Cogent Engineering

ISSN: 2331-1916 (Online) Journal homepage: www.tandfonline.com/journals/oaen20

Hydraulic brake fault detection from acoustic signals using frequency domain filter bank integrated deep learning framework

Saumye Saran Das, Raushan Kumar, Manas Ranjan Prusty, Tapan K. Mahanta & Subhra Rani Patra

To cite this article: Saumye Saran Das, Raushan Kumar, Manas Ranjan Prusty, Tapan K. Mahanta & Subhra Rani Patra (2025) Hydraulic brake fault detection from acoustic signals using frequency domain filter bank integrated deep learning framework, Cogent Engineering, 12:1, 2549368, DOI: 10.1080/23311916.2025.2549368

To link to this article: https://doi.org/10.1080/23311916.2025.2549368

Published online: 21 Aug 2025.

COGENT ENGINEERING
2025, VOL. 12, NO. 1, 2549368
https://doi.org/10.1080/23311916.2025.2549368

MECHANICAL ENGINEERING | RESEARCH ARTICLE

OPEN ACCESS Check for updates

Hydraulic brake fault detection from acoustic signals using frequency domain filter bank integrated deep learning framework

Saumye Saran Das[a], Raushan Kumar[a], Manas Ranjan Prusty[b], Tapan K. Mahanta[c] and Subhra Rani Patra[d]

[a]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India; [b]Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, India; [c]School of Mechanical Engineering, Vellore Institute of Technology, Chennai, India; [d]Information Systems and Operations Management, University of Texas at Arlington, TX, USA

ABSTRACT

Brake failure is a critical safety issue in the automotive industry, often resulting from mechanical wear, hydraulic malfunctions, or electronic failure. Common causes include brake fluid leakage, worn-out brake pads, and damaged brake lines, exacerbated by poor maintenance and ignored warning signs, such as squealing. Effective fault detection is essential to ensure road safety. This study is motivated by the critical need to enhance the accuracy of brake fault detection and classification, which is a key factor in improving road safety and reducing the risk of accidents caused by brake system failures. This paper presents an automated approach for detecting hydraulic brake faults by analyzing five types of brake sound anomalies, Type I to Type V, covering various disk pad wear scenarios. These time-domain sound signals are converted to frequency-domain image representation using a mel spectrogram, and a custom 14 layer Deep Sequential Convolutional Neural Network (CNN) model classifies these images. The model achieved 97.7% accuracy in binary classification and 97.8% in 5-class classification, with perfect results at 20 epochs during hold-out cross-validation. This study significantly enhances the safety of hydraulic brake systems by automating fault detection and classification using a deep sequential CNN framework.

ARTICLE HISTORY
Received 4 December 2024
Revised 22 January 2025
Accepted 25 February 2025

KEYWORDS
Hydraulic disc brake system; multi-class brake sound; frequency domain filter bank; mel-spectrogram; deep sequential models

SUBJECTS
Mechanical Engineering; Acoustical Engineering; Image Processing; Machine Learning; Computer Engineering

1. Introduction

Brake failure poses a significant safety concern in the automotive industry and has the potential for catastrophic consequences. Brake-related defects account for approximately 2% of annual motor vehicle accidents in the United States, leading to approximately 1500 fatalities and 100,000 injuries according to (National Highway Traffic Safety Administration) 2022 data (Stewart, 2021). Brake failure can be attributed to various factors, including mechanical wear and tear, hydraulic system malfunctions, and electronic control failures. Common causes of brake failure include brake fluid leakage, worn-out brake pads, and damaged brake lines. Neglecting warning signs, such as squealing or reduced braking effectiveness, as well as inadequate maintenance, can exacerbate this problem. Global accident statistics reveal that the causes of accidents are distributed as follows: roughly 60–70% of accidents result from human errors, 20–30% are attributed to suboptimal road conditions inconsistent with traffic characteristics, and 10–20% are due to technical malfunctions in cars (Road traffic injuries, 2024). Investigations into the causes of accidents indicate that a portion of accidents are linked to insufficient maintenance, resulting in an inadequate technical state of motor vehicles.

According to the Ministry of Internal Affairs of Russia, 15% of total accidents are a consequence of inadequate technical conditions of motor vehicles, and within this category, 30–40% are specifically related to issues in the brake systems of the cars involved (Bazhenov and Amirseyidov, 2019).

CONTACT Manas Ranjan Prusty manas.ranjan@vit.ac.in Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, India

The repercussions of brake failure are substantial, with studies indicating a 2.5-fold increased risk of fatal injuries in road accidents related to such malfunctions (Elliott et al., 2007). It is imperative for individuals and the automotive industry to prioritize regular maintenance and promptly address warning signs to mitigate the risks posed by brake failure. In India, Maharashtra ranks second after Tamil Nadu in road accidents and faces challenges related to high temperatures during prolonged braking (Nemade et al., 2019). This leads to minimized friction between the disk and friction pads, causing non-operation of the brakes, a reduction in the coefficient of friction, and accelerated wear of the braking pads. Experimental research is required to address the braking problems associated with heat generation during prolonged braking.

The remainder of this paper is organized as follows. In Section 2, we explore previous studies and their statistical data on brake failure to understand the existing knowledge in this area. Section 3 explains the motivation and innovation of this research project. In Section 4, the data acquisition procedure is explained, detailing how the sound is captured. Section 5 explains our proposed architecture, detailing the use of a deep sequential convolutional neural network (CNN) model and how we integrated oversampling techniques to enhance our approach. Section 6 details the results obtained, discusses them, and uses comparison tables to illustrate and support our findings. Finally, Section 7 concludes our study, summarizes the findings, and discusses their implications.

2. Related work

Various methodologies have been explored to enhance the precision and effectiveness of fault detection in the domain of hydraulic system fault diagnoses. Alamelu et al. investigated vibration-based fault diagnosis for an automobile hydraulic brake system using an Artificial Immune Recognition System (AIRS) algorithm. Through experiments on a hydraulic braking system simulating nine fault conditions, they achieved a notable classification accuracy of 97.82% %using the AIRS 2 algorithm [Alamelu et al., 2020). An alternative perspective was presented by An et al., who utilized grey incidence analysis and pattern recognition for diagnosing faults in hydraulic generator units. By simplifying the diagnostic process, they relied on normalized standard fault pattern sets, contributing to an efficient fault diagnosis methodology (An et al., 2007). Similarly, An et al. employed pattern recognition and cluster analysis for fault diagnosis in hydraulic generator units, utilizing Euclidean distance and normalized standard fault pattern sets to enhance the diagnostic accuracy (An et al., 2008).

Shifting the focus to hydraulic pumps, Siyuan et al. suggested a method based on squared estimation fault variations utilizing Principal Component Analysis (PCA). This study established a main element model and employed Q statistics to detect faults, providing insight into different fault types (Siyuan et al., 2011). In the context of hydraulic brake systems, Jayakrishnan et al. applied machine learning techniques along with vibration analysis to predict faults. Achieving a classification accuracy of 83.48%, they utilized classifiers such as J48, LMT, Random Forest, and Random Tree, enabling the effective condition monitoring of brakes using vibration signals (Jayakrishnan et al., 2020). Regarding hydraulic disk brake systems, Lian et al. introduced an online diagnostic approach for identifying faults related to leakage and oil contamination. Utilizing features extracted from a model and an SVM classifier, this study attained an impressive accuracy of 92.5% (Lian et al., 2013). You and Lu presented an innovative empirical neural network model that relied on hybrid particle swarm optimization (You & Lu, 2018). Huang et al. focused on deep learning models with multirate data samples for fault diagnosis. Their proposed model automatically extracted features from multirate sampling data, demonstrating high diagnostic and fault pattern recognition accuracy even in scenarios with a significant imbalance in the sample data (Huang et al., 2022).

Wang et al. presented a new fault diagnosis method called DRSFSI-DL, which utilizes deep learning networks and DRS frequency spectrum images. This study compared this approach with PSD and cepstrum images, demonstrating superior generalization ability and diagnostic accuracy with a specific data segment length of 4096 (Li et al., 2023). Ji et al. proposed a DSmT-based three-layer technique utilizing a multi-classifier approach for fault detection in hydraulic systems. This innovative method employs a layered hybrid technique with a three-layer structure, simplifying diagnostic tasks by breaking them down into subtasks (Ji et al., 2021). Alamelu et al. presented a continuous monitoring system for

vibration-centric fault diagnosis in a hydraulic braking system. Using data mining and rough set theory, statistical parameters were extracted from vibration signals under various fault conditions (T M et al., 2019).

Xu et al. proposed an approach based on a multi-output Support Vector Machine for diagnosing complex faults in hydraulic systems. This approach addressed the issue of simultaneous faults by applying multi-output classification instead of creating new classes (Xu et al., 2021). Liu et al. proposed a data-driven method for estimating wheel cornering power (WCP) in Hydraulic Control Units (HCU). Utilizing a Gated Recurrent Unit (GRU) model with data augmentation (DA), the GRU model outperformed traditional models, such as MLP, LSTM, and transformers. The shifting operation in the DA further improves the generalization ability of the model (Liu et al., 2023). Jegadeeshwaran and Sugumaran explored the diagnosis of faults in a hydraulic brake system using vibration-based signals and machine-learning techniques. The study proposed C4.5 decision tree and SVM classifier algorithms and achieved an impressive 98.72% accuracy. Research has emphasized the significance of selecting appropriate statistical features for accurate fault diagnosis (Jegadeeshwaran & Sugumaran, 2015). Chen (2011) developed a model with a layered architecture based on Bayesian networks for fault diagnosis in hydraulic braking systems. The algorithm implements a statistical approach to eliminate weak causal relationship rules during Bayesian network structure learning, leading to increased accuracy when compared to fuzzy logic methods. Gao and Xiang (2017) proposed a Walsh transform (WT)-based method for denoising noisy vibration signals in the context of compound faults in roller bearings. The WT method demonstrated feasibility through simulation and experimental investigations, showing its potential in enhancing the accuracy of fault diagnosis.

