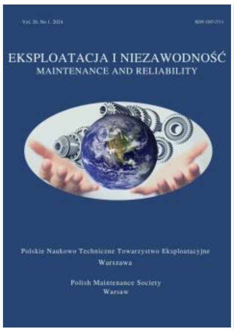


Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 26 (2024), Issue 1

journal homepage: http://www.ein.org.pl

Article citation info:
Balaji P A, Venkatesh S N, V S, Fault Diagnosis of Suspension System Based on Spectrogram Image and Vision Transformer, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2024: 26(1) http://doi.org/10.17531/ein/174860

Fault Diagnosis of Suspension System Based on Spectrogram Image and Vision Transformer

Indexed by: Web of Science Group

Arun Balaji P[a], Naveen Venkatesh S[a], Sugumaran V[a,*]

[a] School of Mechanical Engineering (SMEC), Vellore Institute of Technology, Chennai - 600127, India

Highlights

- This study suggests using image vision for dynamic signal (vibration) pattern recognition.
- By using spectrogram images as input, the model captures both temporal and frequency components for precise fault identification.
- This study introduces a deep learning model for diagnosing multiple faults in automobile suspension systems, addressing a gap in suspension system fault diagnosis.

Abstract

The suspension system in an automobile is essential for comfort and control. Implementing a monitoring system is crucial to ensure proper function, prevent accidents, maintain performance, and reduce both downtime and costs. Traditionally, diagnosing faults in suspension systems has relied on specialized setups and vibration analysis. The conventional approach typically involves either wavelet analysis or a machine learning approach. While these methods are effective, they often demand specialized expertise and time consumable. Alternatively, using deep learning for suspension system fault diagnosis enables faster and more precise real-time fault detection. This study explores the use of vision transformers as an innovative approach to fault diagnosis in suspension systems, utilizing spectrogram images. The process involves extracting spectrogram images from vibration signals, which serve as inputs for the vision transformer model. The test results demonstrate that the proposed fault diagnosis system achieves an impressive accuracy rate of 98.12% in identifying faults.

Keywords
condition monitoring, vision transformer, fault diagnosis, image generation, McPherson suspension system

1. Introduction

Urbanization has transformed transportation into an integral aspect of human life, where people depend on diverse modes of commuting. Cars are the preferred choice among these modes owing to their safety, comfort, and convenience. In the modern era, conventional cars possess the capacity to reach or even exceed speeds of 180 km/h However, when a vehicle is not adequately maintained, traveling at high speeds can result in severe consequences, as the majority of reported road accidents occur because of a lack of control over direction or braking. The suspension system plays a crucial role in maintaining lateral and longitudinal stability because it is interconnected with the steering system, which ensures that the vehicle maintains consistent contact and pressure on the road (1). Any failure in the components of the suspension system can directly affect the performance of the braking and steering systems, leading to a 12% increase in braking time and a 30% increase in braking distance. Consequently, these factors can contribute significantly to the occurrence of potentially fatal accidents. Additionally, when cars are in motion, they are subjected to various forces, such as acceleration, braking, road disturbances, and centrifugal force during cornering. These forces can cause discomfort to car

(*) Corresponding author.
E-mail addresses: A. Balaji P (ORCID: 0000-0001-5008-3252) arunbalaji707@gmail.com, N. Venkatesh S (ORCID: 0000-0002-4034-8859) naveenvenkatesh.s@vit.ac.in, S. V (ORCID: 0000-0002-5323-6418) v_sugu@yahoo.com

occupants and reduce the overall maneuverability of a vehicle (2). A passive suspension system, consisting of springs and dampers, helps suppress and dissipate these unwanted forces, converting them into heat. Furthermore, the extended usage of a vehicle can result in gradual deterioration of suspension components, ultimately leading to suspension failure. Therefore, it is of utmost importance to maintain the functionality of suspension systems to guarantee a safe and comfortable driving experience. Although semi-active and active suspension systems offer dependable real-time monitoring, their high cost and the need for additional control systems and actuators make them impractical for most automobiles (3–5). Consequently, passive suspension systems are extensively utilized due to their uncomplicated structure, reliable performance, and cost-effectiveness.

The McPherson type suspension system is preferred over the double wishbone type due to its simplicity, lightweight, and cost-effectiveness. It consists of components like struts, ball joints, tie rods, and lower control arms, which can wear out over time, especially when exposed to varying road conditions and loads. Factors like wear, lack of lubrication, misalignment, heavy loads, mishandling, improper installation, and corrosion can increase the chances of faults. Detecting these faults early is crucial to maintain suspension performance, minimize maintenance disruptions, and prevent potentially dangerous accidents. Therefore, fault diagnosis is essential for ensuring safety, reliability, and comfort in vehicle operation.

Various techniques have been developed for fault diagnosis, including knowledge-based, data-driven (signal-based), analytical modelling (model-based), and hybrid techniques. Among these techniques, data-driven methods are widely employed owing to their capability to operate in real time. The data or signals acquired during the data-acquisition process display distinct signature patterns for particular fault conditions, enabling effective classification. Parameters such as the vibration, pressure, load, and displacement provide valuable information regarding the state of the suspension system. following Table 1 compares state of art suspension system fault diagnosis.

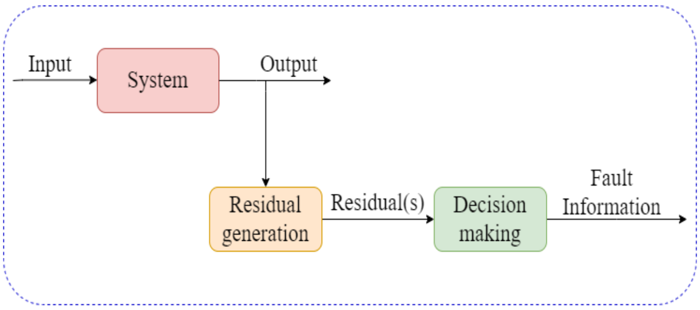


Fig. 1. Fault Diagnosis using data driven approach.

Table 1. Experimental comparison of state of art suspension fault diagnosis study.

S.no	Method	Approach	component	Reference
1	Wavelet analysis	Model based approach	Damper	(6)
2	Acceleration transmissibility method and chamber Pressure difference method	Data driven approach	Damper	(7)
3	Numerical method	Model based approach	Tyre	(8)
4	Difference in Wheel speed	Data driven approach	Tyre	(9)
5	Numerical and finite element analysis method	Model based approach	Ball joint	(10)
6	Time and frequency domain analysis using short term Fourier transform (STFT)Short term Fourier transformer	Data driven approach	Ball joint	(11)
7	Finite element analysis	Model based approach	Strut mount	(12)
6	Proposed method (uses vibration signal acquired from the single sensor to detect multiple faults)	Data driven approach (online)	Lower arm bush fault, lower arm ball joint, tie rod ball joint, strut, strut mount fault and tyre low pressure	

From Table 1, one can ascertain he vibration-based data-driven approach has garnered significant attention in fault diagnosis of automobile system for compelling reasons. Firstly, faults often manifest distinct patterns in vibration signals, making it an effective method for their detection. Moreover, vibration signal acquisition demonstrates remarkable sensitivity, enabling the identification of even minor deviations. Additionally, vibration signals exhibit a higher signal-to-noise ratio compared to acoustic emission, rendering them more valuable for fault diagnosis. Lastly, advancements in integrated circuit technology have enhanced the reliability and cost-effectiveness of accelerometer sensors, further bolstering the practicality of this approach. Table 2 illustrates the research work carried out using vibration-based fault diagnosis in the field of automotive technology in recent times.

Table 2. Research works on automotive technology using vibration-based methods.

