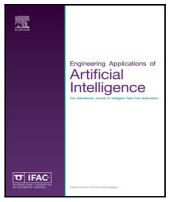
Engineering Applications of Artificial Intelligence 121 (2023) 105965

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



MSRConvNet: Classification of railway track defects using multi-scale residual convolutional neural network

Hakan Acikgoz [a,*], Deniz Korkmaz [b]

[a] Gaziantep Islam Science and Technology University, Faculty of Engineering and Natural Sciences, Department of Electrical and Electronics Engineering, Gaziantep, Turkey

[b] Malatya Turgut Ozal University, Faculty of Engineering and Natural Sciences, Department of Electrical and Electronics Engineering, Malatya, Turkey

ARTICLE INFO

Keywords:
Railway track defects
Rail defect classification
Multi-scale convolutional network
Residual connection

ABSTRACT

The development of an automated rail line defect classification system is of great benefit, as railway tracks must be periodically monitored and inspected to guarantee the safety of rail transportation. In this paper, an effective multi-scale residual convolutional network (MSRConvNet) model is proposed to classify the different types of railway track defects. The skip connections with residual learning blocks are used to increase the effectiveness of the network. The multi-scale convolutions are connected with parallel and two skip connections in the structure to distribute detailed feature maps with each other. Therefore, different scale feature maps can be extracted. The data augmentation method is performed to ensure a balanced class distribution and to eliminate the negative effect of the imbalanced dataset. The proposed model is compared with both benchmark deep learning models and the different variations of the designed network. The results verify that the proposed model can reach superior classification fulfillment, and the MSRConvNet provides an overall accuracy of 99.83%, precision of 99.83%, sensitivity of 99.83%, specificity of 99.94%, F1-score of 99.83%, and Matthew's correlation coefficient of 99.78% for four defect classes.

1. Introduction

Around the world, the railway industry has a major role in the economy and development of a country and directly impresses the transportation quality of living people. The fast growth of the high-speed railways has significantly enlarged the prevalence of railway transportation. However, accidents or problems in railway operations are an issue that needs attention, as they will affect the social and political risks and the economy and reputation of the country (Ghafoor et al., 2022; Ji et al., 2021). The railway systems generally operate in different environments with steel rail tracks with rails, sleepers, ballasts, and fasteners. Therefore, railway tracks are the most critical infrastructure of railway systems to maintain the safety and performance of the railway network (Yang et al., 2022; Ye et al., 2020). Defects and malfunctions in the track can lead to unacceptable results such as derailment, death, injury, economic cost, and loss of trust. In addition, the risk of injury or death to workers during rail maintenance work should be considered (Ji et al., 2021). Therefore, safe railway operations require an effective operating and control structure that relies heavily on periodical inspection, monitoring, and maintenance of rail track conditions (Aydin et al., 2021b,a; Zhang et al., 2018; Zheng et al., 2021).

Due to the long-term effects of the train and the natural environment on the railway systems, many defects occur on the rail, such as deformation of the rail and breakage of the sleeper, which will endanger the safety of the railway (Min et al., 2018; Tu et al., 2021). The lack of rapid inspection systems also delays the repair of defects on the rails after accidents and damage. There are different types of rail defects such as rolling contact fatigue defects, rail corrugations, shatter cracking, squat defects, split head, and wheel burns (Ji et al., 2021; Kishore et al., 2019). These defects can consist of manufacturing faults, delays in maintenance, and rolling contact fatigue. Surface cracks in the rail are complex problems and usually depend on certain criteria such as axle load and loading frequency. Squats are the subsurface laminations that generally occur on the running band of rails (Alvarenga et al., 2021; Kishore et al., 2019; Ye et al., 2020). These rail defects should be rapidly and effectively detected to prevent disasters as a result of rail failures. Therefore, automated rail surface defect detection has become a more and more critical topic in railway transportation.

In railway systems, laser inspection is usually used for two-dimensional profile measurement. Laser sensors are mounted on measurement carts and trains and surface conditions of the rail are evaluated. This method is based on a comparison with previously measured standard profiles or using thresholding methods. However, the laser sensor-based inspection method is quite complex and laborious when environmental conditions are taken into account (Ye et al., 2020). Ultrasonic testing, magnetic particle testing, acoustic

* Corresponding author.
E-mail addresses: hakan.acikgoz@gibtu.edu.tr (H. Acikgoz), deniz.korkmaz@ozal.edu.tr (D. Korkmaz).

https://doi.org/10.1016/j.engappai.2023.105965
Received 14 April 2022; Received in revised form 21 September 2022; Accepted 4 February 2023
Available online 15 February 2023

emission, and eddy current are the other commonly used detection methods. Among these, ultrasonic and eddy current testing methods are frequently preferred non-contact approaches. Ultrasonic testing can detect internal defects and rail cracks using the propagation and attenuation properties of sound waves (Espinosa et al., 2018). The eddy current testing can also detect the rail head cracks using eddy currents (Kishore et al., 2019). Nevertheless, harsh environmental conditions, long operation time, and tiring processes are disadvantages of these non-contact methods (Ji et al., 2021; Ni et al., 2021).

In recent years, the above challenges of conventional detection methods in railway transportation have been started to address with the rapid development of computer vision-based algorithms. Yang et al. (2022) proposed an improved residual network-based feature extraction approach, and constructed a recognition model of track defects. Their model was designed in three stages as; feature extraction, region of interest (ROI) extraction, and defect identification. They achieved an average identification accuracy of 94.45% and the feature identification model accuracy of 97.09%. In another study, a transfer learning strategy using limited training data was presented by Zheng et al. (2021). You only look once-v3 (YOLOv3) and RetinaNet models were used to detect rail track cracks. The dataset used in paper was composed of the collected samples from China Railway Corporation and a publicly available dataset. They indicated that the real cracks detected by the YOLOv3-based model gave confidence scores of less than 0.50. Zhang et al. (2018) also proposed a visual detection structure for railway surface defects. The Gaussian mixture model combined with Markov random field was used to detect accurate and rapid surface defects. Their method performed well with both noisy and railway images, enabling the identification and segmentation of the defects, with an average detection performance of 92% precision and 88.8% sensitivity. Alvarenga et al. (2021) introduced a rail surface defect detection and classification system based on the eddy current approach. In addition, they analyzed the eddy current signals using wavelet transforms and a convolutional neural network (CNN). The results gave a classification efficiency of approximately 98%. Yuan et al. (2019) designed an end-to-end lightweight CNN model for detecting railway surface defects. Their CNN model was developed based on MobileNetV2 and combined with YOLOv3. They achieved better detection accuracy against YOLOv3. Espinosa et al. (2018) designed a classification model for rail breakages in double-track railway lines. The detection of breakages was performed based on the analysis of eight currents. In the classification process, the principal component analysis technique was also used. Jin et al. (2020) designed a system for surface defects. A gaussian mixture model was used for segmentation and a faster recurrent network was utilized for objective location. They achieved well results with 96.74% precision, 94.13% sensitivity, 95.18% overlap, and 0.485 s/frame speed on average. Aydin et al. (2021b,a) developed a fusion model to extract deep CNN features and a support vector machine (SVM) used to classify track defects. In the study, the overall accuracy was obtained as 97.10%. Wu et al. (2020) also designed a classification and detection method. They firstly extracted rail defect features and classified them based on the feature distribution and contour morphological features. Their detection accuracy rate was obtained as 97.3% and the average detection time was observed as 0.2 s/frame. Zhang et al. (2022) proposed a supervised segmentation model for no-service rail surface defects. These defects were defined into three types as strip-shaped, spot-shaped, and block-shaped. A pooling combination module was also applied to use the size attributes of the defect types. However, most studies have still detected rail surface defects with various bounding boxes and segmentation methods. The main difficulties of these methods are that the rail surfaces are quite noisy during the image acquisition phase and the defects do not have a uniform distribution. In addition, the defect detection performances of most studies have not reached high accuracy and there have been limited studies focused on the classification of the rail surface from an image. Therefore, railway track defect classification can still be improved by designing new types of deep learning models to provide robustness and improve accuracy performance.

To deal with the aforementioned problems, an effective multi-scale residual convolutional network, namely MSRConvNet, is designed in this paper. The MSRConvNet classifies the various types of railway track defects using a publicly available image dataset. In the proposed CNN model, three multi-scale blocks are constructed with short residual connections and different sizes of convolutions. The multi-scale convolutions are connected with parallel and two skip connections between the parallel structure to distribute detailed feature maps with each other. Therefore, different scale feature representations with few self-parameters can be extracted. Afterwards, the features are concatenated as a feature vector and fed to the last low-level convolution with reducing the output dimension. This structure does not only give the local features, but also provides the distinctive feature fusion at different scales. Residual connections in each block share feature information with each other that improve the training performance of the network. In addition, five synthetic defect samples are generated with the appropriate augmentation methods to overcome the imbalanced class distribution and rise the generalization ability of the proposed model. In the experiments, the proposed model is divided into three structures for each multi-scale residual convolution block and these variations are separately performed to assess the effectiveness of the proposed model. The augmentation effect on the network, the determination of the most suitable optimizer and activation function are analyzed to improve the designed network performance. Finally, the obtained results are compared with various deep learning methods and a general comparison of the existing state-of-the-art models is given. The scientific contributions and novelty of this paper can be expressed as follows:

- An effective multi-scale residual convolution network is proposed. The proposed model is a simple and applicable network architecture that extracts hierarchical features with different level kernels. This structure gives superior accuracy improvement for the automated railway track defect classification system.
- The residual network-based fusion framework is designed. Incorporating residual skip connections into the network enables the effective transmission, combination, and reuse of features. Therefore, the proposed model's ability to handle complex functions and obtain high-level features is enhanced by avoiding the disappearance and explosion of gradients.
- An augmentation processing is performed to deal with the imbalanced class distribution of the dataset. The obtained results show that the augmented images can well eliminate the issue of insufficient data samples and significantly improve the proposed network performance.
- The experimental results are assessed with variations of the multi-scale blocks and existing state-of-the-art studies. It is confirmed that the proposed model gives superior classification performance and outperformed the existing studies.

