Densenet169-based Deep Transfer Learning Technique for Rail Surface Defect Detection

Noor Fatima Memon[*], Tanweer Hussain[**], Wajiha Shah[***], Ali Akbar Shah[**]

[*]Department of Mechatronics Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan
[**]National Center of Robotics and Automation.-Condition Monitoring Systems Lab, Mehran University of Engineering and Technology, Jamshoro, Pakistan
[***]Department of Electronic Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan

Abstract- Railway transportation plays an essential role in modern transportation infrastructure that requires the utmost safety and reliability. One of the significant concerns in railway management is the precise detection and timely repair of track surface defects, as these anomalies can lead to derailments, incidents, and disruptions in railway operations. Traditional inspection methods often rely on human expertise, which is susceptible to errors and time-consuming. In recent years, a growing interest has emerged in employing deep learning (DL) methods to automate the recognition and classification of faults in railway tracks. This paper aims to enhance the existing knowledge of this field by utilizing a convolutional neural network (CNN) based transfer learning model, DenseNet169, which can classify railway track images into defective and non-defective classes. In defective class, this research focuses on four defects, particularly: corrugation, flaking, shelling, and squats. In addition, data preprocessing and data augmentation techniques are also applied to overcome the challenge of class imbalance. The DenseNet169 results yield in high levels of recall, precision, F1-score, reaching cumulative accuracy of 97%. The developed system can serve as a complementary confirmation mechanism to reduce the likelihood of errors and improve the accuracy of surface fault detection in railway tracks.

Index Terms- Automated Machine Learning, Deep Neural Networks, Deep transfer learning, Densenet169, Rail Surface Defects.

I. Introduction

In many developing countries, including Pakistan, the railway network is a crucial mode of transportation as it contributes to their economic development by facilitating trade and commerce across various regions. However, railway companies worldwide face various challenges in maintaining high productivity levels in today's competitive and rapidly expanding global market [1]. To achieve optimal productivity, companies require maximum availability of assets, and maintenance remains a complex issue in ensuring such availability. Each year, billions of euros are spent on maintenance to meet objectives. Despite this, the costs continue to rise due to the intricate nature of inspections required to detect multiple defects on rail systems, making it challenging. Typical faults in railway tracks are surface faults that occur due to the repetitive passage of railway wagons over components or the impact of broken wheels, causing fatigue [2]–[4].

Surface defects' identification is necessary to maintain the safety and efficiency of railway tracks. However, this task is time-consuming and expensive because it needs regular inspections by qualified professionals operating specialized tools. Moreover, the traditional approach of visual assessments may not be adequate to detect all sorts of surface irregularities, as they may be imperceptible to humans [2].

Railway companies are analyzing the usage of advanced technologies such as Machine Learning (ML) to automate the inspection and maintenance procedures in order to overcome the challenges. These technologies can assist in detecting surface abnormalities with greater accuracy and efficiency [5], thus lessening maintenance time and cost. Moreover, ML can also facilitate predictive maintenance strategies that determine possible defects before they cause severe damage to rail infrastructure. Therefore, to acquire a more precise recognition of surface faults of the railroads, this paper proposes a deep transfer learning model, DenseNet169, to categorize railway track images effectively into five classes: corrugation, flaking, shelling, squats and non-defective, as illustrated in Fig. 1.

Fig 1. Types of classes used in this research

In this research, the dataset used for classification underwent data augmentation techniques such as scaling or flipping, to improve the accuracy of the results. Moreover, to address the issue of class imbalance, pre-trained weights have been assigned to these classes, which has affected the duration of the training process. To expedite the training process and achieve faster convergence, DenseNet169 model has been employed in this study, along with some fine and hyper-parameter tuning. This iterative process helped optimize the model's performance and find the optimal set of hyperparameters for achieving the best results.