Component / System	Techniques	Reference
Ball joint	Transmissibility analysis	(13)
	Inertial sensor with Short term Fourier transformer	(14)
Damper and Tire	Parameter estimation	(15)
Damper	Failure mode analysis	(16)
	Latter Analysis with ANN	(17)
Transmission System	Particle swam optimization with Back propagation approach	(18)
Spring	Average Correlation Signals based Stochastic Subspace Identification (ACS-SSI)	(19)
Suspension system	Bayes learning algorithm with J48 algorithm	(20)
Tyre pressure monitoring system	Dominance-based rough set approach	(21)
	Logistic model tree with J48 algorithm	(22)

From the literature survey, it was understood that most fault diagnosis studies use a data-driven approach, which comprises three key phases: data acquisition, signal processing, and decision-making. During the data acquisition phase, sensors such as accelerometers, ammeters, tachometers, and acoustic sensors collect machine condition data. These signals undergo signal processing techniques, including fast Fourier transform, independent component analysis, and principal component analysis, to extract valuable information.

While these studies have demonstrated the effectiveness of mathematical and machine learning models, they have limitations. Mathematical models can be complex and sensitive to parameter variations. Rigorous feature selection and extraction processes add complexity to machine learning models. Conventional fault diagnosis techniques relying on signal processing and human expertise face challenges. Thus, recent research has focused on deep learning-based data-driven approaches, as they can acquire knowledge without explicit feature extraction. Vibration signals often reveal distinct fault patterns, facilitating early detection. They possess remarkable sensitivity, capable of detecting even subtle irregularities, and exhibit a superior signal-to-noise ratio compared to acoustic emission, enhancing diagnostic capabilities. Advances in integrated circuit technology have made accelerometer sensors and computation devices more reliable and cost-effective, further bolstering the feasibility of this approach.

Researchers have explored machine learning as an alternative. Machine learning and deep learning-based fault diagnosis follow a similar process, with machines autonomously making decisions based on knowledge acquired during training, eliminating the need for human intervention. Amid these challenges, several research gaps remain unanswered in field of suspension system fault diagnosis:

- Limited research on the identification of multiple suspension system faults.
- Lack of studies focused on identifying faults in bushes and tie rods.
- Many studies involving damper and ball joint fault diagnosis require a vibration platform for data collection.
- A shortage of studies applying machine learning and deep learning to suspension system fault diagnosis.

The efficiency of fault-diagnosis techniques based on machine learning depends significantly on feature engineering, which involves the extraction and selection of features. Choosing the right feature extraction method requires in-depth domain knowledge and expertise to achieve accurate fault classification, while minimizing computational requirements and time. Furthermore, these feature extraction techniques are sensitive to variations in the environmental systems and mechanical characteristics. Conventional manual feature extraction methods hinder the exploration of novel features owing to the influence of the existing features and evaluation criteria. Owing to this constraint, researchers have increasingly embraced deep-learning-based fault diagnosis as a viable solution. Although numerous studies have been conducted on machine-learning-based fault diagnosis, the manual intervention required for feature engineering diminishes the robustness of the algorithms and can yield unsatisfactory outcomes in certain applications. To address these challenges, deep learning can be employed to extract features and perform classifications directly from images derived from the vibration signals. This approach increases the accuracy of the fault diagnosis.

Deep learning (DL) is a powerful tool for data processing; however, it requires substantial computational power. Fortunately, recent advancements in processing technology have made DL more accessible and applicable to various fields, including speech recognition, robotics, text classification, and object detection. Convolutional neural networks (CNNs) form the fundamental architecture of DL models, allowing the extraction and learning of intricate features from image datasets. CNNs are particularly useful in speech recognition, pattern recognition, and object detection(23). Despite the extensive applications of DL, only a limited number of studies have explored its potential for the fault diagnosis of suspension systems. This represents a significant opportunity for further research and discovery in this field. In this study, a vision transformer (ViT), a neural network that incorporates the attention mechanism proposed by Vaswani et al., was utilized to address this gap and explore the potential of DL in the fault diagnosis of suspension systems (24). The encoder–decode architecture is utilized to transform one sequence of elements into another sequence. The attention mechanism plays a crucial role in capturing long-distance features in the time-series data. The transformer model has shown remarkable performance in the field of Natural Language Processing (NLP), specifically in tasks such as machine translation and speech recognition. It outperformed cyclic neural networks and short-term memory networks that rely on iterative serial training (25). The Transformer model facilitates parallel training and captures global information by processing natural language processing (NLP) words, leading to significant improvements in training accuracy. Building on the success of the Transformer in NLP, this study proposes its application in fault diagnosis scenarios. To assess the effectiveness of the vision transformer in fault diagnosis, a case study was conducted using spectrogram images derived from the vibration signals acquired from suspension system under different fault conditions.

In this study, the fault diagnosis of a suspension system was evaluated using spectrogram images as inputs to a vision transformer model. By utilizing spectrogram images, the model can effectively learn both the temporal and frequency components of signals, which are crucial for accurate fault classification. Moreover, the conversion from raw images to spectrogram images reduces the dimensional complexity and allows the representation of the frequency components necessary for capturing fault-specific vibration patterns. This conversion process enhances the robustness of the model by minimizing noise and reducing overall complexity.

The current study introduces several novel aspects:

- The utilization of spectrogram images as input enables the model to effectively capture both the temporal and frequency components of the signals, thereby facilitating accurate fault identification.
- By adopting a vision transformer instead of a conventional convolutional neural network (CNN), the model becomes capable of simultaneously learning the temporal and frequency components within the images. This approach enhances the accuracy of fault classification.
- The utilization of a pretrained vision transformer, initially trained on a larger dataset, allows for fine-tuning of the model on specific custom datasets. This process enhances the performance and adaptability of the model to the given fault diagnosis task.

To evaluate the performance of the Vision Transformer (ViT) model in diagnosing suspension faults, an experimental study was conducted. The study encompassed one good condition and seven fault conditions, namely, lower arm (ball joint and bush worn out), strut mount failure, worn out strut, external damaged strut, low tire pressure, and tie rod ball joint worn. The experimental study followed the outlined process below:

- Vibration signals were collected from the sensor and subsequently converted into spectrogram images. These spectrogram images were utilized as input for the Vision Transformer algorithm.
- Hyperparameter tuning was conducted to optimize the performance of the Vision Transformer networks. This process involved adjusting various parameters to find the most suitable configuration for the classifier.
- Based on the outcomes of the hyperparameter tuning, appropriate parameters were recommended for the Vision Transformer classifier. These optimized parameters were then utilized in the fault diagnosis system specifically designed for detecting suspension faults.

The overall process of suspension fault diagnosis using the Vision Transformer is depicted in Figure 2.

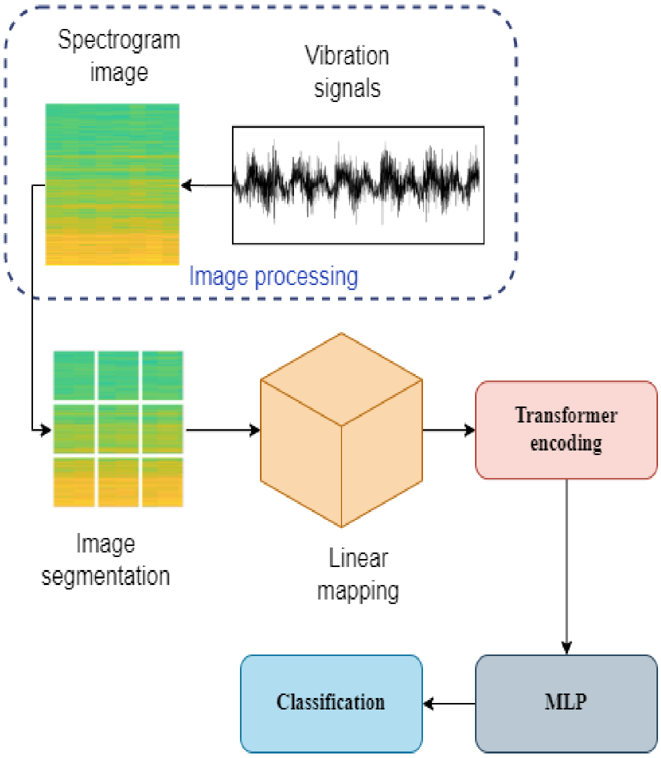


Fig. 2. Work flow of fault diagnosis of suspension system using vision transformer.