The other sections of this paper can be briefly explained as follows: The background theories with the overview of the proposed model, multi-scale residual model, and dataset improvement with the augmentation processing are presented in Section 2. The experiments and various comparison studies are realized in Section 3. The evaluations obtained from the paper are summarized in Section 4.

2. Methodology

Although the expert knowledge-based routines of inspecting rail tracks help to increase the safety of rail transportation, it also causes difficulties that limit control capacity and where human error is inevitable. In this context, the use of CNN-based approaches can bring rapid and robust defect detection with increased representation and generalization abilities (Rawat and Wang, 2017). In general, most of the CNN models are capable of nonlinear feature extraction in image-related problems. Among them, multi-scale convolutional networks are an effective feature extraction method. Multi-scale CNNs can learn

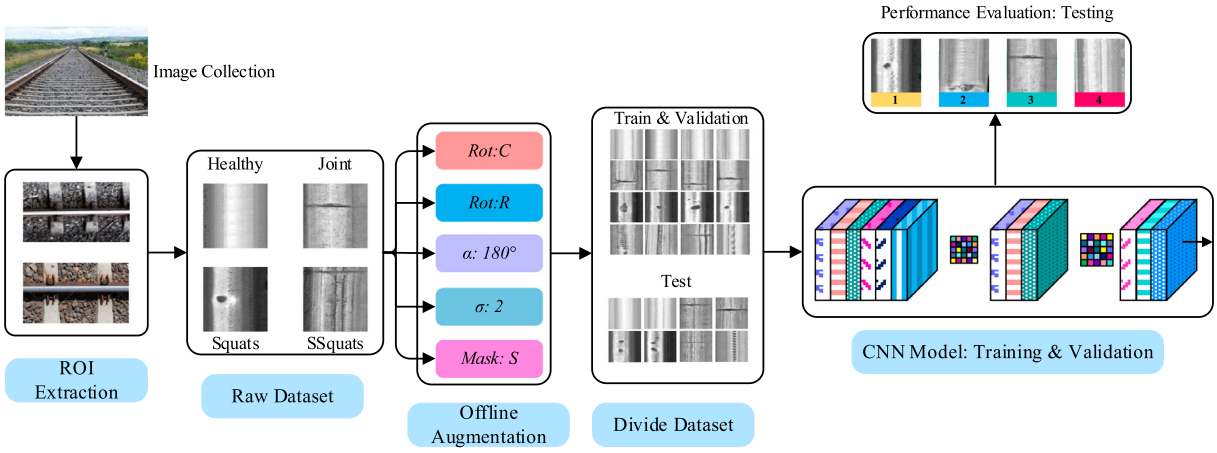


Fig. 1. General structure of the proposed railway surface defect classification method.

different levels of receptive fields and the representational ability of CNN can be improved (Gong et al., 2019). Cao et al. (2021) proposed an end-to-end multi-scale residual network to get an impressive dehazing influence on railway monitoring images. The residual structure was used to fuse multi-scale feature information. Wang et al. (2020) designed a deep network for crack segmentation using fully convolutional networks and multi-scale structured forests. Yu et al. (2022) proposed a multi-scale residual CNN model to extract deep and large-scale features in bearing fault reconstruction diagnosis. Xu et al. (2021) also presented a multi-scale feature learning model based on dual module feature extraction for surface defect classification. These studies have produced effective solutions to different problems by taking advantage of multi-scale feature extraction and residual network. Therefore, a robust CNN-based rail surface defect classification network is proposed in this study. The developed model aims to effectively classify surface defects using residual connections and multi-level convolutions.

In this section, the framework of the proposed classification method is firstly introduced. Later, multi-scale convolutional and residual learning structures are presented in detail. Finally, data augmentation techniques for network improvement are given.

2.1. Overview of the proposed model

The overall framework of the proposed model is given in Fig. 1. The proposed model consists of three parts: preprocessing for the rail position detection, data augmentation for network performance improvement, and MSRConvNet-based classification process. After the collected images of each rail region by a high-speed camera, a filtering algorithm is applied to eliminate the noise problems. Then, the ROI method determines the rail positions from the original images and extracts the rail surfaces. Therefore, a rail defect dataset only including rail surfaces can be obtained. It should be noted that these filtering and ROI extraction algorithms are not performed in this paper since Aydin et al. (2021b,a) arranged the dataset using the above-preprocessing algorithms. Although the used dataset contains 1838 images in total, the class distributions have fewer samples. Hence, the synthetic images are obtained to prevent the imbalanced class distribution and rise the network accuracy. The used augmentation techniques are reversing, rotation columns (C) and rows (R), filtering, and sharpen (S) masking operations, respectively. Also, the dataset is randomly split into training, validation, and test subsets. Then, the designed CNN model is trained with training and validation data. Finally, the trained model is employed to analyze the network performance with testing data.

2.2. Multi-scale residual convolutional network

In many applications, the network depth of plays a critical role in the classification performance and some applications require complex network structures with multiple layers. It is desirable to develop simple and highly applicable CNN structures in design. In addition, gradients vanishing and degradation problems appear when the depth of the architecture is reduced (Shen et al., 2018). He et al. (2016) designed a residual learning approach to avoid the degradation problems and increase training performance. As shown in Fig. 2, the residual network is an effective CNN structure that has residual connections which are utilized to carry extracted features from initial layers to deeper layers. These skip connections operate $H(x) = f(x,\theta) + x$ and provide a convenient path for forward and backward information propagation. Since residual mapping is generally easier, multi-scale convolutions could be easily learned and improved accuracy could be achieved. Therefore, a multi-scale residual CNN architecture is used to effectively classify railway defects in this paper.

In the proposed model, three multi-scale blocks are constructed inspired by Li et al. (2018) with the short residual connections and different sizes of convolutions. The designed MSRConv block is given in Fig. 3. Multiscale convolutions extract multiscale spectral features from the input with a few intrinsic parameters and provide a stronger generalization capableness. Residual connections in each block are performed with shortcut connections that ensure to share and reuse of feature information between different scale features with each other (Qin et al., 2020; Zhu et al., 2021).

Therefore, local features with different levels of visual perception can be utilized. Each MSRConv block consists of two stages as multi-scale convolutions and short residual connections. The multi-scale convolutional structure includes both parallel convolutions and two skip connections between the parallel structure. In this way, each convolution distributes detailed feature maps with each other to obtain different scale feature information. 5×5 and 3×3 convolution filters are used in the inner loop while 1×1 filter at the end of the block is used to reduce the computational complexity. Formally, $P_k(\cdot)$ and L_N can be denoted as the output of the ReLU functions and feature maps of the blocks, respectively.

For each output, it can be computed from the previous convolutions as follows:

$$P_1 = \mu\left(\omega_{1,3} * L_{N-1} + b_{1,3}\right) \tag{1}$$

$$P_2 = \mu\left(\omega_{2,5} * L_{N-1} + b_{2,5}\right) \tag{2}$$

$$P_3 = \mu\left(\omega_{3,3} * \left[P_1 \oplus P_2\right] + b_{3,3}\right) \tag{3}$$

$$P_4 = \mu\left(\omega_{4,5} * \left[P_2 \oplus P_1\right] + b_{4,5}\right) \tag{4}$$

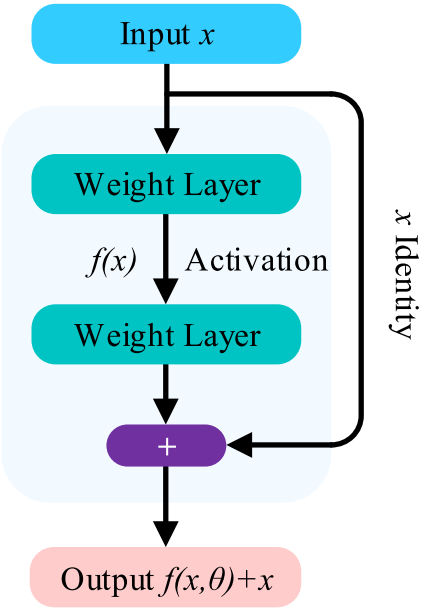


Fig. 2. Basic scheme of the residual block.

In these equations, $\omega_{i,j}$ and $b_{i,j}$ represent the weights and biases, respectively. The subscript (i, j) defines the order of the layer and filter size. $\oplus$ also represents the concatenation operation. The convolutions are enabled with ReLU functions to decompose the noises and increase the learning speed. The ReLU function μ is expressed by:

$$\mu(x) = max(0, x) \tag{5}$$

At the end of the multi-scale block, P_3 and P_4 outputs are concatenated and sent to the $Conv(1, 1)$. Therefore, the output of the convolution is given as:

$$S = \omega_{5,1} * [P_3 \oplus P_4] + b_{5,1} \tag{6}$$

The residual connections reduce the gradient disappearance and increase the propagation of feature maps. As the residual form, mapping of the block output is obtained with element-wise addition and it can be expressed in (7):

$$L_N = S + L_{N-1} \tag{7}$$

In this form, there is a short connection between the layers that allows features to be carried directly from one residual unit to another. L_{N-1} represents the short connection and S represents the function of stacked multi-scale layers. Therefore, L_N gives new feature map and the output of this block obtains distinctive features of the input images and network performance can be improved (McNeely-White et al., 2020; Tekchandani et al., 2020).

Using the aforementioned MSRConv blocks, the general structure of the MSRConvNet model is given in Fig. 4. The initial convolution layer, named as *Block-1* increases the activations of the inputs. *Block-2* also extracts general input features and sends the information to the first MSRConv block. According to the design scheme, the input size of the network is set to 100 × 100 pixels.

At every depth change, one average pooling layer is performed to decrease the dimension of the features with a statistic of nearby outputs, and it operates the $\sum f(x)/|f(x)|$. Before the fully connected layer, a global average pooling is used that is an extreme of average pooling and it can reduce a tensor with the size of $w \times w \times d$ to $1 \times 1 \times d$.

In addition, a dropout layer is utilized to avoid overfitting. Finally, the fully connected layer gives the probability of the defined classes according to the defects and this layer obtains a deep feature vector.

