



Acoustic signal-based misfire detection in internal combustion engines using machine learning techniques

Fahime Salehi¹, Ashkan Moosavian², Jafar Hashemi Daryan³, Hamed Moqtaderi^{1*}

¹ Department of Mechanical Engineering, Faculty of Engineering, Alzahra University, Tehran, Iran

² Department of Mechanical Engineering, Technical and Vocational University (TVU), Tehran, Iran

³ Irankhodro Powertrain Company (IPCo), Tehran, Iran

ARTICLE INFO

Keywords:

Engine Fault Detection

Misfire Detection

Acoustic Signal

Machine Learning

Convolutional Neural Network

ABSTRACT

This study investigates misfire detection in a four-cylinder, four-stroke gasoline engine by combining acoustic signal processing techniques with artificial neural networks. Misfires were induced by sequentially cutting fuel injection for each cylinder while the engine operated at a constant speed of 760rpm, and the resulting acoustic signals were captured under controlled test-cell conditions. Fast Fourier Transform (FFT), Mel-Frequency Cepstral Coefficients (MFCC), and Short-Time Fourier Transform (STFT) techniques were applied for feature extraction. The findings indicate that the artificial neural network (ANN) and the one-dimensional convolutional neural network (1D-CNN), when using FFT-derived features, achieved accuracies of 98.40% and 99.36%, respectively. Additionally, the two-dimensional convolutional neural network (2D-CNN), utilizing features extracted from the STFT, achieved an accuracy of 99.71%. These results demonstrate that the proposed methods—particularly the 2D-CNN perform exceptionally well in distinguishing between healthy and misfiring engine states, proving to be an effective approach for the real-time monitoring and fault diagnosis of gasoline engines.



© 2026 Iranian Society of Engine, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution Noncommercial 4.0 International (CC BY-NC 4.0 license). (<https://creativecommons.org/licenses/by-nc/4.0/>).

* Corresponding author

E-mail address: h.moqtaderi@alzahra.ac.ir (H. Moqtaderi)

Received 21 November 2025; Accepted 28 April 2026

E-ISSN: 2345-4121/ISSN: 1735-5214

Cite this article: Salehi F, Moosavian A, Hashemi Daryan J, Moqtaderi H. Acoustic signal-based misfire detection in internal combustion engines using machine learning techniques. The Journal of Engine Research. 2026 Feb 20;72(4):1-20. doi: [10.22034/ER.2026.2076337.1107](https://doi.org/10.22034/ER.2026.2076337.1107)

1- Introduction

Motor vehicles play a central role in modern transportation, with the internal combustion engine (ICE) acting as the core of the powertrain, significantly influencing overall performance and efficiency. Misfire is a common engine defect that severely disrupts its performance. In automotive manufacturing companies, hot testing represents the final stage of production. Currently performed by human operators, this process is susceptible to human error, which ultimately increases after-sales service costs.

To address this issue, Mechanical Fault Diagnosis based on Audio signal analysis (MFDA) is employed to monitor the engine's condition during hot operation [1].

MFDA utilizes microphones to capture and analyze the acoustic characteristics of the engine. This method offers several advantages: it enables non-contact measurement (eliminating the need for equipment disassembly), demonstrates resilience in harsh conditions (such as high temperatures, corrosive environments, nuclear radiation, and toxic substances), remains insensitive to surface conditions (including oily, slippery, or geometrically irregular surfaces), and operates without disrupting normal equipment function. Consequently, it is highly suitable for early fault detection, remote diagnostics, and online monitoring [2].

Recently, the application of artificial intelligence (AI), particularly machine learning (ML), has accelerated in this domain. The following section reviews recent literature on fault diagnosis utilizing acoustic and vibration signals.

In 2018, Jafarian *et al.* [3] investigated misfire and valve clearance faults in an automobile engine. They acquired vibration data using 4 sensors positioned at various locations on the engine under diverse operating conditions. Feature extraction was conducted via Fast Fourier Transform (FFT), followed by the application of Artificial Neural Networks (ANN), Support Vector Machines (SVM), and k-Nearest Neighbors (kNN) classification algorithms to distinguish between the healthy and faulty states of the engine.

In 2021, Firmino *et al.* [4] performed a comparative study of two intelligent systems employing distinct methodologies for misfire detection in a Ford Zetec 4-cylinder gasoline engine. The first methodology utilized vibration analysis, extracting features from engine block vibration data via FFT, which were then processed by an ANN, achieving an accuracy of 99.30%. The second methodology relied on acoustic analysis, utilizing identical feature extraction techniques on engine sound data, resulting in an accuracy of 98.70%. Although the vibration analysis demonstrated marginally higher accuracy, the acoustic system provided the distinct advantage of non-contact data acquisition.

Furthermore, in 2021, Lima *et al.* [5] introduced a completely non-invasive method for detecting and isolating internal combustion engine faults via sound signal processing. Their system collected engine acoustic data using smartphones, processed the signals, and applied an ANN to classify faults based on fractal dimension features. A comparative method based on wavelet multiresolution analysis was also implemented. This proposed solution facilitated real-time fault diagnosis without physical contact, maintained a low computational cost, and achieved a 99% success rate for the evaluated faults during experimental validation.

In 2022, Rodrigues *et al.* [6] compared two algorithms for identifying misfires in operational internal combustion engines. The first algorithm utilized Wavelet Multiresolution Analysis (WMA), while the second employed chaos-based signal analysis using the Density of Maxima (SAC-DM). Experimental tests were conducted under three conditions: a healthy engine, a single-cylinder misfire, and a dual-cylinder misfire. The results indicated that both WMA and SAC-DM accurately detected misfires, though the SAC-DM method required a shorter temporal window.

Also in 2022, Terwilliger and Siegel [7] developed a system for the acoustic characterization of road vehicles using deep learning to analyze sounds recorded by mobile devices. By implementing cascading architectures (multi-level neural networks),

they extracted vehicle attributes such as fuel type, engine configuration, cylinder count, and aspiration type. This extracted information was subsequently used to improve the detection of engine misfires. Trained on a comprehensive dataset, the system demonstrated a substantial improvement in misfire detection accuracy.

In 2023, Sherib et al. [8] developed a methodology for the early detection of automotive engine malfunctions. By integrating data from vibration and oxygen sensors, they collected precise engine performance metrics. Advanced software analysis identified patterns indicative of anomalies, such as incomplete combustion or knocking, leading to the design of an intelligent system capable of automated engine fault diagnosis.

In 2024, Li et al. [9] proposed MITDCNN, a novel method for diagnosing diesel engine misfires under significant environmental noise and diverse operating conditions. Vibration signals were collected from the cylinder head across various engine speeds to extract 1-dimensional amplitude vector features and 2-dimensional image features. These features were processed through a multi-modal feature extraction network and identified by a cross-channel fusion detection network. The results indicated that the accuracy and noise robustness of MITDCNN were markedly superior to existing methods.

In 2024, Yazdi et al. [10] introduced a novel approach for spark plug fault diagnosis based on acoustic signals. They developed a Convolutional Transformer model that synergized the feature extraction capabilities of Convolutional Neural Networks (CNNs) with the temporal modeling proficiency of Transformers. This architecture effectively captured both local and long-term temporal dependencies within the acoustic signals. In a complex 4-class classification task, the model achieved an impressive accuracy of 97.1%, outperforming existing benchmarks and providing a robust solution for enhancing internal combustion engine reliability.

Similarly, in 2024, Karabacak [11] investigated internal combustion engine fault detection utilizing both classical machine learning and deep learning methodologies. Engine vibration signals were acquired to distinguish between healthy conditions, single-cylinder misfires, and dual-cylinder misfires. Classical signal features were extracted to train ANN and SVM models, whereas Short-Time Fourier Transform (STFT) spectrograms were used to train CNN models. The ANN outperformed the SVM (achieving 97%), while two CNN models achieved a perfect diagnostic accuracy of 100%. The integration of the more complex ResNet-50 model further optimized test accuracy.