Existing research predominantly utilizes sound signals for dataset creation during brake-fault detection. However, these methods may be susceptible to inaccuracies owing to the sensitivity of the model to minor fluctuations and noise in sound signals. Such noise can compromise the reliability of the fault detection and classification. Therefore, it is imperative to investigate alternative methodologies that exhibit a greater resilience to noise. This study aimed to assess the performance of Convolutional Neural Networks (CNNs) in this context, evaluating their potential as a more robust and accurate approach for brake fault detection in the presence of noise.

3. Motivation and innovation

The current methods for brake fault detection are plagued by significant limitations in the accuracy and differentiation of fault types. Many existing models struggle to achieve high accuracy because of their inability to effectively interpret complex audio data and to distinguish between various types of brake faults. This inadequacy leads to less-precise diagnostics and a potential increase in undetected or misclassified faults, which can compromise vehicle safety (Road traffic injuries, 2024). The urgency of this issue is underscored by alarming statistics. According to the World Health Organization, approximately 1.19 million people die annually from motor vehicle crashes, with a significant portion of these accidents attributed to brake system failures. A survey by the National Motor Vehicle Crash Causation Survey (NMVCCS) highlighted that brake system failures are responsible for approximately 22% of defect-related crashes, emphasizing the critical need for more effective fault detection and classification methods (National Motor Vehicle Crash Causation Survey (NMVCCS), 2024). This study seeks to address these challenges by proposing an innovative approach that significantly improves the fault detection accuracy in hydraulic brakes, thereby enhancing overall road safety.

The novelty of this research lies in the integration of cutting-edge techniques to address the challenges of brake fault detection.

- Brake sound signals are transformed into mel spectrogram images, a representation that better captures the frequency and time-domain characteristics of sounds, mimicking human auditory perception.
- Random oversampling was employed to augment the dataset and ensure that the model was well-trained across all fault types.

- The core of this approach is a custom deep sequential Convolutional Neural Network (CNN), meticulously designed with a 14-layer architecture tailored specifically for mel spectrogram analysis.

This research not only advances the accuracy of fault detection but also significantly improves the differentiation between various fault types, representing a substantial step forward in the field of automotive safety.

4. Data acquisition procedure

A hydraulic brake system primarily contains three parts, namely, the caliper, piston, and brake pads, as shown in Figure 1. Of these three parts, brake faults generally occur in the brake pads. The brake pad consisted of an outer and an inner layer. In general, wear and tear to the brake parts occur in an even or uneven manner in the inner disk pad or in combination in the inner and outer disk pads. There are four categories of brake faults: Even Disk Pad Wear – Inner and Outer (Type I), Even Disk Pad Wear – inner (Type II), Uneven Disk Pad Wear – inner (Type III), and Uneven Disk Pad Wear – Inner and Outer (Type IV). Type V represents sounds from brakes without faults. Hence, in our experimental analysis, we considered only the four faults that might occur in the brake pads, along with no faults. This causes the total number of classes considered for the analysis to be five.

A National Instruments USB 6210 data acquisition system was used to capture sound radiation from the hydraulic disk brake system. A vehicle with a front-wheel drive was selected, and the wheel was lifted using a hydraulic lift jack. For an accelerated single-gear vehicle, a brake of pressure 50 bar was applied. The rotational speed of the wheel during braking was in the range of 120-150 rpm. During the braking period, the sound pressure was measured using a pre-polarized 1/4'' ICP microphone placed 0.5 m distance from the front wheel and it is 1 m above the ground. The sensitivity of the microphone is 45.3 mv/Pa. Acoustic signals were recorded for 4 s at a sampling frequency of 44100 Hz. Five types of brake pads were chosen for the experimental acoustic signal capture, as mentioned above. The sample acoustic signal for each of these five brake faults is shown in Figure 2. The time-domain acoustic signals obtained from the microphone were converted to. wav file using MATLAB for our analysis.

This pilot study was conducted using a controlled experimental setup to capture brake sound data with consistency and precision. To maintain the study's focus and reduce its complexity, several influential factors were intentionally not considered. For instance, the type of braking system, such as disk or drum brakes, can produce distinct sound patterns due to differences in their mechanical operation. Brake pad contact patch size and shape, as well as surface conditions, also play a critical role. Larger and evenly distributed patches tend to produce smoother sounds, whereas uneven patches or contaminated surfaces can generate irregular noise and vibrations.

Additionally, the friction coefficient, determined by the brake pad material, can affect both sound amplitude and frequency, with higher coefficients potentially leading to sharper sounds. Environmental

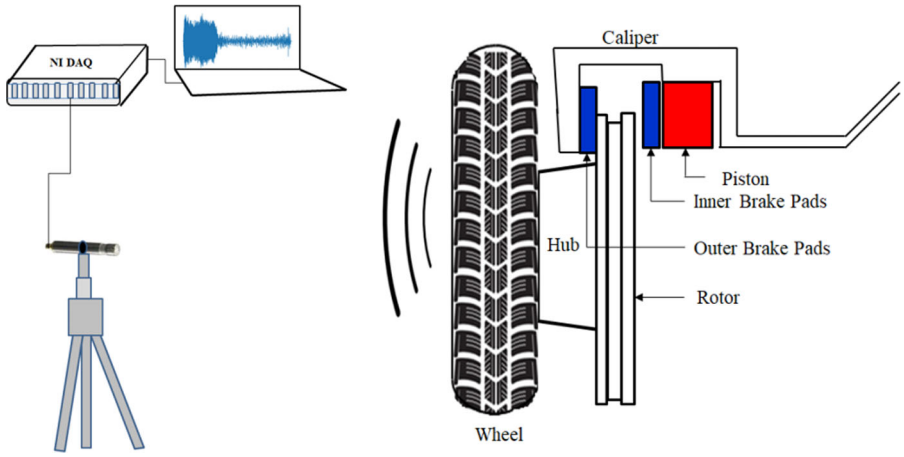

Figure 1. Pictorial representation of the experimental setup.

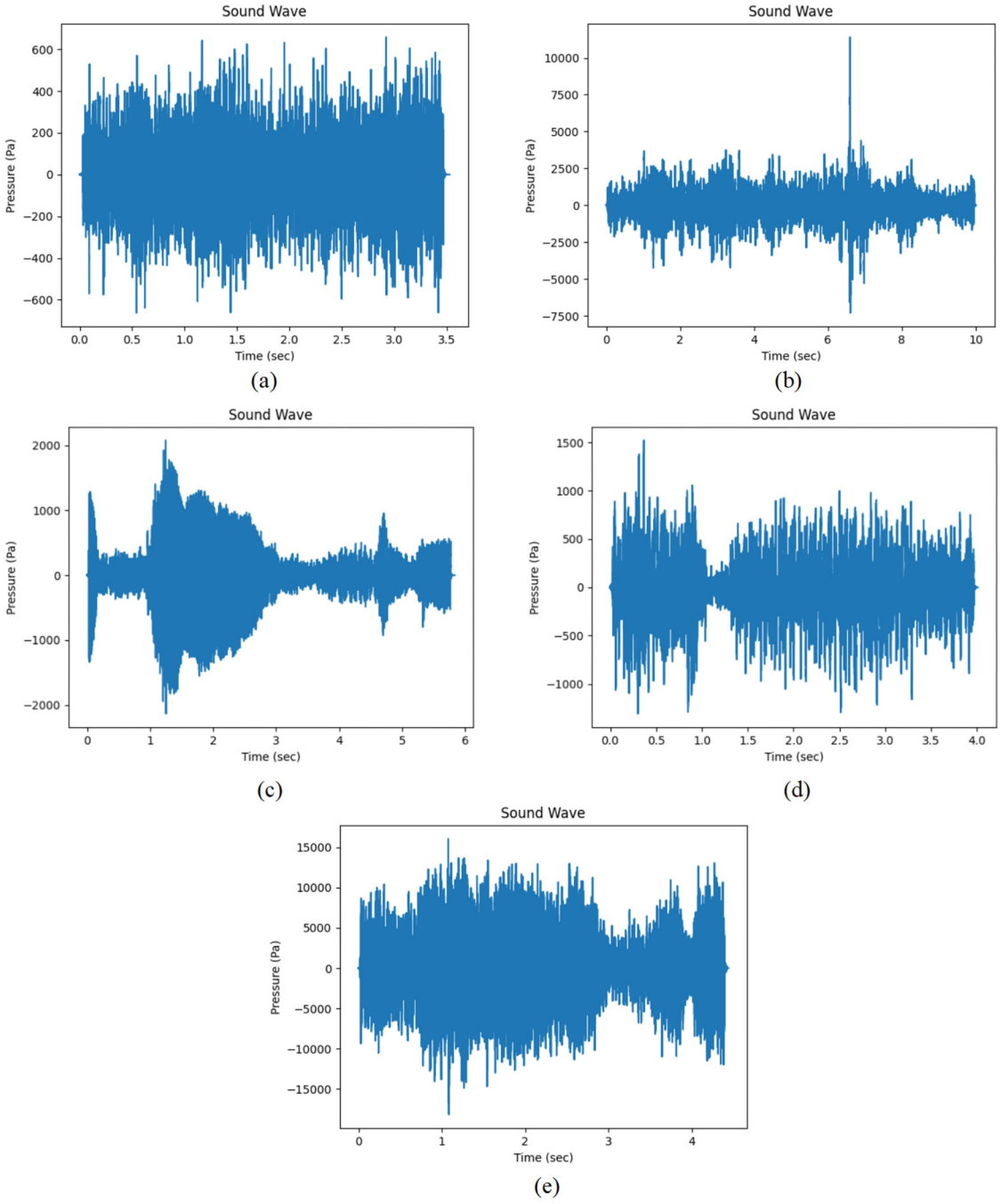

Figure 2. Acoustic signals for 5 classes: (a) Type I (b) Type II (c) Type III (d) Type IV (e) Type V.

