

## Non-contact detection of railhead defects and their classification by using convolutional neural network

Imran Ghafoor<sup>a</sup>, Peter W. Tse<sup>a,\*</sup>, Nauman Munir<sup>b,c</sup>, Amy J.C. Trappey<sup>d</sup>

<sup>a</sup> Croucher Optical Nondestructive Testing Laboratory, Department of Advanced Design and System Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

<sup>b</sup> Department of Mechanical Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea

<sup>c</sup> Department of Industrial Engineering, University of Management and Technology, Lahore, Pakistan

<sup>d</sup> Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 300, Taiwan

### ARTICLE INFO

#### Keywords:

Laser ultrasonic technology

Rayleigh waves

Non-destructive testing

Convolutional neural network (CNN)

Rail defect classification

### ABSTRACT

Railhead defects must be detected and classified intelligently in order for railway transportation systems to operate safely. Rail defect identification and categorization can be automated by using machine learning models to process rail image data (acquired using cameras). However, such an automated method has significant drawbacks: it cannot detect subsurface defects, picture data requires a high-end GPU with a long computational time, and machine learning model training can be influenced by image quality, which is dependent on light intensity and shooting altitude. Rayleigh waves are a potential candidate for rail inspection because they can detect both surface and subsurface defects and travel long distances on curved surfaces (like a rail) at high speed. This article looks into the possibility of combining fully non-contact laser ultrasonic technology (LUT) and a deep learning approach for intelligent detection and classification of railhead surface and subsurface defects. The fully non-contact LUT was used to actuate and capture laser-generated Rayleigh wave signals on railhead specimens in order to create a database of A-scan signals from healthy, surface, subsurface, and edge defect railheads. The classification capabilities of a support vector machine (SVM), a fully connected deep neural network (DNN), and a convolutional neural network (CNN) were examined after they were applied to the preprocessed signals without extracting any statistical/signal processing-based characteristics. The comparative analysis demonstrates that CNN is robust in classifying railhead defects. As a result, when combined with CNN, the laser ultrasonic technology may ensure automatic detection and classification of railhead surface and subsurface flaws.

### 1. Introduction

Because of the rapid development of high-speed railways, the volume of railway transportation has increased significantly in recent years. However, as railway usage grows, the rail tracks are subjected to more stressful cycles with heavy loads, which cause defects and, as a result, train derailment. As a result, research into an efficient and intelligent inspection system that can detect and classify various types of rail defects is critical to ensuring the safety and performance of railway networks.

Non-destructive testing (NDT) techniques are commonly used to detect rail track defects. Visual inspection, ultrasonic testing [1],

\* Corresponding author.

E-mail address: [meptse@cityu.edu.hk](mailto:meptse@cityu.edu.hk) (P.W. Tse).

eddy current testing [2], and the acoustic emission (AE) method [3] are all traditional NDT methods for rail inspection. In most cases, the inspection data is manually analyzed to find defects. However, this is a time-consuming, costly, and error-prone activity. Furthermore, the operator's experience influences the outcome. The lack of an automatic defect classification system is one of the most significant issues with inspecting data. As a result, one of the rail industry's primary concerns is the transition from a manual to a fully automated defect classification process.

In recent years, the automatic detection of defects in railways has become a very popular topic. The image data collected by a high-speed video camera is primarily used for this purpose. The collected images are then subjected to some signal-processing or machine-learning-based methods for automatic defect detection and classification. Previously, signal processing-based methods were primarily used to process pixel values and identify defect regions by defining thresholds [4,5]. Many machine learning methods, on the other hand, have been proposed for rail defect detection and classification using shallow and deep learning approaches. The researchers used complex features extracted manually from images in shallow learning techniques.

Santur et al. [6] used PCA, singular value decomposition (SVD), histogram match (HM), and kernel principal component analysis (KPCA) techniques on a healthy and defective rail track dataset to compare the ability of various feature extraction techniques for inspection of rail track. PCA-created features produced the most accurate results in their study. Tastimur et al. classified rail defects such as head check, fracture, scour, and undulation failures using a Haar-like features-based AdaBoost classifier on real-time camera images [7]. Xiong et al. proposed a novel 3D laser profiling system (3D-LPS) to capture rail surface profile data, which was then classified using a decision tree classifier to classify rail surface flaws such as Abrasion, Corrugation, Scratch, Corrosion, and Peeling [8]. Jiang et al. proposed a hybrid approach combining Wavelet Packet Transform (WPT), Kernel Principal Component Analysis (KPCA), and SVM to classify rail rolling contact fatigue (RCF) defects at different depths [9].

Deep learning is a recent advancement that presents an end-to-end classification approach that can automatically extract features from a given dataset [10–14]. Soukup et al. [15] used a database of rail images to train convolutional neural networks (CNNs) to distinguish between defective and healthy rail. Their study was one of the first to use a deep learning model to classify rail defects, and the CNN performed significantly better than model-based approaches. Gibert et al. developed a fully Convolutional Network for inspecting crossties and fasteners on rail tracks. Their database was made up of single-view line-scan camera images of rails [16]. Kang et al. developed an insulator (between the catenary and the earth) surface defect detection system for automatic inspection of railway catenary using a deep convolutional neural network (CNN) [17]. Faghih-Roohi et al. [18] used three DCNN structures to image data (collected from 700 km long rail tracks located in the Netherlands) to automatically detect and classify rail classes (Normal, Joint, Weld, L-squat, M-squat, S-squat). They compared the classification performances of small, medium, and large DCNN architectures, and their findings showed that the largest DCNN outperformed the smaller DCNN. Santur et al. [19] used a three-layer CNN model to classify healthy and defective rail track images captured by 3D laser cameras.

According to the literature review, automatic detection and classification of rail track defects can be accomplished successfully using rail track image data and machine learning technology [6–9,14–23]. However, there are some drawbacks to using this visual inspection based on image technology. First, because the video camera only captures surface images of the rail track, it can only detect surface defects. However, subsurface defects in rail track can also pose a risk to its integrity, and these defects will go undetected by this technology. Second, the image-based database used to train a machine learning model is heavily reliant on image quality, which can be influenced by a variety of factors such as complex background, light intensity, and shooting altitude while capturing images, as well as defect shape, orientation, and surface properties of the material. Furthermore, the rail surface is usually corroded by an oxide scale that has the same color as certain rail surface defects. This could make defect classification by machine learning models extremely difficult. Finally, the video camera typically generates a large image database, which necessitates the use of a high-end GPU and a lengthy computational time.

Laser ultrasonic inspection could be used to compensate for the shortcomings of video camera-based automatic visual inspection. It offers high sensitivity and penetration for non-contact rail inspection. Furthermore, without the use of a contact-type transducer, laser ultrasonic testing (LUT) can inspect inaccessible and complex geometries. A laser-generated Rayleigh wave-based inspection can also have the advantages of increased inspection speed and coverage [24], as well as the ability to propagate around curved surfaces [25]. This is because the Rayleigh wave experiences little attenuation during propagation, and most of its energy is confined within a wavelength on a specimen surface [26]. Rayleigh waves, in addition to surface defects, have properties that make them ideal for detecting sub-surface defects [27] that a video camera cannot detect.

Laser ultrasonic technology was combined with deep learning (DL) models in this study to aid non-contact and automatic railhead defect detection and classification. On healthy and defective rail samples, a laser emission unit and an optical system [30] were used to actuate narrowband Rayleigh waves. The A-scan Rayleigh wave signals of various frequencies were recorded at railhead specimens (using a 3D scanning laser vibrometer) to create a database of healthy, surface, subsurface, and edge defect signals. Rayleigh wave physics knowledge was used to preprocess these signals before feeding them into machine learning models. This signal preprocessing aids the network's learning process by encoding more accurate feature representations. Because signals recorded in the field can be noisier than those measured in the laboratory, artificial noise with signal-to-noise ratios (SNR of SNR 10, SNR 7, and SNR 5) was added to the signals to anticipate this situation and to test the robustness of the machine learning models. Finally, for the classification of healthy, surface, subsurface, and edge defect signals, a fully connected deep neural network (DNN) with drop out, a convolutional neural network (CNN), and support vector machines (SVM) were used. The results show that the adopted CNN performs significantly better than DNN and SVM, even for signals with low SNR.

The following is the structure of the paper: Section 2 describes the generation of Rayleigh waves signals databases, including details on the experimental setup and test specimens, Section 3 describes signal preprocessing steps, the addition of noise, and the separation of training and testing datasets, Details on networks architecture is given in Section 4. Section 5 presents results and discussions, and

Section 6 concludes the study.