In another 2024 study, Li et al. [12] proposed an enhanced YOLOv8-based algorithm for gasoline engine misfire detection. Normal and misfire sound signals were processed via wavelet transformation into time-frequency images to train, validate, and test a CNN. Within the YOLOv8 architecture, C2f modules were replaced with BiFormer and CBAM modules to augment complex feature extraction capabilities. The optimized model achieved an accuracy of 99.71%, with a per-signal detection time of under 0.1 seconds, rendering it highly suitable for IoT devices and autonomous vehicles.

In 2025, Yu et al. [13] proposed a deep learning approach utilizing a Channel Attention Convolutional Model (SENET) for automotive engine misfire detection. The study evaluated 11 distinct states, including normal operation, single-cylinder, and dual-cylinder misfires. Data collected via high-precision acceleration sensors were partitioned into training and testing sets to ensure model generalization. Utilizing an input sequence of 1000 time steps, the SENET model achieved an accuracy of 99.8% and an F1-score exceeding 0.98 across all fault typologies. Comparative analyses revealed significant data overlap issues in alternative models like LSTM and Transformers, highlighting SENET's superior efficiency and precision.

Also in 2025, Isa et al. [14] introduced a vibration-based fault detection methodology. Using piezo-based sensors (accelerometers, piezo-films, and micro-fiber composites), they measured vibrations from BMW, Toyota, and Proton engines under varied conditions. Fast Fourier Transform (FFT) and the Z-freq statistical method were employed to extract

frequency characteristics corresponding to spark plug faults, valve clearance issues, and valve cracks. 2-dimensional graphical representations and coefficient analyses facilitated precise fault differentiation, establishing the Z-freq approach as a reliable foundation for AI-integrated predictive maintenance systems.

Furthermore, in 2025, Huang *et al.* [15] evaluated misfire detection in multi-cylinder internal combustion engines. They demonstrated that traditional single-harmonic diagnostic methods suffer from low accuracy in engines with numerous cylinders and low crankshaft stiffness, exhibiting high misdiagnosis rates despite phase correction. While dual-harmonic methods improved single-cylinder fault detection, accurately identifying dual-cylinder faults remained challenging. By isolating 3 critical torsional vibration measurement points and deploying a 3-input CNN, they successfully detected dual-cylinder misfires with a misdiagnosis rate of less than 5%.

As evidenced by the reviewed literature, acoustic signals, combined with classical signal processing and AI techniques, have been effectively utilized for engine misfire detection. However, no prior research has conducted a systematic comparison of FFT, STFT, and Mel-Frequency Cepstral Coefficients (MFCC) features across ANN, 1D-CNN, and 2D-CNN models using controlled acoustic recordings from an engine test cell. Furthermore, recording engine signals from 2 distinct locations allows for a comprehensive evaluation of the impact of sensor placement on diagnostic accuracy.

2- Experimental Setup

To record automotive engine acoustic data, a Zoom H4n Pro recorder was utilized. Featuring high recording quality (24-bit / 96 kHz) and internal X/Y microphones with an adjustable angle of 90 to 120 degrees, this device is optimally suited for accurately capturing engine noise under diverse conditions. The recorder was tripod-mounted and configured to save audio files in WAV format on an SD card, facilitating seamless data transfer and processing. Its capacity for simultaneous monitoring and detailed audio capture provided the precise foundational data required for the signal processing and analysis phases of this research. Figure 1 illustrates the recorder used in the experimental setup.



Figure 1 Recorder H4n Pro

Data acquisition was performed on a 4-cylinder, 4-stroke gasoline engine operating at a constant engine speed of approximately 760 RPM, encompassing both healthy and misfire conditions. The recorder was positioned in 2 distinct locations: one above the engine and the other adjacent to the intake manifold. Figure 2 displays the engine under investigation.



Figure 2 Engine under study

It should be noted that this study is limited to a single gasoline engine operating under a constant, low-speed condition (760 RPM). Because misfire acoustic signatures vary significantly with fluctuations in speed, load, and combustion dynamics, the current findings cannot be universally generalized to real-world driving scenarios. Future research should expand upon this work by incorporating experiments across a broader speed range (from idle to 6000 RPM), variable load conditions, and a diverse array of engine models.

Upon starting the engine and achieving steady-state operation, acoustic signals were recorded for 60 seconds. The recording device was deployed at 2 distinct locations: directly above the engine and adjacent to the intake manifold, as depicted in Figures 3 and 4. The audio recorder was configured with a sampling frequency of 48 kHz and a bit depth of 16 bits.

Under healthy engine conditions, data acquisition was repeated 3 times per location. Subsequently, misfire conditions were systematically induced by sequentially deactivating the fuel injection for each cylinder (from Cylinder 1 to Cylinder 4). For each specific cylinder misfire scenario, the recording process was repeated 2 times. All experimental trials were executed on a 4-cylinder, 4-stroke gasoline engine operating at a constant speed of 760 RPM. Ultimately, a total of 10 healthy-class and 16 misfire-class recordings (each lasting 60 seconds) were obtained for both microphone placements, constituting the dataset for algorithmic training and evaluation.

During the preprocessing phase, the recorded audio files, nominally 60 seconds in duration at 48 kHz, were segmented into discrete frames comprising 4096 samples. While a 60-second clip theoretically yields approximately 703 frames, minor variations in actual recording duration (ranging from 59.8 to 60.6 seconds) resulted in 700 to 708 frames per file. Following data aggregation and the application of an 80/20 stratified train-test split, these inherent temporal variations accounted for the specific sample totals observed in the testing phase (e.g., 1441 healthy and 2335 misfire frames).

Conversely, for the neural network experiments, a standardized threshold of 600 segments per file was enforced to eliminate duration variability. This standardized procedure yielded exactly 1200 healthy and 1920 misfire segments within the test dataset.



Figure 3 Installing the recorder on top of the engine



Figure 4 Installing the recorder in front of the intake manifold

3- Theoretical Background

Misfire is a prevalent fault in internal combustion engines, occurring when a cylinder fails to ignite normally due to mechanical or electrical malfunctions. This phenomenon significantly impacts vehicle performance, leading to elevated emissions, increased fuel consumption, and potential engine damage [16]. Specifically, a misfire fault can reduce engine efficiency by up to 25% and substantially increase exhaust gas pollution [4]. Similar to other mechanical anomalies, engine misfire generates unique vibrational and acoustic emission patterns. Various signal processing methods exist for feature extraction; in this study, both frequency-domain and time-frequency-domain techniques were employed.

In traditional signal analysis, features are manually extracted and evaluated. This approach is not only time-consuming but often yields lower accuracy. Conversely, in artificial intelligence-driven signal analysis, deep learning models autonomously learn hierarchical features and can identify complex patterns that traditional methods might overlook. Consequently, the fault detection process is streamlined and achieves higher accuracy. In the signal processing phase of this research, the acoustic signals were segmented into windows of 4096 samples. Subsequently, Fast Fourier Transform (FFT), Mel-Frequency Cepstral Coefficients (MFCC), and Short-Time Fourier Transform (STFT) were utilized for feature extraction. The dataset was split, allocating 80% for training and 20% for testing. This procedure was standardized for a constant engine speed of 760 RPM, and the performance of the respective algorithms is detailed in the Results section.

3-1- Fast Fourier Transform

Given that the engine operates at a constant rotational speed, FFT serves as a highly suitable technique for acoustic analysis in this context [4].

FFT is a fundamental signal processing algorithm that efficiently transforms a time-domain signal into its corresponding frequency-domain representation. This analysis relies on the Fourier series, as expressed in Equations (1)-(3), where a_k and b_k denote the Fourier coefficients.