factors such as tire type (summer or winter) and wheel dimensions might also influence sound propagation due to changes in vibration patterns and contact dynamics. Finally, background noise from the environment or other vehicle components could interfere with the clarity of the brake sound signals, making it harder to isolate meaningful data. These factors, while crucial for understanding real-world conditions, were beyond the scope of this study.

5. Proposed architecture

The proposed architecture, shown in Figure 3, consists of four stages. Stage 1 involves converting sound signals into images using a frequency-domain filter bank. Stage 2 comprises pre-processing techniques implemented on the image dataset, including randomly oversampling the dataset after separating the

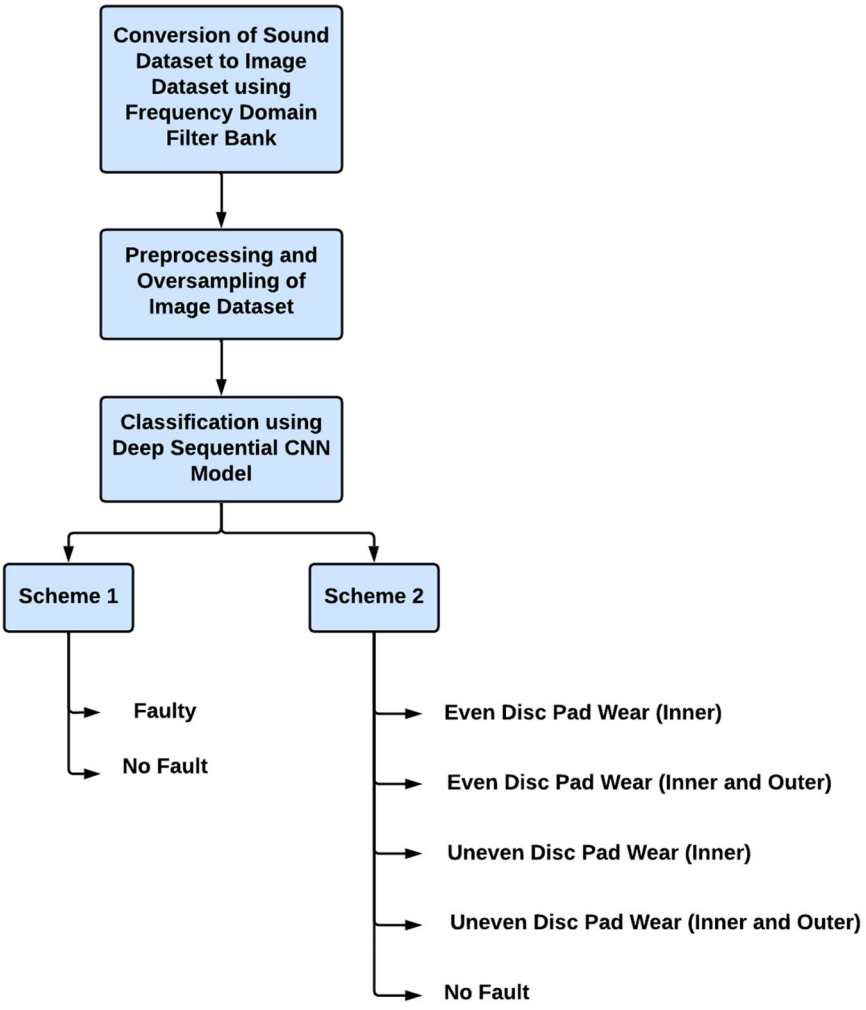

Figure 3. Block diagram of the proposed model.

data into folds or groups. The images are then passed on to stage 3, where they are randomly provided to the Deep Sequential Convolutional Neural Network (CNN) model as input. Finally, in stage 4, performance testing of the proposed model is performed based on binary classification (Scheme 1) as well as multi-class classification (Scheme 2). The first scheme involves the binary classification of the images into two classes: faulty and no-fault. Samples from all four fault classes are included in the faulty class category in the first scheme. In contrast, the second scheme, which is a multi-class classification, categorizes the images across five classes: Type I, Type II, Type III, Type IV, and Type V. Because 30 images were considered for each category, Scheme 1 encompasses 60 images together, whereas Scheme 2 has a total of 150 images. In the realm of signal processing, acoustic signals capture variations in air pressure caused by sound waves. These signals carry essential information regarding the frequency, amplitude, and duration of the sounds they represent.

5.1. Conversion of sound dataset to image dataset using frequency domain filter bank

Sound waves are intricate compositions of waves of various frequencies, each contributing a distinct pitch to the overall auditory experience. The sound signals present in the dataset are time-domain signals composed of various frequencies and amplitudes at each moment. Because human hearing is non-linear and is more sensitive to differences in low frequencies, the Mel scale is used to analyze acoustic signals. The Mel Scale is a perceptual scale of pitches that reflects how humans hear and interpret different frequencies. In other words, the Mel scale describes sound frequencies judged to be equidistant by human listeners. Unlike the linear scale of Hertz, the Mel scale is non-linear, emphasizing the perceptual

Figure 4. Flowchart of conversion of an acoustic signal to a mel spectrogram.

differences in pitch rather than the physical ones. The Mel scale is a logarithmic transformation of signal frequency. A standard formula is used to convert a frequency signal in Hertz (Hz) to the Mel scale, as shown in Eq. (1), where f is the frequency in Hz and $m(f)$ is the frequency in mels.

$$m(f) = \begin{cases} \dfrac{3f}{100}, & f < 1000 \\ 15 + 27\log_{6.4}\left(\dfrac{f}{1000}\right), & f \geq 1000 \end{cases} \tag{1}$$

A mel spectrogram is a frequency-domain filter bank that is a visual representation of the frequency content of a sound signal as it varies with time, using the mel scale to emphasize perceptually relevant pitch differences. It converts a traditional spectrogram into a format that aligns more closely with human auditory perception. By grouping frequencies based on the mel scale, mel spectrograms are valuable tools in speech and music processing, aiding tasks such as speech recognition and sound analysis. As depicted in Figure 4, a mel spectrogram is obtained by first applying a Fourier transform to a short window of the audio signal to convert it from the time to the frequency domain. Then, the resulting spectrum is transformed using a frequency-domain filter bank that mimics the non-linear frequency perception of the human ear according to the mel scale. Finally, the logarithm of the filter bank energies is taken to better match human auditory perception, producing a mel spectrogram that emphasizes relevant pitch differences for a more effective analysis in applications such as speech recognition and music processing. The original sound dataset is converted into an image dataset using a mel spectrogram. The resulting images of the Mel spectrogram graphs were converted to a standard size of $224 \times 224 \times 3$. The resizing of images helps decrease the computational effort required by the model. Figure 5 illustrates the image generated after passing the acoustic signals through the mel spectrogram module for each of the five classes considered.

5.2. Pre-processing and oversampling of dataset

The original sound dataset contains 30 sound signals from each of the five classes of multi-class classification (Scheme 2), which involves categorization across five classes, each consisting of 30 mel spectrograms. Binary classification (Scheme 1) is categorized into two classes: faulty and no-fault. Eight mel spectrograms were randomly selected from the Type II and Type III fault classes and seven spectrograms were chosen randomly from the remaining two classes. This implies that for Scheme 1, both classes comprise a total of 30 mel spectrogram images. Because the size of the dataset was small, a random oversampling technique was used to increase the number of mel spectrogram images used to train the CNN model. In random oversampling, a sample from the dataset is randomly selected, and a new sample is generated from it, which is added to the dataset. The samples are chosen randomly and replaced, meaning that the samples can be chosen and added to the dataset again (Mohammed et al., 2020). This technique involves duplicating the existing samples to augment the dataset, thereby balancing the number of samples across different classes. Specifically, random oversampling involves randomly replicating images from the original dataset and adding them to the training set. This process helps mitigate class imbalance by ensuring that each class has a sufficient number of samples for model training (Yang et al., 2022). This is one of the methods used to overcome the class imbalance problem. The dataset is split into folds or groups before oversampling because the images are duplicated, which ensures that the data used during testing are not used during training. The testing scheme used in this work is 5-fold cross validation; therefore, the dataset is divided into five folds or groups, each containing six image samples of each class. Each fold of the original dataset of size 6 is increased to 400, meaning that each class now has 2000 mel spectrograms in total across folds. Pseudocode 1 elucidates this process of pre-processing and oversampling of the dataset.