## 2. Generation of database

### 2.1. Experimental arrangement

**Fig. 1** depicts a fully non-contact inspection system used for the generation and sensing of Rayleigh waves. It is made up of an excitation laser (Nd: YAG Laser emission unit (SLIII-EX, Continuum Electro-Optics, Inc.)) that emits a laser beam with a wavelength of 532 nm and a pulse duration of 8 ns. The energy per pulse was set to 272 mJ with a 10 Hz repetition rate and a coherent energy meter was used to measure it. For all experiments, the power density was kept less than  $10 \text{ MW/cm}^2$  to ensure that the experiments are performed within thermos-elastic regime. Because of the point source excitation, a typical laser-generated acoustic pulse has a wide bandwidth. However, in practical applications that require frequency-dependent information, narrowband waves are usually preferred [28]. This is due to the fact that a specific narrowband wave must be emitted to the specimen. To avoid the appearance of disturbance and noise, the range of narrowband wave must be selected based on geometry and properties of the material. The selected narrowband wave can be generated when the laser is excited in the form of a specific width of line arrayed pattern. An optical system or a slit mask can be used to create such patterns [29,30].

With the use of a lens configuration termed integrated Sagnac interferometer-based optical system (SIOS), the output point laser beam was transformed into a line arrayed pattern (LAP) [30]. With the use of SIOS mirror M3, the frequency of the generated ultrasonic waves was regulated by modifying the line width of LAP. A 3D SLDV, Scanning Laser Doppler Vibrometer (Polytec, PSV-500-3D-M) was utilized to sense the generated Rayleigh waves to assure non-contact sensing. Three scanning heads, a junction box, and a computer make up the 3D-SLDV. It measures surface motion as the velocity of surface vibrations using the Doppler shift phenomenon. The signals were then sent to a computer for signal preprocessing, machine learning model training, and defect categorization. **Fig. 2** shows a schematic depiction of the entire flaw detection and classification system.

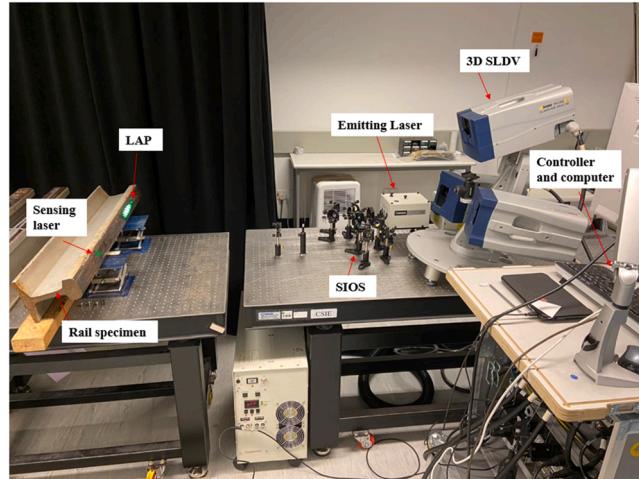
### 2.2. Defects in rail

Rail flaws are classified into two types: geometrical flaws and structural flaws. Geometric defects are related to the geometric conditions of the rail track, whereas structural defects can occur as a result of manufacturing flaws, poor rail handling, rail fatigue, and wear. The presence of non-metallic inclusions is one of the most common sources of manufacturing flaws. Under operative loads, these inclusions can cause local stress concentrations, which can eventually lead to cracks.

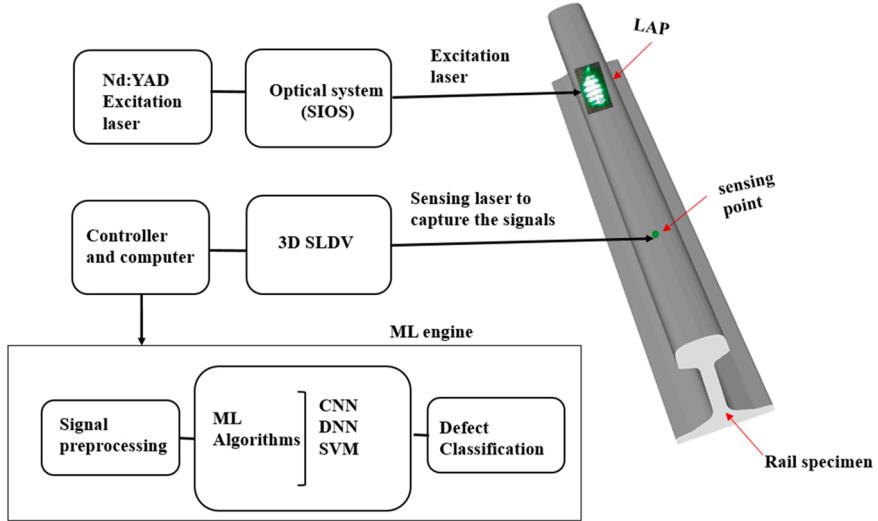
Defects caused by inappropriate rail handling primarily result from unexpected breaks or spinning of the train's wheels. Rail fatigue and wear are caused by the wearing mechanisms of the rolling surface between the rail and the wheel. Rail structural flaws can develop at the rail structure's surface or subsurface, and they can also occur at different locations inside the rail structure. Surface flaws occur as a result of rolling contact fatigue (RCF) or impact from damaged wheels. The direction of propagation of flaws under operative loads determines whether rail structural defects are transverse or longitudinal. **Fig. 3** depicts a few examples of surface and subsurface railhead defects.

### 2.3. Test specimens and experimental procedure

Four steel rail specimens, A, B, C, and D, were utilized in this work to collect laser-generated Rayleigh wave signals. These



**Fig. 1.** Experimental setup for non-contact inspection of the railhead.



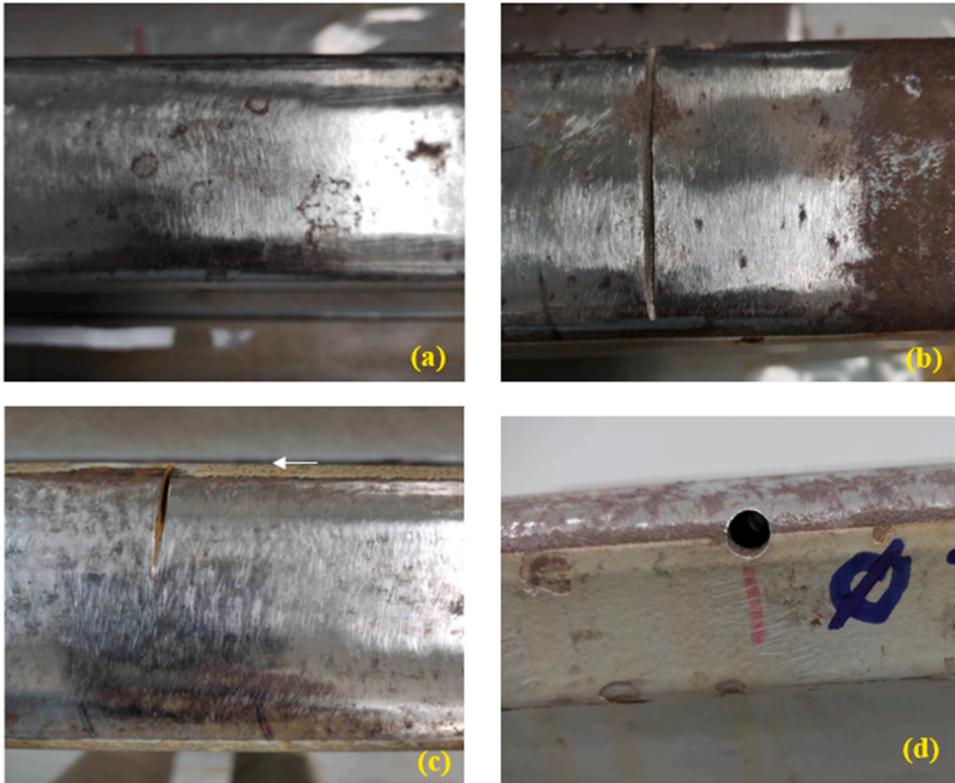
**Fig. 2.** Schematic representation of defect detection and classification system.



**Fig. 3.** Common defect types that occur on railhead; (a) Horizontal stratification (b) Transverse crack (c) Gauge side head checking and flaking (d) Vertical Split Head [31,44]

specimens were provided by Hong Kong MTR Corporation Limited. All specimens were one meter long; however, specimens A, B, and C were greater in size (17 cm height) and weighed 60Kg, while specimen D was short (14.5 cm height) and weighed 45Kg. As illustrated in Fig. 4, specimen A has no flaw, whereas samples B, C, and D have an artificial surface (3 mm deep, 2 mm wide, and approximately 5.5 cm long), subsurface (4 mm deep, 7.5 mm in diameter), and edge (3 mm deep, 1.5 mm wide) defect, respectively.

