Article

Engine Fault Detection by Sound Analysis and Machine Learning

Ferit Akbalık [1,*], Abdulnasır Yıldız [2], Ömer Faruk Ertuğrul [3] and Hasan Zan [4]

1   Social Sciences Vocational School, Batman University, 72040 Batman, Turkey
2   Department of Electrical and Electronics Engineering, Dicle University, 21280 Diyarbakir, Turkey; abnayil@dicle.edu.tr
3   Department of Electrical and Electronics Engineering, Batman University, 72040 Batman, Turkey; omerfaruk.ertugrul@batman.edu.tr
4   Vocational School, Mardin Artuklu University, 47200 Mardin, Turkey; hasanzan@artuklu.edu.tr
*   Correspondence: ferit.akbalik@batman.edu.tr

Abstract: Traditional vehicle fault diagnosis methods rely heavily on the expertise of mechanics or diagnostic tools available at service centers, which can be costly, time-consuming, and may not always provide accurate results. This study presents a comprehensive vehicle fault diagnosis framework, which utilized Mel-Frequency Cepstral Coefficients (MFCCs), Discrete Wavelet Transform (DWT)-based features, and the Extreme Learning Machine (ELM) classifier. To address the limitations of previous works, the proposed framework leverages a large, diverse dataset encompassing various vehicle models and real-world operating conditions. Significantly improved robustness and generalizability of the fault diagnosis system were achieved. The results of the experiments demonstrate the superiority of the MFCC-based features combined with the ELM classifier, achieving the highest performance metrics in terms of accuracy, precision, recall, F1-score, macro F1-score, and weighted F1-score, which are 92.17%, 92.24%, 92.22%, 92.10%, and 92.06%, respectively. Slightly lower performance was obtained while employing the DWT-based features compared to employing MFCC-based features. Additionally, frequency analysis was conducted to identify specific frequency bins, which are the most indicative of different fault types in providing valuable guidance for future diagnostic efforts. Overall, the proposed framework provides a reliable and practical solution for accurate vehicle fault detection, paving the way for future advancements in automotive diagnostics.

Keywords: vehicle fault detection; extreme learning machines; mel-frequency cepstral coefficients; wavelet transform

1. Introduction

Car faults are an unfortunate reality for many drivers, with unexpected vehicle repairs being a common occurrence in the automotive industry. These faults can range from minor issues (e.g., a blown fuse, a flat tire) to more serious problems that can impact vehicle safety and performance. For instance, a faulty brake system or a malfunctioning engine can pose significant risks to drivers and passengers and can result in costly repairs. In fact, according to a report by the American Automobile Association, the average expenditure on unexpected vehicle repairs in the United States was between USD 500 and USD 600 per vehicle in 2017 [1]. However, this figure can vary widely based on several factors, such as the age, and model of the vehicle, as well as driving habits and conditions.

Traditional car fault diagnosis methods have relied heavily on the expertise of mechanics or diagnostic tools, which may be available at service centers [2,3]. These methods often involve manual inspections, computerized scanning, or diagnostic tests to identify and address potential issues in vehicles, such as engine faults, front-end assembly faults, airflow faults, spark plug, and electrical faults. However, these traditional methods are often costly, time-consuming, and may not always provide accurate results [4]. Recently,

there has been a surge of interest in machine learning-based fault diagnosis systems [5–7]. Leveraging advancements in artificial intelligence and data analytics, these systems aim to automate fault detection processes, enhance accuracy, and enable proactive maintenance strategies [8,9].

The general structure of machine learning-based fault diagnosis methods encompasses several stages, i.e., data acquisition, preprocessing, feature extraction, and classification [10,11]. Data acquisition involves gathering relevant information from vehicles, such as sensor readings and diagnostic codes, and recording relevant signals [12]. Preprocessing focuses on cleaning and organizing the data to eliminate noise and inconsistencies [13]. Feature extraction aims to identify key parameters or characteristics indicative of potential faults [14]. Finally, classification algorithms are applied to categorize the data into different fault types or states [15].

Two primary types of signals, which are vibration and sound signals, are employed in fault diagnosis [16]. Vibration signals capture mechanical movements and dynamics within the vehicle and provide insights into structural integrity and component performance. On the other hand, sound signals reflect acoustic emissions associated with engine operation and component interactions and offer valuable clues about the health and functionality of various vehicle parts.

Several studies have explored fault diagnosis using vibration signals and demonstrate the versatility and effectiveness of this approach. For example, Jegadeeshwaran and Sugumaran [17] presented a method of vibration-based continuous monitoring system and analysis using a machine learning approach. Their study focused on fault diagnosis in hydraulic braking systems by acquiring vibration signals from a piezoelectric transducer under both good and faulty brake conditions. They employed decision tree algorithms to identify the most relevant features among different faulty conditions and they achieved a classification accuracy of 97.45%. Similarly, Barbieri et al. [18] aimed to identify damages and diagnose damaged components in automotive gearboxes by comparing vibration signals of damaged and undamaged systems. They employed various signal analysis techniques (e.g., wavelet transform and mathematic morphology) to verify damage presence and used a signal processing technique combining pattern spectrum and selective filtering for component failure identification. Jafarian et al. [19] explored vibration analysis for fault detection in an internal combustion engine and focused on detecting faults related to poppet valve clearance and incomplete combustion. They utilized four accelerometers on the engine body, applied the Principal Component Analysis (PCA) technique for data analysis, and achieved high efficiency in fault classification and detection. These studies collectively highlight the wide applicability and effectiveness of vibration-based fault diagnosis in diverse automotive systems. A summary of studies employing vibration signals is given in Table 1.

Table 1. Summary of studies on vehicle fault detection using vibration signals.