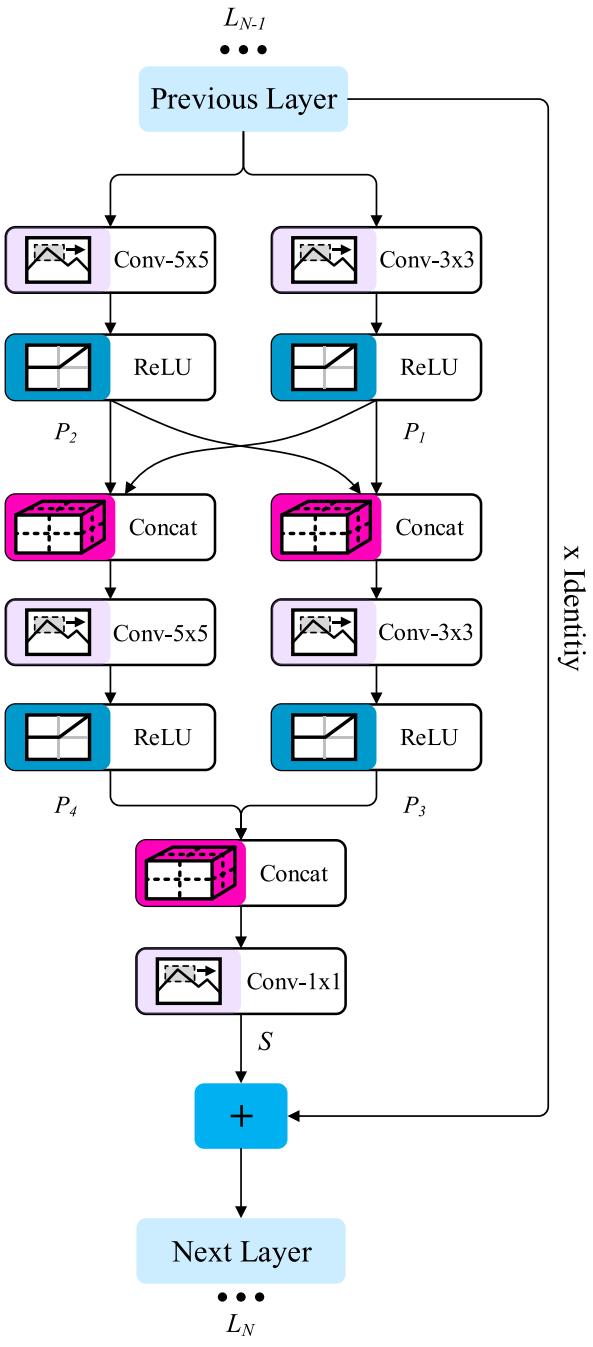


Fig. 3. Structure of the MSRConv block.

As given in Eq. (1), the fully connected layer multiplies the input by a weight matrix w and appends it to a bias vector b:

$$Y_o = w_o * Y_{o,n-1} + b_o \tag{8}$$

Through short residual connections and multi-scale convolutions, the proposed model not only further improves information propagation during the training process, but also better utilizes the multi-scale spectral features. In Table 1, the detailed layer configuration of the proposed CNN architecture is summarized.

2.3. Data augmentation for network improvement

Training convolutional networks with limited data is generally caused to overfitting and the network exhibits lower generalization

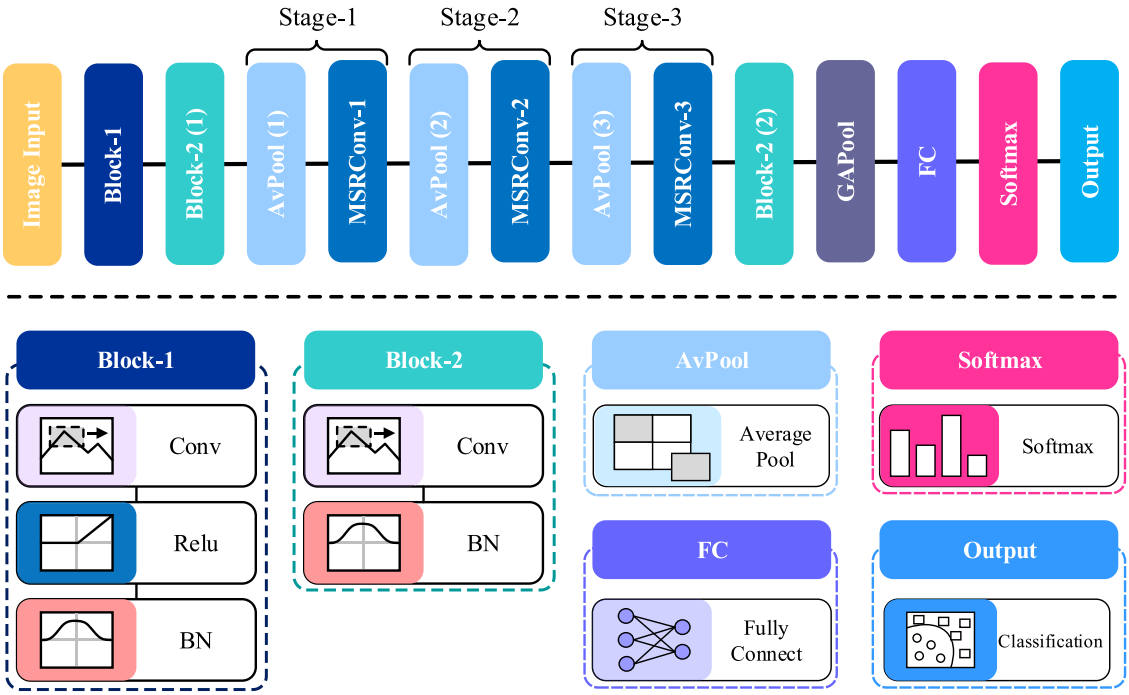


Fig. 4. Details on the complete multi-scale CNN architecture.

Table 1
Detailed layer configuration of the designed CNN architecture.

Layer	Type	#5 × 5	#3 × 3	#1 × 1	Stride	Output
Input	Image input	–	–	–	–	100 × 100 × 3
Block-1	Conv + BN + ReLU	–	32	–	1	100 × 100 × 32
Block-2 (1)	Conv + BN	64	–	–	1	100 × 100 × 64
AvPool (1)	Average Pooling	–	–	–	2	50 × 50 × 64
MSRConv-1	Conv + ReLU Conv	64 –	64 –	– 64	1	50 × 50 × 64
AvPool (2)	Average Pooling	–	–	–	2	25 × 25 × 64
MSRConv-2	Conv + ReLU Conv	64 –	64 –	– 64	1	25 × 25 × 64
AvPool (3)	Average Pooling	–	–	–	2	13 × 13 × 64
MSRConv-3	Conv + ReLU Conv	64 –	64 –	– 64	1	13 × 13 × 64
Block-2 (2)	Conv + BN	64	–	–	1	13 × 13 × 64
GAPool	Global Average Pooling	–	–	–	–	1 × 1 × 64
fc	Fully connected	–	–	–	–	1 × 1 × 4
Softmax	Softmax	–	–	–	–	1 × 1 × 4

capability and poor classification performance. On the contrary, training with a large amount of data can improve network performance. Increasing the quantity, quality, and diversity of the data in the data set can also increase the effectiveness of the developed model. Data augmentation is the most commonly used method that generates synthetic data to satisfy the required training data. Synthetic data generation with data augmentation reduces the overfitting of CNN caused by limited samples and it gives better classification performance in image-related problems (Rafiq et al., 2021; Shorten and Khoshgoftaar, 2019). Therefore, data augmentation minimizes the distance between class distributions and ensures an equal number of samples for whole classes.

In the railway dataset, there are 1838 images representing four defect classes. This dataset is not a large data set and it has fewer samples for the training process. In addition, the healthy class covers approximately one-third of the total dataset with 608 samples and the severe squat samples are less than 400 images. Therefore, the data augmentation method is preferred to overcome the imbalanced class distribution and increase the generalization capability of the proposed model.

The most critical point at this stage is to preserve the overall structure of the dataset. If an inappropriate augmentation is applied, the generated new samples are not representative of the real world and do not contribute to network performance. In the preprocessing, the reversing, rotation, filtering, and masking operations are used to rise the number of samples. For the reversing, images are rotated 180° counter clockwise around the center point. The rotation operation is applied two times by reversing the elements in each column and row. For the filtering, two-dimensional Gaussian filtering is used with selecting the standard deviation (σ) of 2. Masking operation is also applied by subtracting a blurred image from itself. After all, five various samples are generated from an original image. Fig. 5 presents an example of this processing.

3. Experimental studies and results

In this section, the experiments and comparison studies are realized in detail. In the experiments, the number of samples for each class is

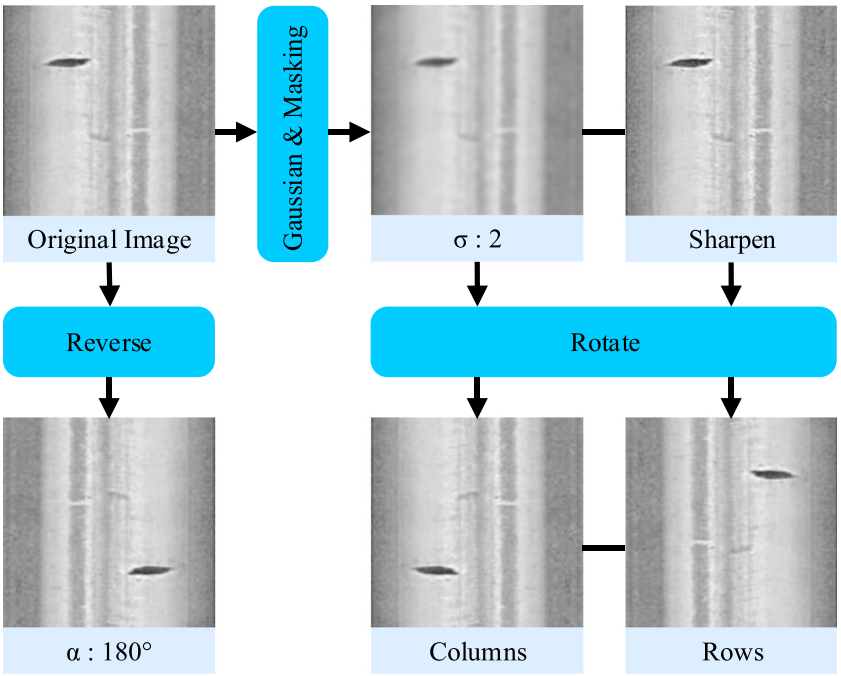


Fig. 5. Data augmentation process for a sample image.

increased by utilizing different data augmentation methods. Therefore, the effect of data augmentation processing is discussed in a separate subsection. The experiments are performed in a workstation with Intel (R) i7-10750H CPU @2.60 GHz, NVIDIA Quadro P620 GPU, and 16 GB RAM memory. All codes developed for the studies are implemented in MATLAB® R2021b software package. The proposed CNN model is trained using the stochastic gradient descent with momentum (Sgdm) optimizer. The mini-batch size, initial learning rate, and dropout rate are set to 16, 0.01, and 0.5, respectively. Moreover, the size of the input image is resized as 100 × 100 pixels.