Line scans were done on all rail specimens utilizing fully non-contact laser-based inspection equipment (shown in Fig. 1). During the experiment, the positions of the excitation and sensing lasers were modified so that the detection range of each type of defect could



**Fig. 4.** Railhead specimens; (a) Healthy (specimen A) and with (b) surface defect-Specimen B (c) edge defect- Specimen C (d) subsurface defect – Specimen D.

be identified, and the length of each line scan was chosen correspondingly. The trials revealed that the maximum inspection ranges for surface, subsurface, and edge faults were 30 cm, 10.7 cm, and 10 cm, respectively. Several line scan measurements were taken for each test specimen, with the laser excitation position, laser excitation energy and line width of the LAP being varied.

This approach aided in the capturing of laser-generated Rayleigh wave signals of varying amplitudes and frequency. 170 data were averaged at each sensing site to produce a clear signal with 2048 sample points. This aided in capturing clean signals in a large window that contained both incident and defect echoes. Furthermore, the 3D SLDV's sampling frequency was adjusted at 5.12 MHz.

### 3. Signals preprocessing, noise addition and segregation of database

The creation of multimode and superimposed modes, mode dispersion, reflections from railhead geometries and boundaries, and unfiltered responses are all issues for laser-generated Rayleigh wave monitoring of rail track [32]. The preprocessing of signals based on physical knowledge of Rayleigh waves can improve the explainability of actual datasets and help networks encode appropriate feature representation. In this context, the Time of Flight (TOF) technique and digital band-pass filtering were utilized to locate defect echoes by removing undesired components from a given signal. Surface waves formed on railheads are well known to be Rayleigh waves [33], and they are non-dispersive up to 40 cm away from the laser excitation region [34]. As a result, the dispersion in the signals was ignored.

#### 3.1. Signals filtration

The signals were initially filtered using a zero phase FIR band-pass filter in the first step. The wavelength and frequency of the produced waves were determined using the LAP line width. The filter's bandwidth was then selected using the approximated frequency of the generated waves. The passband of the band-pass filter was fixed between 300 and 500 kHz to record Rayleigh waves traveling with central frequencies of 380 and 420 kHz. Similarly, to capture Rayleigh with central frequencies of 520 kHz and 680 kHz, respectively, the passband was changed to 400–600 kHz and 550–750 kHz. During the experiment, the band-pass filtering step was carried out in the 3D-SCDV controller and computer system. This band-pass filtering removed unwanted frequency components from the signal, making the incident wave and defect echo more visible.

### 3.2. Time of flight (TOF) analysis and signal windowing

In this study, all of the collected signals had a clear incident (Rayleigh) wave traveling at a velocity of roughly 3045 m/s. As a result, the defect peak for each signal was determined based on the time of flight computation. The recorded signals included, in addition to incident and defect echoes, shock waves, reflections from specimen ends, and additional reflections from head geometry. The existence of these undesired wave packets may have an impact on network learning efficiency and lengthen computing time. Signal windowing is a powerful approach in this regard since it can minimize computing time without removing necessary information from a given signal.

We used a rectangular window with a unit height and a  $T_{win}$  width. The  $T_{win}$  was chosen in such a way that the windowed signal contained clear incident and defect echo (based on TOF of propagating waves information), while removing shock waves and rail end reflections where appropriate. The time plots of the signals captured at test specimens A, B, C, and D are shown in Fig. 5(a), (b), (c), and (d), respectively.

### 3.3. Database

Following signal preprocessing, a database with four different classes was created, including healthy signals, surface defect signals, subsurface defect signals, and edge defect signals. Because each specimen's line scan was different in length, each class has a different number of signals, as described in Section 2.3. Table 1 lists the central frequencies of all the signals that were generated in the narrowband frequency range (with the help of SIOS).

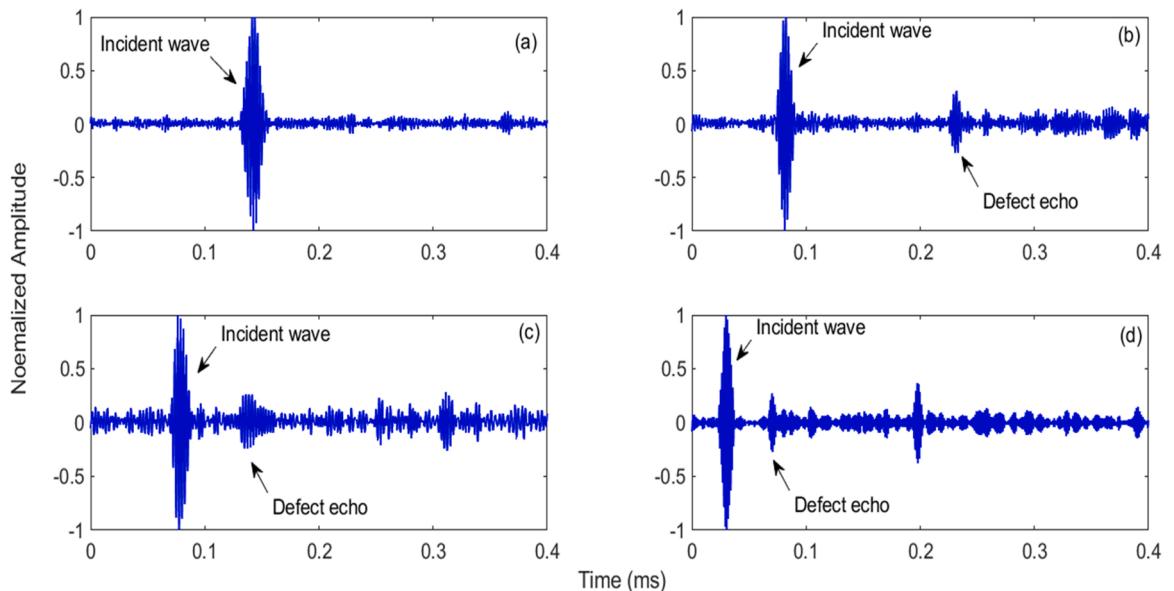
The majority of signals in the healthy, surface, and subsurface defect classes have 420 kHz central frequencies, but all signals in the edge defect class have higher (680 kHz) frequencies. This is because, in comparison to surface and subsurface defects, the edge defect was small in size, necessitating the use of Rayleigh waves with higher frequencies (shorter wavelengths) to detect it. To test the robustness of the ML models on different frequency datasets, healthy signals with 380 kHz and 520 kHz frequencies, as well as surface defect signals with 380 kHz frequencies, were included in the respective datasets. However, each signal in each dataset has the same number of sampling points (2048). The database with each type of defect is given in Table 1.

### 3.4. Noise addition to signals

The noise levels in the laser-generated ultrasonic signals captured in the field may be higher (due to different experimental and environmental conditions) than shown in Fig. 5. Furthermore, a machine learning network should be robust to the various noise levels found in the ultrasonic dataset. In this regard, the signals in the database were corrupted with artificial noises to test the performance of the proposed networks in adverse environments.

Assume the signal  $v(t)$  is a given ultrasonic signal and it is corrupted by a random Gaussian noise  $e(t)$ , then the resulting noisy signal  $\hat{v}(t)$  can be written as:

$$\hat{v}(t) = v(t) + e(t) \quad (1)$$



**Fig. 5.** A-scan signals captured at railhead specimens; (a) Healthy (specimen A) and with (b) surface defect-Specimen B (c) subsurface defect-Specimen C (d) edge defect- Specimen D.

**Table 1**

Number of signals in each class (healthy and defective) of the database.

	Number of signals				
	380KHz	420KHz	520KHz	680KHz	Total
Healthy	130	1344	200	×	1674
Surface defect	150	1543		×	1693
Subsurface defect	×	743	×	×	743
Edge defect	×	×	×	792	792
Total					4902

Where the signal-to-noise ratio (SNR) is defined as

$$SNR = 10 \log_{10} \frac{\sum_{i=1}^N |v(t_i)|^2}{\sum_{i=1}^N |e(t_i)|^2} (dB) \quad (2)$$

The generated database was corrupted with three noise levels of SNR 10, SNR 7 and SNR 5. In this way, in addition to the original (clean) database, three more databases were generated named SNR 10, SNR 7 and SNR 5.

### 3.5. Segregation of database

The training of a neural network with a given dataset, followed by its performance evaluation with the same dataset, can result in very high accuracy. However, this is methodologically incorrect because such practice may cause the network to fail to predict unseen data. However, the actual performance of the neural network can be tested by dividing the given database into two parts: training dataset and testing dataset. This is a common and beneficial practice in the research community. As a result, all clean, SNR 10, SNR 7, and SNR 5 datasets were divided into training and testing datasets.

The number of signals recorded at each specimen, however, was not equal due to experimental constraints. As a result, the experimentally generated database was imbalanced, with certain classes receiving more signals than others, as shown in [Table 1](#). As a result, the datasets for each class were separated in such a way that the parentages of the training and testing datasets for each defect class remained consistent. As shown in [Table 2](#), this is 90% for the training dataset and 10% for the datasets.