Reference	Dataset	Methodology	Key Findings
Jegadeeshwaran and Sugumaran [17]	550 signals recorded from a brake system in a controlled environment, 8 brake fault types	Statistical feature extraction and decision trees	Overall accuracy of 97.5%
Barbieri et al. [18]	Multi-step recording of signals from 13 gearboxes in a lab environment, 3 gearbox fault types	Wavelet transform-based feature extraction and comparison-based classification	Significant differences between signals recorded from gearboxes with and without damage

Table 1. *Cont.*

Reference	Dataset	Methodology	Key Findings
Jafarian et al. [19]	Signals recorded from a car engine in a controlled environment	Statistical features of PCA components and comparison-based classification	High accuracy in fault classification
Ahmed et al. [20]	600 signals recorded from a car engine in a lab environment, 7 engine fault types	Time–frequency domain feature extraction and artificial neural networks	Overall accuracy of 97%
Taghizadeh-Alisaraei and Mahdavian [21]	Signals recorded from a car engine in a lab environment, in cases involving faulty and healthy injectors	Time–frequency analysis	Significant differences between signals recorded from the engine with and without faulty injectors
Wang et al. [22]	411 signals recorded from 8 valve train states in a controlled environment	Time–frequency analysis and probabilistic neural networks	Overall accuracy of 97.6%

Although high success rates have been reported in vibration signals, it is hard to implement them in a real-word system. Therefore, a practical method is required to distinguish the faults. Sound signals offer distinct advantages in fault diagnosis due to their ease of recording using commonly available devices such as cell phones or microphones. Therefore, this makes sound analysis a practical and cost-effective approach to diagnosing vehicle faults. Studies conducted by various researchers further exemplify the potential of sound-based fault diagnosis in automotive systems. For instance, Madain et al. [23] identified distinct sounds associated with specific engine malfunctions and developed an algorithm using sound techniques in diagnosis. They reported high error detection rates through analysis of engine sound samples collected from a laboratory environment. Similarly, Jian-Da Wu and Chiu-Hong Liu [24] developed a fault diagnosis system for internal combustion engines based on the discrete wavelet transform technique applied to sound emission signals and showcased its effectiveness in fault recognition under diverse engine operating conditions. Another study that delved into the application of acoustic signal processing methods for assessing internal combustion engine technical conditions proposed new algorithms for automatic detection of valve clearance issues based on acoustic signal components [25]. Additionally, Mofleh et al. [26] conducted a study aimed at detecting faults in spark-ignition engines using acoustic signals and an Artificial Neural Network (ANN) system. It highlighted the high potential of ANN-based fault detection in internal combustion engines using acoustic signals, particularly in identifying simulated spark plug and misfire faults. These studies collectively underscore the practicality and efficacy of sound-based fault diagnosis methods in the automotive industry and offer valuable insights for developing reliable diagnostic systems. A summary of such studies employing sound signals is provided in Table 2.

Despite recent advancements, current studies in vehicle fault diagnosis often encounter significant drawbacks. These limitations include the reliance on data recorded in controlled laboratory settings, which may not fully represent real-world vehicle conditions. Furthermore, many studies are constrained to specific car models or fault types, limiting the generalizability and applicability of their findings. There is a pressing need for a comprehensive dataset that encompasses diverse vehicle models, real-world vehicle conditions, and a wide range of fault scenarios to enhance the effectiveness and accuracy of fault diagnosis systems.

Table 2. Summary of studies on vehicle fault detection using sound signals.

Reference	Dataset	Methodology	Key Findings
Madain et al. [23]	Few signals recorded from two cars, 2 types of faults (shell bearing and exhaust)	RPM-based dominant frequency and comparison-based classification	100% accuracy for the first car and 90% for the second
Wu and Liu [24]	300 signals recorded from one engine, 5 engine-related faults	Wavelet transform-based feature extraction and artificial neural networks	Overall accuracy of 99%
Figlus et al. [25]	Signals recorded from two engines, one fault type (excessive valve clearance)	Wavelet transform-based feature extraction and comparison-based classification	Efficiency of the algorithm presented
Mofleh et al. [26]	60 signals recorded from one engine, 2 spark-ignition faults	Frequency domain feature extraction and artificial neural networks	Overall accuracy of 73.3%
Nevea and Sybingco [27]	36 signals recorded from the 1996–2000 model Honda Civic, 3 engine-related faults	Fourier transform and power spectrum density for feature extraction and fuzzy logic inference for classification	Overall accuracy of 56%
Wang et al. [5]	140 signal recordings from Santana 2000 model vehicle	Hilbert–Huang transform for feature extraction and support vector machines for classification	Overall accuracy of 90%
Siegel et al. [28]	992 2.5-long signals recorded from 4 cars, one type of engine fault	Fourier, Wavelet, and MFCC-based feature extraction and support vector machines	Overall accuracy of 99%
Yılmaz et al. [29]	100 signals recorded from various cars, 2 engine-related faults	Wavelet transform-based feature extraction and k-nearest neighborhood	Overall accuracy of 91.8%

Our study directly addresses these challenges by leveraging a diverse and extensive dataset comprising real-life vehicle sounds. This dataset captures a wide array of vehicle conditions various vehicle models, and an extensive range of fault scenarios encountered in everyday driving. We employ advanced signal processing techniques, including Mel-Frequency Cepstral Coefficients (MFCCs), Wavelet Transform, and Relief-F methods, for robust feature extraction and feature selection, while using Extreme Learning Machines (ELM) for the classification.

The summary of our study's approach and key contributions is as follows:

- Instead of employing a laboratory-collected dataset, comprehensive data were collected from real-life vehicle conditions and diverse vehicle models.
- In order to increase the success of the proposed approach, advanced signal processing techniques, which are MFCC, Wavelet Transform, and Relief-F, were employed.
- A thorough frequency analysis was conducted in each fault type, and specific frequency components, which are associated with different types of faults, were identified.

This paper is structured as follows. Section 2 gives details about utilized data collection methods and focuses on how sound signals were acquired from vehicular systems under various operating conditions. Section 3 elaborates on employed methodology encompassing employed signal processing techniques and feature extraction methodologies that are used in vehicle fault diagnosis based on sound signals. In Section 4, we present the obtained experimental results, which includes performance evaluations of the employed diagnostic system, comparisons with existing methods in the literature, and a detailed frequency

analysis of each identified fault type. Furthermore, the implications of our findings and insights gained from the experimental outcomes are discussed. Finally, Section 5 serves as the conclusion of this study and summarizes key contributions made in this research, and proposes directions for future research and development in the field of vehicle fault diagnosis by sound signal analysis.