3.1. Dataset description

The dataset used in this study is a publicly available dataset and can be accessed from Aydin et al. (2021b,a). The railway surface images were obtained by the Railways Research and Technology Center (DATEM) with the measurement train on Ankara–Konya and Ankara–Eskisehir lines in Turkey. The dataset consists of four classification classes as Healthy, Joint, Squats, and Severe Squats (SSquats). There are 492, 408, 608, and 330 samples for the Healthy, Joint, Squats, and SSquats classes before the augmentation, respectively. After the augmentation processing, five new samples are obtained for each sample. Then, a rearranged dataset is obtained by combining the raw samples with the augmented samples. Fig. 6 shows the randomly selected image samples from the dataset.

Finally, a total of 3444, 2856, 4256, and 2310 samples are obtained for Healthy, Joint, Squats, and SSquats, respectively. Therefore, a balanced dataset is provided by equating the number of samples in each class with that of the one with the minimum samples. After dataset preparation, 85% of the dataset is used for training and validation while 15% is selected for testing. Moreover, the same training, validation, and testing subsets are used in all experimental studies for a fairer comparison.

3.2. Evaluation metrics

In order to assess and analyze the performance of the proposed model, six evaluation metrics are used; Accuracy (Acc), Precision (Pr), Sensitivity (Sn), Specificity (Sp), F1-score (F1), and Matthew's Correlation Coefficient (MCC). Acc indicates the success of the network in classification. Pr gives the correct proportion of all predicted samples while Sn shows the proportion of true positive results to all samples. Sp measures the ratio of all non-category negative samples. F1 presents the harmonic average of Pr and Sn. MCC also gives the differences between the actual and predicted classes. According to the confusion matrix, the quantitative metrics are described as follows:

$$Acc = \frac{TP_N + TN_N}{TP_N + FP_N + TN_N + FN_N} \tag{9}$$

$$Pr = \frac{TP_N}{TP_N + FP_N} \tag{10}$$

$$Sn = \frac{TP_N}{TP_N + FN_N} \tag{11}$$

$$Sp = \frac{TN_N}{TN_N + FP_N} \tag{12}$$

$$F1 = \frac{2 \times TP_N}{2 \times TP_N + FP_N + FN_N} \tag{13}$$

$$MCC = \frac{(TP_N \times TN_N) - (FP_N \times FN_N)}{\sqrt{(TP_N + FP_N) \times (TP_N + FN_N) \times (TN_N + FP_N) \times (TN_N + FN_N)}} \tag{14}$$

where, the numbers of correctly classified, opposite classified, incorrectly classified, and misclassified defects are symbolized as TP_N, TN_N FP_N, and FN_N, respectively.

3.3. Analyzes and results

In this section, obtaining the most suitable parameters for the proposed model and improving the classification performance of the proposed model are considered as the main goals. In this context, four experiments are carried out to prove the superiority of the proposed model and divided into four sections.

3.3.1. Evaluation of the data augmentation

The first experiment is conducted to investigate the effect of augmentation on the classification performance of the proposed MSRConvNet. In order to obtain higher accuracy classification performance from the proposed model, the number of samples applied to the input

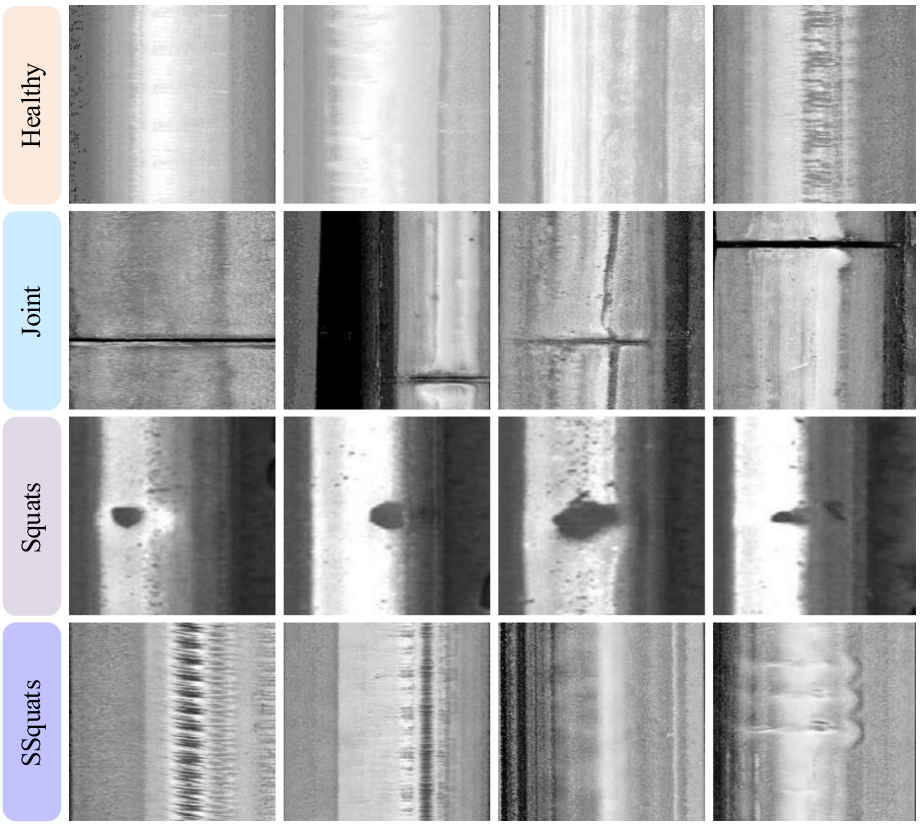


Fig. 6. Randomly selected defect samples from the dataset.

Table 2
Quantitative metric results of the raw and augmented dataset.

Method	Acc	Pr	Sn	Sp	F1	MCC
Proposed model with the raw data	0.9643	0.9650	0.9643	0.9881	0.9643	0.9527
Proposed model with the augmented data	0.9983	0.9983	0.9983	0.9994	0.9983	0.9978

of the network is augmented. As mentioned in the previous section, the five new samples are obtained from the original images in each class and a new dataset is prepared. The metric results achieved for the raw and augmented dataset are listed in Table 2. In addition, the bar graph is presented in Fig. 7 to better observe the metric results obtained from both datasets.

As can be clearly seen from Table 2, the proposed MSRConvNet with the augmented dataset provides better accuracy than the method with the raw dataset. The Acc value of the proposed MSRConvNet based on the augmented dataset is calculated as 0.9983. When both datasets are analyzed in terms of their Pr values, the proposed MSRConvNet with augmented data gives a better value, which is 0.9983. The Pr value of the other dataset is calculated as 0.9650. As a result of these values, it is understood that the difference between the mentioned datasets is at satisfactory values. According to metric results in Table 2, while the Sn and Sp values of the proposed MSRConvNet based on the augmented data are calculated as being 0.9983 and 0.9994, respectively, those of the raw dataset are 0.9643 and 0.9881, respectively. The F1 value of the proposed MSRConvNet with an augmented dataset reaches 0.9983, which is 3.53% higher than the values achieved by the other dataset. When the MSRConvNet structures with raw and augmented datasets are compared in terms of their MCC values, the proposed model-based raw dataset provides a better result, which is 0.9978. As seen in this experimental study, the classification performance of the proposed MSRConvNet is improved when the augmentation processing is applied to the raw dataset. As a general evaluation, using the data augmentation process, the Acc, Pr, Sn, Sp, F1, and MCC values of the proposed MSRConvNet are improved as 3.53%, 3.45%, 3.53%, 1.14%, 3.53%, and 4.73%, respectively.

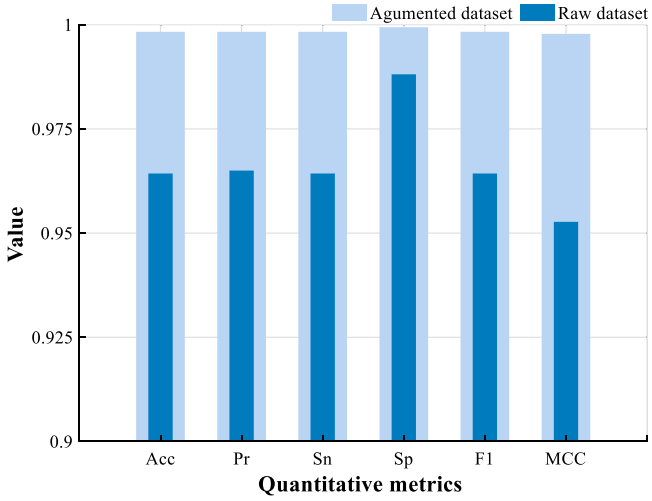


Fig. 7. Comparison of the quantitative metrics for raw and augmented dataset.

Table 3
Quantitative metric results of solvers.

Solver	Acc	Pr	Sn	Sp	F1	MCC
Adam	0.9924	0.9925	0.9924	0.9975	0.9924	0.9899
Rmsprop	0.9933	0.9933	0.9933	0.9978	0.9933	0.9910
Sgdm	0.9983	0.9983	0.9983	0.9994	0.9983	0.9978

Table 4
Quantitative metric results of activation functions.