The article is structured into five sections. Section 2 and Section 3 describe the related work on detection methods of railway track faults and provide details of the methodology used, including used dataset, fine-tuning methods, hyper-parameters, evaluation metrics, and transfer learning model. Section 4 presents the proposed approach's results and validation, while Section 5 includes the conclusion and future work.

II. LITERATURE REVIEW

In the last few years, numerous conventional non-destructive approaches, such as an eddy current and ultrasonic methods, were applied to discover surface damages. But these existing procedures were an exceptional challenge, which involved reliability and system budget. Therefore, automatic machine learning-based techniques were examined on comparatively uncomplicated international data from a predictive modeling contest to explore track geometry fault [6]. Thus, these techniques were implemented to predict the level of the track imperfections by operating the inertial navigation system and laser displacement sensor to get dynamic orbital acceleration and displacement data, respectively, but later, the imbalance wheel and the transmission system influenced that data and generated some abnormal values [7]. In addition, other kinds of irregularities related to lateral, longitudinal level, alignment, and cross-level were detected through roll bogie frame accelerations or car-body vibrations by employing multiple classifiers based on decision tree (DT), linear support vector machine (SVM), and Gaussian SVM algorithms to determine the location of the faults [6], [7]. Though, the researchers did not focus on acquiring the optimized parameters from the used algorithm, and due to the small number of training data clusters, the effectiveness of the utilized approach lessened [8]. On top of that, the work merely concentrated on straight track sections and observed irregularities within the 3–25m wavelength range [9]. Furthermore, vertical track abnormality is another complex and crucial evaluation of the railway's health. Consequently, multi-level evidential reasoning (M-ER) rule model was suggested, which consisted of the optimized referential evidence matrix and fusion specifications to develop a two-level evidence fusion mechanism to acquire samplings per vibration signal fused with their closest neighboring recorded samples received by KNN technique and integrated by the ER rule [10]. For surface faults, a deep multimodal rail inspection system (DM-RIS) was established, which had a spatially constrained Gaussian mixture model based on Markov random field (MRF) for segmentation and Faster-RCNN for objective location in a parallel structure [11]. Similarly, a fusion of two DL models was constructed, which included a contrast adjustment to the actual rail image, and then finding the location to reduce the overall features in order to provide those to Support Vector Machines (SVM) algorithm and consume less time to detect damages [12]. Fasteners are essential components to keep rails in an affixed position, and their state requires steady checking to guarantee safe transportation. Due to this reason, several image processing technologies and DL networks have been presented to detect fastener position and recognition simultaneously [13]–[15]. Hence, traditional image processing and latent Dirichlet allocation methods have been utilized to identify the location [11], [12]. For recognition, Dense-SIFT features, Faster-RCNN, and Support Vector Data Description (SVDD) have been employed, whereas, for fastener form, conditional random fields and Bayesian hierarchical model have been used [13]–[15].

In Pakistan, the current monitoring system for railway infrastructure relies heavily on manual visual inspection, which has already been deemed inaccurate and time-consuming [16], [17]. However, the adoption of advanced technologies like DL can help overcome these challenges by facilitating more precise and efficient detection and maintenance of surface defects. This, in turn, can lead to a reduction in the time and cost required for maintenance and enable predictive maintenance to prevent considerable damage to the rail infrastructure.

III. METHODOLOGY

In this research, a transfer learning approach is adopted using a DenseNet169 model based on CNN as the base model. The presented methodology involves specifying a new upper layer as a fine-tuning model, and the transfer learning pipeline follows a standard sequence of feature extraction from the source dataset, followed by optimizing the model for the target dataset. The following subsections will deliver further details.

A. Data Collection and Data Pre-processing

The dataset employed in this study was collected from Kotri railway junction station and is split into training and test sets with an 80:20 ratio. The distribution of the dataset in each category is provided in Table 1.

Table 1 Original Dataset of the Railway Track.