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kt) + b_k \sin(kt)) \quad (1)$$

$$a_k = \int_0^{2\pi} f(t) \cos(kt) dt \quad (2)$$

$$b_k = \int_0^{2\pi} f(t) \sin(kt) dt \quad (3)$$

Equations (1) to (3) represent a stationary signal in terms of sines and cosines in the time domain, the transformation of the time-domain signal into a frequency-domain signal is performed by Eq. (4). Where $f(t)$ is the time-domain signal, i is the imaginary number, and ω is the angular velocity, and $e^{i\theta} = \cos(\theta) + i\sin(\theta)$.

$$f(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (4)$$

3-2- Mel-Frequency Cepstral Coefficients (MFCC)

Mel-Frequency Cepstral Coefficients (MFCC) represent one of the most prevalent methods for audio feature extraction. These coefficients are extensively utilized in speech recognition, speaker identification, sound classification, and various other audio signal processing applications due to their robustness in extracting features from audio samples with minimal noise. MFCCs are derived by passing Fast Fourier Transform (FFT) coefficients through a bandpass filter bank. The mathematical expression for the Mel-scale conversion is presented in Equation (5) [17]:

$$Freq_{mel} = \frac{x \log \left[\frac{(c+f)}{x} \right]}{\log(2)} \quad (5)$$

In Equation (5), $Freq_{mel}$ denotes the logarithmic scale of the conventional frequency (f). The coefficient x plays a critical role in the MFCC computation, as it facilitates the mapping of high-frequency sounds to lower-frequency representations, thereby enhancing the detection accuracy of variations within the audio sample.

The optimal range for x typically falls between 250 and 350, corresponding to the number of triangular filters employed within the 200 to 1200 Hz frequency band. This specific range is critical as it encompasses the dominant audio information. A comprehensive illustration of a complete filter bank is provided in Figure 5 [17].

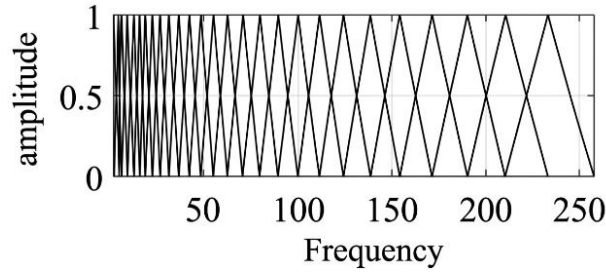


Figure 5 Basic full filter bank

In the final stage of the process, MFCCs are calculated as depicted in Equation (6) and are subsequently denoted as 'Feature MFCC [17].

$$F_{MFCC} = \sqrt{\frac{2}{N}} \sum_{K=1}^N (\log S_k) \cos \left[n(k - 0.5) \frac{\pi}{N} \right] \quad (6)$$

where, S_k is the output of the filter bank where k varies from 1 to N , where N is the length of the DFT.

3-3- Short-Time Fourier Transform (STFT)

The Short-Time Fourier Transform (STFT) is a fundamental and highly effective technique in time-frequency signal processing. As an extension of the standard Fourier Transform, the STFT addresses its primary limitation: the inability to localize frequency content in time. In this approach, a non-stationary signal is partitioned into short, overlapping segments, within which the signal is assumed to be pseudo-stationary. This segmentation process is known as windowing. By applying the Fourier Transform to each of these localized windows, the frequency content of the signal is extracted across discrete time intervals. Consequently, the STFT provides a simultaneous representation of both time and frequency information. Figure 4 illustrates the conversion of a time-domain signal into its frequency representation using this method [17]. The mathematical formulation for calculating the STFT is presented in Equation (7).

$$Y(t, f) = STFT[y(t)] = \int_{-\infty}^{\infty} y(u) h * (u - t) e^{-2j\pi fu} du \quad (7)$$

where $y(t)$ is the original audio signal and $h(t)$ is an STFT window function.

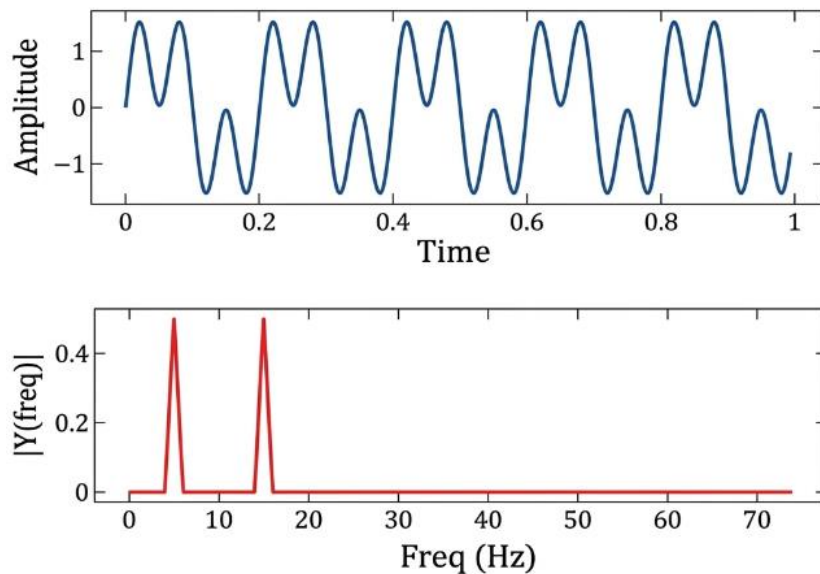


Figure 6 Equivalent Fourier transformation of the provided digital signal [11]

3-4- ANN, 1D-CNN, and 2D-CNN

An artificial neuron is a computational model inspired by its biological counterpart, designed to receive multiple inputs and produce a single output. Artificial Neural Networks (ANNs) comprise numerous such neurons interconnected across multiple layers, working collaboratively to solve complex computational problems [18]. Within the context of this supervised learning study, the recorded audio data was labeled to construct an ANN model for classifying engine acoustic signals. The proposed network architecture consists of 1 input layer, 2 hidden layers, and 1 output layer containing a single neuron. Although binary classification tasks often employ two output neurons, utilizing a single neuron with a sigmoid activation function offers a more streamlined and easily interpretable approach. The output of this neuron represents the probability of an instance belonging to the positive class; if this probability exceeds a threshold of 0.5, the data is classified as positive; otherwise, it is deemed negative. Regarding the network's activation functions, the Rectified Linear Unit (ReLU) was implemented in the hidden layers, while the sigmoid function was applied to the output layer.

Convolutional Neural Networks (CNNs) are deep learning architectures highly effective in analyzing structured data such as images, audio, and text. These networks fundamentally rely on convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for the final classification. A One-Dimensional Convolutional Neural Network (1D-CNN) is a type of feedforward artificial neural network that optimizes its parameters via the backpropagation algorithm. Compared to a Two-Dimensional Convolutional Neural Network (2D-CNN), the 1D-CNN features a simpler and more compact structure, enabling effective training and classification even with limited datasets [19]. The architectural configurations of the implemented 1D-CNN and 2D-CNN are summarized in Table 1 and Table 2, respectively.