2. Experimental Studies

The following section provides detailed information on the experimental studies conducted in three categories: (a) development of the experimental setup, (b) considered faults in the suspension system, and (c) data acquisition process.

To simulate real-time McPherson suspension system operation in front-wheel drive vehicles, a quarter-car model was used as an experimental setup. Signals were collected with a vibration sensor (accelerometer) attached to the suspension system control arm using adhesive. Various faults were introduced by systematically replacing suspension components, resulting in unique vibration signals for each fault condition. Additionally, vibration signals from a healthy suspension system were obtained for comparison. The experimental setup was meticulously designed to accurately represent different suspension system faults, enabling thorough analysis and evaluation.

2.1 Experimental Setup

This study used the suspension system of a commercially available Hyundai i10 model to establish an experimental setup. The resulting suspension setup, as shown in Figure 3, comprises components such as a strut, lower arm, tie rod, wheel, drive shaft, motor, idle roller, and loader. The primary objective of this setup was to evaluate the performance of the passive suspension system when the tire operated at a constant speed on a smooth surface. The setup was designed and fabricated to ensure accurate positioning of the suspension system, including the wheel (rim and tire), above the two idle rollers, enabling seamless rotation. To minimize the presence of undesirable vibrations, the torque generated by the motor is transmitted to the wheel through the utilization of a constant-velocity joint (CV) and belt drives. The height of the idle rollers was adjustable based on the load requirements, which were determined using a pressure gauge and controlled through a hydraulic jack and a guided pillar assembly. This flexibility enabled the setup to accommodate various load conditions during experimental testing.

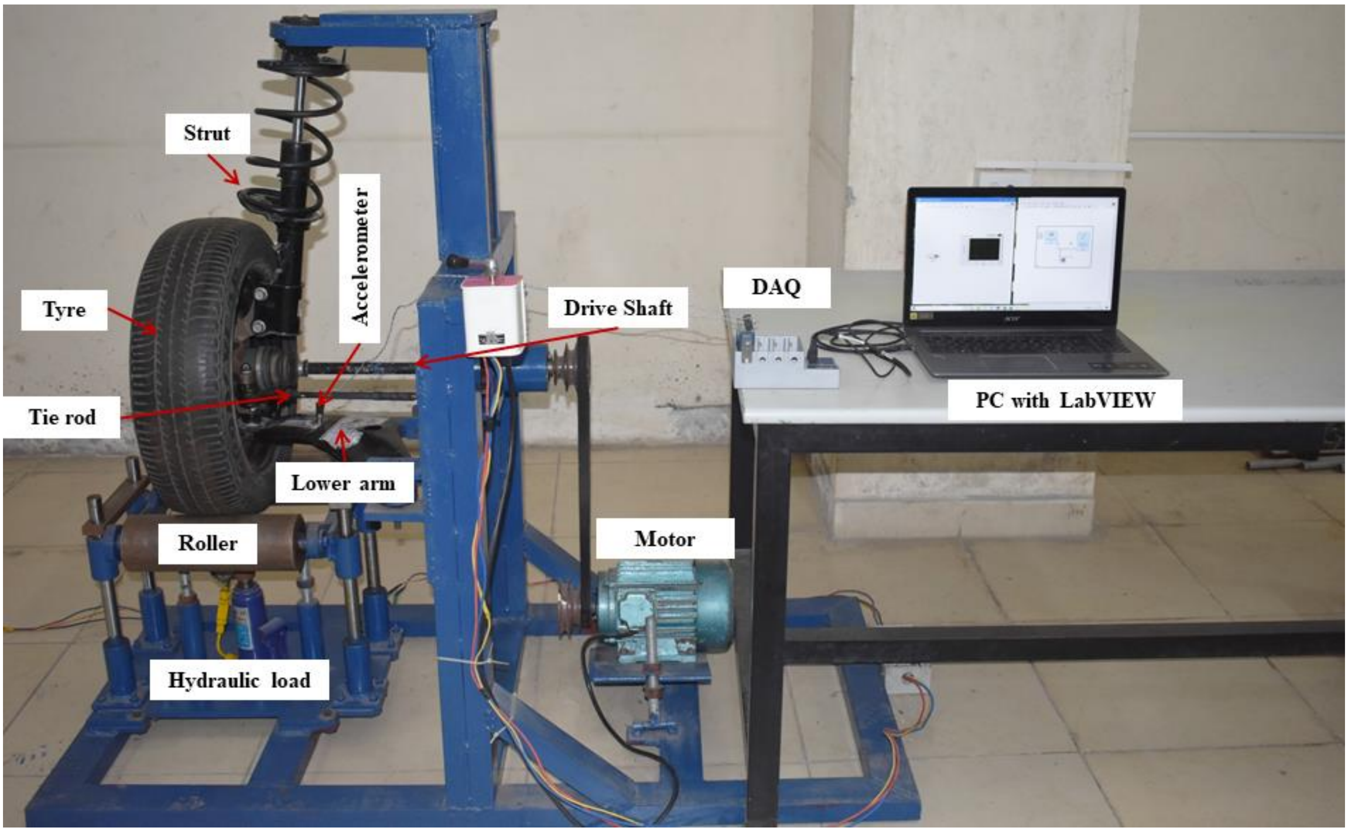


Fig. 3. Experimental setup of suspension system.

2.2 Data Acquisition

The process of converting real-world phenomena into digital values, which can be stored, visualized, and analysed on a computer, is known as data acquisition (DAQ). In this study, fault diagnosis of the suspension system is carried out by acquiring vibration signals using an accelerometer. The accelerometer used in the study is a piezoelectric sensor with a sensitivity of 10.26mV/g and is mounted on the lower arm of the suspension system using adhesive. To convert the analog vibration signals into a digital format, the Ni9234 DAQ is utilized. This DAQ module is connected to a USB chassis. The data acquisition process is facilitated by the NI LabVIEW software, which supports the DAQ system. During the signal collection process, the following parameters are considered:

- Sampling length: 10,000 samples
- Sampling frequency: 25 kHz
- Number of instances for each condition: 100 signals

By adhering to these parameters, a sufficient amount of data is collected for each fault condition and used for further processing.

2.3 Faults in Suspension System

A suspension system is crucial for ensuring the safety and comfort of vehicle occupants. It consists of various components, such as the strut (comprising a damper and coil spring), lower arm, tie rod, strut mount, and knuckle. Throughout the operational lifespan of a suspension system, it is exposed to dynamic loading conditions. Factors such as prolonged usage, rough road conditions, gradual wear and tear of internal components, and the impact of moisture and corrosion can contribute to faults in individual components of the suspension system. The presence of faults in a suspension system can significantly affect its performance, reliability, and longevity. Figure 4 provides a visual representation of the different types of faults that can occur in suspension components, and the following sub section describes various fault considered in the study along with their causes and their symptoms. It is crucial to understand and diagnose these faults accurately to ensure effective maintenance and optimal functioning of the suspension system.

The proper identification and timely rectification of these faults are essential for maintaining vehicle safety and enhancing the overall driving experience.