2. Dataset

Audio signal recordings were collected from vehicles, which were serviced at official Ford or Toyota service centers and ensured a diverse range of cars from these reputable brands. A cellphone served as the recording device, which captured sounds, while the cars were stationary, and their engines were idling at ideal operating temperatures. As seen in Figure 1, the cellphone was positioned 15 cm above the hood and centered, with the hood closed to mimic real-world conditions. Engine sounds were recorded for 30 s each, sampling at a frequency of 48 kHz to capture detailed acoustic information.
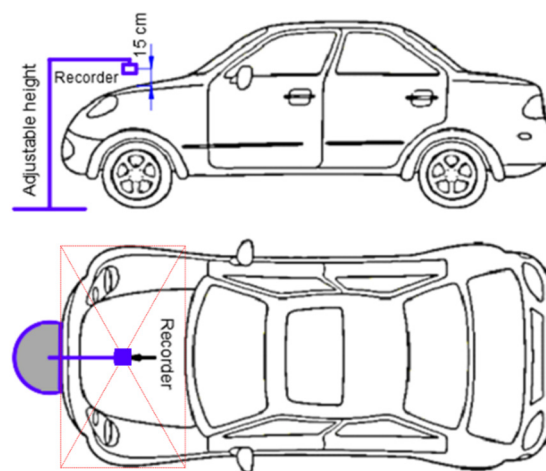

Figure 1. Setup for audio signal recording.

Professional mechanics diagnosed the cars as either healthy or with one of the following faults: spark plug issues, airflow irregularities, electrical malfunctions, engine/turbo problems, or front-end problems. The distribution of each diagnostic class is outlined in Table 3.

Table 3. Distribution of vehicle diagnostic classes.

Diagnostic Class	Number of Cases
Healthy	50
Spark Plug Issues	40
Airflow Irregularities	52
Engine/Turbo Problems	44
Front-End Problems	48
Electrical Malfunctions	46

Spark Plug Issue: Typically related to ignition problems, which result in misfires, rough idling, and decreased engine performance.

Airflow Irregularities: Pertaining to issues with the air intake system that affect engine combustion and efficiency.

Electrical Malfunctions: Encompassing faults within the vehicle's electrical system that directly impact engine performance. This may include issues with sensors, wiring, or other electrical components that affect engine operation and efficiency.

Engine/Turbo Problems: Referring to issues within the engine or turbocharger system that impact power delivery and overall engine performance.

Front-End Problems: Including issues with steering, suspension, or other components affecting the vehicle's front-end operation.

The collected dataset comprises audio signals, which were collected from a wide range of gasoline-engine vehicles such as Ford Focus (2014–2021), Ford Kuga (2020–2021), Ford Ecosport (2021), Ford Mondeo (2016), Toyota Corolla (2015), and Toyota Auris (2010). It was aimed to ensure diversity in vehicle models to capture a comprehensive range of engine sounds and fault types.

In addition, an example signal for each vehicle diagnostic class is provided in Figure 2 in order to demonstrate the characteristics of the recorded audio signals for each class.
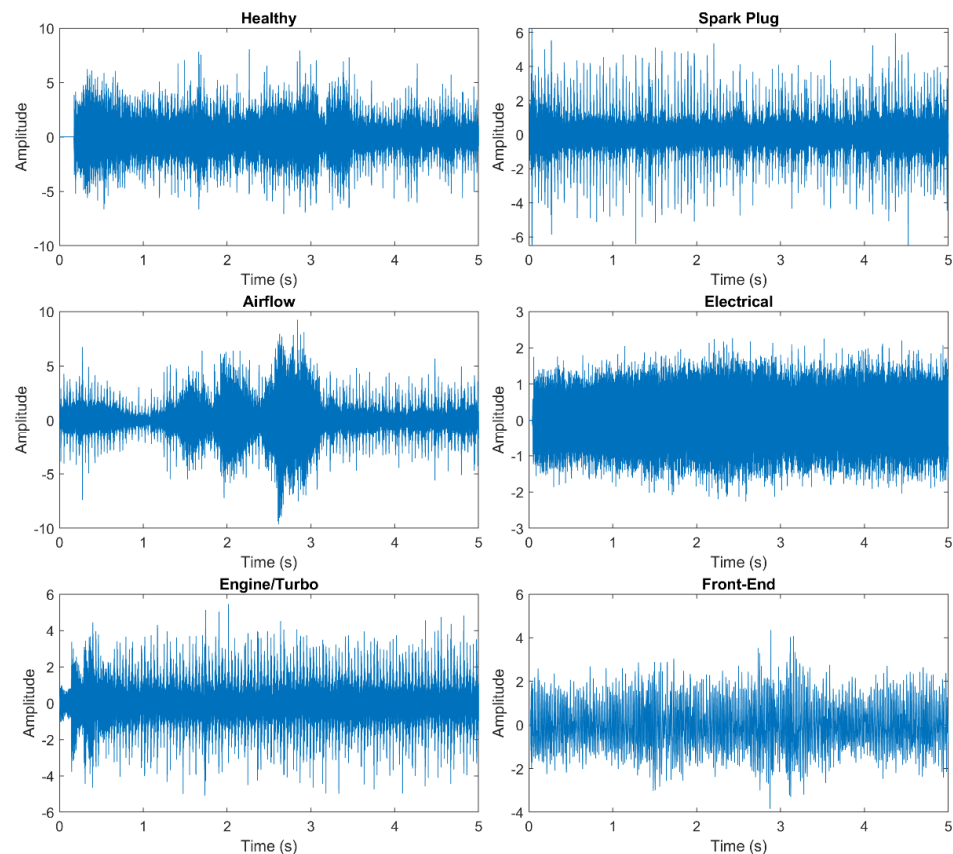

Figure 2. Example audio signals for each vehicle diagnostic class.