S. No.	Classes	Training Set	Test Set
1.	Corrugation	320	80
2.	Flaking	320	80
3.	Non-Defective	320	80
4.	Shelling	320	80
5.	Squats	320	80
	Total	**1600**	**400**

B. Data Augmentation

A pre-processing approach that applies various transformations to existing images in a dataset, aimed to expand dataset size and diversity, combat overfitting, and enhance the DL models' performance that requires significant training data, is called data augmentation [18]. This process aims to generate additional training examples that are similar but not identical to the original images in the dataset to help the model understand more robust and generalizable segments that apply to a wide range of input images. Therefore, data augmentation techniques have been utilized in this research using a built-in function, Image Data Generator, to improve the learnability of the dataset. The used techniques are shown in Table 2.

Table 2 Utilized Data Augmentation Techniques.

Data Augmentation Techniques	Values
Rescale	1./255
Shear range	0.2
Zoom range	0.2
Horizontal flip	True or 0.5

C. Class Imbalance

Class imbalance can occur due to various reasons such as sampling bias, data collection bias, or distribution of data [19].

Although the class weights are not used in this analysis explicitly, the categorical cross-entropy loss function is still applied during model training. This function is commonly used for multi-class classification problems and implicitly accounts for class distribution in the dataset. This function penalizes the model for making incorrect predictions and aids to balance the contribution of each class to the overall loss.

D. Transfer Learning

Transfer learning has been utilized in different fields to enhance the precision of classification, tackle overfitting issues, and boost the model's ability to generalize [20]. In this study, the application of transfer learning involves taking a pre-trained model that has learned a large set of features from a source dataset, and then adapting it to a target dataset because it can serve as a launching pad for building a new model customized to perform a distinct task from the original model.

E. DenseNet169

DenseNet169 was chosen for its relatively low parameter count despite having a depth of 169 layers, and its ability to effectively handle the vanishing gradient problem, particularly when working with limited datasets. To tackle these problems, the model employs a densely connected architecture that facilitates the reuse of features from preceding layers, allowing for more direct information flow between layers and promoting faster convergence [21]. In addition, it uses bottleneck and MBConv blocks to reduce parameters without sacrificing capacity. The bottleneck blocks compress input feature maps, while the MBConv blocks leverage depth-wise separable convolutions and squeeze-and-excitation blocks to decrease computational costs [22]. The model's combination of feature reuse and gradient flow promotes faster convergence and higher accuracy, surpassing traditional CNN architectures and resulting in an exceptional performance on various image classification benchmarks, and validating its effectiveness in solving complex real-world problems. The network architecture presented in Fig. 2 was used in this research, and the pre-trained model was modified by adding a Flatten layer and a Dropout layer at the top.

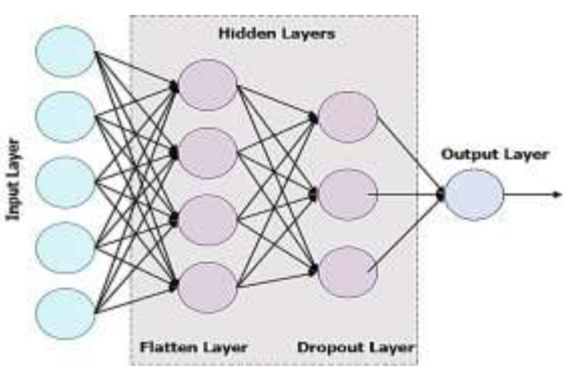

Fig. 2. Illustration of an Implemented Neural Network with Two Hidden Layers Fine Tuning

f. Fine Tuning

Fine-tuning a model involves freezing the upper layers and extracting the fundamental parameters [21]. In CNNs, the higher layers typically learn low-level features and can be applied to several types of images. As the railway track dataset used to train the model becomes diverse, it acquires distinct characteristics, which is the objective of fine-tuning to adapt unique attributes that better suit the track dataset. The parameters' details of the used model are given in Table 3.