Activation function	Acc	Pr	Sn	Sp	F1	MCC
Leaky ReLU	0.9949	0.9950	0.9949	0.9983	0.9950	0.9933
ELU	0.9975	0.9975	0.9975	0.9992	0.9975	0.9966
ReLU	0.9983	0.9983	0.9983	0.9994	0.9983	0.9978

3.3.2. Influences of optimizer and activation function

In order to improve the classification accuracy of the proposed MSRConvNet, it is aimed to obtain the most suitable optimizer and activation function in the second experiment. First, the performance of the proposed MSRConvNet is tested on different solvers such as adaptive moment estimation (Adam), Sgdm, and root mean squared propagation (Rmsprop). Table 3 shows the effect of solver performances in the testing dataset. As can be clearly seen from Table 3 the quantitative metric results of all the solvers are very close to each other.

The accuracy values obtained from Adam, Rmsprop, and Sgdm solvers are calculated as being 0.9924, 0.9933, and 0.9983, respectively. the performance of the Sgdm solver is relatively higher than the others. From these values, the Sgdm offers the most reliable classification performance among all solvers. According to Table 3, the Sgdm slightly outperforms the other solvers in terms of their Pr and Sn values. The results of Sgdm are calculated as 0.9983 and 0.9983, respectively. While those of Adam are 0.9925 and 0.9924, those of the other are obtained as 0.9933 and 0.9933, respectively. When the solver is selected as Adam, the Sp value goes up to 0.9994. When all solvers are analyzed in terms of F1 values, the Sgdm gives the best value, which is 0.9983. Those of Adam and Rmsprop are calculated as being 0.9924, and 0.9933, respectively. For the values in the MCC, as in the previous metrics, Sgdm provides the best results. From the results of this experimental study, it is clearly seen that Sgdm has the best metric results. The performance of Leaky ReLU, ELU, and ReLU, which are among the most powerful activation functions, are also tested in this experiment. The metric results of the experiment performed using the testing dataset to investigate the effectiveness of activation functions on classification performance are given in Table 4.

As can be seen from Table 4, it can be stated that ReLU contributes to a relatively small improvement in all metric results. When all activation functions are investigated in terms of their Acc values, the ReLU has the best value, which is 0.9983. Those of the ELU and Leaky ReLU are calculated as being 0.9975, and 0.9949, respectively. On the other hand, when the ReLU is selected as an activation function, the maximum classification performance is achieved with satisfactory values, which are 0.9983 for Pr, 0.9983 for Sn, 0.9994 for Sp, 0.9983 for F1, and 0.9978 for MCC, respectively. While the ELU provides the second-best results, Leaky ReLU shows the worst metric results, which are 0.9950 for Pr, 0.9949 for Sn, 0.9983 for Sp, 0.9950 for F1, and 0.9933 for MCC, respectively.

In addition, the bar graphs in Fig. 8 are presented for a better analysis of the quantitative metric results obtained from the solvers and activation functions. From the presented bar graphs, it is clearly seen that the proposed MSRConvNet with the selection of Sgdm and ReLU provides more accurate results in the classification of defects in the railway tracks.

3.3.3. Comparison of the multi-scale residual block variations for network structure

The last experimental study presents the performance of different structures of the proposed model. There are three multi-scale residual convolutional block variations in the proposed model. The proposed model is divided into three structures for each multi-scale residual convolution block, and a model name is defined for each structure. The network with one multi-scale residual block is called as MSRConvNet-1, while the networks with two and three multi-scale residuals are designated MSRConvNet-2 and MSRConvNet-3, respectively.

First of all, the confusion matrices of MSRConvNet variations are separately presented in Fig. 9. As can be seen from the confusion matrices, the proposed model (MSRConvNet-3) correctly classifies all images in Joint and Squats classes.

On the other hand, one image in the Healthy class is misclassified into the SSquats class, and one image in SSquats class is misclassified into the Healthy class. the proposed model (MSRConvNet -3) correctly predicts 1186 out of 1188 railway track images. While MSRConvNet-2 only classifies all images in the healthy class, MSRConvNet-1 could not classify any class correctly. As can be seen from the confusion matrix of MSRConvNet-2, six images of the SSquats class are misclassified into the Joint class, three images are misclassified into the SSquats class, and two images are misclassified into the Health class. As can be clearly observed from the confusion matrix presented for MSRConvNet-1, one image in Healthy class, eight images in Joint class, five images in SSquats class, and two images in Squats class are misclassified into SSquats class, SSquats class, SSquat-Joint-Healthy classes, and SSquats class, respectively. Moreover, the quantitative metric values obtained from all networks are presented in Table 5 where the best results are in bold. In addition, Fig. 10 is presented to visually analyze the classification results.

As can be clearly seen from Table 5, the MSRConvNet-3 achieves the overall Acc value of 0.9983. As observed, the MSRConvNet-3 correctly predicts all railway track faults in Joint and Squats classes, whereas the images in other classes are predicted with 0.9966 accuracy values. The overall Acc values of MSRConvNet-1 and MSRConvNet-2 are identified to be 0.9865 and 0.9891, respectively. These values can prove that the proposed model is superior to other multi-scale residual-based methods.

When all MSRConvNet variations are analyzed for their Pr values, the proposed model obtains the best values for all classes. The average Pr values of MSRConvNet-1, MSRConvNet-2, and MSRConvNet-3 are calculated as being 0.9866, 0.9893, and 0.9983, respectively. It is clearly seen that the classification performance can be improved through the multi-scale residual blocks in the proposed model. When all MSRConvNet models are evaluated for their average Sn and Sp values, the MSRConvNet-3 has the best values, which are 0.9983 and 0.9994, respectively. These results show that the MSRConvNet-3 outperforms the other architectures. The F1 values of MSRConvNet-1, MSRConvNet-2, and MSRConvNet-3 are identified to be 0.9865, 0.9891, and 0.9983, respectively. From Table 5, it is shown that the MSRConvNet-3 has a more satisfactory classification effect than other architectures. In summary, the MSRConvNet-3 shows a competitive performance due to its powerful residual blocks. The MSRConvNet-3 achieves a better MCC value as compared to other residual-based methods.

The average MCC values of methods are found as being 0.9821, 0.9855, and 0.9978, respectively. From these results, MSRConvNet-3

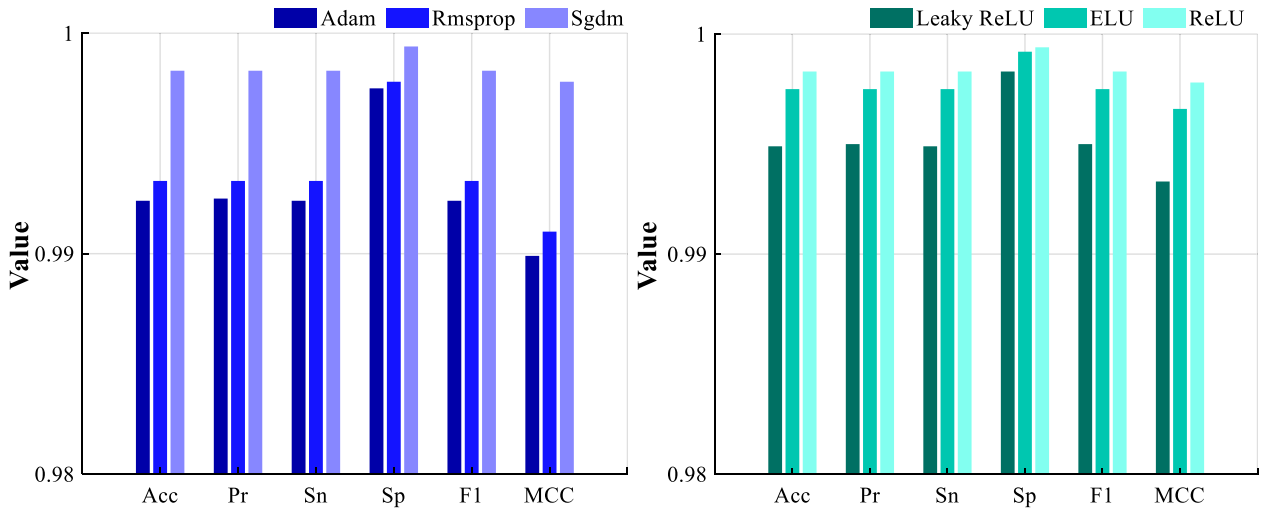


Fig. 8. Comparisons of the classification performance of solvers and activation functions.

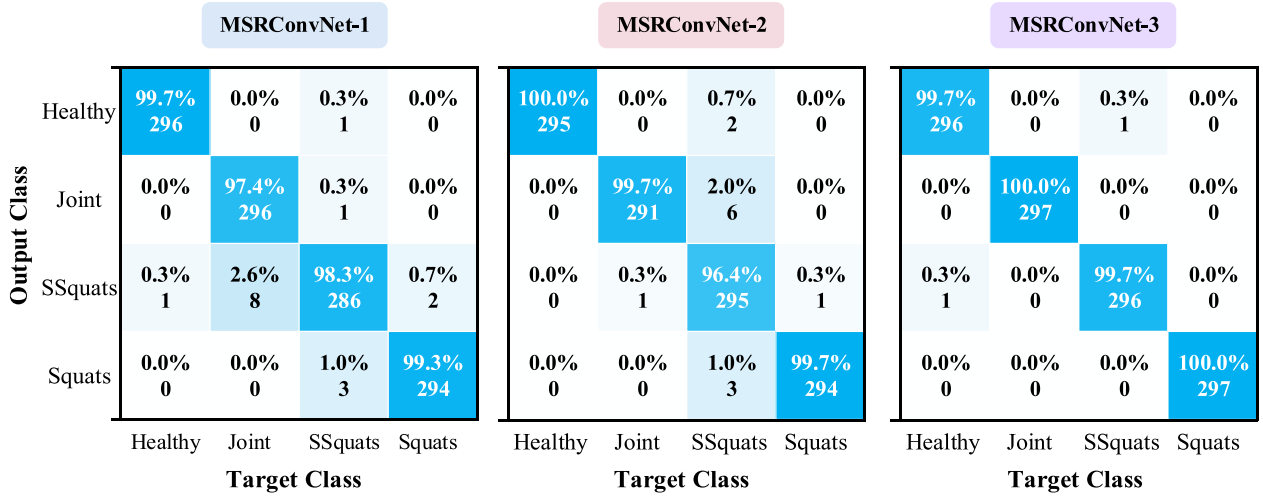


Fig. 9. Confusion matrices of the MSRConvNet variations.

