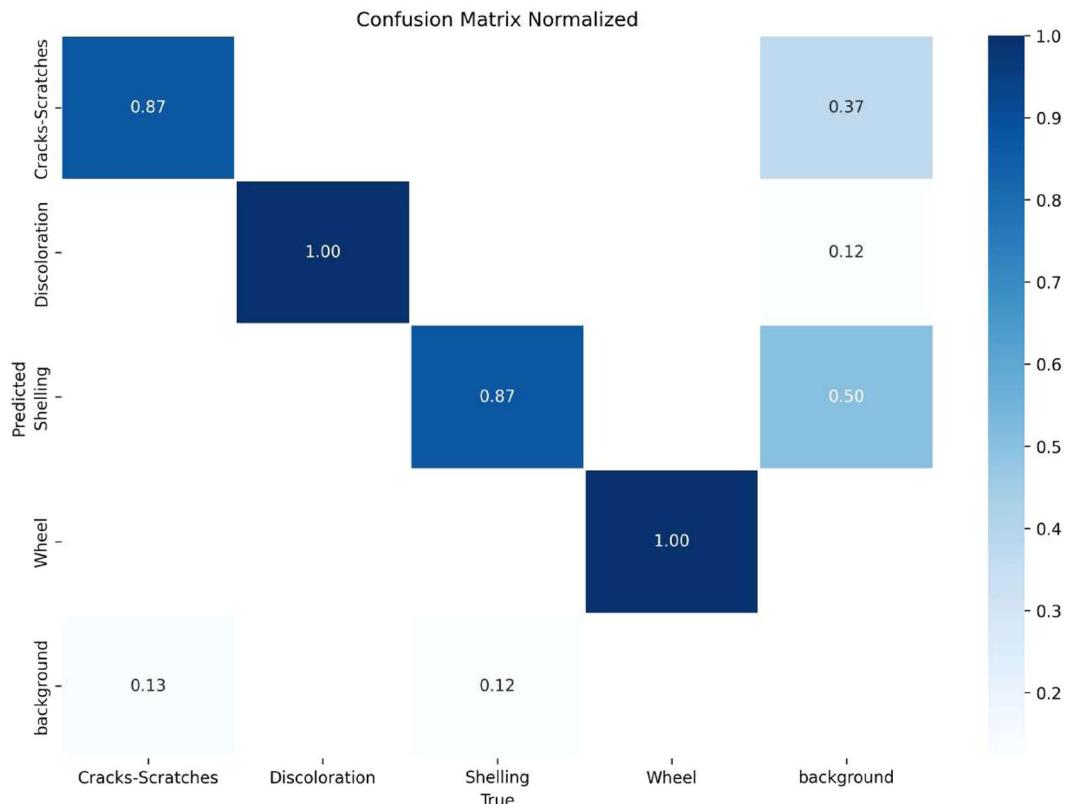
Technical validation was a necessary step to assess the potential of the FaultSeg dataset. The annotations were verified manually by annotators with the assistance of railway inspection team to verify that the annotations were correct and free of error. To further validate the dataset, it was used to train the YOLOv9 instance segmentation model.

**Data usability validation.** The main purpose of the FaultSeg dataset is to train AI-based, state-of-the-art fault object detectors. To create a baseline segmentation benchmark, the dataset was split into three subsets for

Model Parameter	Value
Epochs	100
Batch size	8
Workers	16
Optimizer	Adam
IOU	0.7
Learning rate	0.01

**Table 5.** Hyper Parameters utilized for the Technical Evaluation of Faults.

Class	Images	Instances	Box Precision	Box Recall	Box F1 Score	Box mAP50	Box mAP50-95
All	20	65	0.907	0.932	0.917	0.949	0.729
Wheel	15	15	0.798	0.867	0.830	0.883	0.495
Cracks/Scratches	4	6	1	0.986	0.995	0.995	0.733
Discoloration	19	24	0.831	0.875	0.852	0.924	0.714
Shelling	20	20	1	1	1	0.995	0.974

**Table 6.** Class-Wise Evaluation of YOLOv9-seg on the FaultSeg Dataset.**Fig. 8** Confusion Matrix of YOLOv9.

training, testing, and validation. After splitting the dataset, 799 images along with their respective labels were used to train the instance segmentation model i.e., YOLOv9-seg model. The training configuration was set as shown in Table 5.