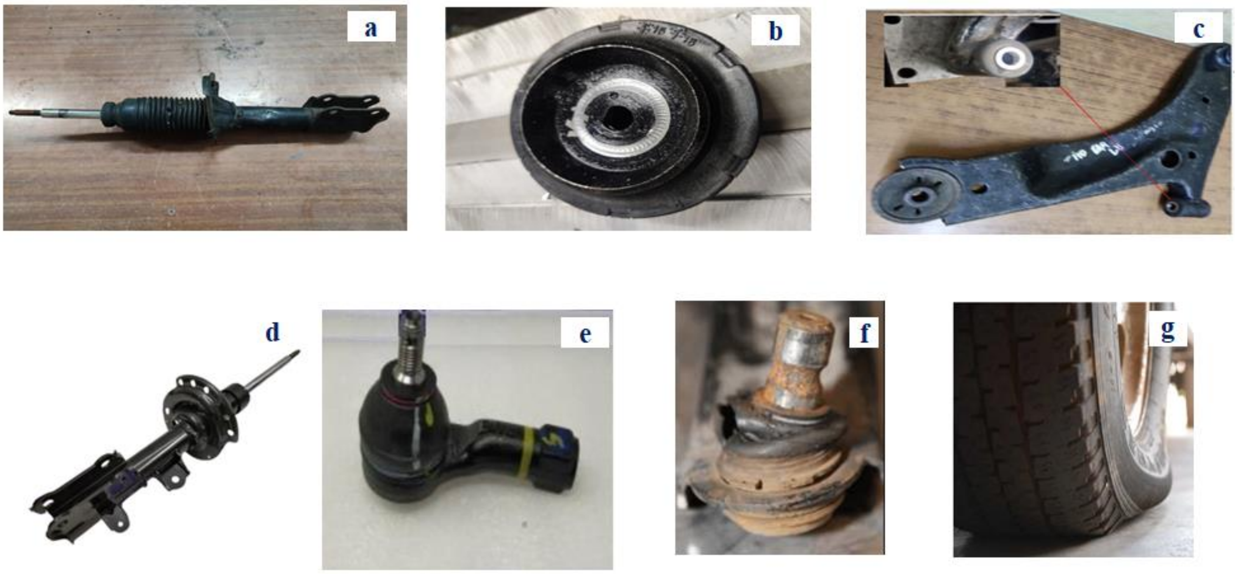


Fig. 4. Faults in the McPherson suspension system
(a) STED, (b) STMF, (c) LABW, (d) STWO, (e) TRBJ, (f) LABJ and (g) LWP

2.3.1 Faults considered in the study are:

1. The lower arm ball joint fault (LABJ) may fail due to factors such as driving in rough terrain, loss of lubrication, heavy loads, and frequent braking. This can manifest as wheel wobbling and steering wheel vibration (26).
2. The tie rod ball joint fault (TRBJ) can experience failure due to aggressive driving, uneven tire wear, wheel misalignment, improper fitting, and lack of lubrication. Symptoms may include a lack of steering control and uneven tire wear.
3. Strut mount failure (STMF) failure can occur due to worn-out struts, broken springs, and fatigue load. This condition may present as minor cracks (12).
4. Strut external damage (STED) is typically caused by physical damage from foreign objects like stones, resulting in dents and a rigid suspension.
5. A worn-out strut (STWO) can result from prolonged use, loss of lubrication, and rupture. This condition may lead to different ride heights, a bouncy ride, steering pull to one side, and increased braking distance (7).
6. Lower arm bush worn-out (LABW) can occur due to misalignment, rash driving, and timely degradation. Symptoms include cracking noise and loss of directional control (27).
7. Low wheel pressure (LWP) may result from check valve failure or punctures. This can lead to stiff or hard steering, uneven tire wear, and related issues (28).

2.4 Vision Transformers

The Vision Transformer (ViT) is a neural network architecture that overcomes certain limitations of Convolutional Neural Networks (CNNs) in image processing tasks. Unlike CNNs, which process input images using convolutional layers, the ViT model employs an attention mechanism to handle image patches. These patches are small, fixed-size crops of the input image, and they are treated as a sequence of vectors, similar to how transformer models process sequences of text. By using the attention mechanism, the Vision Transformer can effectively capture global patterns in the image, rather than being restricted to local regions. This allows the model to have a broader understanding of the image content and improves its ability to recognize complex visual patterns.

Another advantage of the Vision Transformer is its ability to achieve good performance with smaller amounts of data. CNN typically demand significant quantities of labelled data for effective training due to their reliance on learning hierarchical features through convolutional and pooling layers. In contrast, ViT utilize self-attention mechanisms to capture global data relationships. This makes the Vision Transformer particularly beneficial in scenarios where data availability is limited or costly to obtain. Overall, the Vision Transformer introduces a new approach to image processing tasks, leveraging the power of attention mechanisms and enabling effective learning from smaller datasets.

3. Image Generation and Image Processing

In this study, the utilization of spectrogram images as input for the model offers several benefits. Spectrogram images provide a visual representation of the frequency content of the signals, making it easier for the model to identify patterns in signals belonging to specific classes. To generate these spectrogram images, a MATLAB program was employed to plot the spectrograms of the signals. The process of converting vibration signal into spectrogram image is explained in the following section

3.1 Spectrogram image

The fast Fourier transform (FFT) is applied to the entire vibration dataset consisting of 10,000 sample points. The spectral content of each sample in this dataset is then normalized to a range between 0 and 1. This normalized amplitude is utilized as the intensity value for each pixel in the resulting spectral image, which has dimensions X × Y. The conversion from the normalized amplitude of each sample to the corresponding pixel intensity is described by Eq. (1):

P[i, j] = A[i-1] * X + j, where i = 1 to Y and j = 1 to X (Eq.1).

In this equation, P[i, j] represents the intensity (colour intensity) of the pixel (i, j) in the FFT image generated from the entire 10,000-point vibratory signal, where both dimensions have a size of X × Y (with X = Y). A[(i-1) × X + j] signifies the normalized amplitude of each sample in the FFT of the complete vibratory signal. The number of pixels in the spectrogram image corresponds to the 10,000 samples in the FFT of the entire vibratory signal (29). Each spectrogram image comprises X × Y pixels, and Figure 5 shows the sample image of faults converted from the vibration signals extracted for each test conditions.

For the current study considered eight distinct test conditions: STED, STMF, LABW, TRBJ, LABJ, STWO, LWP and good condition. By utilizing the recorded vibration signals, this study generated a comprehensive dataset of 800 images with 100 images for each test condition. This approach provides a diverse and extensive dataset for the model to learn from and generalize, ultimately leading to enhanced accuracy and reliability in fault diagnosis under different test conditions. Figure 5 illustrates a sample spectrogram representing various test conditions. These spectrogram images were subsequently employed as inputs for the Vision Transformer (ViT) model to evaluate its performance in fault diagnosis.

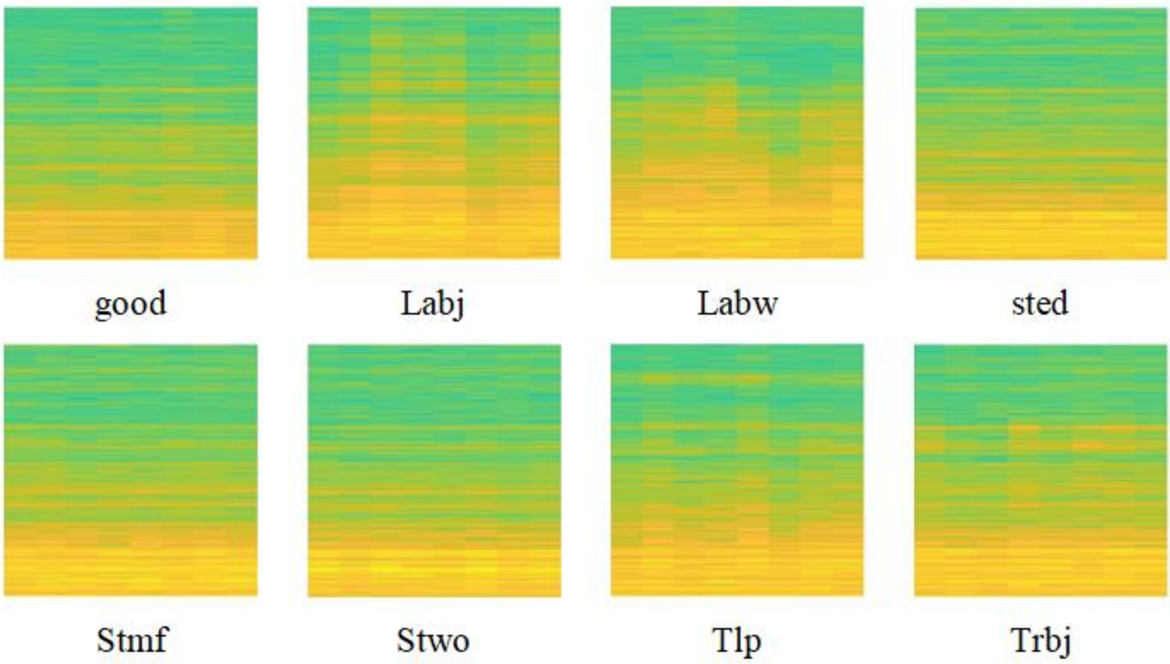


Fig. 5 Sample spectrogram images of considered faults conditions.