Table 3 Number of parameters in proposed DenseNet169 model

Total params	Trainable params	Non-trainable params
13,175,365	13,016,965	158,400

g. Hyperparameter Tuning

Hyperparameters refer to the settings that are specified prior to the commencement of the model training process and cannot be determined from the data itself, unlike model parameters which are learned through the training process [23]. Although pre-trained models have already undergone training with specific hyperparameters, it is important to recognize that these hyperparameters are fundamental to the model's performance. During fine-tuning, adjusting hyperparameters may be necessary to achieve optimal results [24]. The applied model is pre-trained and does not explicitly specify hyperparameters. However, the detail of the utilized hyperparameters is provided in Table 4.

Table 4 Utilized Hyperparameters

Hyperparameters	Values
Learning/Training rate	0.001
Training batch size	32
Training iterations	35
Patience	10

By tuning the above hyperparameters, we can optimize the model's performance by reducing overfitting, and training time.

h. Evaluation or Performance Metrics

The assessment of the model's performance during training and testing is crucially dependent on evaluation metrics [25] as it provides a quantitative measure of how competently the model can solve a given task. Furthermore, we can impartially assess the performance of the suggested model by using these metrics to compare it with other models or benchmarks [26]–[28] so that later decisions can be made about its suitability for a specific task and pinpoint improvement areas. The present study employs a few evaluation metrics, such as precision, accuracy, recall and f1-score, to gauge the proposed model's performance.

IV. RESULTS & VALIDATION

Experiments were conducted on the DenseNet169 model, which was selected due to its small size and robust performance, to observe the implementation of the proposed strategy. To achieve the desired results, key parameters were modified including stack size and learning rate. Specifically, the stack size was increased to enhance the model's capacity to capture complex patterns and the learning rate was fine-tuned periodically to improve the model's responsiveness to changes in the data. The optimizer "Adam," having a default value of a learning rate of 0.001, was used to optimize the model's performance and constrain the binary cross-entropy loss function. The top layer of the model was modified, as this has been shown to have a significant impact on the model's performance.

The evaluation of the proposed strategy employed rigorous evaluation metrics to measure the model's performance, including accuracy, precision, recall, and F1-score. Through experimentation, it was found that the modified DenseNet169 achieved significantly better results than the baseline model. Fig. 3 shows the implemented model's confusion matrix, while Table 4 illustrates the correlation between the model's iterations and the bias-variance tradeoff. The training vs. validation accuracy and loss curves are depicted in Figs. 4 and 5, respectively.

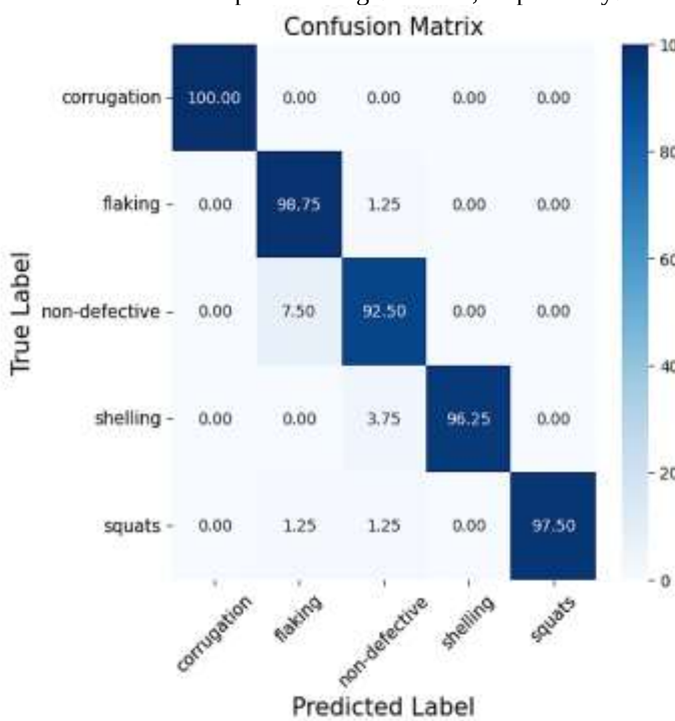

Fig. 3. Confusion Matrix of the DenseNet169 Model

Table 4. Epochs vs. Bias-Variance Tradeoff

Epochs	Bias-Variance Tradeoff
29	Under-fitting
50	Optimal or Appropriate-fitting
51	Over-fitting