The training was validated using 20 unseen sample images of wheels. YOLOv9 is a single stage detector model that processes the input image using an advanced CSPDarknet53 backbone, which efficiently extracts hierarchical features which is key aspect of contextual learning in images. The YOLOv9 neck architecture's feature aggregation method aids in the more accurate integration of multi-scale features and helps to pinpoint the exact location and attributes of various objects in varying sizes. The head architecture of YOLOv9 contains detection layers with bounding-box prediction, object-ness score prediction, and class probability prediction. The usage of an anchor box can improve the accuracy of localization.

Reference	Usage	Availability	Data Collection Method	Year of Publication	Source	Techniques	Image Resolution	Number of Samples	Defect Types	Dataset
<sup>36</sup>	Model training, performance evaluation	Public	Sensor data	2018	IEEE Xplore	Neural Networks, Support Vector Machines	Not Specified	1600 (flat spots), 6070 (shelling), 920 (non-roundness)	Flat spots, Shelling, Non-roundness	Wheel Defect Detection Dataset
<sup>37</sup>	Model training, defect prediction	Public	Sensor data	2020	ResearchGate	Neural Networks, Wavelet Transform	Not Applicable	Not Specified	Wheel flats	Hilbert Transform Dataset
<sup>38</sup>	Condition diagnosis, maintenance actions	Public	Smartphone images	2021	ResearchGate	Convolutional Neural Networks	Smartphone images	4600	Multiple defects	Rail Wheel Tread Images
<sup>39</sup>	Real-time classification	Public	Sensor data	2023	ResearchGate	1D Convolutional Neural Networks	800 × 600 pixels	2000	Tire wear	1D-CNN Tire Wear Dataset
<sup>40</sup>	Defect detection, model training	Public	Simulation data	2022	ResearchGate	Wavelet Transform, Neural Networks	1024 × 768 pixels	1500	Wheel flats	Wavelet Packet Decomposition Dataset
<sup>41</sup>	Model training, performance evaluation	Private	Simulation data	2023	MDPI Sensors	Hybrid Deep Learning	Not Applicable	Not specified	Wheel flat defects	Wheel Defect Detection
<sup>42</sup>	Defect identification, model training	Public	Image data and sensor data	2017	TU Delft Research Portal	Machine Learning and deep learning	Not Specified	Not Specified	Various wheel defects (including cracks, flats, and shelling)	Railway Wheel Defect Identification
<sup>34</sup>	Defect identification and model training	Public	Image data	2024	Zenodo	Deep learning	2704 × 1520 pixels	1872	Various wheel defects (Cracks/ Scratches, Discoloration, and Shelling)	Proposed technique (FaultSeg)

**Table 7.** Comparative Analysis of Faultseg Data Set with other existing data sets and studies.

**Model validation.** The model was validated using precision, recall, f1 score, as well as confusion matrix shown in Table 6 and Fig. 8. A brief description of their concept and calculation is given below:

- Confusion Matrix.** A confusion matrix is a metric used to visualize the result of a classification, represented as true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). It helps to identify the type of errors made by the model. For multiclass classification scenario, the values of TP, FP, TN, and FN are calculated independently for each target class.

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

- Precision.** The proportion of true positive detections among all positive detections. It is computed for Class<sub>A</sub> using formula given below:

$$Precision_{ClassA} = \frac{TP_{ClassA}}{TP_{ClassA} + FP_{ClassA}}$$

- Recall.** The proportion of true positive detections among all actual defects. It is computed using formula given below:

$$Recall_{ClassA} = \frac{TP_{ClassA}}{TP_{ClassA} + FN_{ClassA}}$$

- F1-Score.** F1-Score is the harmonic mean of precision and recall. It's a balanced measure of the quality of an individual model. It is computed using formula given below:

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

The confusion matrix shows excellent performance of the model for all individual classes, especially wheel and discoloration.

**Comparative analysis against previous studies.** The study was compared with previous studies based on image resolution, number of samples, techniques, applications, types of defects and usage shown in Table 7.

The proposed method<sup>34,35</sup> employs DL for defect detection and classification. It provides high-resolution images ( $2704 \times 1520$  pixels) and 3 different defect types found in wheels: Cracks/Scratches, Discoloration and Shelling. The primary asset of this technique is the inclusion of 1872 images. In contrast to other defect detection approaches which were based on sensor data or simulations with fewer defect-type classes and lower image quality, it creates detailed visual data and extensive manually annotated images. This could result in more accurate and robust defect detection in real-world railway maintenance.