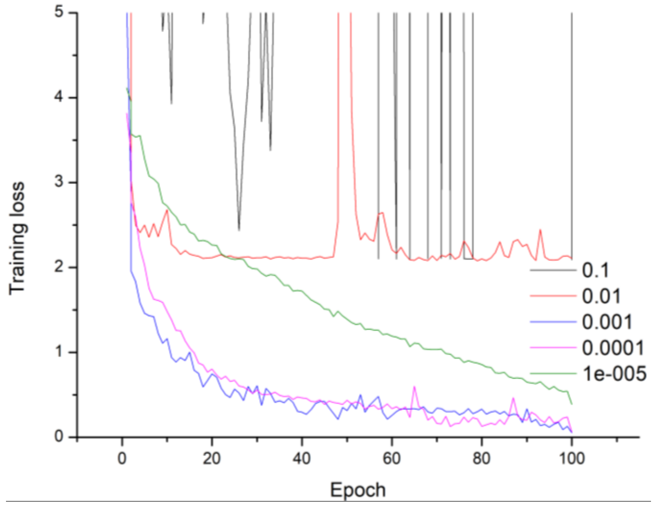
4. Result and Discussion

In this section, a thorough investigation is carried out to assess the performance of the Vision Transformer (ViT) model in diagnosing suspension system faults. The evaluation is conducted through five distinct experiments, which involve modifying key hyperparameters: the learning rate, patch size, batch size, number of heads, and number of MLP layers. The identification of optimal values for these hyperparameters is crucial in achieving the highest possible classification accuracy while effectively utilizing computational resources and minimizing processing time. As a result, the proposed ViT model emerges as an efficient solution for fault diagnosis in the suspension system.

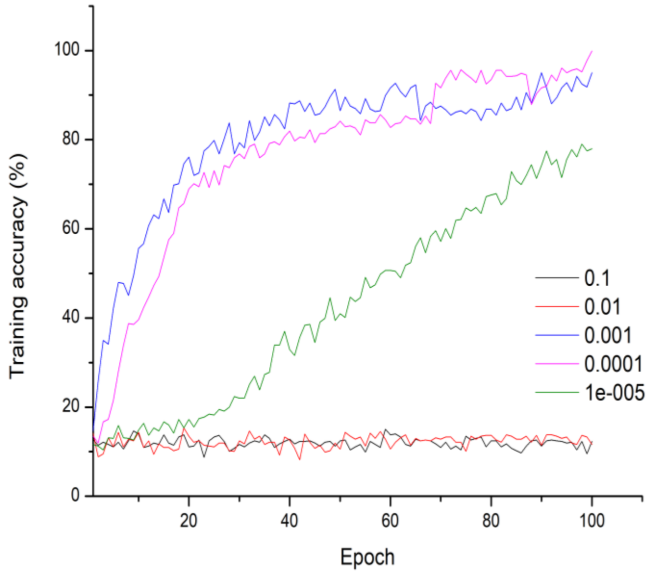
The insights obtained from this study hold significant value in optimizing the performance of the ViT model in real-world scenarios where efficient resource utilization is paramount. By understanding the impact of different hyperparameter settings on classification accuracy, researchers and practitioners can enhance the accuracy and effectiveness of fault diagnosis in suspension systems.

4.1 Effects of learning rate

In the current study, the learning rate is varied from 0.00001 to 0.1, and the corresponding classification accuracy of the ViT model is presented in Table 3. The learning rate is a critical hyperparameter that determines how quickly the model's loss value converges to the minimum. A large learning rate can cause the network's loss gradients to increase rapidly, leading to poor model performance. Conversely, a low learning rate leads to slower convergence as the loss gradients gradually update. Hence, it is important to determine the optimum learning rate that suits for this particular application at hand. From Table 3, it is evident that the ViT model achieves the highest classification accuracy of 99.39% when the learning rate is set to 0.0001. Once the optimal learning rate is identified, it is used to fine-tune the other hyperparameters. The performance of the ViT model across different learning rates is summarized in Table 3, while Figure 6 illustrates the training loss and training accuracy curves for different learning rate.



a)



b)

Fig. 6. Comparison curve for different learning rate (a) training loss, (b) training accuracy.

Table 3. Performance comparison of ViT model with different learning rate.

Learning rate	0.00001	**0.0001**	0.001	0.01	0.1
Test Accuracy (%)	84.66	**99.39**	98.16	11.04	11.04
Computational Time (s)	271	**284**	257	246	253

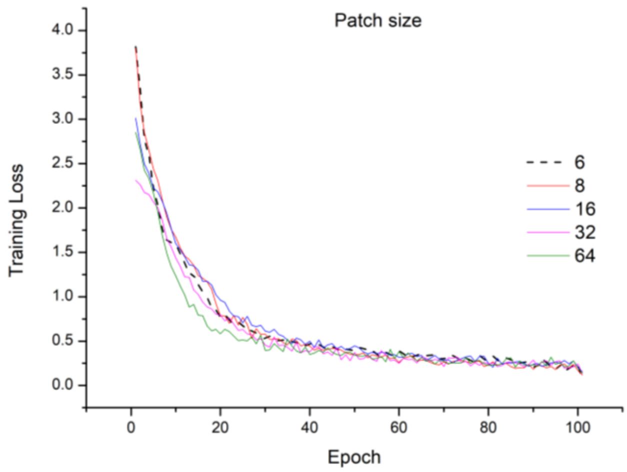
4.2 Effect of patch size

In the Vision Transformer (ViT) model, images are divided into non-overlapping patches and processed as sequences. The self-attention mechanism of the ViT model combines information from different patches and mitigates the loss caused by segmentation. To determine the optimal patch size, a random search method is employed. This method compares the classification accuracy of the ViT model for different patch sizes to select the best-performing one. For the current study, patch sizes of 6x6, 8x8, 16x16, 32x32 and 64x64 were investigated. The performance of the ViT model for each patch size is presented in Table 4.

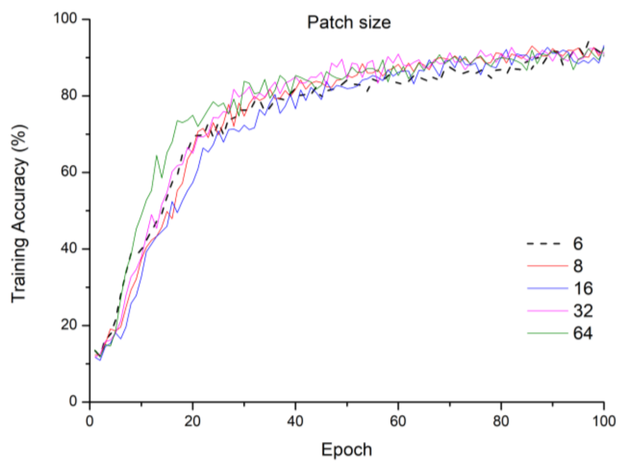
Table 4. Performance comparison of ViT model with different patch size.