Table 1 Information of 1D CNN Layers

Layer name	CNN Models (Kernel Filter, Activation Function, Strides, Padding)
C1	Conv 1D (64*3*1, ReLU, 1, None)
C2	Conv 1D (64*3*1, ReLU, 1, None)
MXP3	Maxpooling 1D (2,2,1)
FC1	Fullyconnect (100, ReLU)
Output	Sigmoid, Classification

Table 2 Information of 2D CNN Layers

Layer name	CNN Models (Kernel Filter, Activation Function, Strides, Padding)
C1	Conv 2D (32*3*3, ReLU, 1, None)
C2	Conv 2D (48*3*3, ReLU, 1, None)
C3	Conv 2D (128*3*3, ReLU, 1, None)
MXP3	Maxpooling 2D(2,2,1)
FC1	Fullyconnect (128, ReLU)
FC2	Fullyconnect (64, ReLU)
Output	Sigmoid, Classification

4- Results

4-1- ANN Using FFT

The results obtained from the Artificial Neural Network (ANN) model, utilizing feature extraction via the Fast Fourier Transform (FFT), demonstrate high accuracy in distinguishing between healthy and misfiring engine states. Figure 7 illustrates the training and testing accuracy of the model, indicating a consistent improvement across epochs, with the final testing accuracy reaching 98.40%. Figure 8 presents the training and testing loss; the concurrent reduction in both metrics signifies stable convergence and the

mitigation of overfitting. Furthermore, the confusion matrix depicted in Figure 9 demonstrates the model's precise classification capabilities, correctly identifying 1187 out of 1200 healthy samples and 1883 out of 1920 misfiring samples.

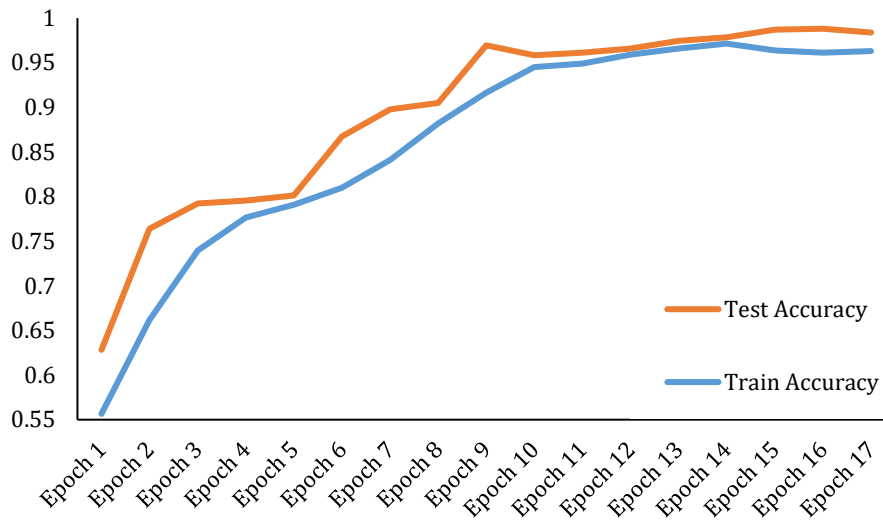


Figure 7 Training and Testing accuracy of ANN

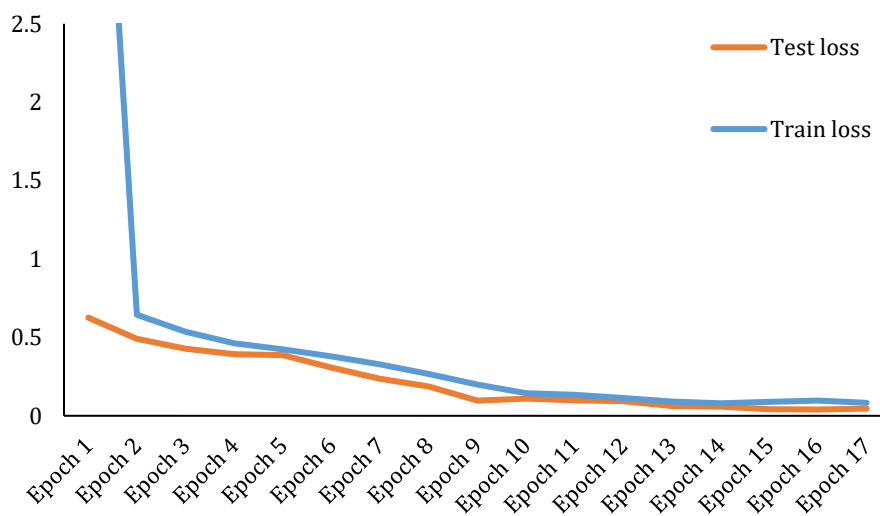


Figure 8 Training and Testing loss of ANN

True label	Healthy	1187	13
	Misfire	37	1883
		Healthy	Misfire
		Predicted label	

Figure 9 Confusion matrix

4-2- ANN Using MFCC

This section presents the results obtained from the Artificial Neural Network (ANN) model utilizing Mel-Frequency Cepstral Coefficients (MFCCs) for feature extraction. Figure 10 illustrates the training and testing accuracy of the model; it can be observed that both metrics consistently improved throughout the epochs, with the final testing accuracy reaching 96.73%. Figure 11 displays the training and testing loss, where the concurrent reduction in both curves indicates stable learning and the effective prevention of overfitting. Finally, Figure 12 presents the confusion matrix, demonstrating that the model correctly classified 1177 out of 1200 healthy samples and 1841 out of 1920 misfiring samples.

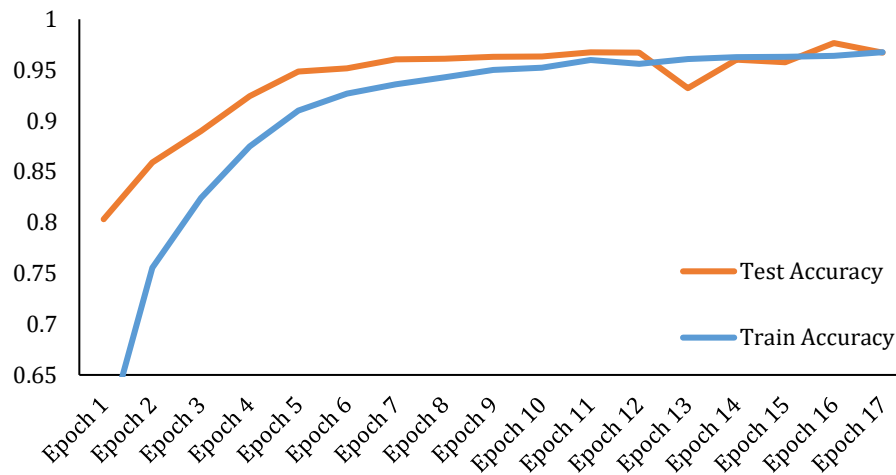


Figure 10 Training and Testing accuracy of ANN

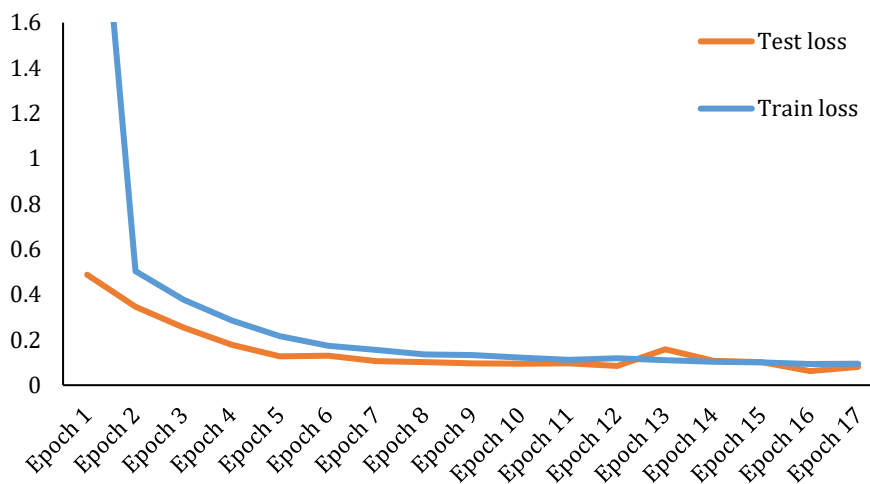


Figure 11 Training and Testing loss of ANN

True label	Healthy	1177	23
	Misfire	79	1841
		Healthy	Misfire
		Predicted label	

Figure 12 Confusion matrix

4-3- 1D CNN using STFT

This section presents the results of the 1D Convolutional Neural Network (CNN) using feature extraction through the Short-Time Fourier Transform (STFT). Figure 13 illustrates the accuracy variations for the training and testing datasets over the training epochs, showing that both training and testing accuracies improve throughout the epochs, with the test accuracy reaching 98.01%. Figure 14 shows the loss values for the training and testing datasets, indicating that as the model training progresses, the loss decreases significantly and approaches near zero, demonstrating effective learning. Figure 15 depicts the model's confusion matrix, where the model successfully classifies 1415 healthy samples out of 1441 and 2286 out of 2335 misfiring samples as misfiring.