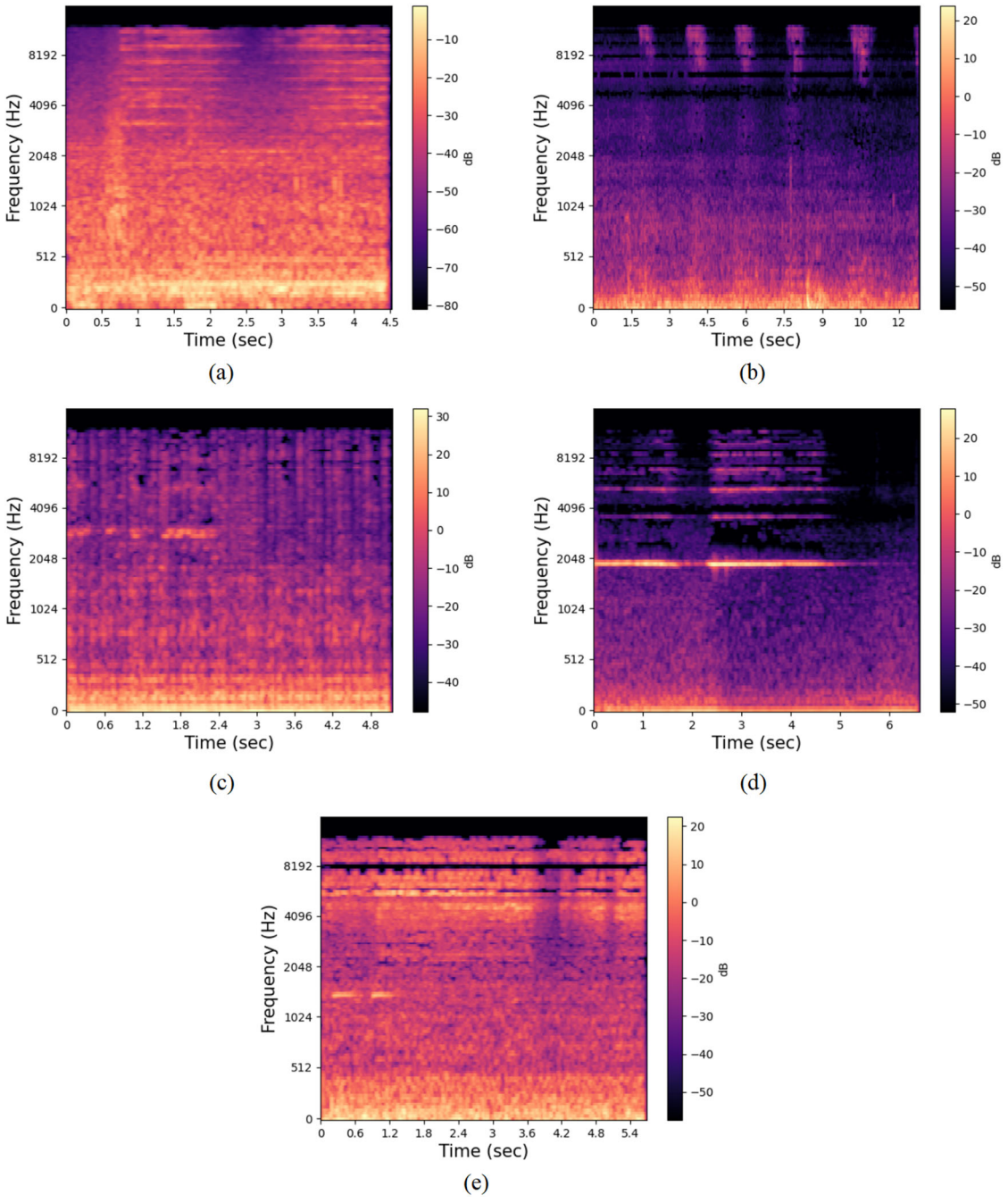

Figure 5. Mel Spectrogram generated for the 5 classes from the acoustic signals for (a) Type I (b) Type II (c) Type III (d) Type IV (e) Type V.

Pseudocode 1: Pre-processing and random oversampling of dataset

Input: Dataset with five classes (C1, C2, …, C5) containing 30 image instances.
Output: Final dataset with 5 folds (Fold 1, Fold 2, …, Fold 5) for 5-fold cross validation
Procedure:

1: **for** each class Ci of the dataset **do**
2: Each class Ci is divided into five groups (G1, G2, …, G5), each containing six imageinstances
3: **for** each group Gj of class Ci **do**
4: Randomly oversample the 6 instances to create 400 images in each group
5: **end for**
6: **end for**

7: **for** each class Ci **do**
8: for each group Gj do
9: Add the group Gj of Ci to fold Fj to form the final dataset
10: **end for**
11 **end for**

5.3. Classification using deep sequential CNN model

Convolutional Neural Networks (CNNs) are a class of deep neural networks developed for the analysis of structured grid data, particularly in contexts such as image processing. CNNs have been extensively used for image recognition and classification. The CNN learns object features and local patterns present in the images and is invariant to shifts in the position of the features. CNN can also reduce the adverse effects of noise. CNNs provide unmatched performance, but are computationally demanding. Convolutional Neural Network are typically composed of several different types of layers: convolution, Max Pooling, flattened, dense (Fully Connected) layer and dropout layers, as shown in Figure 6. In the first layer of the CNN model, the convolution layer is the heart and soul of the CNN, and each layer can have multiple filters with weighted parameters. The Convolution layer applies the convolution operation to the filters and a local group of pixels defined by the filter size. Convolution layers can learn the kernel filters that help extract the hierarchical features present anywhere in the input over a group of local pixels. The output is known as a feature map. The same weights and biases are used for each output filter. Hyper parameters, such as filter size, stride, and padding, vary depending on the task at hand.

The convolution operation is defined using Eq. (2) where h represents the convolution filter, $f(x,y)$ is the image pixel, and (k, l) is the filter size. The output of the convolution layer is fed to the Rectified Linear Unit activation function (ReLU). ReLU is a piecewise linear function that outputs the input directly if it is positive; otherwise, it outputs zero. It introduces nonlinearity to the feature map output, requires minimal computation, and is therefore preferred over other activation functions. The formula for calculating the ReLU function is given in Eq. (3), where x and $f(x)$ represents the input and output pixel values, respectively.

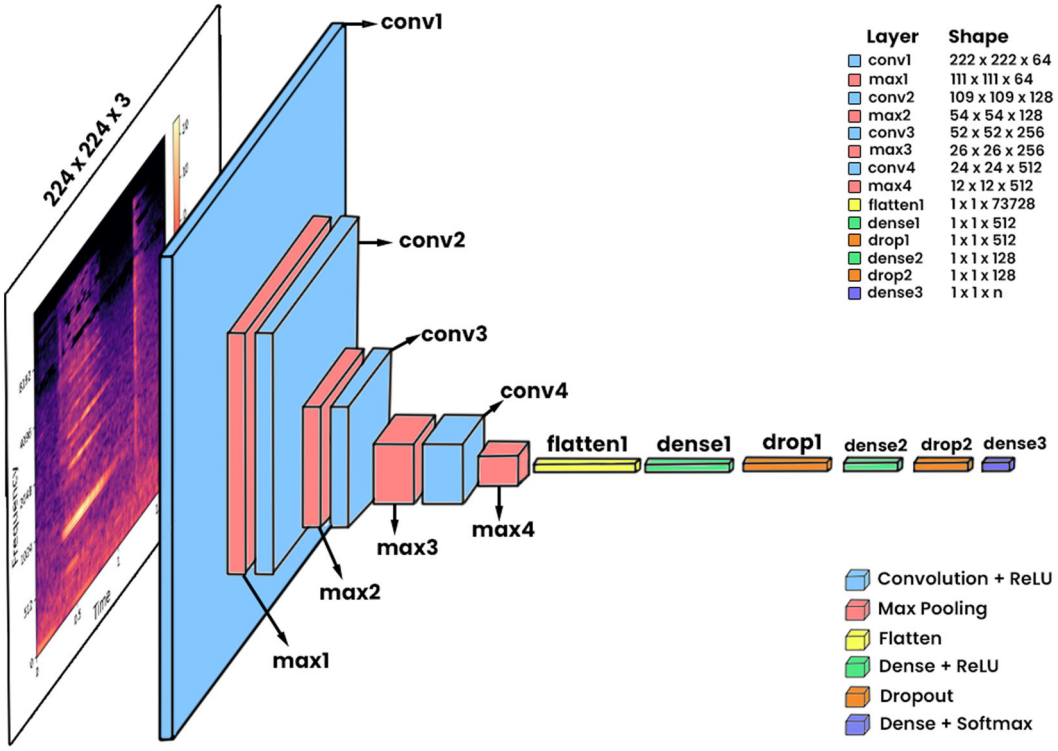

Figure 6. Layer-wise visualization of the proposed Deep Sequential CNN model.