### Usage Notes

This FaultSeg dataset will be a very useful resource for railway scientists, engineers, data scientists, and computer vision researchers working on automated fault identification and railway maintenance solutions. These high-resolution annotated images are essential building blocks to create and evaluate DL models that address potential applications, restrictions, and suggestions for further research. The data has also been acquired by our own indigenous Way Side inspection System that has been developed in our own Laboratory of the University and tested multiple times on the Railway Tracks and on Railway Stations.

Moreover, the utility of the FaultSeg dataset is not limited to the YOLO models. It can be used to train and evaluate multiple other state-of-the-art object detection and segmentation models as well including Detectron2, FastInst, SparseInst, etc. Applications, Limitations and recommendations for future work are described below

**Applications.** The FaultSeg can be used for several applications within the railway maintenance domain:

1. **Training Detection Models.** The annotated data from FaultSeg can be used to train DL models, including one or two-stage detection models. These models can automate the detection of defects in train wheels, such as cracks or shelling, thereby reducing the need for manual inspections and lowering associated costs.
2. **Automated Inspection Systems.** FaultSeg can be utilized to develop real-time automated inspection systems that identify train wheel defects. These systems can help prevent failures and accidents by alerting authorities before defects exceed critical thresholds.
3. **Resource Allocation.** By analyzing the defect, maintenance resources can be allocated more efficiently. This ensures that issues are addressed on time leading to the optimization of available resources.
4. **Regulatory Compliance and Safety Protocols.** The insights gained from FaultSeg dataset can be used to develop and refine safety protocols by ensuring that maintenance practices comply with industry regulations and standards. Real-time knowledge about the occurrence of defects can help in maintaining adherence to safety and quality guidelines.
5. **Benchmarking.** FaultSeg can serve as a benchmark dataset for evaluating the performance of various object detection and instance segmentation models. This includes comparing algorithms for train wheel defect detection in terms of processing speed, dataset size, accuracy, and training duration.

**Limitations.** The FaultSeg is comprehensive dataset, but it has some limitations:

1. **Class Imbalance.** The defective classes may have a varying number of instances, leading to a skew towards instances like the Wheel class. This imbalance between training instances will affect model performance in some way. One approach can be to oversample the minority class or use data augmentation.
2. **Viewpoint Constraint.** The images were captured from a consistent viewpoint and under a limited range of conditions at a specific location. The dataset was collected at different intervals throughout the day to encompass various lighting conditions.
3. **Resolution Limitations.** Although Images are high-resolution, some defects may not be detectable. Moreover, it was noticed that some defects are only detected with thermal or ultrasonic imaging techniques.

**Recommendations for Future Work.** To further enhance the FaultSeg and its applications, the following recommendations are proposed:

1. **Sampling a wider range of conditions, locations and types of defects.** Larger dataset, which include more conditions, locations, or types of defects. It will increase the generalizability of models trained on them.
2. **Data Sources.** Using additional data from a thermal or ultrasonic camera to reveal surface defects that normal images cannot acquire.
3. **Standardized benchmarks.** Establishing standardized benchmarks against which different model-based approaches to automate defect detection can be compared. It would help researchers assess their merits.
4. **Collaboration and open access.** Collaboration among researchers and making the FaultSeg freely accessible will promote innovation and accelerate the development of reliable railway maintenance solutions.

## Code availability

The raw video data was collected and stored on a GPU-enabled machine. Subsequently, the videos were processed to extract individual frames using a simple Python script.

Received: 12 August 2024; Accepted: 29 January 2025;