## 4. Machine learning

Without explicit instruction, a machine learning (ML) algorithm can learn the underlying pattern from given data. The classification of a given set of data is a predictive modeling problem in which the class labels in the input data are predicted. The use of machine learning (ML) in the rail industry is growing in popularity, owing to its ability to deal with large amounts of data. Furthermore, advances in ML algorithms and computational hardware (e.g., GPU acceleration) have improved their performance in performing tasks like image classification to human-level. For classification of a given data, a variety of ML approaches (for example, decision trees (DTs), K nearest neighbors (KNNs), support vector machines (SVM), artificial neural networks, and random forests (RFs)) are available, each with different potentials, drawbacks, and applications.

One of the most popular machine learning (ML) models is neural networks. They are used to map an input space (Rayleigh wave time-series signal) to a continuous (regression) or discrete (classification) target space as a numerical function approximator. A simple neural network (NN) typically consists of three layers: an input layer, a hidden layer, and an output layer, as in a fully connected single hidden layer neural network. Deeper NNs typically have multiple hidden layers between the input and output layers. Every node in the current layer is connected (with a specific weight) to every node in the previous layer in a fully connected NN.

Another type of deep neural network is the convolutional neural network (CNN) proposed by Lecun et al. [35]. A CNN has convolutional layers and pooling layers in addition to fully connected layers. Convolution layers, unlike fully connected layers, are connected to specific regions of the preceding layers (based on defined filters/convolutional kernels). The input data is downsampled by the pooling layer, which saves time and prevents overfitting [36]. These convolution and pooling layers allow the CNN to learn hierarchical features from the raw or featured data. Due to weight sharing, a CNN requires fewer parameters to learn a problem than a

**Table 2**

Segregation of database into training and testing datasets.

	Training Dataset	Testing Dataset
Healthy (H)	1506	168
Surface defect (SD)	1524	169
Subsurface defect (SSD)	669	74
Edge defect (ED)	713	79
Total Signals	4412	490

fully connected deep neural network.

For the classification of laser generated ultrasonic signals, a fully connected deep neural network (DNN) and a convolutional neural network (CNN) were used in this study. The Support Vector Machine (SVM) [37–39] classifier was used to classify clean and defect signals present in the given datasets in order to compare the defect classification performances of the under investigation deep learning models with other popular machine learning models. The architectures and parameters of the model are provided into subheadings.

Accuracy, precision, and recall scores were calculated and confusion matrices were developed to evaluate the performance of under investigation ML models on testing datasets. The accuracy of a dataset is defined as the number of correctly predicted observations divided by the total number of predictions made. Precision is the ratio of true positives (True positives) to overall predictive positives (True Positive+False Positive), while recall is the number of true positive class predictions made from true positive examples (True Positive+False negative). Both precision and recall scores are useful in predicting the outcome of an imbalanced data set. Precision score is extremely useful when the cost of a false positive is high. Conversely, when the cost of a false negative is high, recall score becomes important. Eqs. (3), (4), and (5) give the mathematical forms of accuracy, precision and recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

where, TP, TN, FP, FN stands for true positive, true negative, false positive and false negative, respectively.

#### 4.1. Support vector machine

The support vector machine is a machine learning model that is based on kernels and can be used for both regression and classification. SVM, which is based on strong theoretical foundations, performs better in nonlinear and high-dimensional feature spaces. The number of inputs and outputs given to the model in this study were 2048 and 4, respectively. The polynomial function with degree

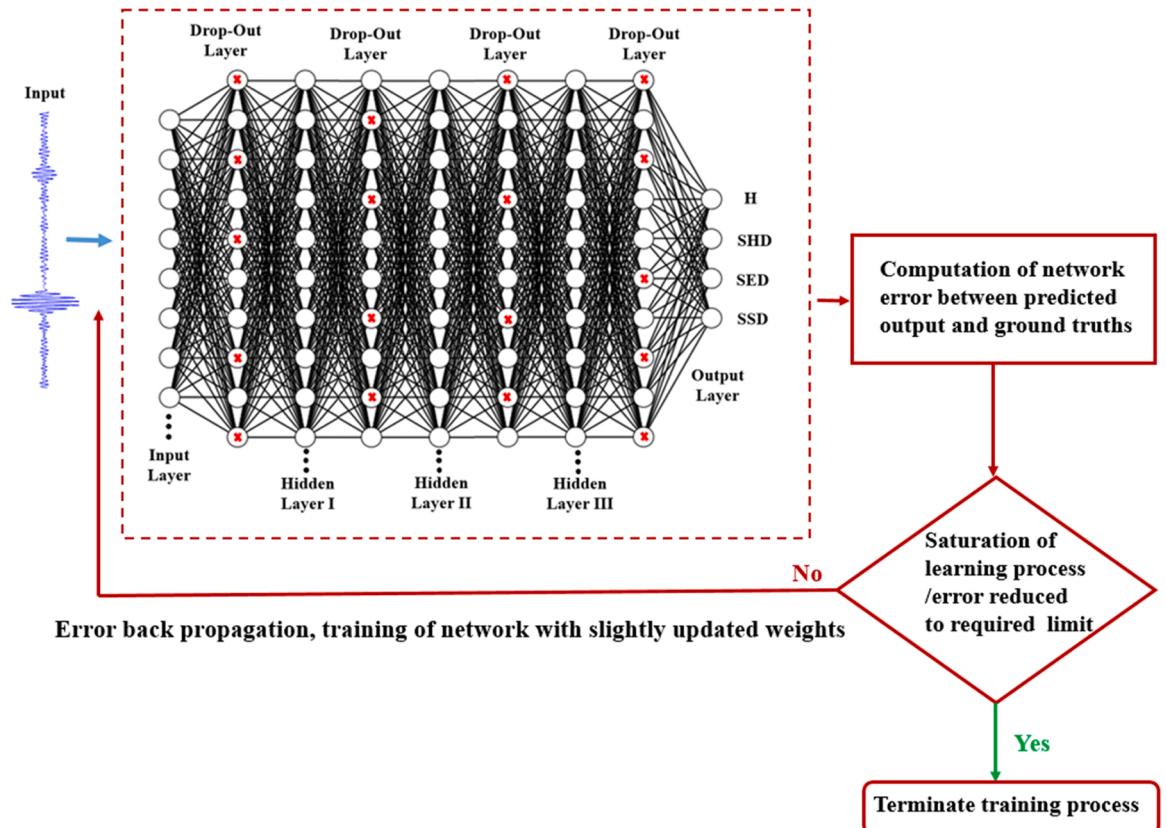


Fig. 6. Architecture and flow diagram of training process of fully connected deep neural network.

4 was chosen as the kernel function. This kernel maps the input space into a higher dimensional space and produces the best results for acoustic signals [40]. Eq. 6 gives the mathematical expression for a polynomial function.

$$K(X_i, X_j) = (X_i \cdot X_j + 1)^d \quad (6)$$

Where K is function, X is input vector and d is degree of polynomial. The hardware used for training consisted of i7 9700 processor with 64 Gb of RAM. During execution, the model was kept running until it converged.

#### 4.2. The architecture of the proposed DNN

The fully connected deep neural network (DNN) used in this study had one input layer, three hidden layers, four 0.5 probability dropout layers, and an output layer. Tensorflow 2.0, a deep learning open-source software developed by Google, was used to create this fully connected DNN. The training hardware included an i7 9700 processor with 64 GB of RAM. The numbers of nodes in the input layer were chosen based on the number of sampling points of the Rayleigh wave signals. The number of nodes in the output layer, on the other hand, was kept equal to the types of defect classes, i.e. healthy, surface, subsurface, and edge. The number of nodes in hidden layers, on the other hand, was determined based on the best results obtained after running several trials.

Due to the state of the art performance of rectified linear unit (Relu),  $\text{Relu}(x) = \max[0, x]$  in deep learning [41], it was used as an activation function of the hidden layers. The sparse softmax cross-entropy loss was used as a loss function to calculate the error. This loss function gives excellent classification results for a multi-class problem. It is a two-stage function in which the softmax activation function (shown in Eq. 7) is applied first and then the cross-entropy [36] is calculated by Eq. 8.

$$\sigma(y)_i = \frac{e^{y_i}}{\sum_{k=1}^K e^{y_k}} \quad (7)$$

$$H_y'(y) = - \sum_i y'_i \log(y_i) \quad (8)$$

where y is a vector from inputs to the output layers, i ( $i = 1, 2, \dots, K$ ) is the index of the output units.  $y_i$  and  $y'_i$  represent the predicted probabilities and true labels, respectively. Fig. 6 depicts the entire network's architecture as well as a flow diagram of the training process. Table 3 shows the model's parameters, including the layer type, output shape of the data in each layer, and number of calculated parameters.