$$g(x,y) = h * f(x,y) = \sum_{i=0}^{k} \sum_{j=0}^{l} f(x+i, y+j) * * h(i,j) \tag{2}$$

$$f(x) = \max(0, x) \tag{3}$$

The output of the activation function was then used as the input for the pooling layer. The key features present in the different neighborhoods of the feature maps are captured by the pooling layer, which applies a pooling operation of a specified pool size. This reduced the spatial dimensions (height and width) of the feature map. The Max Pooling layer was used in this study, and it calculates the maximum value present over a pooling window of size defined by the pool size for each input channel. The pooling window is then moved to the next patch of the feature map. The Convolution layer, together with the pooling layer, helps achieve shift invariance. The max-pooling operation was calculated using Eq. (4) where (*a, b*) represents the pool size, *f*(*x*, *y*) represents the value of the image pixel, and *h*(*x*, *y*) represents the output of the max pooling operation.

$$h(x,y) = \max_{i=0..a, j=0..b}(f(x+i, y+j)) \tag{4}$$

The flattened layer receives the feature map from the pooling layer as the input and flattens the multidimensional map into a 1-dimensional input for the dense layers of the ANN to process. This is the bridge between the convolutional and pooling layers of the CNN and the dense layers of the ANN. The 1-dimensional output array generated from the flattened layer was used as the input to the dense (or fully connected) layer. The neurons in the dense layer are connected to every neuron in the preceding layer. These layers can effectively grasp intricate patterns within the data owing to the interconnectedness among each neuron in the current layer and all neurons in the preceding layer. These layers are computationally expensive to train, and suffer from overfitting. The convolutional and pooling layers help mitigate these factors. The output of the dense layers is then passed to an activation function (generally a ReLU). The dropout layer randomly eliminates or nullifies the contributions of some neurons in the preceding layer during the training. The number of neurons or nodes that dropped out is defined by the drop rate. This is a regularization technique that helps reduce overfitting and improves the generalization error. The dropout layer is generally used after the dense layer, which makes the model more robust and less dependent on specific neurons. The last layer of the CNN model is a dense layer, with the number of nodes determined by the type of classification. The activation function used by this layer is the softmax activation function, which outputs a probability distribution over classes. The softmax function is the most suitable for mutually exclusive multi-class classification and is calculated using Eq. (5), where w_t represents the weight vector of the output layer's t^{th} neuron of the output layer and v represents the feature vector for the fully connected layer before it.

$$J(t) = \frac{e^{w_t v}}{\sum_{t=1}^{T} e^{w_t v}} \tag{5}$$

The CNN architecture suggested in this study is a deep sequential model with 14 layers. The dimensions of the input images were $224 \times 224 \times 3$. The model follows a sequential pattern of a 2D-Convolutional layer with a ReLU activation function, followed by a 2D-MaxPooling layer. This pattern was reproduced four times, after which a flattened layer was used to transform the resulting 3D array into a 1D array. The obtained 1D array serves as the input for a dense (fully connected) layer equipped with a ReLU activation function, followed by a dropout layer with a drop rate of 0.1. This sequence of layers was replicated. A dense layer with a softmax activation function serves as the last layer in the proposed model, and the number of nodes in this layer depends on the specific classification task. A batch size of 32 images was used during training to enable parallel processing and a faster and more generalized learning of the internal parameters of the model. The hyper parameters and their respective values are shown in Figure 7.

The proposed model uses the Adaptive Moment Estimation (or Adam) optimization algorithm, which is a variation of the stochastic gradient descent used to learn the optimal network weights and biases. Adam accelerates the optimization process by calculating an automatically adapted step size for each input parameter, based on the gradients for each variable. The default learning rate of 0.001 was used

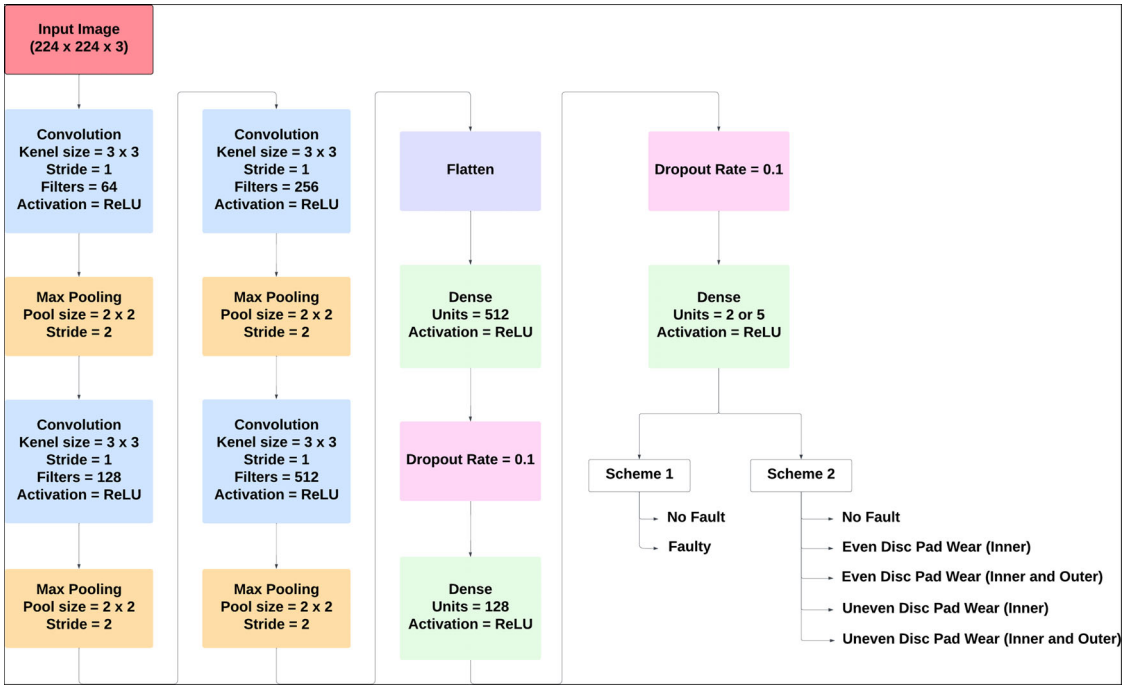

Figure 7. Hyper parameters and their respective values for the proposed Deep Sequential CNN model.

in this study. The Adam algorithm implements Eq. (6), where α represents the learning rate, $\widehat{v_t}$ is the sum of squares of past gradients at time t, ε is a small positive constant, $\widehat{m_t}$ is the sum of gradients at time t, and w_t and w_{t+1} are the weights at times t and $t+1$, respectively. The loss function used is the categorical cross-entropy loss function and is calculated using Eq. (7) where y_i denotes the true value of the i^{th} class label, $\widehat{y_i}$ is the predicted value of the i^{th} class label, and CE is the categorical cross-entropy loss function.

$$w_{t+1} = w_t - \widehat{m_t}\left(\frac{\alpha}{\sqrt{\widehat{v_t} + \varepsilon}}\right) \tag{6}$$

$$CE = -\sum_{i=1}^{N} y_i.\log\left(\widehat{y_i}\right) \tag{7}$$

Table 1 presents the essential hyperparameters for the proposed Deep Sequential CNN model for fault diagnosis in brake systems. As described in the table below, 2000 instances of each class category comprise the entire dataset. The images were resized to a resolution of $224 \times 224 \times 3$ before being passed on to the model for training and testing. The model used the Adam optimization algorithm with a learning rate of 10^{-3}. Learning of the model was achieved in three highlights in terms of epochs: 5, 10, and 20. A categorical cross-entropy loss function was used as the loss function for the model.

The selected hyper parameters for the model included an image resolution of $224 \times 224 \times 3$, batch size of 32, learning rate of 0.001, Adam optimizer, and categorical cross-entropy loss function. These parameters are widely recognized as standards in related research domains, providing a solid foundation for model performance. The number of epochs was set to 5, 10, and 20, as the model demonstrated optimal performance within this range, with the risk of overfitting increasing with additional training. The model architecture, composed of 14 layers, was chosen to maintain a lightweight design that reduced the computational overhead. To ensure that these parameters were optimally selected, the study employed a rigorous approach, including grid search, cross-validation, and empirical testing, which allowed for the fine-tuning of critical hyper parameters such as the learning rate, batch size, and number of layers. Extensive hyper parameter tuning is crucial for enhancing the model's performance in fault classification. Throughout the training and validation, accuracy and loss metrics were closely monitored, allowing for iterative refinement of the model, ultimately achieving the high accuracy reported in the results.

Table 1. Hyper-parameter table of the proposed deep Sequential CNN model.