3. Methodology

3.1. Overview

This paper presents a framework comprising data acquisition, preprocessing, feature extraction, fine-tuning, and classification, as illustrated in Figure 3. Data acquisition details are explained in Section 2. In the preprocessing phase, each audio signal was normalized and trimmed to 5 s to ensure consistency. Feature extraction was performed using Mel-Frequency Cepstral Coefficients (MFCCs) and Discrete Wavelet Transform (DWT) techniques, capturing important acoustic characteristics for fault diagnosis. The fine-tuning process involves Grid Search optimization to find optimal hyperparameters for machine learning models, while Extreme Learning Machines (ELM) are utilized for efficient and accurate classification of vehicle fault diagnostic classes. Additionally, Fourier transform-based feature extraction and Relief-f feature selection algorithm are employed for frequency analysis of each fault type, aimed at uncovering crucial frequency components associated with different fault types.
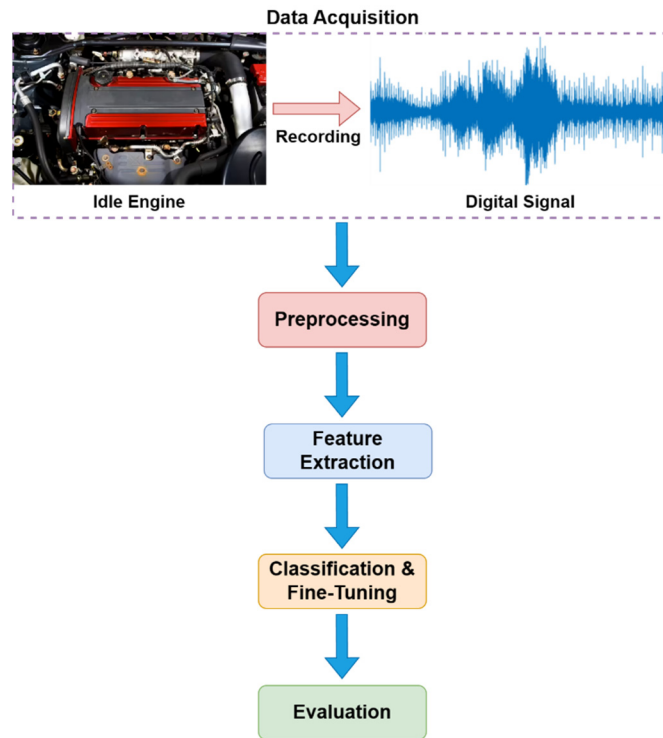
Figure 3. Methodology block diagram for vehicle fault diagnosis using sound signals and machine learning.

3.2. Feature Extraction

Feature extraction is a critical step in audio signal processing for fault diagnosis in vehicle systems. This section outlines two widely used techniques: MFCC and DWT.

MFCC is a prominent technique in audio signal processing. The process involves several steps [30]:

- Windowing the audio signal into short segments using Equation (1).

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \qquad N - 1 \geq n \geq 0 \qquad (1)$$

- Applying the Discrete Fourier Transform (DFT) to each frame to convert the time-domain signal into the frequency domain. DFT can be defined for a frame X comprising of N samples as in Equation (2).

$$X_n = \sum_{k=0}^{N-1} X_k e^{-\frac{2\pi jkn}{N}} \ , \qquad n = 0, 1, 2, 3 \ldots, N-1 \qquad (2)$$

- Mapping the frequency spectrum to the Mel-scale to approximate human auditory perception using Equation (3).

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f_{lineer}}{700}\right) \qquad (3)$$

- Applying logarithmic compression to the Mel spectrum.
- Computing the Discrete Cosine Transform (DCT) to obtain the MFCC coefficients as follows:

$$C_k = \sum_{n=0}^{N-1} \log(S_n) \cos\left[k\pi \frac{(2n+1)}{2N}\right] \qquad (4)$$

where C_k represents the k-th MFCC coefficient, N represents the total number of mel-frequency filters, S_n represents the energy of the n-th mel-frequency filter bank, and k is the index of the MFCC coefficient, usually ranging from 0 to $k-1$.

These MFCC coefficients capture essential spectral features of the audio signal, such as pitch, timbre, and formants, which are crucial for fault diagnosis in vehicle systems. In this study, the mean of each coefficient over each frame was calculated as a feature. Details regarding MFCC parameters are provided in Section 4.

DWT is a powerful tool for analyzing signals in both time and frequency domains simultaneously [31]. The process involves the following:

- Decomposing the audio signal into different frequency bands using wavelet functions, such as Daubechies and Symlets wavelets. For an input signal $x(t)$, the approximation coefficients A_j and detail coefficients D_j at level j are computed using:

$$A_{j+1}(k) = \sum_n h(n - 2k) A_j(n) \tag{5}$$

$$D_{j+1}(k) = \sum_n g(n - 2k) A_j(n) \tag{6}$$

where h and g are the low-pass and high-pass filter coefficients, respectively, corresponding to the wavelet function.
- Extracting features from each level of decomposition.

These features provide insights into the time–frequency characteristics of the signal, aiding in fault detection and classification in vehicle systems. In this study, we calculated the energy, standard deviation, and entropy of each coefficient to serve as features for the classification task. Details regarding the decomposition levels and wavelets employed are given in Section 4.

3.3. Classification and Fine-Tuning

For the classification of vehicle fault diagnostic classes, we employed ELM, a machine learning algorithm based on a single hidden layer feedforward neural network architecture [32–35]. In the ELM algorithm, the input layer weights and thresholds are assigned randomly, while the output layer weights are calculated based on these assignments. ELM training consists of two parts: (1) generating random hidden layer parameters from a predefined range, and (2) calculating the generalized inverse output weight matrix [36]. ELM is popular due to its fast learning speed, generalization ability, and simplicity. As illustrated in Figure 4, the ELM inputs map the features to the hidden layer, which are then passed on to the output layer. The output from ELM learning can be used for various tasks such as classification, regression, and clustering.