Patch Size	**6**	8	16	32	64
Test Accuracy (%)	**99.39**	97.55	93.87	95.71	96.32
Computational Time (s)	**284**	247	247	241	241

Additionally, Figure 7 illustrates the training loss and training accuracy curves for different patch sizes. Based on the results in Table 4, it can be observed that a patch size of 6x6 yields the highest classification accuracy compared to the other patch sizes. Therefore, for further tuning processes, a learning rate of 0.0001 and a patch size of 6x6 are used in the ViT model.



a)



b)

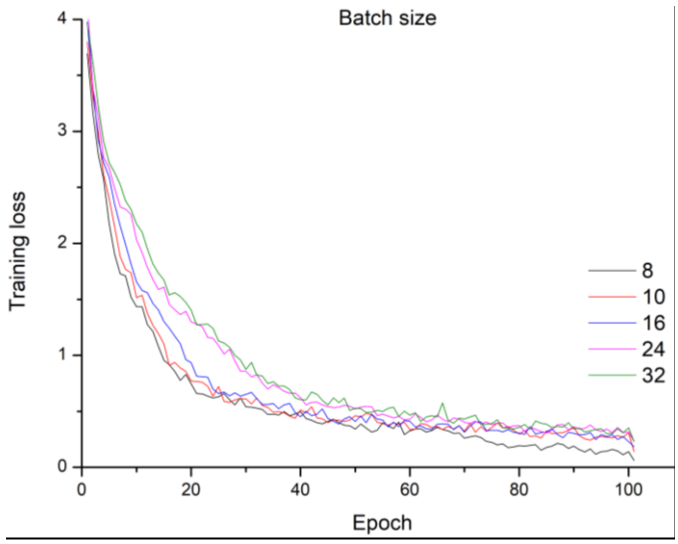
Fig. 7. Comparison curve for different patch size (a) training loss, (b) training accuracy.

4.3 Effect of batch size

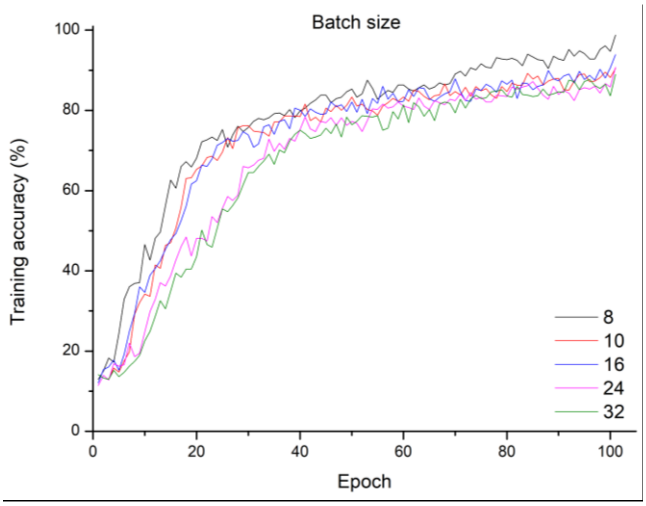
In the Vision Transformer (ViT) model, the batch size determines the number of images processed in parallel during training. Finding the optimal batch size allows the model to work more efficiently. When the batch size is large, the training process tends to be more stable due to the averaging of gradients across multiple images. However, if the batch size is too large, it can lead to memory limitations and reduced generalization performance. Thus, it is crucial to determine a suitable batch size for the ViT model in this specific application. In this study, batch sizes of 8, 10, 16, 24, and 32 were experimented with, and it was found that the proposed ViT model performed well when a batch size of 8 was used. With this batch size, the model achieved a fault classification accuracy of 98.77%. Therefore, for further parameter tuning, such as the number of heads and number of transformer layers, a batch size of 8 was utilized. Table 5 compares the performance of the ViT model with respect to different batch sizes, and the corresponding training loss and training accuracy are presented in Figure 8.

Table 5. Performance comparison of ViT model with different batch size.

Batch Size	**8**	10	16	24	32
Test Accuracy (%)	**98.77**	96.25	93.87	90.81	88.96
Computational Time (s)	**231**	260	231	167	144



a)



b)

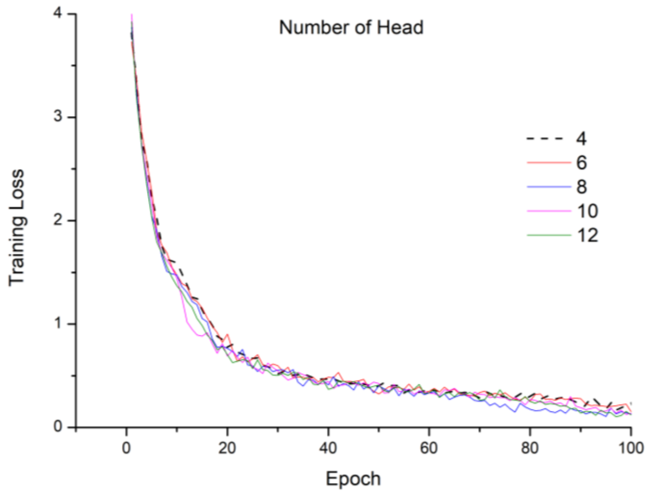
Fig. 8. Comparison curve for different batch size (a) training loss, (b) training accuracy.

4.4 Effect of number of head

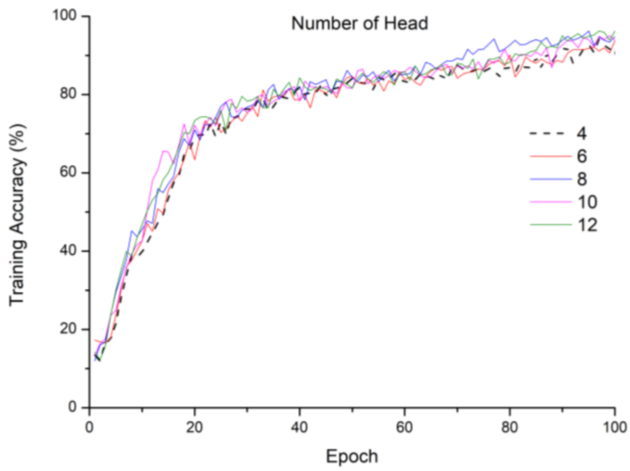
The number of heads in a ViT model determines the number of attention mechanisms used in the self-attention mechanism of the transformer. Each head learns to focus on a different part of the input, allowing the model to capture multiple patterns or relationships simultaneously. Increasing the number of heads can potentially improve performance on complex datasets, with the cost of increases in computational complexity and memory usage. In addition to that, increasing the number of heads also increase the risk of over fitting, particularly when working with smaller datasets. Therefore, it is essential to identify the optimal number of heads for a particular dataset. In this study, the performance of the ViT model was evaluated by varying the number of heads from 4, 6, 8, 10, to 12. The results showed that the model achieved the maximum classification accuracy when the number of heads was set to 4. Table 6 provides a comparison of the ViT model's performance with respect to the number of heads, while Figure 9 presents the corresponding training loss and training accuracy for different values of number of heads.

Table 6. Performance comparison of ViT model with respect to number of heads.

No. of Heads	**4**	6	8	10	12
Test Accuracy (%)	**99.39**	96.93	97.55	95.09	95.09
Computational Time (s)	**284**	325	227	235	244



a)



b)

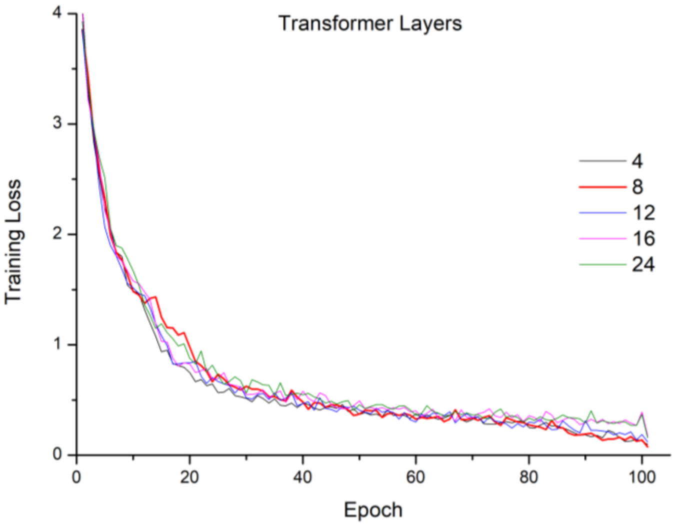
Fig. 9. Comparison curve for variation in number of head (a) training loss, (b) training accuracy.