Table 5
Quantitative metric results of the MSRConvNet variations for each defect class.

Model variation	Class	Acc.	Pr	Sn	Sp	F1	MCC
MSRConvNet-1	Healthy	0.9966	0.9966	0.9966	0.9988	0.9966	0.9955
	Joint	0.9966	0.9736	0.9966	0.9910	0.9850	0.9800
	SSquats	0.9629	0.9828	0.9629	0.9943	0.9727	0.9639
	Squats	0.9898	0.9932	0.9898	0.9977	0.9915	0.9887
	Overall	0.9865	0.9866	0.9865	0.9955	0.9865	0.9821
MSRConvNet-2	Healthy	0.9932	1.00	0.9932	1.00	0.9966	0.9955
	Joint	0.9797	0.9965	0.9797	0.9988	0.9881	0.9842
	SSquats	0.9932	0.9640	0.9932	0.9876	0.9784	0.9713
	Squats	0.9898	0.9966	0.9898	0.9988	0.9932	0.9910
	Overall	0.9891	0.9893	0.9891	0.9964	0.9891	0.9855
MSRConvNet-3	Healthy	0.9966	0.9966	0.9966	0.9988	0.9966	0.9955
	Joint	1.00	1.00	1.00	1.00	1.00	1.00
	SSquats	0.9966	0.9966	0.9966	0.9988	0.9966	0.9955
	Squats	1.00	1.00	1.00	1.00	1.00	1.00
	Overall	**0.9983**	**0.9983**	**0.9983**	**0.9994**	**0.9983**	**0.9978**

not only has satisfactory performance but also provides a more reliable classification capability.

To further analyze the classification performance of the MSRConvNet-3, the improvement percentages are obtained over the other methods. When evaluating the improvement percentages in average Acc values, the proposed model improves the performance of MSRConvNet-1 by 1.20%, and MSRConvNet-2 by 0.93%. While the improvement percentages are obtained as 1.20%, 1.20%, 0.39%, 1.20%, and 1.60% for Pr, Sn, Sp, F1, and MCC in MSRConvNet-1, these

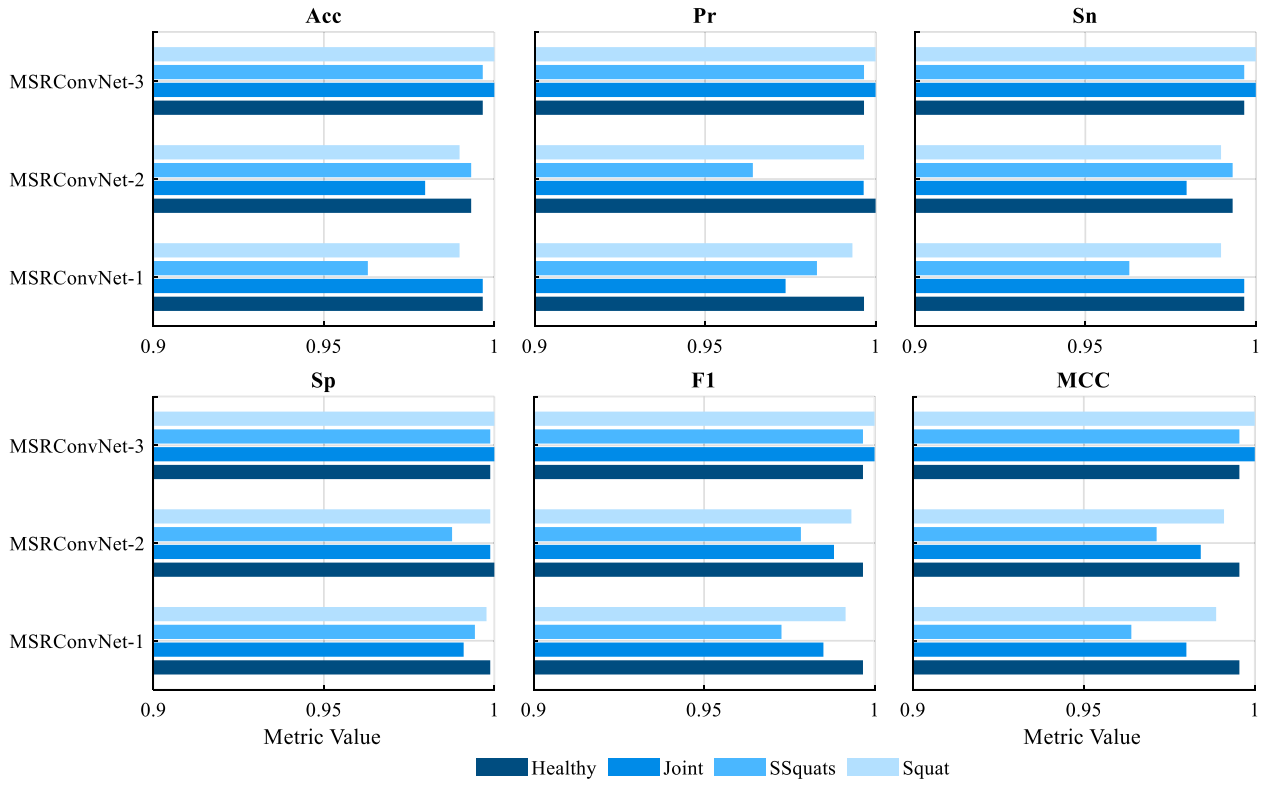


Fig. 10. Comparison of the metric results obtained from all MSRConvNet variations.

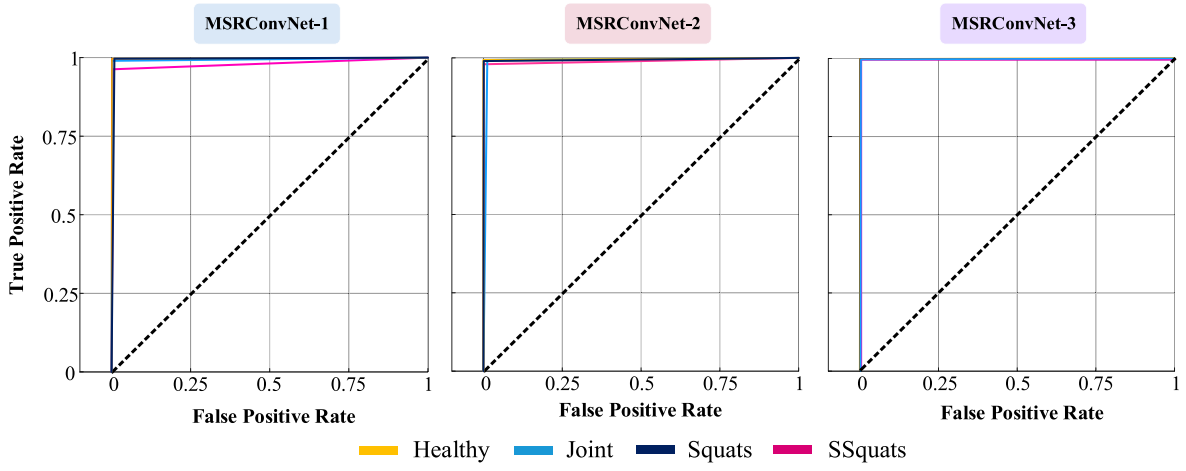


Fig. 11. ROC curves of the all MSRConvNets for each defect class.

values are calculated as 0.91%, 0.93%, 0.3%, 0.93%, and 1.25% in MSRConvNet-2.

Moreover, the receiver operating characteristic (ROC) curves of all MSRConvNet variations are shown in Fig. 11. As can be seen from this figure, ROC curves are presented separately for each class. The ROC curves obtained for MSRConvNet-3 rise immediately along the true positive rate axis compared to other methods. These curves are proof that the proposed model has superior performance.

The t-distributed stochastic neighbor embedding (t-SNE) algorithm is used to visually investigate the features extracted from the different layers of the MSRConvNet-3. The obtained features are presented separately for each layer in Fig. 12.

Each class in this figure is shown in a different color. From Fig. 12, it can be observed that the classes are not exactly at separate points in the first layers of MSRConvNet-3. As is known, satisfactory classification performance is achieved when each class is clustered at different points. When the features collected from the next layers are analyzed, it is clearly seen that a heterogeneous classification for each fault obtains at a different point. It is observed that each class of faults is very well separated at the fc layer of the proposed model, which is the greatest evidence of the superior ability of the MSRConvNet-3 to classify.

In addition, the gradient-weighted class activation mapping (Grad-CAM) method is used to visualize the point at which the proposed model took the classification decision. This method allows the proposed model to analyze which parts of the image are more important for classifying the railway surface defects. In Fig. 13, the outputs of Healthy, Joint, SSquats, and Squats classes are visually presented by using the Grad-CAM method. As can be seen in Fig. 13, original images are randomly selected for each defect class. By applying these pictures