ELM transforms input vector $x = [x_1, x_2, \ldots, x_n]$ into the hidden layer representation using the weight matrix W and the bias vector b. Each neuron in the hidden layer uses an activation function $g(.)$. The connections between the hidden layer outputs and the output layer are represented by the weights β. In ELM, the weights and bias values for the hidden layer are randomly assigned and kept fixed. The weights β for the output layer are learned using the least squares method. The relationship between the input and output of the ELM is calculated as summarized in Equations (7)–(9).

$$H = g(W \cdot X + b) \tag{7}$$

$$Y = H \cdot \beta \tag{8}$$

$$\beta = H^+ \cdot T \tag{9}$$

where H is the matrix of hidden layer outputs, Y denotes the output vector, H^+ represents the Moore–Penrose pseudoinverse of the matrix H, and T is the target vector.
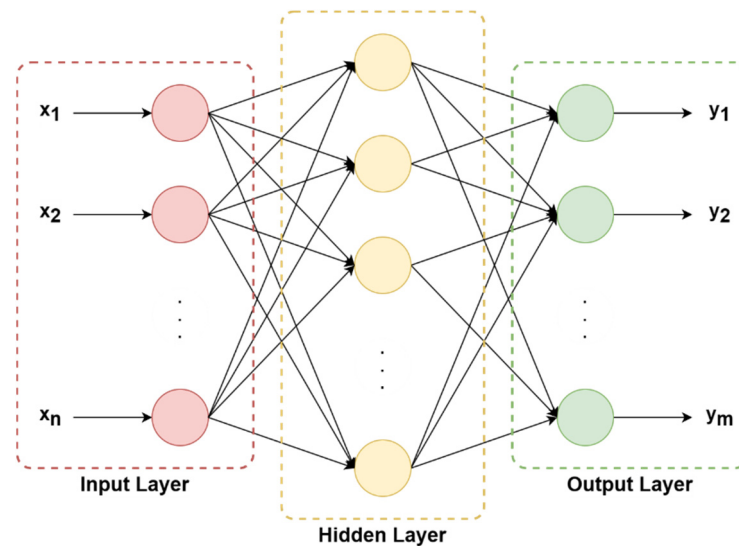

Figure 4. Structure of an ELM model.

To optimize the performance of our ELM classifier, we utilized Grid Search, a hyperparameter optimization technique [37]. Grid Search systematically explores a predefined set of hyperparameters, evaluating each combination using 5-fold cross-validation to determine the optimal parameters that yield the highest classification accuracy. This method allows for a thorough examination of the model's performance across various settings, ensuring the selection of the most effective parameter set. The specific parameters used for Grid Search, such as the number of hidden neurons and activation functions, are outlined in Table 4. This exhaustive search process ensures that our model is finely tuned, enhancing its robustness and reliability in delivering accurate fault diagnosis results.

Table 4. Grid search parameters for optimizing ELM classifier.

Hyperparameter	Values
Number of hidden neurons	2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100
Activation Function	sigmoid, sine, triangular basis, radial basis, and hard limit function

3.4. Performance Evaluation

The performance of the classification model is assessed using standard metrics, including accuracy, precision, recall, macro, and weighted averaged F1-scores. These metrics provide a comprehensive understanding of the model's effectiveness in correctly diagnosing vehicle faults. Accuracy measures the overall correctness of the model, precision indicates the proportion of true positive diagnoses out of all positive diagnoses, recall (or sensitivity) assesses the model's ability to identify true positives, and F1-score is the harmonic mean of precision and recall, offering a balanced evaluation metric [38]. Each metric is calculated as outlined in Equations (10)–(14). Five-fold cross-validation is employed in the performance analysis to provide a reliable estimate of the model's effectiveness across different subsets of the data. This method helps in ensuring that the evaluation is not biased by any particular partitioning of the data.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{10}$$

$$Precision\ (PR) = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \tag{11}$$

where N is the number of classes, TP_i is the number of true positives for class i and FP_i is the number of false positives for class i.

$$Recall\ (RE) = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \tag{12}$$

where FN_i is the number of false negatives for class i.

$$Macro\ F1\text{-}Score = \frac{1}{N}\sum_{i=1}^{N}\frac{2 \cdot PR_i \cdot RE_i}{PR_i + RE_i} \tag{13}$$

$$Weighted\ F1\text{-}Score = \sum_{i=1}^{N}w_i\frac{2 \cdot PR_i \cdot RE_i}{PR_i + RE_i} \tag{14}$$

where PR_i and RE_i are the precision and recall for class i, respectively. Additionally, confusion matrices, which display the model's classification results, were constructed to provide a detailed view of how well the classifier distinguishes between fault classes. This visual representation is crucial for understanding the classifier's strengths and areas needing improvement.

3.5. Relief-F Algorithm and Frequency Analysis

In this section, we detail the process of feature selection using the Relief-F algorithm and the subsequent frequency analysis conducted to identify the most relevant frequency components for each fault type.

The Relief-F algorithm is a powerful feature selection method that helps identify and rank the most relevant features in a dataset, making it particularly useful for high-dimensional data [39]. Relief-F works by iteratively sampling instances from the dataset and comparing the sampled instance with its nearest neighbors of the same and different classes. For each feature, it increases the relevance score if the feature value of the instance is similar to that of its nearest neighbor from the same class and decreases the score if the feature value is similar to that of its nearest neighbor from a different class. This process effectively highlights features that are consistently good at distinguishing between classes while ignoring irrelevant or redundant features.