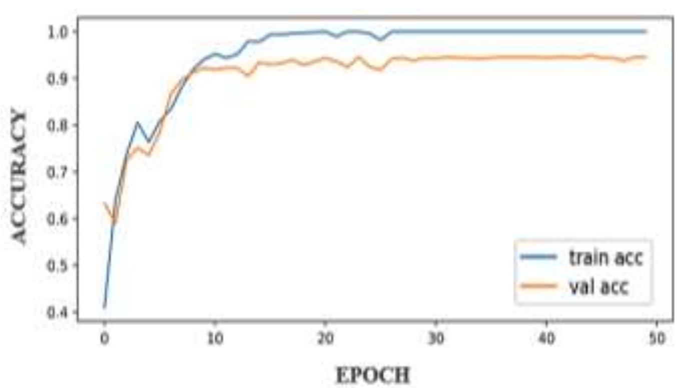

Fig. 4. Accuracy Graph of DenseNet 169

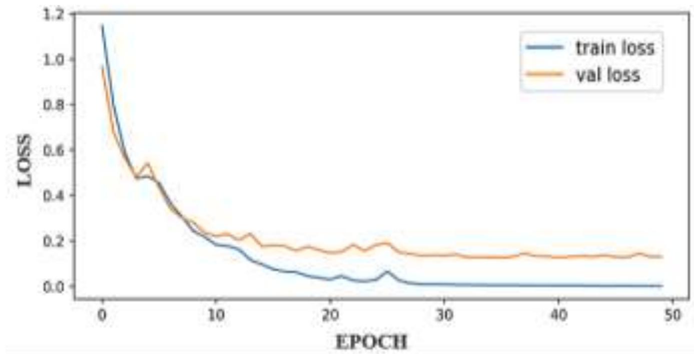

Fig. 5. Loss Graph of DenseNet169

From the above results, it has been observed that despite some fluctuations during training, the DenseNet-169 model achieved accurate classification results even with an imbalanced dataset, highlighting DL's potential to improve fault detection. These findings reinforce the idea that DL can complement conventional inspection methods to reduce errors in fault detection on railway tracks. The comparison results of precision, recall, and f1-score for each class, including non-defective and four types of surface faults, are showcased in Fig. 6.

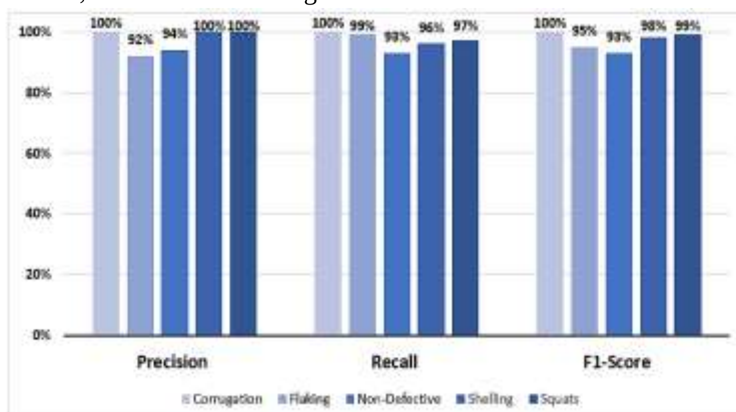

Fig. 6. Comparative results of Precision, Recall and F1-Score per Class

The model validation using a random track image showcased reliable and accurate defect detection and categorization, as depicted in Fig. 7, affirming its value in enhancing railway safety and maintenance.