Published online: 20 February 2025

## References

- Trepáčová, M., Kurečková, V., Zámečník, P. & Řezáč, P. Advantages and disadvantages of rail transportation as perceived by passengers: A qualitative and quantitative study in the Czech Republic. <https://doi.org/10.5507/tots.2020.014>.
- Alkomy, H. & Shan, J. Modeling and validation of reaction wheel micro-vibrations considering imbalances and bearing disturbances. *J Sound Vib* **492**, 115766 (2021).
- Mageł, E. & Kalousek, J. Identifying and interpreting railway wheel defects. (1996).
- Liu, X. Z., Xu, C. & Ni, Y. Q. Wayside Detection of Wheel Minor Defects in High-Speed Trains by a Bayesian Blind Source Separation Method. *Sensors* **19**, 3981 (2019).
- 537 train accidents reported during last five years - Pakistan Observer. <https://pakobserver.net/537-train-accidents-reported-during-last-five-years/>.
- Srinivasarao, G. *et al.* Deep learning based condition monitoring of road traffic for enhanced transportation routing. *Journal of Transportation Security* **17**, 1–23 (2024).
- Serradilla, O., Zugasti, E., Rodriguez, J. & Zurutuza, U. Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. *Applied Intelligence* **52**, 10934–10964 (2022).
- Jamwal, A., Agrawal, R. & Sharma, M. Deep learning for manufacturing sustainability: Models, applications in Industry 4.0 and implications. *International Journal of Information Management Data Insights* **2**, (2022).
- Tachtatzis, C. *et al.* Condition Monitoring and Predictive Maintenance of Assets in Manufacturing Using LSTM-Autoencoders and Transformer Encoders. *Sensors* **24**, 3215 (2024).
- Skarlatos, D., Karakasis, K. & Trochidis, A. Railway wheel fault diagnosis using a fuzzy-logic method. *Applied Acoustics* **65**, 951–966 (2004).
- Bai, Y., Yang, J., Wang, J. & Li, Q. Intelligent Diagnosis for Railway Wheel Flat Using Frequency-Domain Gramian Angular Field and Transfer Learning Network. *IEEE Access* **8**, 105118–105126 (2020).
- Bernal, E., Spiriyagin, M. & Cole, C. Ultra-Low Power Sensor Node for On-Board Railway Wagon Monitoring. *IEEE Sens J* **20**, 15185–15192 (2020).
- Li, H., Wang, H., Xie, Z. & He, M. Fault diagnosis of railway freight car wheelset based on deep belief network and cuckoo search algorithm. **236**, 501–510, <https://doi.org/10.1177/09544097211029155> (2021).
- Fu, W. *et al.* Recent Advances in Wayside Railway Wheel Flat Detection Techniques: A Review. *Sensors* **23**, 3916 (2023).
- Araín, A. *et al.* Railway track surface faults dataset. *Data Brief* **52**, 110050 (2024).
- Shaikh, M. Z. *et al.* State-of-the-Art Wayside Condition Monitoring Systems for Railway Wheels: A Comprehensive Review. *IEEE Access* **11**, 13257–13279 (2023).
- Shaikh, M. Z. *et al.* Predictive Maintenance in Urban Railway Systems Using Machine Learning Models. *2024 Global Conference on Wireless and Optical Technologies (GCWOT)* 1–5, <https://doi.org/10.1109/GCWOT63882.2024.10805699> (2024).
- Shaikh, M. Z. *et al.* AI-Powered Real-Time Detection of Wheel Defects in Railways Using Yolov8. *2024 Global Conference on Wireless and Optical Technologies (GCWOT)* 1–7, <https://doi.org/10.1109/GCWOT63882.2024.10805632> (2024).
- Zhang, Z., Shao, S. & Gao, Z. A novel method on wheelsets geometric parameters on line based on image processing. *2010 International Conference on Measuring Technology and Mechatronics Automation, ICMTMA* **1**, 257–260 (2010).
- Karaköse, M., Yaman, O. & Akin, E. Real time implementation for fault diagnosis and condition monitoring approach using image processing in railway switches. *International Journal of Applied Mathematics, Electronics and Computers* **307**–307, <https://doi.org/10.18100/IJAMEC.270627> (2016).
- Tastimur, C., Yaman, O., Karakose, M. & Akin, E. A real time interface for vision inspection of rail components and surface in railways. *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)* <https://doi.org/10.1109/IDAP.2017.8090267> (2017).
- Bernal, E., Spiriyagin, M. & Cole, C. Wheel flat detectability for Y25 railway freight wagon using vehicle component acceleration signals. *Vehicle System Dynamics* **58**, 1893–1913 (2020).
- Karakose, E., Gencoglu, M. T., Karakose, M., Aydin, I. & Akin, E. A new experimental approach using image processing-based tracking for an efficient fault diagnosis in pantograph-catenary systems. *IEEE Trans Industr Inform* **13**, 635–643 (2017).
- Santur, Y., Karakose, M. & Akin, E. An adaptive fault diagnosis approach using pipeline implementation for railway inspection. *Turkish Journal of Electrical Engineering and Computer Sciences* **26**, 987–998 (2018).
- Choudhary, A. K. & Ahmad Khan, D. Introduction to Conditioning Monitoring of Mechanical Systems. *Advances in Intelligent Systems and Computing* **1096**, 205–230 (2020).
- Sikora, P. *et al.* Artificial Intelligence-Based Surveillance System for Railway Crossing Traffic. *IEEE Sens J* **21**, 15515–15526 (2021).
- Li, Y., Zuo, M. J., Lin, J. & Liu, J. Fault detection method for railway wheel flat using an adaptive multiscale morphological filter. *Mech Syst Signal Process* **84**, 642–658 (2017).
- Torabi, M., Mohammad Mousavi, S. & Younesian, D. A High Accuracy Imaging and Measurement System for Wheel Diameter Inspection of Railroad Vehicles. *IEEE Transactions on Industrial Electronics* **65**, 8239–8249 (2018).
- Huang, Y., Lin, J., Liu, Z. & Huang, C. A Morphological Filtering Method Based on Particle Swarm Optimization for Railway Vehicle Bearing Fault Diagnosis. *Shock and Vibration* **2019**, 2593973 (2019).
- Kim, G., Kim, H. & Koo, J. A Study on Cepstrum Analysis for Wheel Flat Detection in Railway Vehicles. *Journal of the Korean Society of Safety* **31**, 28–33 (2016).
- Shaikh, M. Z. *et al.* Design and Development of a Wayside AI-Assisted Vision System for Online Train Wheel Inspection. *Engineering Reports* **e13027**, <https://doi.org/10.1002/ENG2.13027> (2024).
- Shaikh, M. Z., Ahmed, Z., Baro, E. N., Hussain, S. & Milanova, M. Deep learning based identification and tracking of railway bogie parts. *Alexandria Engineering Journal* **107**, 533–546 (2024).
- 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 Appl Sci* **3**, 1–10 (2021).
- Shaikh, M. Z. *et al.* FaultSeg: A Dataset for Train Wheel Defect Detection. <https://doi.org/10.5281/ZENODO.13162335>.
- Shaikh, M. Z., Jatoi, S., Baro, E. N., Das, B., Hussain, S. & Chowdhry, B. S. FaultSeg: A Dataset for Train Wheel Defect Detection. *figshare*. <https://doi.org/10.6084/m9.figshare.27996866> (2025).
- Krummenacher, G., Ong, C. S., Koller, S., Kobayashi, S. & Buhmann, J. M. Wheel Defect Detection with Machine Learning. *IEEE Transactions on Intelligent Transportation Systems* **19**, 1176–1187 (2018).
- Kim, E., Jayaprakasham, N., Cui, Y. & Martin, U. Defect Prediction of Railway Wheel Flats based on Hilbert Transform and Wavelet Packet Decomposition. (2020).