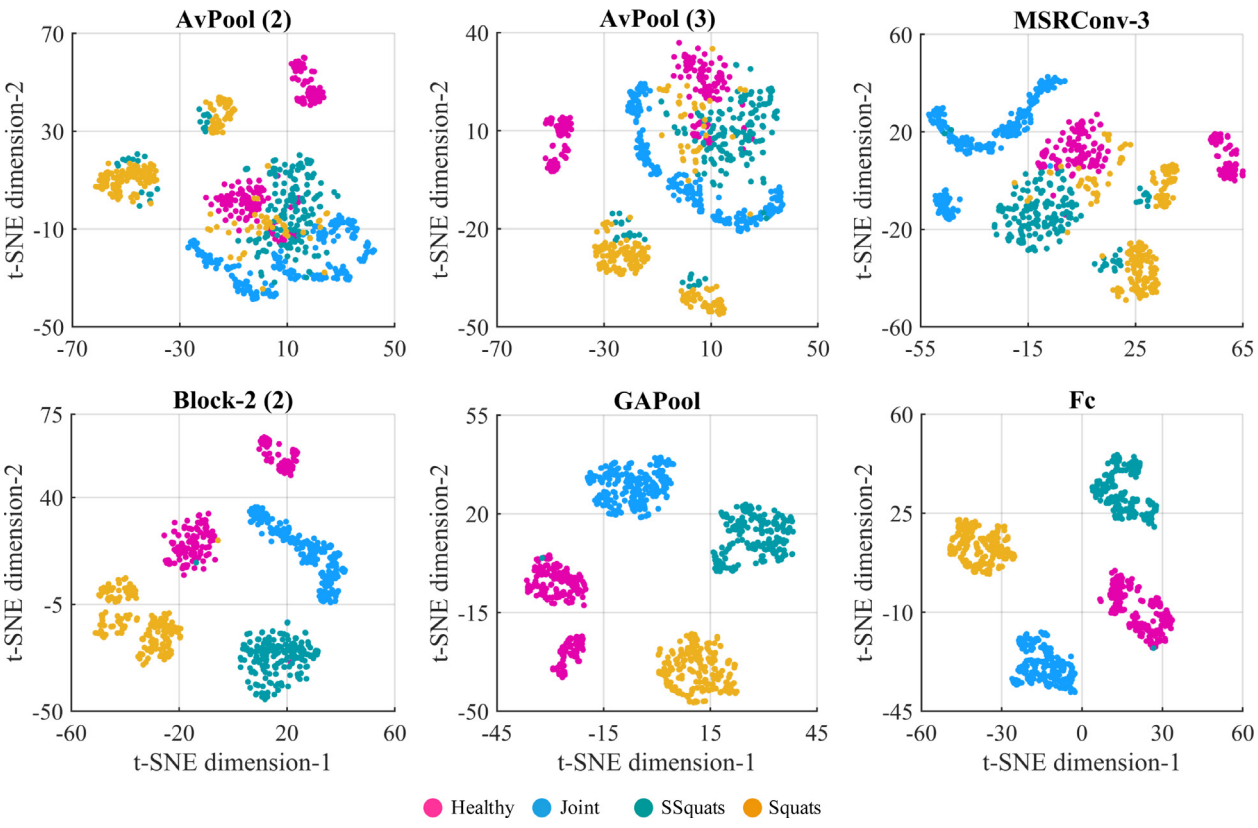


Fig. 12. Visualization of the features obtained from the different layers of the proposed model.

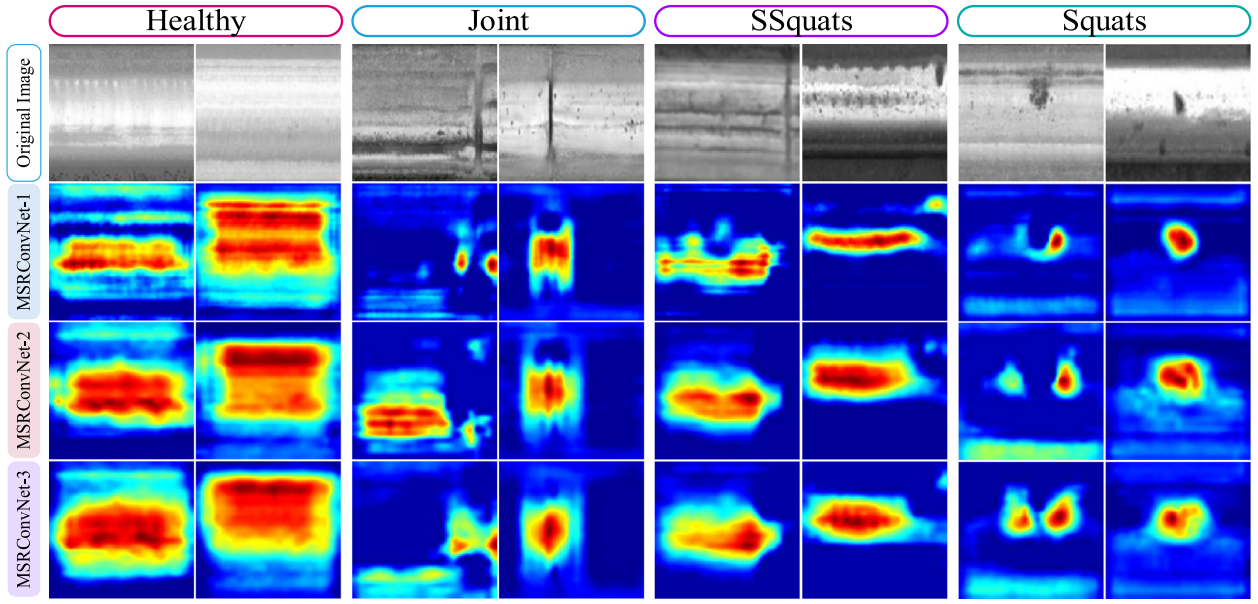


Fig. 13. Heat maps generated by Grad-CAM for each defect.

as inputs to each network, a feature map of the parts affecting the classification score is separately obtained.

3.3.4. Comparison of proposed model and pre-trained deep learning methods

To evaluate the efficacy of the proposed model, a comparison study is realized by applying different pre-trained deep learning methods such as SqueezeNet, AlexNet, GoogLeNet, ShuffleNet, and ResNet-18. For evaluation of classification performance of all methods, the quantitative metrics, namely Acc, Pr, Sn, Sp, F1, and MCC are utilized. The values of these metrics are tabulated in Table 6.

The comparison studies verify that the proposed model performs much better than the pre-trained deep learning methods in the railway track defects classification task. From Table 6, it can be observed that the Acc value of the proposed model is 0.9983, whilst those of SqueezeNet, AlexNet, GoogLeNet, ShuffleNet, and ResNet-18 are calculated as being 0.9596, 0.9630, 0.9745, 0.9815, and 0.9832, respectively. These results clearly show that the proposed model not only

Table 6
Performance evaluation comparison of proposed model and pre-trained deep learning methods.

Method	Acc	Pr	Sn	Sp	F1	MCC
SqueezeNet	0.9596	0.9613	0.9596	0.9865	0.9597	0.9469
AlexNet	0.9630	0.9635	0.9630	0.9877	0.9625	0.9508
GoogLeNet	0.9745	0.9756	0.9745	0.9915	0.9747	0.9665
ShuffleNet	0.9815	0.9822	0.9815	0.9838	0.9836	0.9756
ResNet-18	0.9832	0.9834	0.9832	0.9944	0.9832	0.9777
MSRConvNet-3	**0.9983**	**0.9983**	**0.9983**	**0.9994**	**0.9983**	**0.9978**

Table 7
Comparison of the classification results of the existing studies of the railway track defects.

Study	Year	Method	Class	Acc	Pr	Sn	Sp	F1	MCC
Ye et al. (2020)	2020	SVM using laser-based model	3	0.9670	–	–	–	–	–
Aydin et al. (2021a)	2021	Deep feature combining of CNNs and SVM classifier	4	0.9710	0.9600	0.9625	–	0.9625	–
Proposed model	2022	MSRConvNet	4	**0.9983**	**0.9983**	**0.9983**	**0.9994**	**0.9983**	**0.9978**

gives a more satisfactory result but also proves to have a more reliable ability to classify the railway track defects.

When all methods are compared in terms of their Pr values, the proposed model record a higher value than pre-trained deep learning methods. In comparing the Sn values of models, the best result is obtained from the proposed model, at 0.9983, whereas for SqueezeNet, AlexNet, GoogLeNet, ShuffleNet, and ResNet-18 it is 0.9596, 0.9630, 0.9745, 0.9815, and 0.9832, respectively. According to the metric results given in Table 6, the Sp and F1 values of the proposed model correspond to 0.9994 and 0.9983, respectively. On the other hand, ResNet-18 has the second best Sp and F1 values, which are 0.9944 and 0.9832. SqueezeNet gives the worst performance results, which are 0.9865 and 0.9597, respectively. When all methods are evaluated in terms of their MCC results, the values of the proposed model are significantly higher than the other methods. The results obviously demonstrate that the proposed model is more effective in classifying railway track defects compared to the other methods.

Furthermore, the comparison results also show that the proposed model receives better classification results, in which the 1.54%, 1.71%, 2.44%, 3.67%, and 4.03% average Acc improvements in comparison to SqueezeNet, AlexNet, GoogLeNet, ShuffleNet, and ResNet-18, respectively.

The proposed model improves the Pr values by 1.52% to 3.85%, the Sn values by 1.54% to %4.03, the Sp values by 0.50% to 1.59%, the F1 values by 1.49% to 4.02%, and the MCC values by 2.06% to 5.38%, respectively.

3.4. Comparison between the proposed model and state-of-the-art methods

To provide a rapid and automated railway track defect classification, a powerful deep learning method is proposed and evaluated in this paper. Based on the obtained results, the proposed MSRConvNet gives superior classification performance. In Table 7, a general performance comparison is given between the proposed model and current studies in the literature.

According to Table 7, Ye et al. (2020) proposed a three-dimensional laser model-based railway track surface defect classification method. A set of geometrical features were extracted from the three-dimensional profile model to define a distinguishable pattern for each defect category. In the experiments, SVM, multi-class nearest-neighbor classifier (KNN), and a two-layer feed-forward neural network (FNN) were used for Crack, Squat, and Shelling classes. The best results were achieved with the Gaussian SVM and the overall accuracy was obtained as 96.70%. Aydin et al. (2021a) proposed a feature fusion model by combining the features of the MobileNetV2 and SqueezeNet. After the features were extracted from both CNNs, the feature selection was performed on these features and the new feature set was concatenated to obtain the inputs of the classifier. The ReliefF method was used for feature selection. Finally, the reduced features after the feature selection process were classified with the SVM. In the study, the same dataset with this paper was used and their proposed model was obtained as 97.10% of Acc, 96.00% of Pr, 96.25% of Sn, and 96.25% of F1, respectively. When quantitative metric results given in Table 7 are analyzed, the proposed model outperforms the existing studies in the classification of railway track surface defects. The proposed model achieves better classification accuracy with 99.83% of Acc. As can be seen from these results, the MSRConvNet effectively determines the types of defects and provides superior performance in terms of all metric results.

4. Conclusion

In this study, a multi-scale residual convolutional neural network, namely MSRConvNet, is proposed to classify rail surface defects. A publicly available dataset consisting of the most common defects such as Healthy, Joint, SSquats, and Squats is used. Due to the small number of defect images in the dataset, the augmentation process is performed to obtain more synthetic images. Then, the numbers of images in each fault class are equalized to each other for a balanced dataset. In the proposed network, three multi-scale blocks are constructed with short residual connections and different sizes of convolutions. The multi-scale convolutions distribute detailed feature maps with each other. Residual connections provide feature information sharing and improve network performance. Therefore, different scale feature maps can be extracted.