#### 4.3. Architecture of the proposed CNN

The CNN was also created in Tensorflow, a Deep Learning open-source software developed by Google. The network's input layer had 2048 nodes (equal to sampling points of the signal). The network also had four convolutional layers, with 32 filters in the first layer and 64 filters in the following three layers. For the first layer, a large filter size ( $25 \times 1$ ) was chosen because it performs well in noisy conditions [42]. The exponential linear unit (Elu) (shown in Eq. 9) was used as an activation function in the convolutional layers to avoid vanishing gradient problems [43].

$$\text{ELU}_g(h) = \begin{cases} h & h \geq 0 \\ g(\exp(h) - 1) & h < 0 \end{cases} \quad (\text{Where } g = -1) \quad (9)$$

After the second, third, and fourth convolutional layers, three Maxpooling layers with  $(2 \times 1)$  filter/ stride sizes were used to downsample the convolutional layer output. After flattening, two fully connected layers were used, one with 2048 nodes and the other with 256 nodes and the ReLU activation function. The number of nodes in the hidden layers was also determined by running several trials until the best results were obtained. Dropout layers were also used before and after hidden layers to avoid overfitting. The output layer was the final layer, which had four nodes with the Softmax Cross-Entropy function. Fig. 7 and Table 4 show the architecture, as well as a flow diagram of the training process and the CNN parameters for rail defect classification.

**Table 3**  
Parameters of fully connected deep neural network.

Layer (type)	Output Shape	Parameters
Input layer	(2048)	0
Dropout layer 1	(2048)	0
Hidden layer 1	(500)	1,024,500
Dropout layer 2	(500)	0
Hidden layer 2	(50)	25,050
Dropout layer 3	(50)	0
Hidden layer 3	(50)	2550
Dropout layer 4	(50)	0
Output layer	(4)	255

Total parameters 1,052,355; Trainable parameters 1,052,355; Non-trainable parameters 0.

## 5. Results and discussions

### 5.1. Performance evaluation of DNN

The fully connected DNN's performance was evaluated using the original database (clean signals) and noisy databases with artificial noise of SNR 10, SNR 7, and SNR 5. The total number of training parameters in the DNN model was 1.1 million. Fig. 8 depicts the network accuracy curves. According to Fig. 8, the testing accuracy of the DNN with clean signals database is approximately 57% at the start. The DNN reaches saturation around the 57th epoch and remains nearly constant for the remainder of the epochs. A stable learning accuracy indicates that the model has reached saturation and is no longer learning. As a result, the learning process was terminated after 600 epochs. However, in the case of noisy databases, the network's initial testing accuracy is approximately 41%, and as the number of iterations increased, the accuracy gradually improved and reached saturation points at the 65th, 86th, and 105th epochs with SNR 10, SNR 7, and SNR 5, respectively. These findings show that as the noise level in the signals increases, the learning rates decreases.

This pattern in results can be attributed to the fact that noise in the Rayleigh wave signal can complicate its waveform. The average accuracy achieved with fully connected DNN in 10 trials is 96.94%, 93.10%, 88.86%, and 85.29% for datasets with clean, SNR 10, SNR 7, and SNR 5 signals, respectively. The results show that when ultrasonic signals are cleaned using physical knowledge, the DNN performs well and achieves 96.87% accuracy. When signals are immersed in high noise (SNR 5), the DNN achieves only 85.47% accuracy. These accuracy results need to be improved because high-level noise in a laser-generated ultrasonic signal is to be expected in practice.

### 5.2. Performance evaluation of CNN

The performance of CNN was then assessed using datasets with clean and noisy signals with SNR 10, SNR 7, and SNR 5, respectively. Fig. 9 shows the resulting accuracy curves with 600 epochs. Fig. 9 shows that the testing accuracy of CNN started at around 37% for all datasets. The clean dataset, on the other hand, reached saturation much sooner (at epoch 60th) than the noisy datasets, where the network reached stable learning after about the 300th epoch. The learning rate of the CNN is the fastest for the clean dataset, and it decreases as the dataset's noise level increases, similar to DNN.

For clean, SNR 10, SNR 7, and SNR 5 datasets, the average accuracies achieved (in 10 trials) with CNN are 97.99%, 96.61%, 94.94%, and 92.74%. These findings show that, when compared to DNN, CNN performed significantly better, especially for noisy signals with lower SNRs. This indicates the robustness of CNN for the industrial ultrasonic database, where the recoded signals can have very low SNRs.

### 5.3. Comparison of classification performance

Table 5 and Fig. 10 compare the classification performance of SVM, DNN, and CNN, displaying average accuracies, precisions, and recalls at various SNRs. The accuracy scores show that all models perform well for clean datasets; however, as the noise level in the datasets increases, the performance of SVM and DNN decreases significantly; however, CNN performs better even at SNR 5. For clean dataset. CNN (97.99%) performed 2.69% better than SVM (95.30); however, this performance gap increased to 12.88% at SNR 5,

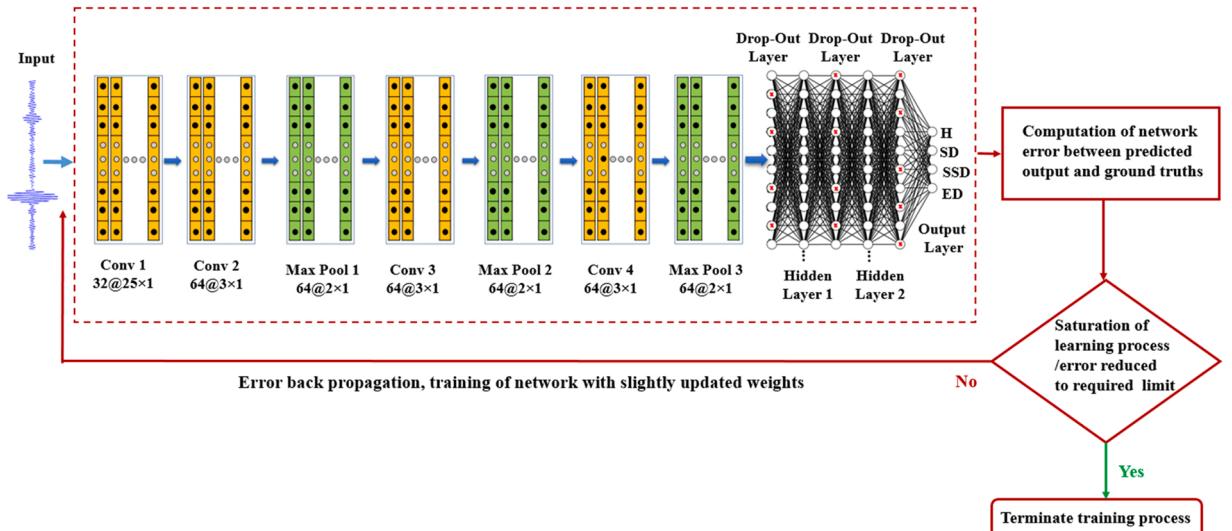
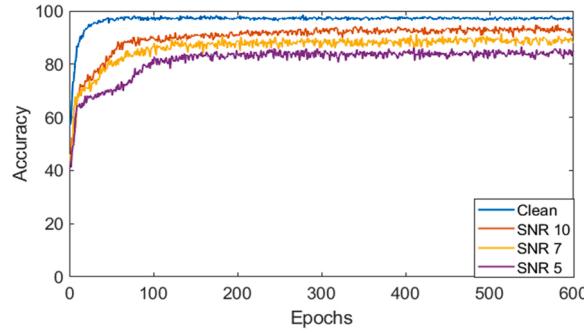
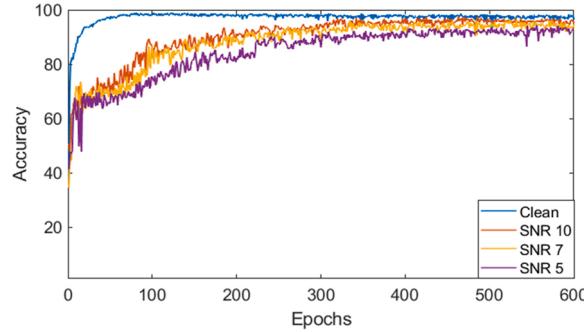


Fig. 7. Architecture and flow diagram of the training process of convolutional neural network.

**Table 4**

Parameters of adopted convolutional neural network.