To understand the frequency characteristics of the audio signals and identify the most diagnostically relevant frequencies for each fault type, we conducted a detailed frequency analysis using the following steps:

1. Frequency Spectrum Calculation: The frequency spectrum of each audio signal was calculated using the Fast Fourier Transform (FFT). FFT transforms the time-domain signal into its frequency-domain representation, providing insight into the signal's frequency components.
2. Spectrum Binning: The resulting frequency spectrum was divided into bins, each 100 Hz wide. This binning process organizes the frequency data into manageable segments. As there are very limited frequency components beyond 15 kHz, a frequency range of 0–15 kHz was considered.
3. Power Calculation: For each 100 Hz bin, the total power was calculated. This step involves summing the squared magnitudes of the frequency components within each bin, giving a measure of the signal's energy within that frequency range.
4. Relevance Determination: For each fault type, the Relief-F feature selection algorithm was applied to the binned frequency spectrum. By evaluating the relevance scores of the bins, Relief-F identified the 5 most relevant bins that contributed most significantly

to distinguishing between faulty and healthy conditions. These relevant bins highlight specific frequency components that are key indicators of each fault type.

4. Results and Discussion

In this section, the vehicle fault diagnosis framework was presented and discussed. The performance metrics and analysis are provided for models utilizing MFCC-based features and DWT-based features. Additionally, a detailed frequency analysis to identify the most relevant frequency components for each fault type, was given. Obtained results are also compared with existing studies to highlight the effectiveness and improvements achieved by the employed methodology.

4.1. Results for MFCC-Based Features

The performance of the vehicle fault diagnosis model using MFCC-based features is evaluated and summarized in Table 5. Various configurations of MFCC parameters, including the number of coefficients and window length, were tested to identify the optimal settings. The table also presents the best hyperparameters found through Grid Search for the ELM classifier, such as the activation function and number of neurons, alongside the resulting precision, recall, F1-score, and accuracy metrics. A 50% window overlap was used in all experiments to enhance feature extraction.

Table 5. Performance results for MFCC-Based features. Values given for window length are in seconds. 50% window overlap is used for all experiments. The bold indicates the highest score.

MFCC Parameters		Best ELM Hyperparameters		Performance Results				
Number of Coefs.	Window Length	Activation Function	Number of Neurons	Precision	Recall	Macro F1-Score	Weighted F1-Score	Accuracy
5	0.02	tribas	100	86.19	85.60	85.70	85.72	85.71
10	0.02	sig	100	90.09	89.24	89.34	89.31	89.29
20	0.02	sin	100	89.95	90.02	89.91	89.88	90.00
30	0.02	sin	100	89.81	89.73	89.51	89.45	89.64
5	0.03	radbas	80	85.74	84.59	84.73	84.73	84.64
10	0.03	sig	90	89.71	89.83	89.69	89.65	89.64
20	0.03	sin	100	**92.24**	**92.22**	**92.10**	**92.06**	**92.14**
30	0.03	sin	90	90.43	89.77	89.65	89.59	89.64
5	0.04	tribas	100	88.37	87.67	87.87	87.98	87.86
10	0.04	sig	90	91.45	91.12	91.11	91.11	91.07
20	0.04	sig	100	91.14	90.79	90.69	90.61	90.71
30	0.04	sig	80	89.67	89.37	89.16	89.00	89.29

The results, which are given in Table 5, indicate that using 20 MFCC coefficients with a window length of 0.03 s and a *sine* activation function yielded the highest performance with a precision of 92.24%, recall of 92.22%, F1-score of 92.10%, and accuracy of 92.14%. This demonstrates that the choice of MFCC parameters and ELM hyperparameters significantly impacts the classification performance.

From the table, it can be observed that the number of MFCC coefficients and window length significantly affect the performance of the proposed method up to a certain point. Specifically, increasing the number of MFCC coefficients from 5 to 20 generally leads to an improvement in precision, recall, F1-score, and accuracy. However, further increasing the number of coefficients to 30 does not seem to result in any significant improvement. Similarly, increasing the window length from 0.02 to 0.03 or 0.04 generally leads to an improvement in performance, with the best results achieved at a window length of 0.03 for most MFCC parameter configurations. However, the choice of activation function and the number of neurons in the ELM classifier also seem to play a role in achieving the best performance.

The confusion matrix for the model with 20 MFCC coefficients and a window length of 0.02 s, shown in Figure 5, provides additional insights into the model's performance. Each cell in the matrix represents the number of instances for which the true class is represented by the row and the predicted class is represented by the column.
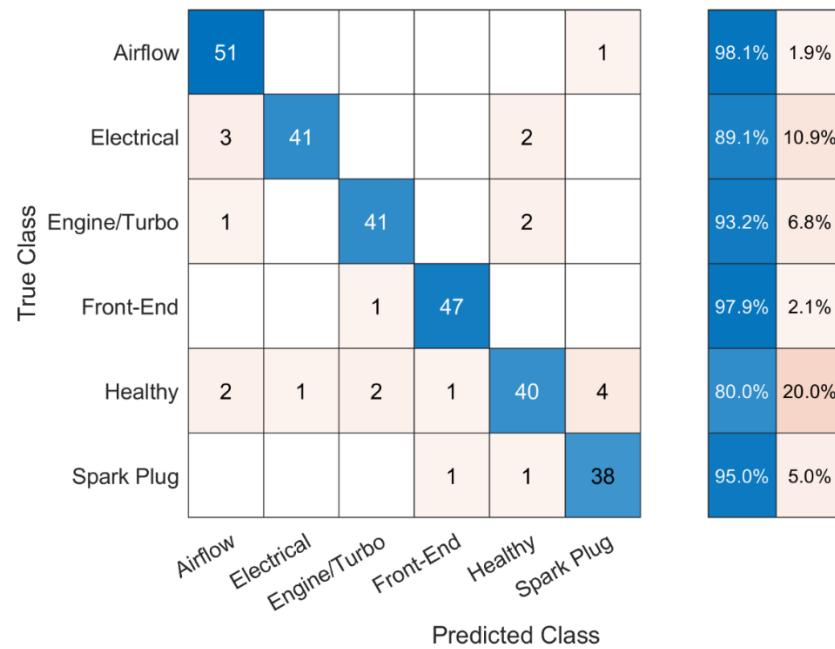

Figure 5. Confusion matrix for 20 MFCC coefficients and window length of 0.02 s using MFCC-based features.