Hyper parameters	Values
No Fault instances	2000
Faulty instances	2000
Type I instances	2000
Type II instances	2000
Type III instances	2000
Type IV instances	2000
Image resolution	224 x 224 x 3
Learning rate	10−3
Epochs	5, 10, 20
Optimizer	Adaptive moment estimation (Adam)
Loss function	Categorical cross-entropy loss function

6. Results and discussion

This study focused on the analysis of five distinct types of brake sound issues encountered in automobiles using a deep CNN framework. The research not only delved into the identification of these sound anomalies but also introduced a robust evaluation methodology that incorporates both hold-out and 5-fold cross-validation approaches. The experiment consistently utilized a batch size of 32, emphasizing the importance of uniformity in model testing. The analysis was performed for both the class classification and five-class classification of brake sounds. In each case, three epoch sizes (5, 10, and 20) were considered. Performance metrics such as precision, recall, F1 score, ROC (receiver operating characteristic (ROC) curve, and AUC (area under the ROC curve) values were employed to gauge the accuracy of the model (Prusty et al., 2017). The training accuracy and loss curves for the proposed model depicted in Figure 8 consistently show rapid learning, achieving 0.95 accuracy within the first 5–10 epochs. The training loss converged, and the validation loss decreased, highlighting the model's effective adaptation and generalization to new data. These findings confirm the proficiency and correctness of the proposed model in capturing intricate patterns during the training phase. The efficacy of the model was initially assessed using hold-out cross-validation for binary classification tasks, which revealed substantial improvements with increasing epochs. Notably, at 20 epochs, the model achieved perfection across all metrics considered, as shown in Table 2. The evaluation was then extended to hold-out cross-validation for a 5-class classification task, where the model exhibited commendable performance with an accuracy of 1.0 for all epochs.

Figure 9 presents the confusion matrix plots for the proposed model, trained using hold-out cross-validation with a sample size of 2000 for each class in the binary and multi-class scenarios. The model was trained on 10,000 spectrograms across epochs of 5, 10, and 20. This approach aimed to ensure that the predictions closely aligned with the average value, and these predicted values exhibited minimal variance. Testing was performed with 800 images for two-class classification and 2000 images for five-class classification, ensuring a robust evaluation across each class label. Across all scenarios, the observed accuracy consistently surpassed 0.98, affirming the high performance and robustness in classifying diverse instances. Figure 9 shows the performance of the proposed model using the metrics from the confusion matrix. A curve closer to the Y-axis indicates superior model performance. As shown in Figure 10, the ROC curve for the binary classification achieved a score of 1.0. This high value underscores the reliability and robustness of the model for distinguishing between positive and negative instances. In the context of multi-class classification, the ROC curve attains a perfect score of one, emphasizing the model's exceptional performance across multiple classes. These results confirm the model's effectiveness in classification tasks and exhibit highly reliable and robust performance.

A k-fold cross-validation was carried out to check the robustness of the model. In this approach, k-fold cross-validation involves the random division of the dataset into k subsets. One subset was designated as the test set, whereas the remaining subsets collectively constituted the training set. The model was trained on the training set and assessed on the test set. This process was repeated until each subset was used as the test set. In our analysis, k is considered to be 5. This iterative method offers a systematic assessment of the performance of the model across various training test splits. Cross-validation is generally preferred because of its capability to expose the model to multiple train-test scenarios, thereby enhancing our comprehension of its performance on unseen data. In contrast, hold-out validation relies

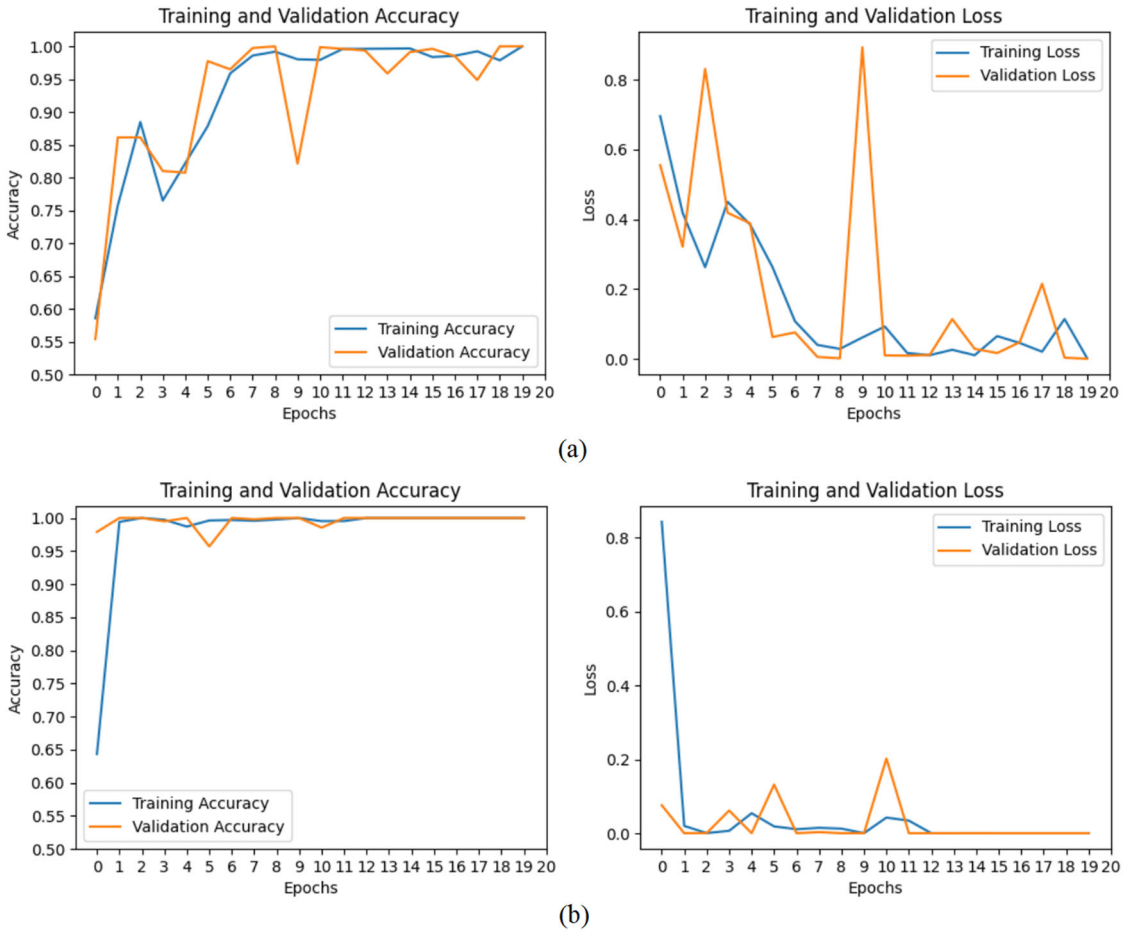

Figure 8. Training and validation accuracy and loss curves of proposed model for 20 epochs for (a) Binary classification (b) Multi-class classification.

Table 2. Performance of proposed deep sequential CNN model using hold-out cross-validation.

	Binary classification					5- class classification				
No of epochs	Accuracy	Precision	Recall	F1 score	AUC	Accuracy	Precision	Recall	F1 score	AUC
5	0.989	0.99	0.99	0.99	0.99	1.0	1.0	1.0	1.0	1.0
10	0.983	0.98	0.98	0.98	0.98	1.0	1.0	1.0	1.0	1.0
20	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

on a single train-test split, making its evaluation score contingent on a specific data partition. Subsequently, the assessment encompassed 5-fold cross-validation for binary classification, consistently showing high-performance metrics. At 20 epochs, the model attained optimal scores for all metrics considered, as shown in Table 3. The evaluation was further extended to a 5-class classification task, which revealed noteworthy results. The accuracy of the model peaked at 0.978 for both 10 and 20 epochs, along with high precision, recall, F1 score, and AUC values (Table 3).

Table 4 provides a comparative analysis of the various models for the detection and identification of brake faults. The existing models, employing both binary and multi-class classification schemes, demonstrated high accuracy, ranging from 0.72 to 0.99, across various cross-validation (CV) methods, including hold-out and 10-fold validation. Other models such as AIRS, DSmT fusion, and SVM classifiers also showed competitive accuracy rates. Notably, the CNN Model with multirate data achieved an impressive accuracy of 0.9997 for multi-class classification. However, some models, including random forest and fuzzy logic, exhibited lower accuracy rates. This comprehensive overview highlights the diverse performances of models in identifying and detecting break faults.