In order to analyze the performance of the proposed model in classifying the defects in railway tracks, comparison studies are carried out with different variations of multi-scale residual convolution blocks. When the quantitative metric results of all MSRConvNets are evaluated, the accuracy values of the proposed model are obtained as 99.66% for Healthy, 100% for Joint, 99.66% for SSquats, and 100% for Squats. The Pr, Sn, Sp, F1, and MCC values of the proposed model are calculated as being 99.83%, 99.83%, 99.94%, 99.83%, and 99.78% respectively. These results clearly show the superiority of the proposed model in classifying faults in the railway tracks. Moreover, when the average improvement percentages for all metrics are analyzed, the proposed model improves the performance of MSRConvNet-1 variation by 1.20% for Pr, 1.20% for Sn, 0.39% for Sp, 1.20% for F1, and 1.60% for MCC. On the other hand, the improvement percentages for MSRConvNet-2 variation range from 0.91% to 1.25%. Furthermore, the classification performance of the proposed model is compared to pre-trained deep learning methods. The obtained classification results show that the proposed model outperforms pre-trained deep learning models. In future

works, it is considered that the proposed model could be combined with different structures in order to be applicable to the classification of defects with a different number of classes.

CRediT authorship contribution statement

Hakan Acikgoz: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Hardware, Visualization, Roles/Writing – original draft. **Deniz Korkmaz:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Hardware, Visualization, Roles/Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Alvarenga, T.A., Carvalho, A.L., Honorio, L.M., Cerqueira, A.S., Filho, L.M.A., Nobrega, R.A., 2021. Detection and classification system for rail surface defects based on eddy current. Sensors 21, 1–15. http://dx.doi.org/10.3390/s21237937.

Aydin, I., Akin, E., Karakose, M., 2021a. Defect classification based on deep features for railway tracks in sustainable transportation. Appl. Soft. Comput. 111, 107706. http://dx.doi.org/10.1016/j.asoc.2021.107706.

Aydin, A., Salur, M.U., Aydin, I., 2021b. Fine-tuning convolutional neural network based railway damage detection. In: EUROCON 2021-19th IEEE Int. Conf. Smart Technol. Proc. pp. 216–221. http://dx.doi.org/10.1109/EUROCON52738.2021.9535585.

Cao, Z., Qin, Y., Jia, L., Xie, Z., Liu, Q., Ma, X., Yu, C., 2021. Haze removal of railway monitoring images using multi-scale residual network. IEEE Trans. Intell. Transp. Syst. 22, 7460–7473. http://dx.doi.org/10.1109/TITS.2020.3003129.

Espinosa, F., García, J.J., Hernández, A., Mazo, M., Ureña, J., Jiménez, J.A., Fernández, I., Pérez, C., García, J.C., 2018. Advanced monitoring of rail breakage in double-track railway lines by means of PCA techniques. Appl. Soft. Comput. J. 63, 1–13. http://dx.doi.org/10.1016/j.asoc.2017.11.009.

Ghafoor, I., Tse, P.W., Munir, N., Trappey, A.J.C., 2022. Non-contact detection of railhead defects and their classification by using convolutional neural network. Optik (Stuttg) 253, 168607. http://dx.doi.org/10.1016/j.ijleo.2022.168607.

Gong, Z., Zhong, P., Yu, Y., Hu, W., Li, S., 2019. A CNN with multiscale convolution and diversified metric for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 57, 3599–3618. http://dx.doi.org/10.1109/TGRS.2018.2886022.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. pp. 770–778. http://dx.doi.org/10.1109/CVPR.2016.90.

Ji, A., Woo, W.L., Wong, E.W.L., Quek, Y.T., 2021. Rail track condition monitoring: a review on deep learning approaches. Intell. Robot. 1, 151–175. http://dx.doi.org/10.20517/ir.2021.14.

Jin, X., Wang, Y., Zhang, H., Zhong, H., Liu, L., Wu, Q.M.J., Yang, Y., 2020. DM-RIS: Deep multimodel rail inspection system with improved MRF-GMM and CNN. IEEE Trans. Instrum. Meas. 69, 1051–1065. http://dx.doi.org/10.1109/TIM.2019.2909940.

Kishore, M.B., Park, J.W., Song, S.J., Kim, H.J., Kwon, S.G., 2019. Characterization of defects on rail surface using eddy current technique. J. Mech. Sci. Technol. 33, 4209–4215. http://dx.doi.org/10.1007/s12206-019-0816-x.

Li, J., Fang, F., Mei, K., Zhang, G., 2018. Multi-scale residual network for image super-resolution. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Comput. Vis. – ECCV 2018. ECCV 2018. In: Lect. Notes Comput. Sci., 11212, pp. 1–16. http://dx.doi.org/10.1007/978-3-030-01237-3_32.

McNeely-White, D., Beveridge, J.R., Draper, B.A., 2020. Inception and ResNet features are (almost) equivalent. Cogn. Syst. Res. 59, 312–318. http://dx.doi.org/10.1016/j.cogsys.2019.10.004.

Min, Y., Xiao, B., Dang, J., Yue, B., Cheng, T., 2018. Real time detection system for rail surface defects based on machine vision. Eurasip J. Image Video Process. 2018, 1–11. http://dx.doi.org/10.1186/s13640-017-0241-y.

Ni, X., Liu, H., Ma, Z., Wang, C., Liu, J., 2021. Detection for rail surface defects via partitioned edge feature. IEEE Trans. Intell. Transp. Syst. 1–17. http://dx.doi.org/10.1109/TITS.2021.3058635.

Qin, J., Huang, Y., Wen, W., 2020. Multi-scale feature fusion residual network for Single Image Super-Resolution. Neurocomputing 379, 334–342. http://dx.doi.org/10.1016/j.neucom.2019.10.076.

Rafiq, H., Shi, X., Zhang, H., Li, H., Ochani, M.K., Shah, A.A., 2021. Generalizability improvement of deep learning-based non-intrusive load monitoring system using data augmentation. IEEE Trans. Smart Grid. 12, 3265–3277. http://dx.doi.org/10.1109/TSG.2021.3082622.

Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput. 29, 2352–2449. http://dx.doi.org/10.1162/neco_a_00990.

Shen, L., Jia, X., Li, Y., 2018. Deep cross residual network for HEp-2 cell staining pattern classification. Pattern Recognit. 82, 68–78. http://dx.doi.org/10.1016/j.patcog.2018.05.005.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6, 1–48. http://dx.doi.org/10.1186/s40537-019-0197-0.

Tekchandani, H., Verma, S., Londhe, N., 2020. Performance improvement of mediastinal lymph node severity detection using GAN and Inception network. Comput. Methods Programs Biomed. 194, 105478. http://dx.doi.org/10.1016/j.cmpb.2020.105478.

Tu, Z., Wu, S., Kang, G., Lin, J., 2021. Real-time defect detection of track components: Considering class imbalance and subtle difference between classes. IEEE Trans. Instrum. Meas. 70, 1–12. http://dx.doi.org/10.1109/TIM.2021.3117357.

Wang, S., Wu, X., Zhang, Y., Liu, X., Zhao, L., 2020. A neural network ensemble method for effective crack segmentation using fully convolutional networks and multi-scale structured forests. Mach. Vis. Appl. 31, 1–18. http://dx.doi.org/10.1007/s00138-020-01114-0.

Wu, F., Li, Q., Li, S., Wu, T., 2020. Train rail defect classification detection and its parameters learning method. Meas. J. Int. Meas. Confed. 151, 107246. http://dx.doi.org/10.1016/j.measurement.2019.107246.

Xu, P., Guo, Z., Liang, L., Xu, X., 2021. MSF-net: Multi-scale feature learning network for classification of surface defects of multifarious sizes. Sensors 21, 1–18. http://dx.doi.org/10.3390/s21155125.

Yang, H., Bi, Q., Yao, Z., Wang, Y., 2022. Accurate and effective framework for identifying track defects. Meas. J. Int. Meas. Confed. 190, 110625. http://dx.doi.org/10.1016/j.measurement.2021.110625.

Ye, J., Stewart, E., Zhang, D., Chen, Q., Roberts, C., 2020. Method for automatic railway track surface defect classification and evaluation using a laser-based 3D model. IET Image Process 14, 2701–2710. http://dx.doi.org/10.1049/iet-ipr.2019.1616.

Yu, H., Miao, X., Wang, H., 2022. Bearing fault reconstruction diagnosis method based on ResNet-152 with multi-scale stacked receptive field. Sensors 22. http://dx.doi.org/10.3390/s22051705.

Yuan, H., Chen, H., Liu, S., Lin, J., Luo, X., 2019. A deep convolutional neural network for detection of rail surface defect. In: 2019 IEEE Veh. Power Propuls. Conf. VPPC 2019 - Proc. pp. 15–18. http://dx.doi.org/10.1109/VPPC46532.2019.8952236.

Zhang, H., Jin, X., Wu, Q.M.J., Wang, Y., He, Z., Yang, Y., 2018. Automatic visual detection system of railway surface defects with curvature filter and improved Gaussian mixture model. IEEE Trans. Instrum. Meas. 67, 1593–1608. http://dx.doi.org/10.1109/TIM.2018.2803830.

Zhang, D., Song, K., Xu, J., Dong, H., Yan, Y., 2022. An image-level weakly supervised segmentation method for No-service rail surface defect with size prior. Mech. Syst. Signal Process 165, 108334. http://dx.doi.org/10.1016/j.ymssp.2021.108334.

Zheng, Z., Qi, H., Zhuang, L., Zhang, Z., 2021. Automated rail surface crack analytics using deep data-driven models and transfer learning. Sustain. Cities Soc. 70, 102898. http://dx.doi.org/10.1016/j.scs.2021.102898.

Zhu, M., Jiao, L., Liu, F., Yang, S., Wang, J., 2021. Residual spectral–spatial attention network for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 59, 449–462. http://dx.doi.org/10.1109/TGRS.2020.2994057.