Layer Type	Kernel Size/Stride	Features Maps	Output Size	Padding	Parameters
1 Input layer	–	–	2048 × 1	–	0
2 Drop Out	–	–	2048 × 1	–	0
3 Conv 1	25 × 1/8 × 1	32	256 × 32	Same	832
4 Conv 2	3 × 1/2 × 1	64	128 × 64	Same	6202
5 Max Pool 1	2 × 1/2 × 1	–	64 × 64	Valid	–
6 Conv 3	3 × 1/2 × 1	64	32 × 64	Same	12,352
7 Max Poll 2	2 × 1/2 × 1	–	16 × 64	Valid	–
8 Conv 4	3 × 1/2 × 1	64	8 × 64	Same	12,352
9 Max Poll 3	2 × 1/2 × 1	–	4 × 64	Valid	–
10 Flattening			256		
11 Drop Out	0.70	–	–	–	–
12 Hidden layer	2048	–	–	–	12,352
13 Drop Out	0.5	–	–	–	–
14 Hidden layer	256	–	–	–	12,352
15 Drop Out	0.5	–	–	–	–
16 Output (Softmax Cross-Entropy)	4	–	4	–	–
Total parameters		164,356			
Trainable Parameters		164,356			
Non-trainable Parameters		0			

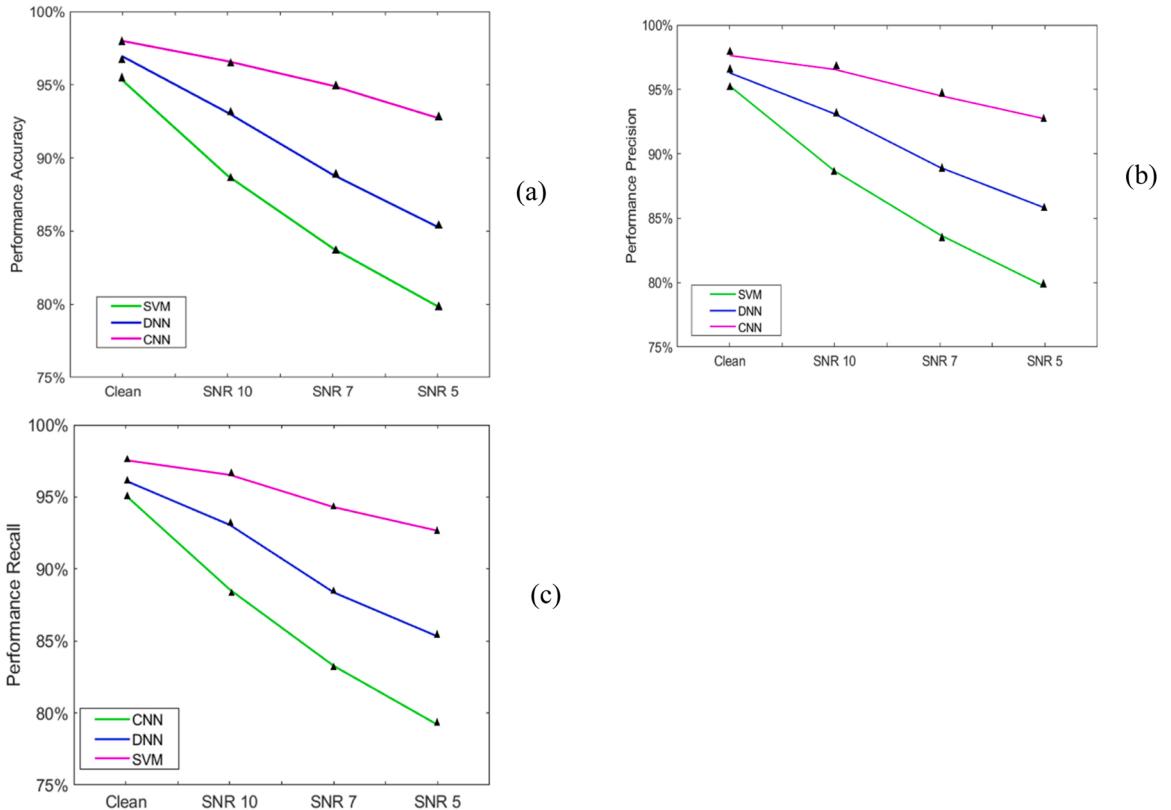
**Fig. 8.** The learning curve of DNN.**Fig. 9.** The learning curve of CNN.

where SVM and CNN have 79.86% and 92.74% accuracies, respectively.

This demonstrates that CNN outperformed SVM across all datasets. Similarly, in the case of the clean dataset, DNN (96.94%) has a small performance gap (1.1%) with CNN (97.99%), but this gap increased to 7.45% at a high noise level (SNR 5). These results show that although the DNN performed better than SVM for all datasets but its performance is significantly less than CNN, particularly for noisy datasets (SNR7 and SNR 5). The similar trend holds for precision and recall results which show that CNN outperforms over both SVM and DNN.

**Table 5**  
Average performance accuracies of the SVM, DNN and CNN.

	SVM	DNN	CNN
<b>Clean</b>			
Accuracy	95.30	96.94	97.99
Precision	95.28	96.29	97.63
Recall	95.10	96.12	97.55
<b>SNR 10</b>			
Accuracy	88.80	93.10	96.61
Precision	88.66	93.10	96.55
Recall	88.57	93.06	96.53
<b>SNR 7</b>			
Accuracy	83.84	88.86	94.94
Precision	83.72	88.95	94.52
Recall	83.27	88.37	94.29
<b>SNR 5</b>			
Accuracy	79.86	85.29	92.74
Precision	79.69	85.80	92.71
Recall	79.18	85.31	92.65



**Fig. 10.** Comparison of SVM, DNN and CNN Performance in terms of (a) accuracy, (b) precision, and (c) recall scores in %age.

#### 5.4. Comparison of training and testing times

The training time was the longest for CNN (220.8 s) followed by DNN (86.24 s) and SVM (19.5 s). However, training time is hardware dependent and most importantly, training of deep neural networks is a one-time process, after training the network to the required accuracy level, its weights and all the other parameters can be saved. The saved model then can be run on any other hardware regardless of its specification provided suitable software tools are used. The testing (model defect detection) time was less than one second for all under investigation ML models i.e SVM, DNN and CNN. This shows that the testing can be performed in a real time fashion.

### 5.5. Defect classification

The ((Number of correctly classified signals)/(Total number of signals)) determines the overall accuracy. Assume that a network correctly identifies the defects that have a high number of signals than those defects that have a low number of signals. Overall, it can still achieve high classification accuracy in this case, according to the definition. To avoid these misleading results, confusion matrices are created to ensure that the network accurately predicts each defect class. Confusion matrices were created with clean and noisy datasets to examine the defect classification accuracy of SVM, DNN, and CNN.

[Fig. 11](#) depicts the confusion matrices (in %ages) generated by SVM, DNN, and CNN using clean and noisy signals (at different SNR values). In addition, [Fig. 12](#) compares SVM, DNN, and CNN to classify healthy, surface, subsurface, and edge defects at various noise levels. The results show that SVM has excellent classification accuracy for each type of signal class in the clean signals dataset, particularly for healthy signals. However, as the noise level in the signals increases, its classification accuracy decreases majorly.

The classification accuracies for clean signals are 95.04%, 95.57%, 95.96%, and 93.27% for healthy, surface, subsurface, and edge defect signals, respectively. The SVM can still classify healthy subsurface and edge defect signals with 91.64%, 94.19%, and 91.39% accuracy at SNR 10, but surface defect classification accuracy drops to 81.16%. At SNR 7, the SVM becomes more confused in classifying defects, and its accuracy suffers even more. It incorrectly predicts surface (8.48%) and edge (0.60%) defects as healthy signals. Surface defect accuracy was reduced to 71.06% because the SVM misclassified 23.45% healthy signals, 1.12% subsurface defect signals, and 4.37% edge defect signals as surface defect signals. The SVM correctly classified 91.30% of subsurface defects, but edge defects were only correctly classified 85.73% of the time. At SNR 5, the SVM provides the poorest performance, with classification accuracy of healthy, surface, subsurface, and edge defects reduced to 87.93%, 65.84%, 89.27%, and 78.20%, respectively.