Fig. 7. Detected Faults in Railway Track

This model was also validated in a real-world scenario using a motorized track recording vehicle (TRV) device, as depicted in Fig. 8, developed by the National Center of Robotics and Automation (NCRA) at Mehran University of Engineering and

Technology, Jamshoro. These findings emphasize that DL techniques have immense potential to improve fault detection accuracy on railway track surfaces.

Fig. 8. DenseNet169 Validation through Motorized Track Recording Vehicle

V. CONCLUSIONS & FUTURE RECOMMENDATIONS

In this research, we developed a system that utilizes a transfer learning model, DenseNet169, based on a CNN, to classify railway track images into defective and non-defective classes. This system further classifies defective images into four specific faults: corrugation, flaking, shelling, and squats. To overcome the class imbalance challenge, we implemented data preprocessing and data augmentation techniques, and during the training process, we optimized the model's performance through fine-tuning and hyperparameter adjustments to achieve faster convergence.

The model was validated using images acquired from the Kotri railway junction station, Sindh, Pakistan, and conducted both simulated testing using input images and field testing using a motorized TRV device. The results demonstrated that the system was effective in specifying and categorizing surface defects on railway tracks, highlighting its potential to lead to a higher standard of maintenance and safer railway operations.

Although the developed system provides a promising solution by achieving a remarkable classification accuracy of 97% even with imbalanced data, some limitations and challenges remain. The system's performance is highly dependent on the quality of input images, and the transfer learning model's effectiveness may vary based on the specific railway infrastructure. Therefore, future work could focus on improving the system's robustness to various environmental conditions and railway systems, expanding the dataset, detecting other faults, and incorporating other transfer learning models.

ACKNOWLEDGMENT

This work received support from the 'Haptics, Human Robotics, and Condition Monitoring Lab' at Mehran University of Engineering and Technology, Jamshoro, under the National Center of Robotics and Automation.

REFERENCES

[1] F. Guo, Y. Qian, D. Rizos, Z. Suo, and X. Chen, "Automatic rail surface defects inspection based on mask r-cnn," *Transp. Res. Rec.*, vol. 2675, no. 11, pp. 655–668, 2021, doi: 10.1177/03611981211019034.

[2] A. Arain, T. Hussain, and S. M. Ujjan, "Rail Surface Faults Identification from Low Quality Image Data Using Machine Learning Algorithms," no. March, pp. 10–21, 2021, doi: 10.21058/gjecs.2021.61002.

[3] S. Faghih-Roohi, S. Hajizadeh, A. Nunez, R. Babuska, and B. De Schutter, "Deep convolutional neural networks for detection of rail surface defects," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016-Octob, pp. 2584–2589, 2016, doi: 10.1109/IJCNN.2016.7727522.

[4] Y. Wu, L. Jia, and F. Guo, "Hybrid deep learning architecture for rail surface segmentation and surface defect detection," pp. 227–244, 2022, doi: 10.1111/mice.12710.

[5] L. Wen, Y. Wang, and X. Li, "A New Cycle-consistent Adversarial Networks with Attention Mechanism for Surface Defect Classification with Small Samples," vol. 3203, no. c, 2022, doi: 10.1109/TII.2022.3168432.

[6] S. Kocbek and B. Gabrys, "Automated machine learning techniques in prognostics of railway track defects," *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 2019-Novem, pp. 777–784, 2019, doi: 10.1109/ICDMW.2019.00115.

[7] B. Ting, Y. L. Chen, T. Kai, X. J. Huang, Z. F. Cao, and M. L. Yang, "Research on Track Irregularity Based on K Nearest Neighbor Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 688, no. 4, 2019, doi: 10.1088/1757-899X/688/4/044002.

[8] H. Tsunashima, "Condition monitoring of railway tracks from car-body vibration using a machine learning technique," *Appl. Sci.*, vol. 9, no. 13, 2019, doi: 10.3390/APP9132734.

[9] A. De Rosa *et al.*, "Monitoring of lateral and cross level track geometry irregularities through onboard vehicle dynamics measurements using machine learning classification algorithms," *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 235, no. 1, pp. 107–120, 2021, doi: 10.1177/0954409720906649.