4.5 Effect of transfer layer

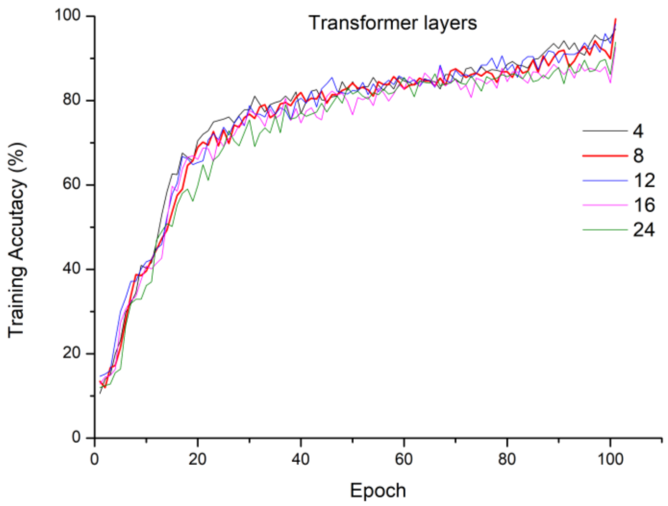
The number of transformer layers in a Vision Transformer (ViT) has a significant impact on its performance. Generally, increasing the number of transformer layers improves the ability of the model to capture complex representations of the input image, thereby enhancing its performance. However, there is a point beyond which increasing the number of layers may lead to degradation in performance. In this study, the performance of the ViT model was evaluated by varying the number of transformer layers. The results showed that when the number of layers was set to 12, the ViT model achieved the maximum classification accuracy of 98.16%. This suggests that 8 layers provide an optimal balance between capturing complex image representations and avoiding potential performance degradation. Table 7 presents the performance of the ViT model for different numbers of transformer layers, highlighting the corresponding classification accuracy.

Table 7. Performance comparison with respect to number of transformer layers.

No. of transformer layers	4	8	**12**	16	24
Test Accuracy (%)	96.93	97.55	**98.16**	92.64	93.87
Computational Time (s)	235	284	**498**	599	874



a)



b)

Fig. 10. Comparison curve for variation in number of transformer layers a) training loss, (b) training accuracy.

4.6 Optimum hyper parameter values

Based on the computational results obtained in the previous sections, this study successfully identified the optimal hyperparameters that significantly enhanced the performance of the ViT model. The best hyperparameters along with their corresponding values are listed in Table 8. These hyperparameters were carefully selected to improve the overall performance and accuracy of the ViT model in diagnosing faults in suspension systems. Furthermore, the evaluation of the ViT model using the identified optimal hyperparameters is shown in Figures 11, 12, and 13. Figure 11 shows the confusion matrix, providing insights into the classification performance of the model across different fault conditions. The training loss and accuracy curves are presented in Figure 12 and 13, respectively, illustrating the learning progress and performance of the ViT model throughout the training process.

The inclusion of these visual representations offers a comprehensive understanding of the performance of the model and reinforces the effectiveness of the identified optimal hyperparameters to achieve improved accuracy and reliability in fault diagnosis for the suspension system.

Table 8. Optimum value of chosen hyper parameter for the suspension fault diagnosis.

Parameter	Value
Learning rate	0.0001
Patch size	6
Batch size	8
Number of head	4
Number of transformer layers	12
Classification accuracy	**98.12%**

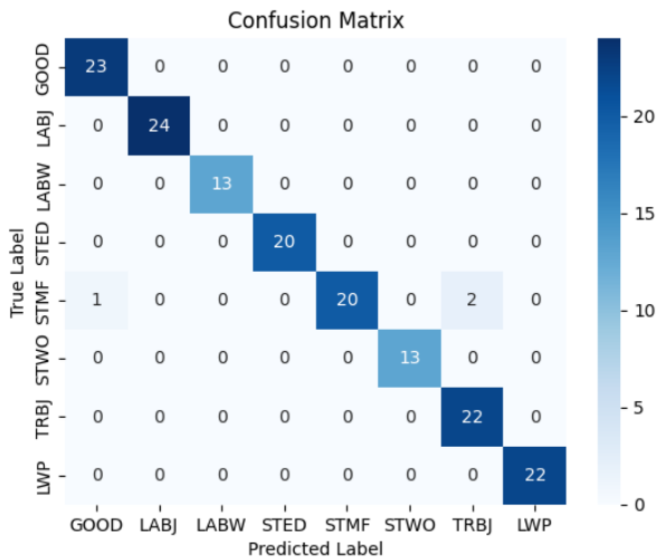


Fig. 11. Confusion matrix of ViT model.

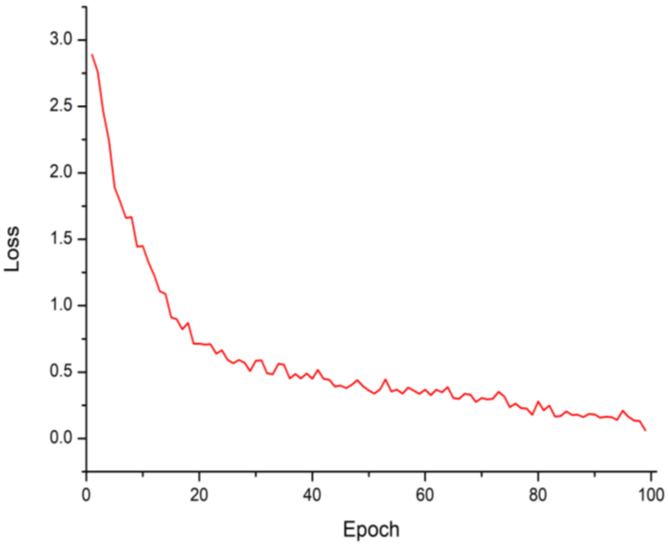


Fig. 12. Training loss plot of ViT model

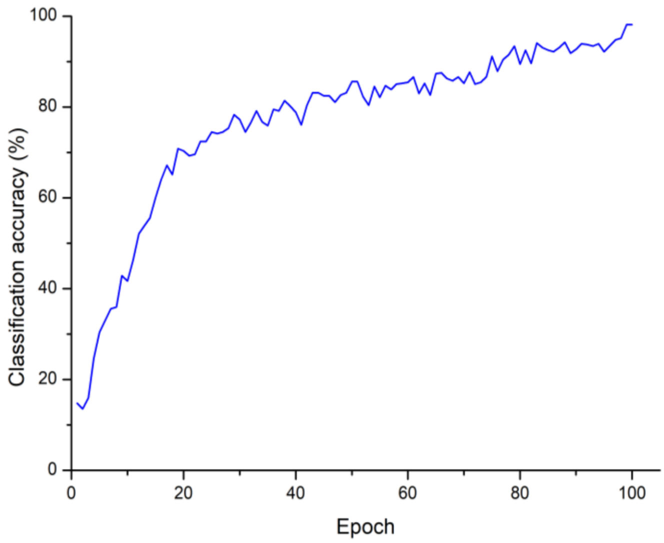


Fig. 13. Training accuracy plot of ViT model.

5. Conclusion

This study introduces an innovative approach to suspension system fault diagnosis, departing from traditional methods that directly analyse vibration signals. Instead, our proposed method harnesses the power of the Vision Transformer (ViT) model, incorporating a self-attention mechanism, to classify various faults, including bushings, ball joints, struts (damper and spring), and tires within the suspension system. This classification is based on spectrum images derived from vibration signals. To optimize the ViT model's performance, key parameters, such as learning rate, patch size, batch size, heads, and transformer layers, were tuned. The best ViT model achieved impressive 98.12% accuracy with a learning rate of 0.0001, patch size 6, batch size 8, 4 heads, and 12 transformer layers. Additionally, it balances performance and computational efficiency, enhancing system reliability by monitoring suspension system component faults. While the proposed model has shown strong performance on the author's specific dataset, its suitability for other datasets remains unverified. Additionally, real-time implementation demands high-end computational resources, which could hinder practical deployment.