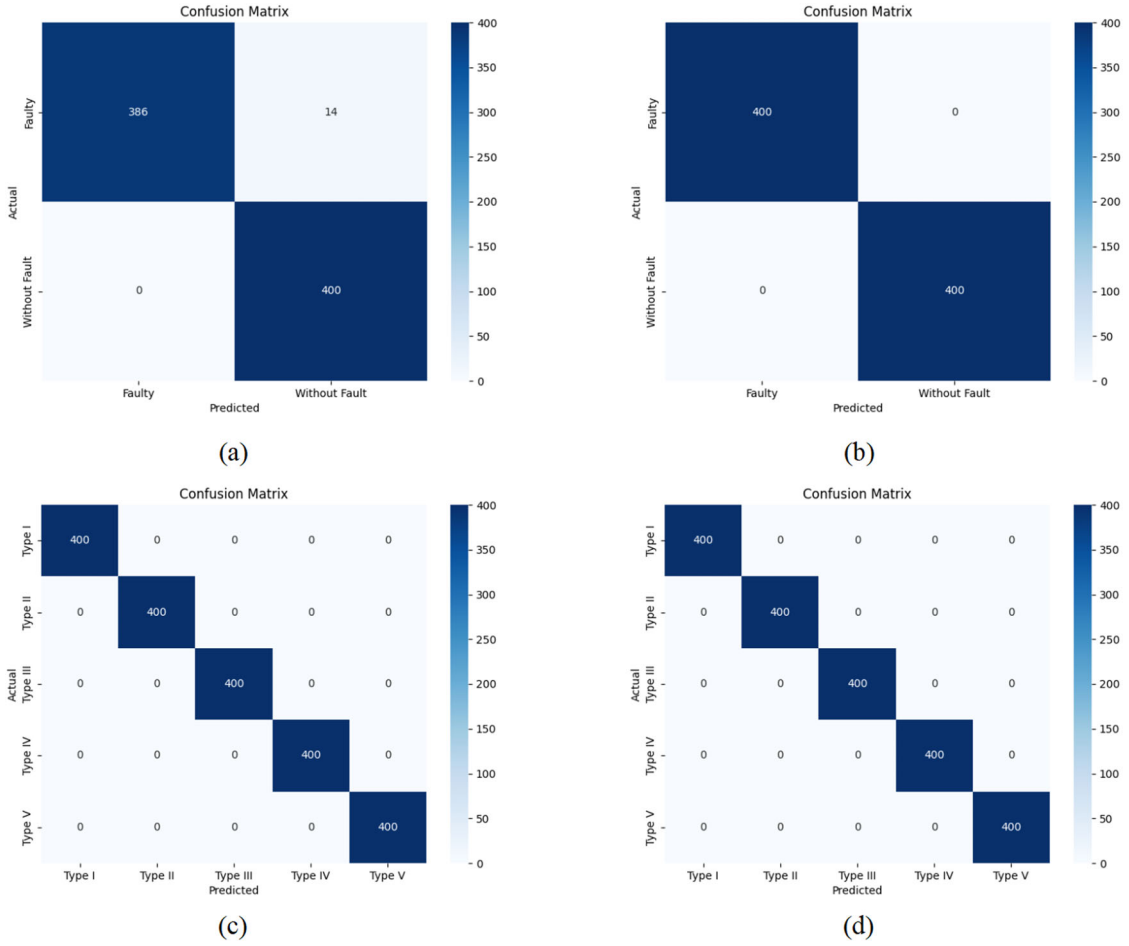

Figure 9. Confusion matrix of (a) 10 epochs for binary classification; (b) 20 epochs for binary classification; (c) 10 epochs for multi-class classification; (d) 20 epochs for multi-class classification.

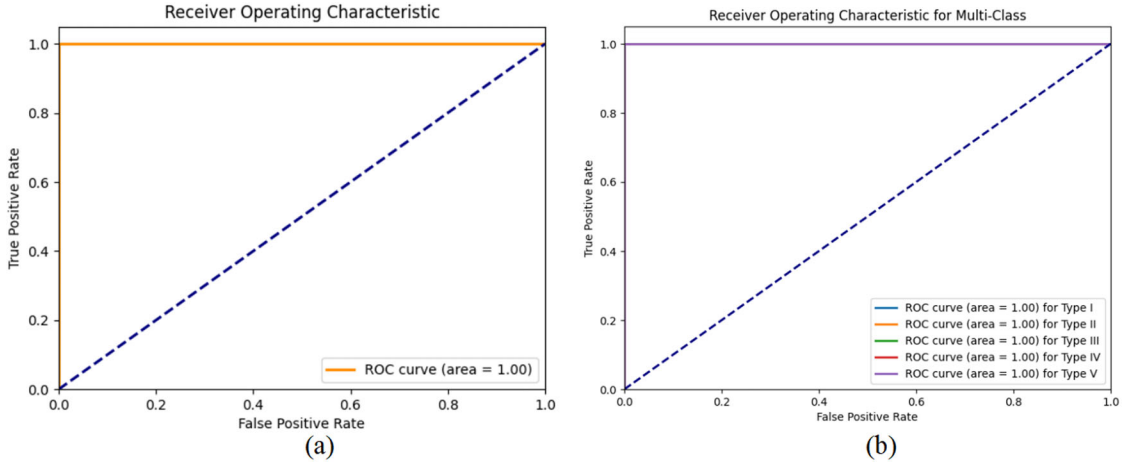

Figure 10. ROC curve of the proposed CNN model for 20 epochs for (a) Binary classification (b) Multi-class classification.

Table 3. Performance of proposed deep sequential CNN model using 5 fold cross-validation.

	Binary classification					5-Class classification				
No of epochs	Accuracy	Precision	Recall	F1 score	AUC	Accuracy	Precision	Recall	F1 score	AUC
5	0.963	0.966	0.964	0.962	0.962	0.975	0.982	0.976	0.972	0.982
10	0.976	0.978	0.976	0.974	0.974	0.978	0.986	0.978	0.976	0.983
20	0.977	0.98	0.976	0.974	0.974	0.978	0.986	0.978	0.976	0.983

Table 4. Comparison of existing models to identify and detect break faults.

	Model	Classification scheme	Accuracy	CV
	Proposed model	Binary	100.0%	Hold-out
		Binary	97.70%	5-fold
		Multi-class	100.0%	Hold-out
		Multi-class	97.80%	5-fold
[6]	AIRS 1	Multi-class	96.91%	–
	AIRS 2	Multi-class	97.82%	–
	AIRS Parallel	Multi-class	96.36%	–
[15]	DSmT Fusion	Multi-class	98.10%	–
	RF	Multi-class	73.30%	–
	CNN	Multi-class	69.90%	–
	LSTM	Multi-class	65.50%	–
[13]	SVM	Binary	95.32%	–
	CNN model with multirate data	Binary	98.12%	–
	SVM	Multi-class	91.40%	–
	CNN model with multirate data	Multi-class	99.97%	–
[19]	C4.5 decision tree	Multi-class	97.45%	–
	C-SVM	Multi-class	98.72%	–
	Nu-SVM	Multi-class	98.36%	–
[10]	Random forest	Multi-class	83.49%	–
	Random tree	Multi-class	75.88%	–
	LMT Classifier	Multi-class	80.33%	–
	J48 Classifier	Multi-class	79.59%	–
[11]	SVM Classifier	Multi-class	92.75%	–
[16]	Rough Set	Multi-class	96.90%	10-fold
	FRNN	Multi-class	95.09%	10-fold
[17]	Multi-class SVM Classifier	Multi-class	81.40%	Hold-out
	Multi-output SVM Classifier	Multi-label	97.35%	Hold-out
[20]	Fuzzy Logic	Multi-class	72.08%	–
	Bayesian Network	Multi-class	80.34%	–

The proposed model exhibited exceptional performance, achieving 0.977 accuracy in binary classification and 0.978 in multi-class classification using deep sequential CNNs, demonstrating its robustness and effectiveness. In contrast, traditional models, such as Random Forest and Fuzzy Logic, show lower accuracy, indicating that they may not capture complex fault patterns effectively. Among other methods, SVM models and CNNs with multi rate data achieve high accuracy, with CNNs in particular showing superior results in multi-class classification. Models such as AIRS and DSmT Fusion also perform well, but lack detailed cross-validation data, which is crucial for assessing their reliability. Overall, advanced deep-learning approaches, particularly CNNs, provide accurate and reliable results for brake fault detection, highlighting their significant advantages over traditional methods.

7. Conclusion and future work

In conclusion, this study addresses the critical safety concerns of brake failure in the automotive industry, emphasizing its potential catastrophic consequences. This study identified various factors contributing to brake failure, such as mechanical wear, hydraulic system malfunctions, and electronic control failures. Brake fluid leakage, worn-out brake pads, and damaged brake lines have been identified as common culprits, emphasizing the importance of effective fault detection and prevention measures. This project introduced a pioneering approach to the detection and classification of faults in hydraulic brake systems, specifically focusing on the analysis of five distinct types of brake sound issues encountered in automobiles. By leveraging a CNN sequential model, this study revolutionized the traditional methods of brake system monitoring. The approach demonstrated exceptional results, achieving remarkable 97.7% accuracy in binary classification and 97.8% accuracy in 5-class classification.

Furthermore, this research explored the impact of sample size on classification accuracy, revealing the significance of larger sample sizes in achieving more accurate results. The robustness of the model was demonstrated through rigorous testing across varying numbers of epochs. The results consistently showed high-performance metrics, underscoring the model's effectiveness in both binary and multi-class classification scenarios. The proposed approach not only offers a novel solution to fault diagnosis but also demonstrates the potential for broader applications in the automotive industry. As brake failure

remains a critical safety concern, the findings of this study contribute significantly to ongoing efforts to improve road safety and prevent catastrophic incidents associated with brake system malfunctions.

The primary limitation of the study is the modest size of the dataset, comprising only 30 samples for each class. In machine learning, larger datasets often contribute to enhanced model performance and generalizability. Future validations should prioritize expanding the dataset to encompass a more comprehensive array of sound signals for each class, thereby ensuring a more robust and adaptable model. Another aspect to consider is the oversampling technique employed, which involves the random duplication of data samples for class balance. Although effective to some extent, random oversampling has the potential drawback of overfitting. Alternative methods to oversample these data can be explored to address the risk of overfitting and increase the model's overall stability. Moreover, this study could benefit from exploring advanced techniques such as Generative Adversarial Networks (GANs). By generating synthetic samples based on the original data, GANs offer a pathway for further enriching and diversifying the dataset. These nuanced considerations underscore the importance of ongoing refinement in model development, ensuring its applicability and effectiveness across diverse sound-signal scenarios.