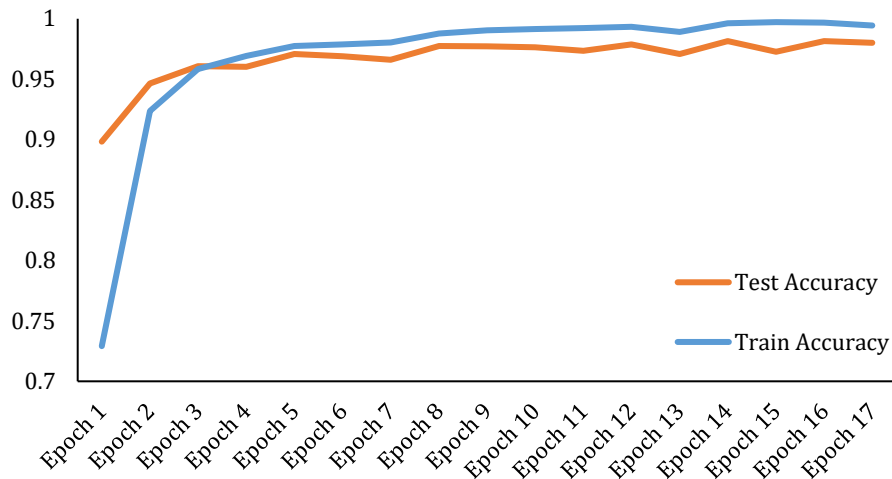


Figure 13 Training and Testing accuracy of 1D CNN

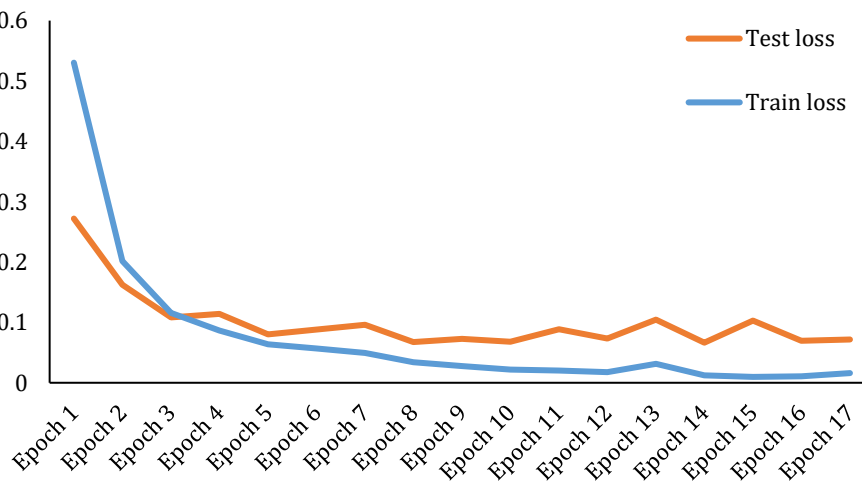


Figure 14 Training and Testing loss of 1D CNN

True label	Healthy	1415	26
	Misfire	49	2286
		Healthy	Misfire
		Predicted label	

Figure 15 Confusion matrix

4-4- 1D CNN using MFCC

In this section, the results for a one-dimensional convolutional neural network (1D CNN) using feature extraction through Mel-frequency cepstral coefficients (MFCCs) are presented. Figure 16 illustrates the accuracy changes for training and testing data based on the number of training epochs. Both training and testing accuracies increased over the epochs, with the testing accuracy reaching 97.91%. Figure 17 shows the loss values for the training and testing data, where the model's error decreased progressively during training and approached near zero. Figure 18 presents the confusion matrix of the model, demonstrating its classification performance. The model correctly identified 1365 healthy samples out of 1441 and 2332 faulty spark ignition samples out of 2335.

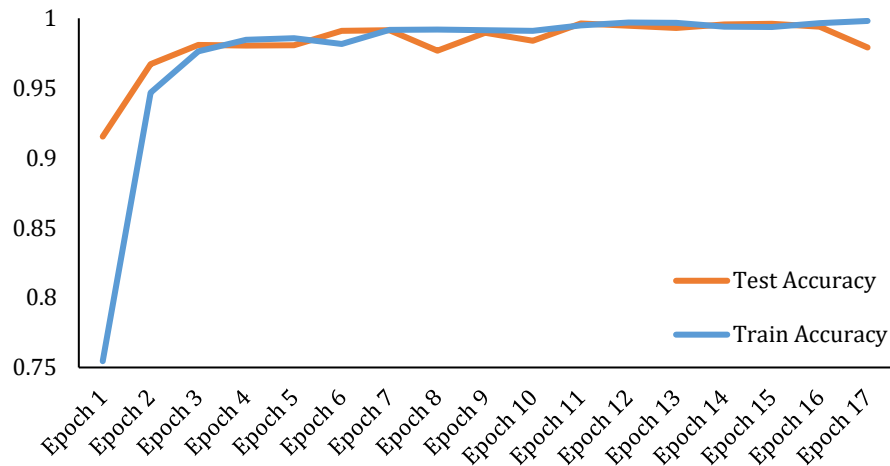


Figure 16 Training and Testing accuracy of 1D CNN

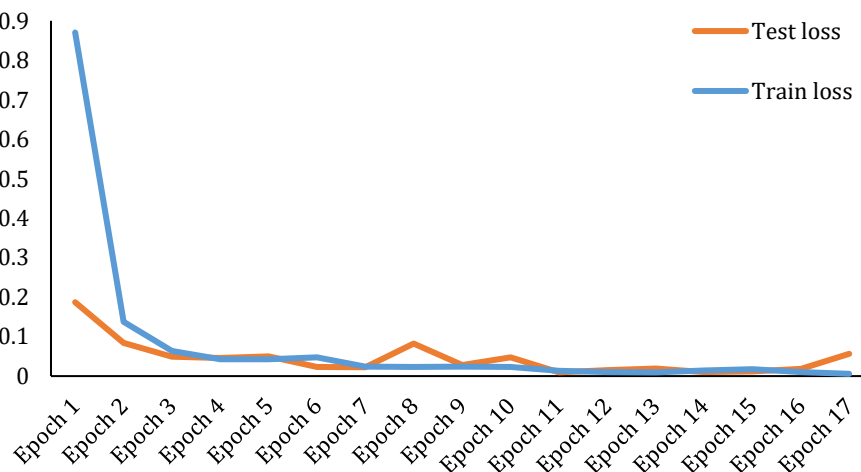


Figure 17 Training and Testing loss of 1D CNN

True label	Healthy	1365	76
	Misfire	3	2332
		Healthy	Misfire
		Predicted label	

Figure 18 Confusion matrix

4-5- 1D CNN using FFT

In this section, the results for a one-dimensional convolutional neural network (1D CNN) using feature extraction through Fast Fourier Transform (FFT) are presented. Figure 19 illustrates the accuracy changes for training and testing data based on the number of training epochs. Both training and testing accuracies increased over the epochs, with the testing accuracy reaching 99.36%. Figure 20 shows the loss values for the training and testing data, where the model's error decreased progressively during training and approached near zero. Figure 21 presents the confusion matrix of the model, demonstrating its classification performance. The model correctly identified 1426 healthy samples out of 1441 and 2326 faulty spark ignition samples out of 2335.