[10] Z. Zhang *et al.*, "Intelligent identification for vertical track irregularity based on multi-level evidential reasoning rule model," *Appl. Intell.*, no. 0123456789, 2022, doi: 10.1007/s10489-021-03114-7.

[11] X. Jin *et al.*, "DM-RIS: Deep multimodel rail inspection system with improved MRF-GMM and CNN," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1051–1065, 2020, doi: 10.1109/TIM.2019.2909940.

[12] I. Aydin, E. Akin, and M. Karakose, "Defect classification based on deep features for railway tracks in sustainable transportation," *Appl. Soft Comput.*, vol. 111, p. 107706, 2021, doi: 10.1016/j.asoc.2021.107706.

[13] X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, and Y. Li, "Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study," *Eng. Appl. Artif. Intell.*, vol. 80, no. September 2018, pp. 66–81, 2019, doi: 10.1016/j.engappai.2019.01.008.

[14] Y. Ou, J. Luo, B. Li, and B. He, "A classification model of railway fasteners based on computer vision," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 9307–9319, 2019, doi: 10.1007/s00521-019-04337-z.

[15] T. Bai, J. Yang, G. Xu, and D. Yao, "An optimized railway fastener detection method based on modified Faster R-CNN," *Meas. J. Int. Meas. Confed.*, vol. 182, no. June, p. 109742, 2021, doi: 10.1016/j.measurement.2021.109742.

[16] I. Alisjahbana, "Using In-Service Train Vibration for Detecting Railway Maintenance Needs," no. 2, 2017.

[17] R. V. Vibrations, "The Rail Surface Defects Recognition via Operating Service Rail Vehicle Vibrations," 2022.

[18] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[19] L. Jiang, S. Jiang, L. Gong, Y. Dong, and Q. Yu, "Which Process Metrics Are Significantly Important to Change of Defects in Evolving

Projects: An Empirical Study," *IEEE Access*, vol. 8, pp. 93705–93722, 2020, doi: 10.1109/ACCESS.2020.2994528.

[20] X. Li, H. Jiang, K. Zhao, and R. Wang, "A deep transfer nonnegativity-constraint sparse autoencoder for rolling bearing fault diagnosis with few labeled data," *IEEE Access*, vol. 7, pp. 91216–91224, 2019, doi: 10.1109/ACCESS.2019.2926234.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[22] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 10425–10433, 2020, doi: 10.1109/CVPR42600.2020.01044.

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning An MIT Press Book*, vol. 29, no. 7553. 2016.

[24] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 2242–2251, 2017, doi: 10.1109/ICCV.2017.244.

[25] Aurélien Géron, *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems*. 2019.

[26] G. H. de Rosa and J. P. Papa, "A survey on text generation using generative adversarial networks," *Pattern Recognit.*, vol. 119, 2021, doi: 10.1016/j.patcog.2021.108098.

[27] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," pp. 1–17, 2020, [Online]. Available: http://arxiv.org/abs/2008.05756.

[28] H. Wu, X. Li, A. T. Liu, Z. Wu, H. Meng, and H. Y. Lee, "Improving the Adversarial Robustness for Speaker Verification by Self-Supervised Learning," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 30, pp. 202–217, 2022, doi: 10.1109/TASLP.2021.3133189.

AUTHORS

First Author–Noor Fatima Memon, BEng., Mehran University of Engineering & Technology, Jamshoro, Pakistan,
Second Author–Dr. Tanweer Hussain, Ph.D., Mehran University of Engineering & Technology, Jamshoro, Pakistan,
Third Author–Dr. Wajiha Shah, Ph.D., Mehran University of Engineering & Technology, Jamshoro, Pakistan,
Fourth Author–Engr. Ali Akbar Shah, MEng., Mehran University of Engineering & Technology, Jamshoro, Pakistan.

Correspondence Author–Noor Fatima Memon