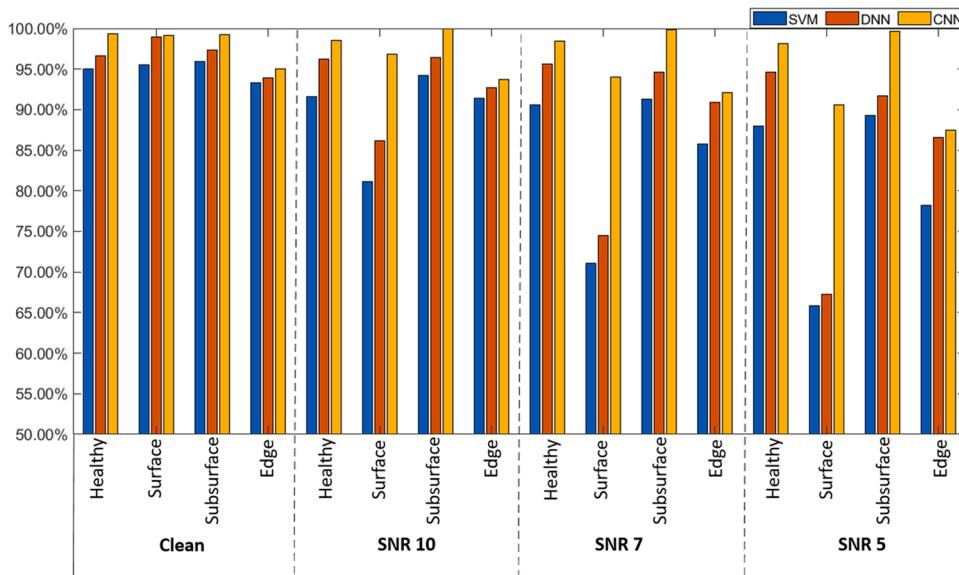
When compared to SVM, DNN's classification accuracy is relatively higher, especially when classifying healthy and subsurface defects immersed in high noise. For clean, SNR 10, SNR 7, and SNR 5 datasets, the classification accuracies of healthy signals are 96.59%, 96.19%, 95.58%, and 94.62%, respectively. Similarly, the DNN correctly classified 97.35%, 96.46%, 94.57%, and 91.67% of clean, SNR 10, SNR 7, and SNR 5 subsurface signals. With a clean dataset, the DNN correctly classified 98.94% of surface defect signals, but its performance degraded with noise.

It correctly classified 86.12% of surface defects at SNR 10, but incorrectly predicted healthy (10.24%), subsurface (1.12%), and edge defect (2.12%) in this class. Because it misclassified 19.31% healthy, 1.89% subsurface, and 4.25% edge defect signals at SNR 7, its accuracy in classifying surface defect fell to 74.54%. At SNR 5, the DNN performs the worst, correctly classifying 67.22% of surface defects and incorrectly classifying 24.75% healthy, 2.36% subsurface, and 5.67% edge defects as surface defects. The classification accuracies of DNN for edge defect remained above 90% for clean (93.94%), SNR 10 (92.70%), and SNR 7 (90.85%) datasets; however, at SNR 5, the classification accuracy dropped to 86.84%.

The defect classification accuracy of CNN is significantly better than both SVM and DNN, as shown in [Figs. 11 and 12](#). With 99.34%, 99.17%, 99.24%, and 95.02% accuracy on the clean dataset, CNN successfully classified healthy, surface, subsurface, and edge defects. Similarly, CNN classified 98.51% of the signals in the SNR 10 dataset as healthy, 96.87% as surface, 100% as subsurface, and 93.67% as edge defect signals. Although the CNN correctly classified 98.45% of healthy, 93.98% surface, and 99.87% subsurface defect at SNR 7, it was a little confused when it came to edge defects (92.06%). Nonetheless, this efficiency outperforms both SVM and DNN. The CNN outperforms SVM and DNN at SNR 5, especially for surface defects, where classification accuracy improved from 65.84% to 90.61%.

	SVM				DNN				CNN			
					Clean							
	Healthy	Surface	Subsurface	Edge	Healthy	Surface	Subsurface	Edge	Healthy	Surface	Subsurface	Edge
<b>Healthy</b>	95.04	4.18	0.00	0.78	96.59	3.23	0.06	0.12	99.34	0.66	0.00	0.00
<b>Surface</b>	2.30	95.57	0.30	1.83	0.47	98.94	0.30	0.30	0.53	99.17	0.24	0.06
<b>Subsurface</b>	0.51	3.41	95.96	0.13	0.13	2.27	97.35	0.25	0.76	0.00	99.24	0.00
<b>Edge defect</b>	0.40	6.33	0.00	93.27	0.54	5.52	0.00	93.94	0.54	4.44	0.00	95.02
SNR 10												
<b>Healthy</b>	91.64	7.65	0.18	0.54	96.19	3.27	0.06	0.12	98.51	0.90	0.30	0.30
<b>Surface</b>	14.71	81.16	0.65	3.48	10.24	86.12	1.12	2.12	0.89	96.87	0.35	1.89
<b>Subsurface</b>	4.55	1.01	94.19	0.25	2.03	1.77	96.46	0.00	0.00	0.00	100.00	0.00
<b>Edge</b>	0.81	7.67	0.13	91.39	0.95	6.49	0.27	92.70	0.13	5.38	0.81	93.67
SNR 7												
<b>Healthy</b>	90.56	8.48	0.36	0.60	95.58	2.75	0.84	0.84	98.45	0.72	0.54	0.30
<b>Surface</b>	23.45	71.06	1.12	4.37	19.31	74.54	1.89	4.25	1.77	93.98	1.36	2.89
<b>Subsurface</b>	5.55	2.77	91.30	0.38	1.89	1.77	94.57	1.77	0.13	0.00	99.87	0.00
<b>Edge</b>	2.42	11.71	0.13	85.73	0.54	7.13	1.48	90.85	0.67	4.58	2.69	92.06
SNR 5												
<b>Healthy</b>	87.93	10.87	0.48	0.72	94.62	3.94	0.66	0.78	98.15	0.84	0.78	0.24
<b>Surface</b>	28.25	65.84	1.24	4.67	24.75	67.22	2.36	5.67	4.43	90.61	2.78	2.19
<b>Subsurface</b>	6.69	3.79	89.27	0.25	4.17	2.27	91.67	1.89	0.38	0.00	99.62	0.00
<b>Edge</b>	4.04	17.09	0.67	78.20	0.94	10.63	1.88	86.54	0.94	7.00	4.58	87.48

[Fig. 11](#). Confusion matrix (in percentage) of SVM, DNN and CNN at different noise levels.



**Fig. 12.** Comparison of SVM, DNN and CNN for defect classification at different SNR values.

## 6. Conclusion

This study looks into non-contact and intelligent railhead defect detection and classification using laser ultrasonic technology and deep learning models. By generating and sensing Raleigh waves on railhead specimens, a completely non-contact laser-based inspection system was used to detect surface and subsurface defects. The signals were preprocessed based on physical knowledge of Rayleigh waves, and a database of healthy, surface, subsurface, and edge defect signals was created. Furthermore, to anticipate noise in field measurements and to test the performance of deep learning models at high noise levels, artificial noise was added to the recorded signals, and new databases with SNR 10, SNR 7, and SNR 5 were generated.

The databases were subjected to two deep learning models, DNN and CNN, and their performance was assessed using accuracy curves and a confusion matrix. The learning rate for both networks was fastest for clean datasets and decreased as the noise level in the signal increased. However, when compared to DNN, CNN achieved significantly improved average accuracies of 96.94–97.99%, 93.10–96.61%, 88.86–94.94%, and 85.29–92.74% for clean, SNR 10, SNR 7, and SNR 5 datasets, respectively. CNN outperformed DNN in terms of railhead defect classification accuracy even at high noise levels (SNR 5) and correctly classified 98.15%, 90.61%, 99.62%, and 87.48% healthy surface, subsurface, and edge defects, respectively.

Furthermore, the defect classification accuracies of these deep learning models were compared to other machine learning approaches (SVM); however, at all noise levels, SVM had a lower classification preference than both DNN and CNN. As a result, combining a laser ultrasonic system with the CNN model ensures non-contact inspection and automatic classification of railhead defects.

## 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.

## Acknowledgments

The work described in this paper is fully supported by a grant from the Innovation and Technology Commission (ITC) (Project No. ITS-205-18FX) of the Government of the Hong Kong Special Administrative Region (HKSAR), China and a grant from the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China (Project No. [T32-101/15-R]). Any opinions, findings, conclusions, or recommendations expressed in this material (or by members of the project team) do not reflect the views of the Government of the HKSAR, ITC, RGC, or Panel of the Assessors for the Innovation and Technology Support Programme of the Innovation and Technology Fund.

## References

- [1] I. Bartoli, F.L. di Scalea, M. Fateh, E. Viola, Modeling guided wave propagation with application to the long-range defect detection in railroad tracks, *NDT & E Int.* 38 (5) (2005) 325–334.