Overall, the confusion matrix confirms that the model performs well across most fault types, with particularly high accuracy for airflow, front-end, and spark plug faults. Specifically, the model achieves the highest performance in detecting airflow faults with an accuracy of 98.1%, correctly identifying 51 out of 52 cases. However, the model shows the lowest performance in identifying healthy cases, with an accuracy of 80.0%, correctly predicting 40 out of 50 instances. This suggests that healthy cases are more prone to being misclassified as faults, indicating an area for further refinement.

The results of the experiments using MFCC-based features demonstrate the effectiveness of the proposed method for fault diagnosis in vehicles. The method can accurately diagnose different types of faults using the extracted MFCC features and the ELM classifier. The results also provide insights into the optimal combination of MFCC parameters and ELM hyperparameters, which can be used to improve the performance of the method in future studies. These findings demonstrate the model's effectiveness in vehicle fault diagnosis and highlight areas for further optimization.

4.2. Results for DWT-Based Features

The performance of the vehicle fault diagnosis model using DWT-based features is evaluated and summarized in Table 6. Various configurations of DWT parameters, including the decomposition level and widely used wavelets such as db4, db8, db20, sym3, and sym8, were tested to identify the optimal settings [24,40]. The table also presents the best hyperparameters found through Grid Search for the ELM classifier, such as the activation function and number of neurons, alongside the resulting precision, recall, F1-score, and accuracy metrics.

Table 6. Performance results for DWT-based features. The bold indicates the highest score.

DWT Parameters		Best ELM Hyperparameters		Performance Results				
Level	Wavelet	Activation Function	Number of Neurons	Precision	Recall	Macro F1-Score	Weighted F1-Score	Accuracy
2	db4	sig	80	82.11	81.09	81.30	81.69	81.43
3	db4	sin	90	82.84	80.63	81.11	81.43	81.07
4	db4	sin	100	78.46	77.44	77.24	77.25	77.50
5	db4	sin	100	81.72	80.95	80.79	80.98	81.07
2	db8	sig	90	82.22	80.99	81.19	81.47	81.43
3	db8	sin	70	83.51	82.42	82.64	82.77	82.86
4	db8	sin	100	77.57	77.16	76.94	77.09	77.14
5	db8	sin	100	79.62	78.05	78.29	78.25	78.21
2	db20	sin	100	**84.88**	83.27	83.54	83.52	83.57
3	db20	sin	80	83.00	82.37	82.45	82.46	82.50
4	db20	sin	100	77.68	75.93	76.00	76.04	76.07
5	db20	sin	100	80.57	79.93	79.97	80.21	80.00
2	sym3	sig	80	81.13	80.58	80.48	80.63	80.71
3	sym3	sin	90	84.36	83.32	83.46	83.52	83.57
4	sym3	sin	100	74.76	74.39	74.19	74.05	74.29
5	sym3	sig	50	78.48	77.46	77.43	75.55	77.50
2	sym8	sin	80	83.19	82.33	82.42	82.45	82.50
3	sym8	sin	70	84.17	**83.60**	**83.72**	**83.86**	**83.93**
4	sym8	sin	100	76.20	75.77	75.37	75.52	75.71
5	sym8	sin	100	77.10	75.98	76.07	76.26	76.07

From the table, it is clear that different combinations of DWT parameters and ELM hyperparameters result in varying levels of performance. The highest accuracy was achieved using a decomposition level of 3 and a *sym8* wavelet with a *sine* activation function, yielding a precision of 84.17%, recall of 83.60%, F1-score of 83.72%, and accuracy of 83.93%. This suggests that the choice of wavelet and decomposition level significantly influences the classification performance.

The confusion matrix in Figure 6 provides further insights into the classification performance for the best configuration (decomposition level of 3 and sym8 wavelet). The model achieves high accuracy for electrical faults (93.5%) and front-end faults (91.7%), indicating strong performance in these categories. However, the model shows lower accuracy for spark plug faults (65.0%), suggesting that distinguishing spark plug faults from other fault types remains a challenge. Additionally, healthy cases are identified with an accuracy of 74.0%, indicating some misclassification into fault categories, which highlights an area for further improvement.

Overall, the results of the experiments using DWT-based features demonstrate the effectiveness of the proposed method for fault diagnosis in vehicles. While the method shows strong performance in certain fault categories, it achieved lower performance overall compared to MFCC-based features. These findings provide insights into the optimal combination of DWT parameters and ELM hyperparameters and highlight areas for further optimization to enhance the performance of the method in future studies.
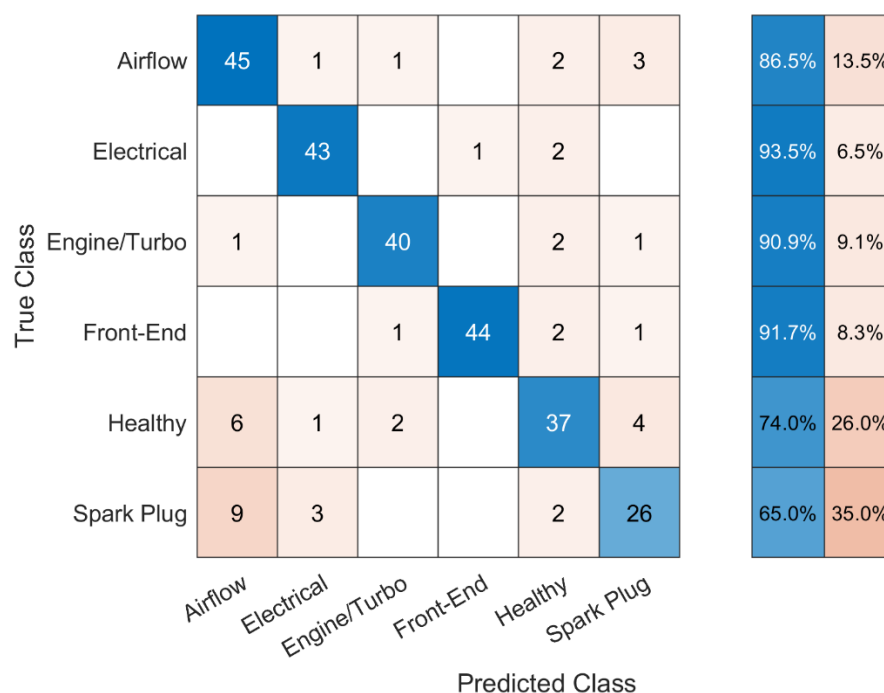
Figure 6. Confusion matrix for decomposition level of 3 and sym8 wavelet using DWT-based features.