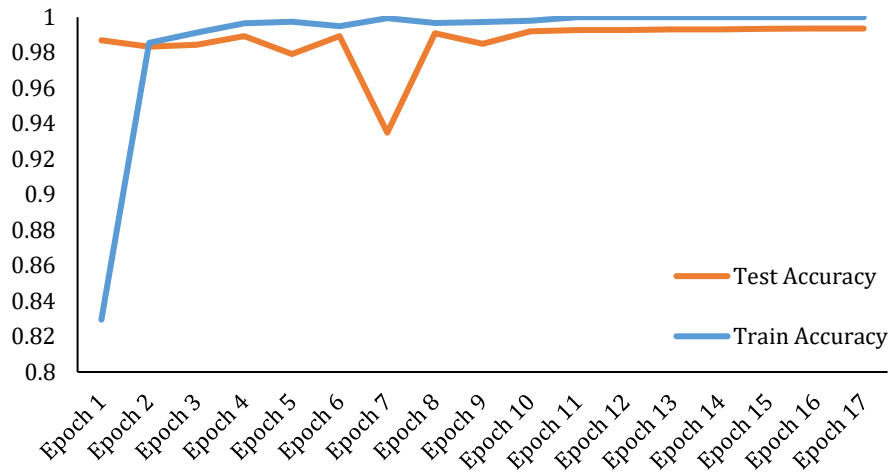


Figure 19 Training and Testing accuracy of 1D CNN

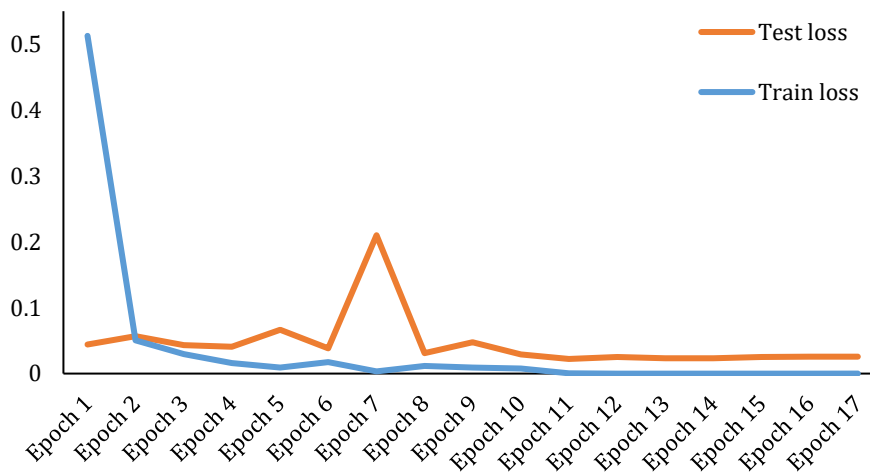


Figure 20 Training and Testing loss of 1D CNN

True label	Healthy	1426	15
	Misfire	9	2326
		Healthy	Misfire
		Predicted label	

Figure 21 Confusion matrix

4-6- 2D CNN using STFT

In this section, the results for a Two-dimensional convolutional neural network (2D CNN) using feature extraction through Short-Time Fourier Transform (STFT) are presented. Figure 22 illustrates the accuracy changes for training and testing data based on the number of training epochs. Both training and testing accuracies increased over the epochs, with the testing accuracy reaching 99.71%. Figure 23 shows the loss values for the training and testing data, where the model's error decreased progressively during training and approached near zero. Figure 24 presents the confusion matrix of the model, demonstrating its classification performance. The model correctly identified 1435 healthy samples out of 1441 and 2330 faulty spark ignition samples out of 2335.

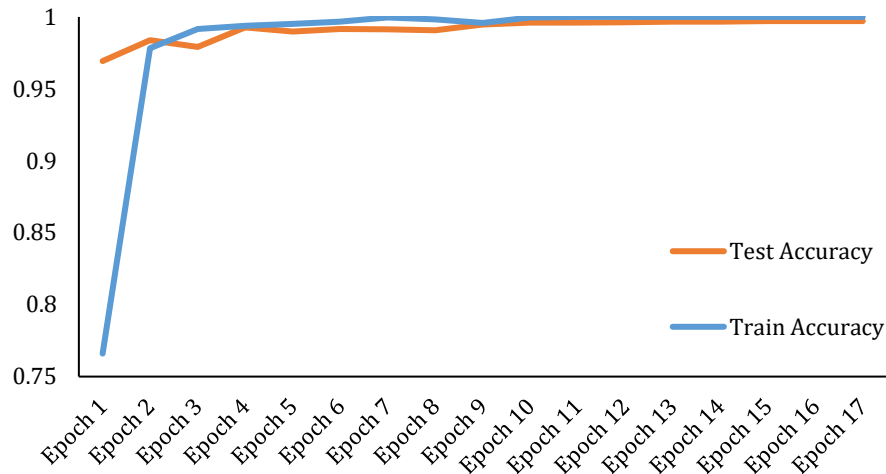


Figure 22 Training and Testing accuracy of 2D CNN

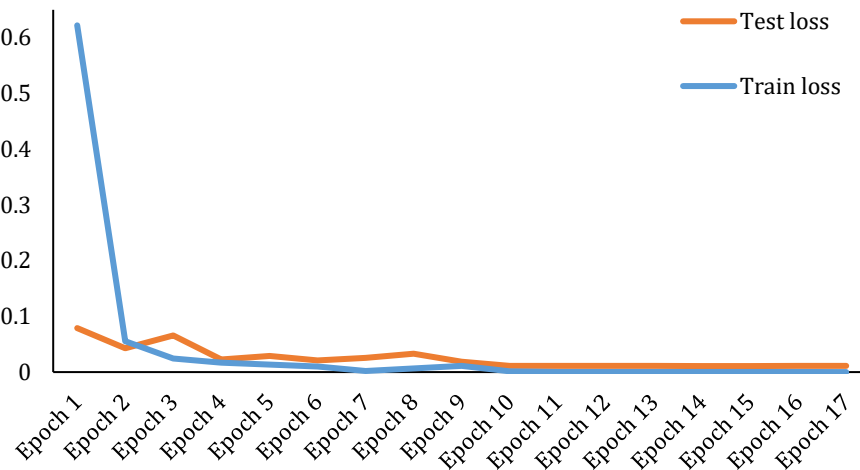


Figure 23 Training and Testing loss of 2D CNN

True label	Healthy	1435	6
	Misfire	5	2330
		Healthy	Misfire
		Predicted label	

Figure 24 Confusion matrix

4-7- 2D CNN using MFCC

In this section, the results for a Two-dimensional convolutional neural network (2D CNN) using feature extraction through Mel-frequency cepstral coefficients (MFCC) are presented. Figure 25 illustrates the accuracy changes for training and testing data based on the number of training epochs. Both training and testing accuracies increased over the epochs, with the testing accuracy reaching 98.68%. Figure 26 shows the loss values for the training and testing data, where the model's error decreased progressively during training and approached near zero. Figure 27 presents the confusion matrix of the model, demonstrating its classification performance. The model correctly identified 1393 healthy samples out of 1441 and 2333 faulty spark ignition samples out of 2335.