- [2] X. Li, B. Gao, W.L. Woo, G.Y. Tian, X. Qiu, L. Gu, Quantitative surface crack evaluation based on eddy current pulsed thermography, *IEEE Sens. J.* 17 (2) (2017) 412–421.
- [3] X. Zhang, Y. Cui, Y. Wang, M. Sun, H. Hu, An improved AE detection method of rail defect based on multi-level ANC with VSS-LMS, *Mech. Syst. Signal Process.* 99 (2018) 420–433.
- [4] E. Resendiz, J.M. Hart, N. Ahuja, Automated visual inspection of railroad tracks, *IEEE Trans. Intell. Transp. Syst.* 14 (2) (2013) 751–760, <https://doi.org/10.1109/TITS.2012.2236555>.
- [5] Q. Li, S. Ren, A visual detection system for rail surface defects, *IEEE Trans. Syst. Man Cybern. Part A Appl. Rev.* 42 (6) (2012) 1531–1542, <https://doi.org/10.1109/TSMCC.2012.2198814>.
- [6] Y. Santur, M. Karakose, E. Akin, Random forest based diagnosis approach for rail fault inspection in railways, in: 2016 National Conference on Electrical, Electronics and Biomedical Engineering (ELECO), 2016, pp. 745–750.
- [7] C. Tastimur, H. Yetis, M. Karaköse, E. Akın, Rail defect detection and classification with real time image processing technique, *Int. J. Compt. Sci. Softw. Eng. (IJCSSE)* 5 (12) (2016).
- [8] Z. Xiong, Q. Li, Q. Mao, Q. Zou, A 3D laser profiling system for rail surface defect detection, *Sensors* 17 (2017) 1791, <https://doi.org/10.3390/s17081791>.
- [9] Y. Jiang, H. Wang, G. Tian, Q. Yib, J. Zhao, K. Chen, Fast classification for rail defect depths using a hybrid intelligent method, *Optik* 180 (2019) 455–468.
- [10] R. Ye, C.-S. Pan, M. Chang, Q. Yu, Intelligent defect classification system based on deep learning, *Adv. Mech. Eng.* 10 (3) (2018) 1–7.
- [11] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, X. Chen, A sparse auto-encoder-based deep neural network approach for induction motor faults classification, *Measurement* 89 (2016) 171–178.
- [12] Z. Wu, S. Wang, T. Liu, A deep learning vision and laser-assisted method for the automatic assembly and positioning of shield pipe pieces, ISSN 0030-4026, *Optik* 167055 (2021), <https://doi.org/10.1016/j.ijleo.2021.167055>.
- [13] O.A. Shawky, A. Hagag, El-Sayed A. El-Dahshan, M.A. Ismail, Remote sensing image scene classification using CNN-MLP with data augmentation, *Optik* 221 (2020), 165356, <https://doi.org/10.1016/j.ijleo.2020.165356>.
- [14] T. He, Y. Liu, Y. Yu, Q. Zhao, Z. Hu, Application of deep convolutional neural network on feature extraction and detection of wood defects, *Measurement* 152 (2020), 107357.
- [15] D. Soukup, R. Huber-Mörk, Convolutional neural networks for steel surface defect detection from photometric stereo images, in: G. Bebis (Ed.), *Advances in Visual Computing. ISVC 2014*, Lecture Notes in Computer Science, Springer, 2014, p. 8887, [https://doi.org/10.1007/978-3-319-14249-4\\_64](https://doi.org/10.1007/978-3-319-14249-4_64).
- [16] X. Gibert, V.M. Patel, R. Chellappa, Deep multitask learning for railway track inspection, *IEEE Trans. Intell. Transp. Syst.* 18 (1) (2017) 153–164.
- [17] G. Kang, S. Gao, L. Yu, D. Zhang, Deep architecture for high-speed railway insulator surface defect detection: denoising autoencoder with multitask learning, *IEEE Trans. Instrum. Meas.* 68 (8) (2019) 2679–2690.
- [18] S. Faghih-Roohi, S. Hajizadeh, A. Nunez, R. Babuska, B. De Schutter, Deep convolutional neural networks for detection of rail surface defects, *Proc. Int. Jt. Conf. Neural Netw. (IJCNN)* (2016) 2584–2589, <https://doi.org/10.1109/ijcnn.2016.772752>.
- [19] Y. Santur, M. Karakose, E. Akin, An adaptive fault diagnosis approach using pipeline implementation for railway inspection, *Turk. J. Electr. Eng. Comput. Sci.* 26 (2018) 987–998, <https://doi.org/10.3906/elk-1704-214>.
- [20] T. Bai, J. Yang, G. Xu, D. Yao, An optimized railway fastener detection method based on modified Faster R-CNN, *Measurement* 182 (2021), 109742.
- [21] Y. Min, B. Xiao, J. Dang, B. Yue, T. Cheng, Real time detection system for rail surface defects based on machine vision, *EURASIP J. Image Video Process.* 3 (2018), <https://doi.org/10.1186/s13640-017-0241-y>.
- [22] H. Zhang, X. Jin, Q.M.J. Wu, Y. Wang, Z. He, Y. Yang, Automatic visual detection system of railway surface defects with curvature filter and improved gaussian mixture model, *IEEE Trans. Instrum. Meas.* 67 (7) (2018) 1593–1608, <https://doi.org/10.1109/TIM.2018.2803830>.
- [23] L. Wang, L. Zhuang, Z. Zhang, Automatic detection of rail surface cracks with a superpixel-based data-driven framework, *J. Comput. Civ. Eng.* 33 (1) (2019) 04018053-1–04018053-9.
- [24] S. Mariani, T. Nguyen, R.R. Phillips, P. Kijanka, F.L. di Scalea, W.J. Staszewski, Noncontact ultrasonic guided wave inspection of rails, *Struct. Health Monit.* 12 (5–6) (2013) 539–548.
- [25] F. Jin, Z. Wang, K. Kishimoto, Basic properties of Rayleigh surface wave propagation along curved surfaces, *Int. J. Eng. Sci.* 43 (2005) 250–261.
- [26] X. Jian, S. Dixon, N. Guo, R. Edwards, Rayleigh wave interaction with surface-breaking cracks, *J. Appl. Phys.* 101 (2007) 064906-1–064906-7.
- [27] W. Zeng, Y. Yao, S. Qi, L. Liu, Finite element simulation of laser-generated surface acoustic wave for identification of subsurface defects, *Optik* 207 (2020), 163812.
- [28] S. Choi, T. Nam, J. Kyung-Young, S.K. Chung, Frequency response of narrowband surface waves generated by laser beams spatially modulated with a line-arrayed slit mask, *J. Korean Phys. Soc.* 60 (2012) 26–30.
- [29] J. Chen, P. Witse, H. Zhang, Integrated optical Mach-Zehnder interferometer-based defect detection using a laser-generated ultrasonic guided wave, *Opt. Lett.* 42 (2017) 4255.
- [30] K. Ng, P.W. Tse, Design of a remote and integrated Sagnac interferometer that can generate narrowband guided wave through the use of laser and effective optics to detect defects occurred in plates, *Opt. Laser Technol.* 123 (2020), 105923.
- [31] Rail defects: classification, codes, types, groups and instructions for defectiveness of railway tracks, <<https://zen.yandex.ru/media/id/5da46baeba281e00b3fa94b4/defekty-relsov-klassifikaciia-kody-vidy-gruppy-i-instrukciia-po-defektostnosti-jeleznodorozhnyh-putei-5ddce8dcfa4bb2f43c81941>>, 2019 (Accessed 13.03.2021).
- [32] D. Hesse, P. Cawley, Surface wave modes in rails, *J. Acoust. Soc. Am.* 120 (2006) 733–740.
- [33] Y. Fan, S. Dixon, R.S. Edwards, X. Jian, Ultrasonic surface wave propagation and interaction with surface defects on rail track head, *NDT & E Int.* 40 (2007) 471–477.
- [34] G. Imran, P.W. Tse, J. Rostami, N. Kim-Ming, Non-contact inspection of railhead via laser-generated rayleigh waves and an enhanced matching pursuit to assist detection of surface and subsurface defects, *Sensors* 21 (9) (2021) 2994, <https://doi.org/10.3390/s21092994>.
- [35] Y. Lecun, L. Bottou, Y. Benjio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (1998) 2278–2324.
- [36] A. Géron, *Hands on Machine Learning with Scikit-Learn and Tensorflow*, O'Reilly Media, Inc., 2017.
- [37] J. Cervantes, F. García-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: applications, challenges and trends, *Neurocomputing* Volume 408 (2020) 189–215.
- [38] J.D. McNamara, D. Scalea, F. Lanza, M. Fateh, Automatic defect classification in long-range ultrasonic rail inspection using a support vector machine-based smart system, *Insight* 46 (2004) 6.
- [39] Y. Chen, H.-W. Ma, G.-M. Zhang, A support vector machine approach for classification of welding defects from ultrasonic signals, *Non-dstr. Test. Eval.* 29 (2014) 243–254.
- [40] M. Achirul Nanda, K. Boro Seminar, D. Nandika, A. Maddu, A comparison study of kernel functions in the support vector machine and its application for termite detection, *Information* 9 (2018) 5.
- [41] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inform. Process. Syst.* (2012).
- [42] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, *Mech. Syst. Signal Process.* 100 (2018) 439–453.
- [43] D.A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), 2015, arXiv preprint, [arXiv:1511.07289](https://arxiv.org/abs/1511.07289).
- [44] *Track Inspector Rail Defect Reference Manual*, Federal Railroad Administration, US, 2015.