Future research holds promise in several areas. Firstly, optimizing the sample length for spectrogram image generation could reduce computational demands. Secondly, enhancing model performance is possible through parameter tuning using techniques like grid search. Lastly, conducting comprehensive training with data under different conditions can improve its applicability and robustness in real-world scenarios.

Reference

1. Fedotov AI, Kuznetsov NY, Lysenko A V, Vlasov VG. Car suspension system monitoring under road conditions. AIP Conf Proc. 2017;1915(December 2017):1–7. https://doi.org/10.1063/1.5017362
2. Ferrari L, Cattini S, Rovati L, Bosi A. A simple measuring system for automotive damper wear estimation. In: 2012 IEEE I2MTC - International Instrumentation and Measurement Technology Conference, Proceedings. 2012. p. 539–43. https://doi.org/10.1109/I2MTC.2012.6229266
3. Lozoya-Santos J, Tudón-Martínez JC, Morales-Menendez R, Ramírez-Mendoza R, Gutierrez AM. Fault detection for an automotive MR damper. In: IFAC Proceedings Volumes (IFAC-PapersOnline). 2012. p. 1023–8. https://doi.org/10.3182/20120523-3-RO-2023.00247
4. Fischer D, Schöner H-P, Isermann R. Model based Fault Detection for an Active Vehicle Suspension. IFAC Proc Vol [Internet]. 2004 Apr;37(22):403–8. Available from: https://doi.org/10.1016/S1474-6670(17)30377-4
5. Arana C, Evangelou SA, Dini D. Series Active Variable Geometry Suspension application to comfort enhancement. Control Eng Pract [Internet]. 2017;59(November 2016):111–26. Available from: http://dx.doi.org/10.1016/j.conengprac.2016.11.011

6. Azadi S, Soltani A. Fault detection of vehicle suspension system using wavelet analysis. Veh Syst Dyn. 2008;47(4):403–18. https://doi.org/10.1080/00423110802094298
7. Ferreira C, Ventura P, Morais R, Valente ALG, Neves C, Reis MC. Sensing methodologies to determine automotive damper condition under vehicle normal operation. Sensors Actuators, A Phys. 2009;156(1):237–44. https://doi.org/10.1016/j.sna.2009.03.035
8. Hamed M, Tesfa B, Gu F, Ball AD. Effects of Tyre Pressure on Vehicle Suspension Performance. Int Lett Chem Phys Astron. 2015;55:102–11. https://doi.org/10.56431/p-l6t1lc
9. Isermann R, Wesemeier D. Indirect Vehicle Tire Pressure Monitoring with Wheel and Suspension Sensors [Internet]. Vol. 42, IFAC Proceedings Volumes. IFAC; 2009. 917–922 p. Available from: http://dx.doi.org/10.3182/20090630-4-ES-2003.00151
10. Zheng J, Ying W, Pan H, Feng K, Tang X, Xu Y, et al. Advantages of using statistical models for detecting faulty components in railway bogies against using simple criteria as defined in standards. Mech Syst Signal Process [Internet]. 2019 Jul 26 [cited 2022 Nov 8];1(1):1–6. Available from: https://www.mdpi.com/2075-1702/11/8/778
11. Miletiev R, Simeonov I, Iontchev E, Yordanov R. Time and frequency analysis of the vehicle suspension dynamics. Int J Syst Appl Eng Dev. 2013;7(5):287–94.
12. Hardikar N, Kulkarni A, Telang A, Kanthale VS. Analysis and Simulation of Elastomeric Strut Mount and its applications for Failure Investigation. Int J Curr Eng Technol. 2011;4(4):356–60.
13. Ferreira C. A New Methodology for Detection of a Loose or Worn ball joint used in vehicles suspension system. Gomes JFS, editor. Conf Theor Exp Mech Mater / 11th Natl Congr Exp Mech Porto/Portugal. 2018;10(November):1–6.
14. Iontchev E, Miletiev R, Bashev V, Simeonov I. Study of the dynamic response and status of the vehicle suspension elements. 2013;3(1):45–51. https://doi.org/10.1016/S1474-6670(17)58392-5
15. Weispfenning T, Leonhardt S. Model-Based Identification of a Vehicle Suspension Using Parameter Estimation and Neural Networks. IFAC Proc Vol. 1996;29(1):4510–5.
16. Jiao N, Guo J, Liu S. Hydro-Pneumatic Suspension System Hybrid Reliability Modeling Considering the Temperature Influence. IEEE Access. 2017;5:19144–53. https://doi.org/10.1109/ACCESS.2017.2751505
17. Melnik R, Koziak S, Dižo J, Kuźmierowski T, Piotrowska E. Feasibility Study of a Rail Vehicle Damper Fault Detection By Artificial Neural Networks. Eksploat i Niezawodn. 2023;25(1):0–3. https://doi.org/10.17531/ein.2023.1.5
18. Wu J, Cui J, Shu Y, Wang Y, Chen R, Wang L. Multi-level health degree analysis of vehicle transmission system based on PSO-BP neural network data fusion. Eksploat i Niezawodn. 2023;25(1):0–2. https://doi.org/10.17531/ein.2023.1.4
19. Liu F, Gu F, Zhao Y, Ball A. A validation study of ACS-SSI for online condition monitoring of vehicle suspension systems. In: Vibroengineering Procedia. 2016. p. 369–75.
20. Arun Balaji P, Sugumaran V. A Bayes learning approach for monitoring the condition of suspension system using vibration signals. IOP Conf Ser Mater Sci Eng. 2021; https://doi.org/10.1088/1757-899X/1012/1/012029
21. Egaji OA, Chakhar S, Brown D. An innovative decision rule approach to tyre pressure monitoring. Expert Syst Appl [Internet]. 2019;124:252–70. Available from: https://doi.org/10.1016/j.eswa.2019.01.051
22. Anoop PS, Sugumaran V. Classifying machine learning features extracted from vibration signal with logistic model tree to monitor automobile tyre pressure. SDHM Struct Durab Heal Monit. 2017;11(2):191–208.
23. Karabacak Yunus Emre. Deep learning-based CNC milling tool wear stage estimation with multi-signal analysis. Eksploat i Niezawodn – Maint Reliab. 2023;25(3):0–2. https://doi.org/10.17531/ein/168082
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;2017-Decem(Nips):5999–6009.
25. Wang Q, Li B, Xiao T, Zhu J, Li C, Wong DF, et al. Learning deep transformer models for machine translation. ACL 2019 - 57th Annu Meet Assoc Comput Linguist Proc Conf. 2020;1810–22. https://doi.org/10.18653/v1/P19-1176
26. Ferreira C. A New Methodology for Detection of a Loose or Worn ball joint used in vehicles suspension system. Conf Theor Exp Mech Mater / 11th Natl Congr Exp Mech Porto/Portugal. 2018;(November):1–6.
27. Khode SS. A Review on Independent Suspension System of Light Commercial Vehicle. IOSR J Mech Civ Eng. 2017 Mar;17(10):14–9. https://doi.org/10.9790/1684-17010061419

28. Muturatnam AB, Sridharan NV, Sreelatha AP, Vaithiyanathan S. Enhanced Tyre Pressure Monitoring System for Nitrogen Filled Tyres Using Deep Learning. Machines. 2023;11(4). https://doi.org/10.3390/machines11040434
29. Youcef Khodja A, Guersi N, Saadi MN, Boutasseta N. Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. Int J Adv Manuf Technol. 2020;106(5–6):1737–51. https://doi.org/10.1007/s00170-019-04726-7