38. Trilla, A., Bob-Manuel, J., Lamoureux, B. & Vilasis-Cardona, X. Integrated Multiple-Defect Detection and Evaluation of Rail Wheel Tread Images using Convolutional Neural Networks. *Int J Progn Health Manag* **12** (2021).
39. Lee, J.-H. *et al.* A Study on Wheel Member Condition Recognition Using 1D-CNN. *Sensors* **23**, 9501 (2023).
40. Mosleh, A., Meixedo, A., Ribeiro, D., Montenegro, P. & Calçada, R. Early wheel flat detection: an automatic data-driven wavelet-based approach for railways. *Vehicle System Dynamics* **61**, 1644–1673 (2023).
41. Wang, H. *et al.* Wheel Defect Detection Using a Hybrid Deep Learning Approach. *Sensors* **23**, 6248 (2023).
42. Alemi, A. Railway wheel defect identification. <https://doi.org/10.4233/981EDD2C-1674-4CBA-8146-CF097B29C4F1> (2019).

## 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. 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 Capacity building and ExchaNgE towards attaining Technological Research and modernizing Academic Learning- CENTRAL Project, ID: 598914, Programme: EPLUS - Erasmus+ Capacity Building in Higher Education, Grant Agreement Number: EAC/A05/2017- Project, Reference: 598914-EPP-1-2018-1-DK-EPPKA2-CBHE-J. We would also acknowledge Institute of Oceanic Engineering Research of the University of Málaga for their support.

## Author contributions

Muhammad Zakir Shaikh 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. Sahil Jatoi contributed to writing the original draft, methodology, visualization, validation, investigation and software. Enrique Nava Baro was involved in writing the original draft, conceptualization, funding acquisition, methodology, supervision, validation, reviewing and project administration. Bhagwan Das was involved in collaboration, validation, project administration and reviewing. Samreen Hussain was involved in the management of project, investigation, collaboration, reviewing and administration of the project. Bhawani Shankar Chowdhry handled project administration, funding, formal analysis, and supervision. All authors read and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04557-0>.

**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](http://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