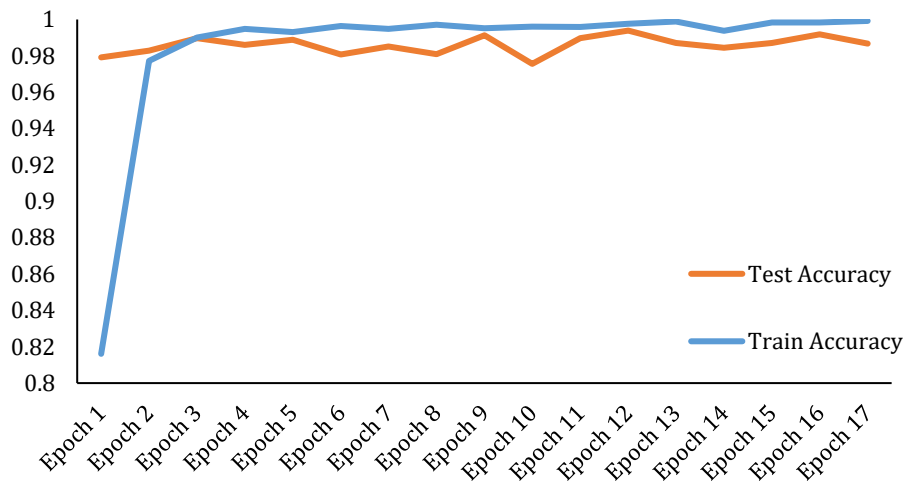


Figure 25 Training and Testing accuracy of 2D CNN

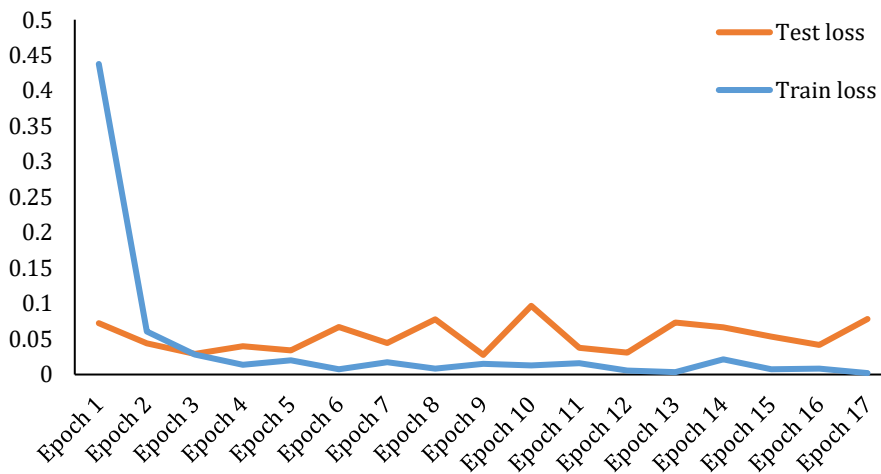


Figure 26 Training and Testing loss of 2D CNN

True label	Healthy	1393	48
	Misfire	2	2333
		Healthy	Misfire
		Predicted label	

Figure 27 Confusion matrix

4-8- Models Comparison

Table 3 lists all the models together to illustrate which gives the best result for fault diagnosis in the engine.

Table 3 Comparison between different methods

Classifier	Class	Preprocessing	Accuracy
ANN	Binary	FFT	98.40
ANN	Binary	MFCC	96.73
1D-CNN	Binary	STFT	98.01
1D-CNN	Binary	MFCC	97.91
1D-CNN	Binary	FFT	99.36
2D-CNN	Binary	STFT	99.71
2D-CNN	Binary	MFCC	98.68

Based on Table 3, the accuracy of the Two-dimensional convolutional neural network (2D CNN) method using feature extraction with short-time Fourier transform (STFT) is higher than the other methods. Additionally, the artificial neural network (ANN) and the one-dimensional convolutional neural network (1D CNN) using feature extraction with fast Fourier transform (FFT) also show significant accuracy.

5- Conclusion

This research focused on the detection of misfire faults in internal combustion engines utilizing acoustic signals and machine learning techniques. To ensure comprehensive data collection, a recorder was positioned at various orientations around the engine, capturing precise acoustic information from multiple angles. All recordings were conducted within a dedicated engine test cell to eliminate ambient noise interference and isolate the pure engine acoustic signature. The faulty condition was induced by manually simulating a misfire via fuel cut-off in each cylinder. Acoustic data for both healthy and misfiring states were acquired at a sampling rate of 48 kHz.

The acquired acoustic signals were subsequently processed using feature extraction techniques, including the Fast Fourier Transform (FFT), Short-Time Fourier Transform (STFT), and Mel-Frequency Cepstral Coefficients (MFCC). For data classification, three neural network architectures were employed: an Artificial Neural Network (ANN), a One-Dimensional Convolutional Neural Network (1D-CNN), and a Two-Dimensional Convolutional Neural Network (2D-CNN).

The results demonstrated that the ANN and the 1D-CNN, utilizing FFT-based features, achieved classification accuracies of 98.40% and 99.36%, respectively. Furthermore, the 2D-CNN, employing STFT-based feature extraction, attained an accuracy of 99.71%. These findings indicate that the proposed methodologies, particularly the 2D-CNN approach, exhibit exceptional performance in distinguishing between healthy and faulty engine states, presenting a highly effective tool for the condition monitoring and fault diagnosis of internal combustion engines.

It should be noted that the high accuracy rates achieved in this study reflect performance strictly under constant-speed, no-load conditions. Since real-world engines operate across a dynamic range of speeds (800–6000 RPM) and varying loads, misfire acoustic signatures can change significantly. Therefore, future research will focus on evaluating the models under variable speeds and loads, testing across multiple engine types, and incorporating baseline comparisons with traditional machine learning algorithms (e.g., SVM, kNN, and Random Forest).

Acknowledgment

The authors would like to express their sincere gratitude to Iran Khodro Powertrain

Company (IPCo) for providing technical support regarding the experimental setup and facilitating the engine tests conducted in this research.