To reduce the complexity of this pilot study, certain factors that may influence the sound samples were intentionally excluded from consideration. Data collection was conducted in a controlled environmental setup, which did not account for variables such as brake type, brake pad contact size and shape, surface conditions, friction coefficient, tire types, wheel dimensions, or background noise. These factors can affect the characteristics of brake sounds and could have a significant impact on their behavior in real-world scenarios. Incorporating these variables in future research would provide a more comprehensive understanding of their influence and enhance the generalizability and applicability of the proposed methodology.

The success of this research in addressing brake failure concerns and implementing a pioneering fault-diagnosis approach opens avenues for future work. One promising direction is the practical implementation of a Convolutional Neural Network (CNN) sequential model in automobile systems. Integrating the developed model into real-time monitoring systems can enhance the safety and reliability of hydraulic braking systems on roads. Moreover, efforts can be directed toward improving the efficiency of the model by exploring advanced neural network architectures, fine-tuning hyperparameters, and incorporating additional features that may further enhance the accuracy of fault detection. Continuous refinement and optimization of the model will contribute to its adaptability to diverse real-world scenarios.

Acknowledgement

The authors would like to thank the Vellore Institute of Technology, Chennai, for providing the infrastructure and motivation to conduct this research.

Authors' contribution

Saumye Saran Das - Data Curation, Implementation and Manuscript writing; Raushan Kumar - Data Curation, Implementation and Manuscript writing; Manas Ranjan Prusty - Conceptualization, Methodology, Supervision; Tapan Kumar Mahanta - Methodology and Manuscript Correction; Subhra Rani Patra - Validation and Visualization; All authors have read and approved the final version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study received no funding.

About the authors

Saumye Saran Das has completed his Bachelors in Technology from the School of Computer Science and Engineering at the Vellore Institute of Technology, Chennai, India. His research interest includes machine learning, IoT and Signal Processing.

Raushan Kumar has completed his Bachelors in Technology from the School of Computer Science and Engineering at the Vellore Institute of Technology, Chennai, India. His research interest includes machine learning, deep learning and image Processing.

Manas Ranjan Prusty is working as Associate Professor in the Research Centre for Cyber Physical Systems at the Vellore Institute of Technology, Chennai, India. His research interest includes deep learning, image processing and signal processing.

Tapan K. Mahanta is working as Assistant Professor in the School of Mechanical Engineering at the Vellore Institute of Technology, Chennai, India. His research interest includes vibration, acoustic and composite materials.

Subhra Rani Patra is working as Clinical Assistant Professor in the University of Texas, Arlington USA. Her research interest includes machine learning, image processing and natural language processing.

Data availability statement

The data used for this research will be made available on reasonable request to Tapan Kumar Mahanta at tapan.mahanta@vit.ac.in.

References

An, X. L., Zhou, J. Z., Liu, L., Yang, J. J., Li, C. S., & Xiang, X. Q. (2008). Vibration fault diagnosis for hydraulic generator units with pattern recognition and cluster analysis. In *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, Dalian, China, 2008, (pp. 1–4) IEEE. https://doi.org/10.1109/WiCom.2008.3037

An, X., Zhou, J., Liu, L., Fang, R., Xiang, X., & Peng, B. (2007). Vibration fault diagnosis of hydraulic generator units based on pattern recognition and grey incidence analysis. In *2007 Second International Conference on Bio-Inspired Computing: Theories and Applications* Zhengzhou, 2007, (pp. 48–51). IEEE. 10.1109/BICTA.2007.4806416

Bazhenov, Y., & Amirseyidov, S. (2019). Statistical analysis of road accidents associated with technical condition of brake systems. *Journal of Physics: Conference Series*, *1177*, 012026. https://doi.org/10.1088/1742-6596/1177/1/012026

Chen, Y. (2011). Applications of Bayesian network in fault diagnosis of braking system. In *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, Hangzhou, China, 2011 (pp. 234–237). IEEE. https://doi.org/10.1109/IHMSC.2011.63

Elliott, M. A., Baughan, C. J., & Sexton, B. F. (2007). Errors and violations in relation to motorcyclists' crash risk. *Accident Analysis and Prevention*, *39*(3), 491–499. https://doi.org/10.1016/j.aap.2006.08.012

Gao, Q., & Xiang, J. (2017). A Walsh transform method to detect faults in mechanical systems. In *2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, Shanghai, China, 2017 (pp. 613–616). IEEE. https://doi.org/10.1109/SDPC.2017.120

Huang, K., Wu, S., Li, F., Yang, C., & Gui, W. (2022). Fault diagnosis of hydraulic systems based on deep learning model with multirate data samples. *Ieee Transactions on Neural Networks and Learning Systems*, *33*(11), 6789–6801. https://doi.org/10.1109/TNNLS.2021.3083401

Jayakrishnan J., Alamelu Manghai, T. M., & Jegadeeshwaran, R., (2020). Real-time condition monitoring on brakes using machine learning techniques. In *2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati*, India, 2020 IEEE. (pp. 1–6). https://doi.org/10.1109/AISP48273.2020.9073173

Jegadeeshwaran, R., & Sugumaran, V. (2015). Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. *Mechanical System Signal Processing*, *52-53*, 436–446. https://doi.org/10.1016/j.ymssp.2014.08.007

Ji, X., Ren, Y., Tang, H., & Xiang, J. (2021). DSmT-based three-layer method using multi-classifier to detect faults in hydraulic systems. *Mechanical System. Signal Processing*, *153*, 107513. https://doi.org/10.1016/j.ymssp.2020.107513

Li, Z., Wang, H., Chen, J., Zhou, Z., & Chen, W. (2023). Research on rolling bearing fault diagnosis based on DRS frequency spectrum image and deep learning. *The International Journal of Acoustics and Vibration*, *28*(2), 211–219. https://doi.org/10.20855/ijav.2023.28.21942

Lian, R., Xu, Z., & Lu, J. Nov (2013). Online fault diagnosis for hydraulic disc brake system using feature extracted from model and an SVM classifier. In *2013 Chinese Automation Congress*, Changsha, China, 2013 (pp. 228–232). IEEE. https://doi.org/10.1109/CAC.2013.6775733

Liu, Y., Chen, Z., Wei, L., Wang, X., & Li, L. (2023). Braking sensor and actuator fault diagnosis with combined model-based and data-driven pressure estimation methods. *Ieee Transactions on Industrial Electronics*, *70*(11), 11639–11648. https://doi.org/10.1109/TIE.2022.3231287

Alamelu, M. T. M., Jegadeeshwaran, R., Sakthivel, G., S., & D., I. R. (2020). Vibration-based fault detection in a hydraulic brake system using artificial immune recognition system with statistical features. In *2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)*, Chennai, India, 2020 (pp. 190–195). IEEE. https://doi.org/10.1109/iSES50453.2020.00049

Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020) "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results".

National Motor Vehicle Crash Causation Survey (NMVCCS). National Highway Traffic Safety Administration, May 01 Accessed: Aug. 23, 2024 [Online]. Available: https://catalog.data.gov/dataset/national-motor-vehicle-crash-causation-survey-nmvccs

Nemade, A. W., Telang, S. A., & Chel, A. L. (2019) "Assessment of brake failure in leading small cars on indian roads and it's critical investigation".

Prusty, M. R., Jayanthi, T., & Velusamy, K. (2017). Weighted-SMOTE: A modification to SMOTE for event classification in sodium cooled fast reactors. *Progress in Nuclear Energy*, *100*, 355–364. https://doi.org/10.1016/j.pnucene.2017.07.015

Road traffic injuries. *Accessed: Mar*, *01* [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Siyuan, L., Linlin, D., & Wanlu, J. (2011). Study on application of principal component analysis to fault detection in hydraulic pump. In *Proceedings of 2011 International Conference on Fluid Power and Mechatronics*, Beijing, China, 2011 (pp. 173–178). IEEE. https://doi.org/10.1109/FPM.2011.6045752

Stewart, T. (2021) . Overview of Motor Vehicle Traffic Crashes in.

T M, A. M., J, R., S, G., R, S., & S. K, D. (2019). Condition monitoring of hydraulic brake system using rough set theory and FUZZY rough nearest neighbor learning algorithms. In *2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), Rourkela, India, 2019* (pp. 229–232). IEEE. https://doi.org/10.1109/iSES47678.2019.00058

Xu, Z., Yu, H., Tao, J., & Liu, C. (2021). Compound fault diagnosis in hydraulic system with multi-output SVM. In *CSAA/IET International Conference on Aircraft Utility Systems (AUS 2020)*,. (pp. 84–89). Institution of Engineering and Technology. https://doi.org/10.1049/icp.2021.0470

Yang, W., Pan, C., & Zhang, Y. (2022). An oversampling method for imbalanced data based on spatial distribution of minority samples SD-KMSMOTE. *Scientific Reports*, *12*(1), 16820. https://doi.org/10.1038/s41598-022-21046-1

You, Z., & Lu, C. Aug (2018). A heuristic fault diagnosis approach for electro-hydraulic control system based on hybrid particle swarm optimization and Levenberg–Marquardt algorithm. *Journal of Ambient Intelligence and Humanized Computing*. *14*, 14873–14882. https://doi.org/10.1007/s12652-018-0962-5