References

- [1] Tang L, Tian H, Huang H, Shi S, Ji Q. A survey of mechanical fault diagnosis based on audio signal analysis. *Measurement*. 2023 Oct 1;220:113294. doi: [10.1016/j.measurement.2023.113294](https://doi.org/10.1016/j.measurement.2023.113294)
- [2] Goli Karimabadi A, Ghassabzadeh Saryazdi M, Ohadi A, Moosavian A. Diagnosing the clearance fault of the camshaft in the hot test stage at the end of the production line by using of audio signal processing. *NDT Technology*. 2024 Feb 20;3(4):41-9. doi: [10.30494/jndt.2024.471587.1147](https://doi.org/10.30494/jndt.2024.471587.1147)
- [3] Jafarian K, Mobin M, Jafari-Marandi R, Rabiei E. Misfire and valve clearance faults detection in the combustion engines based on a multi-sensor vibration signal monitoring. *Measurement*. 2018 Nov 1;128:527-36. doi: [10.1016/j.measurement.2018.04.062](https://doi.org/10.1016/j.measurement.2018.04.062)
- [4] Firmino JL, Neto JM, Oliveira AG, Silva JC, Mishina KV, Rodrigues MC. Misfire detection of an internal combustion engine based on vibration and acoustic analysis. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. 2021 Jul;43(7):336. doi: [10.1007/s40430-021-03052-y](https://doi.org/10.1007/s40430-021-03052-y)
- [5] de V. Lima TL, Filho AC, Belo FA, Souto FV, Silva TC, Mishina KV, Rodrigues MC. Noninvasive methods for fault detection and isolation in internal combustion engines based on chaos analysis. *Sensors*. 2021 Oct 19;21(20):6925. doi: [10.3390/s21206925](https://doi.org/10.3390/s21206925)
- [6] Rodrigues NF, Brito AV, Ramos JG, Mishina KD, Belo FA, Lima Filho AC. Misfire detection in automotive engines using a smartphone through wavelet and chaos analysis. *Sensors*. 2022 Jul 6;22(14):5077. doi: [10.3390/s22145077](https://doi.org/10.3390/s22145077)
- [7] Terwilliger AM, Siegel JE. Improving misfire fault diagnosis with cascading architectures via acoustic vehicle characterization. *Sensors*. 2022 Oct 12;22(20):7736. doi: [10.3390/s22207736](https://doi.org/10.3390/s22207736)
- [8] bin Md Sharib MK, bin Abu Bakar E, bin Hawary AF, Akhtar MN, Abdullah MN. Adaptive system on engine misfire and knocking features for IC engine. *Materials Today: Proceedings*. 2023 Apr 18. doi: [10.1016/j.matpr.2023.03.677](https://doi.org/10.1016/j.matpr.2023.03.677)
- [9] Li W, Liu X, Wang D, Lu W, Yuan B, Qin C, Cheng Y, Căleanu C. MITDCNN: A multi-modal input Transformer-based deep convolutional neural network for misfire signal detection in highnoise diesel engines. *Expert Systems with Applications*. 2024 Mar 15;238:121797. doi: [10.1016/j.eswa.2023.121797](https://doi.org/10.1016/j.eswa.2023.121797)
- [10] Yazdi MH, Aliyari-Shoorehdeli M, Moosavian A. Convolutional transformer approach for engine spark plug fault diagnosis using acoustic signal. *The Journal of Engine Research*. 2024;70(4):56-68. doi: [10.22034/ER.2024.2025036.1036](https://doi.org/10.22034/ER.2024.2025036.1036)
- [11] Karabacak YE. Condition monitoring of internal combustion engines with vibration signals and fault detection by using machine learning techniques. *International Journal of Automotive Engineering and Technologies*. 2024 Dec 12;13(4):191-200. doi: [10.18245/ijaet.1251886](https://doi.org/10.18245/ijaet.1251886)
- [12] Li Z, Qin Z, Luo W, Ling X. Gasoline engine misfire fault diagnosis method based on improved YOLOv8. *Electronics*. 2024 Jul 9;13(14):2688. doi: [10.3390/electronics13142688](https://doi.org/10.3390/electronics13142688)
- [13] Yu F, Huang Y, Chen G, Yang X, Du C, Gong Y. Engine Misfire Fault Detection Based on the Channel Attention Convolutional Model. *Computers, Materials & Continua*. 2025 Jan 1;82(1). doi: [10.32604/cmc.2024.058051](https://doi.org/10.32604/cmc.2024.058051)
- [14] Isa MM, Ngatiman NA, Ahmad MI, Saad SM, Nuawi MZ. Monitoring Valve and Spark Plug Failures using Z-Freq Statistical Analysis. *Journal of Acoustics and Vibration Research*. 2025 Jun 30;3(1):14-21.
- [15] Huang Y, Zou A, Xu K, Yan B, Hua C, Dong D. Misfire Fault Diagnosis in Multi-Cylinder Internal Combustion Engine with Multi-Input Neural Networks. 2025. Available at SSRN 4965428. doi: [10.2139/ssrn.4965428](https://doi.org/10.2139/ssrn.4965428)
- [16] Zheng T, Zhang Y, Li Y, Shi L. Real-time combustion torque estimation and dynamic misfire fault diagnosis in gasoline engine. *Mechanical Systems and Signal Processing*. 2019 Jul 1;126:521-35. doi: [10.1016/j.ymssp.2019.02.048](https://doi.org/10.1016/j.ymssp.2019.02.048)
- [17] Gourisaria MK, Agrawal R, Sahni M, Singh PK. Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques. *Discover Internet of Things*. 2024 Jan 3;4(1):1. doi: [10.1007/s43926-023-00049-y](https://doi.org/10.1007/s43926-023-00049-y)
- [18] Mishra M, Srivastava M. A view of artificial neural network. In 2014 international conference on

advances in engineering & technology research (ICAETR-2014) 2014 Aug 1 (pp. 1-3). IEEE. doi: [10.1109/ICAETR.2014.7012785](https://doi.org/10.1109/ICAETR.2014.7012785)

- [19] Wang H, Liu C, Du W, Wang S. Intelligent diagnosis of rotating machinery based on optimized adaptive learning dictionary and 1DCNN. Applied Sciences. 2021 Nov 30;11(23):11325. doi: [10.3390/app112311325](https://doi.org/10.3390/app112311325)

تشخیص عیب عدم تشکیل شعله در موتور درونسوز با صدابرداری و روش یادگیری ماشین

فهیمه صالحی^۱، اشکان موسویان^۲، جعفر هاشمی داریان^۳، حامد مقتدری^{۱*}

^۱ گروه مهندسی مکانیک، دانشکده فنی و مهندسی، دانشگاه الزهراء، تهران، ایران

^۲ گروه مهندسی مکانیک، دانشگاه ملی مهارت، تهران، ایران

^۳ شرکت تحقیق، طراحی و تولید موتور ایران خودرو (ایپکو)، تهران، ایران

چکیده

این مقاله، ترکیب روش‌های پردازش صدا و شبکه‌های عصبی را در یک موتور بنزینی برای تشخیص نبود شعله بررسی می‌کند. احتراق ناقص با قطع متوالی تزریق سوخت برای هر استوانه در حالی که موتور با سرعت ثابت ۷۶۰ دور بر دقیقه کار می‌کرد، ایجاد شد و داده‌های صوتی تحت شرایط آزمون ثبت شدند. روش‌های تبدیل فوریه سریع، ضرایب کپسترال بسامد مل و تبدیل فوریه زمان-کوتاه برای استخراج ویژگی به کار گرفته شدند. یافته‌ها نشان می‌دهد که شبکه عصبی مصنوعی و شبکه عصبی پیچشی یک بعدی، هنگام استفاده از ویژگی‌های مشتق شده از تبدیل فوریه سریع، بترتیب به دقت ۹۸.۴۰ درصد و ۹۹.۳۶ درصد دست یافتند. علاوه بر این، شبکه عصبی پیچشی دو بعدی، با استفاده از ویژگی‌های استخراج شده از تبدیل فوریه زمان کوتاه، به دقت ۹۹.۷۱ درصد دست یافت. این نتایج نشان می‌دهد که روش‌های پیشنهادی -بویژه شبکه عصبی پیچشی دو بعدی- در تشخیص حالت‌های موتور سالم و احتراق ناقص عملکرد فوق‌العاده‌ای دارند و می‌توانند به عنوان یک رویکرد مؤثر برای نظارت و تشخیص عیب در زمان واقعی موتورهای بنزینی عمل کنند.

اطلاعات مقاله

کلیدواژه‌ها:

تشخیص عیب موتور

تشخیص عدم تشکیل شعله

سیگنال آکوستیک

یادگیری ماشین

شبکه عصبی پیچشی



© 2026 Iranian Society of Engine, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution Noncommercial 4.0 International (CC BY-NC 4.0 license). (<https://creativecommons.org/licenses/by-nc/4.0/>).

* نویسنده مسئول

پست الکترونیکی: h.moqtaderi@alzahra.ac.ir (حامد مقتدری)

دریافت ۳۰ آبان ۱۴۰۴؛ پذیرش ۸ اردیبهشت ۱۴۰۵

شاپای الکترونیکی: ۴۱۲۱-۲۳۴۵ / شاپای چاپی: ۵۲۱۴-۱۷۳۵

Cite this article: Salehi F, Moosavian A, Hashemi Daryan J, Moqtaderi H. Acoustic signal-based misfire detection in internal combustion engines using machine learning techniques. The Journal of Engine Research. 2026 Feb 20;72(4):1-20. doi: 10.22034/ER.2026.2076337.1107