4.3. Results for Frequency Analysis

In this section, the results of the frequency analysis, which was conducted to identify the most relevant frequency components for each fault type, are presented. Different types of engine faults often exhibit distinct frequency components. These components emerge due to the engine's physical structure and operational principles, with each fault generating unique sounds or vibrations at specific frequency ranges. Therefore, examining these frequency components is crucial for accurate fault diagnosis using sound analysis. Table 7 comprehensively outlines the relationship between the fault categories and their associated frequency groups.

Table 7. The most relevant frequencies for each fault type.

Fault Type	Most Relevant Frequency Bins (kHz)				
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Spark Plug Issues	5.2–5.3	9.4–9.5	12.4–12.5	13.2–13.3	14.1–14.2
Airflow Irregularities	3.0–3.1	9.1–9.2	9.2–9.3	10.0–10.1	12.2–12.3
Engine/Turbo Problems	0.7–0.8	4.8–4.9	9.5–9.6	12.1–12.2	12.5–12.6
Front-End Problems	6.6–6.7	8.5–8.6	8.9–9.0	11.6–11.7	13.5–13.6
Electrical Malfunctions	2.1–2.2	11.2–11.3	11.4–11.5	12.1–12.2	14.3–14.4

For spark plug issues, the most relevant frequency bin is 5.2 to 5.3 kHz, indicating that monitoring this high-frequency range is crucial for accurate detection. Airflow irregularities are most prominently indicated by the 3.0 to 3.1 kHz bin, highlighting the need for precise analysis in this low-frequency range. Engine/turbo problems are best identified by the 0.7 to 0.8 kHz bin, suggesting these faults manifest through specific low-frequency sounds. Front-end problems are primarily associated with the 6.6 to 6.7 kHz bin, essential for identifying issues related to components such as the suspension or chassis. Electrical malfunctions affecting the engine show the highest relevance at the 2.1 to 2.2 kHz bin, possibly due to distinctive noise patterns. These findings underscore the importance of focusing on these key frequencies for accurate fault detection.

Overall, the frequency analysis confirms that different fault types are associated with specific and most relevant frequency bins. The identification of these relevant frequency bins provides valuable guidance for future diagnostic efforts, suggesting that including these specific frequencies in the analysis can lead to better and more reliable fault detection outcomes. This insight into the frequency components of various faults enhances our understanding and ability to diagnose vehicle issues more accurately and efficiently.

4.4. Comparison with Other Studies

Numerous publications have addressed the diagnosis of vehicle malfunctions through engine sound analysis, employing a diverse array of methods. For instance, Navea and Sybingco [27] used Fourier transform and power spectral density to detect engine starting issues, drive belt problems, and valve-related faults, achieving a detection accuracy of 56% using recordings from the 1996–2000 Honda Civic model. Siegel et al. [28] focused on misfire faults in four vehicles, using Fourier transform, wavelet transform, and MFCC with SVM, resulting in a 99% accuracy rate. Wang et al. [5] recorded audio from a Santana 2000's engine cylinder head, using Hilbert–Huang transform and SVMs, achieving up to 90% accuracy. Kemalkar and Bairagi [41] studied lubrication, chain, crank, and valve faults in Honda Unicorn and Bajaj Pulsar motorcycles, employing MFCC and achieving accuracy rates between 50% and 75%. The reported results in the literature are summarized in Table 8.

Table 8. Reported results in the literature.

Reference	Dataset	Methodology	Key Findings
Wang et al. [5]	140 signal recordings from Santana 2000 model vehicle	Hilbert–Huang transform for feature extraction and support vector machines for classification	Overall accuracy of 90%
Nevea and Sybingco [27]	36 signals recorded from the 1996–2000 model Honda Civic, 3 engine-related faults	Fourier transform and power spectrum density for feature extraction and fuzzy logic inference for classification	Overall accuracy of 56%
Siegel et al. [28]	992 2.5-long signals recorded from 4 cars, one type of engine fault	Fourier, wavelet transform, and MFCC-based feature extraction and support vector machines	Overall accuracy of 99%
Kemalkar and Bairagi [41]	Sound samples of motorcycles under idling conditions are recorded using a voice recorder with 44.1-kHz sampling frequency and 16-bit quantization.	Liner predictive coding, hidden Markov model, artificial neural network	Accuracy rates between 50% and 75

A significant drawback of previous works is their reliance on data from controlled laboratory settings, often using the same brand of vehicles or a limited number of fault types. These controlled conditions do not adequately capture the variability and complexity of real-world scenarios, leading to models that may not perform well outside the specific conditions under which they were trained. Furthermore, the homogeneity of vehicle models in these studies limits the generalizability of their findings, as the diagnostic methods may not be applicable to different vehicle makes and models.

This dataset includes various types of faults and different environmental settings, thereby enhancing the robustness and generalizability of our fault diagnosis framework. By utilizing data from multiple vehicle brands and real-world conditions, the employed approach is designed to reflect the true complexity and variability of vehicle faults.

5. Conclusions

The proposed study presents a comprehensive vehicle fault diagnosis framework utilizing MFCC-based features and DWT-based features. The results demonstrate the efficacy of MFCC features combined with an ELM classifier, achieving the highest performance metrics. DWT-based features, while effective, showed slightly lower performance compared to MFCC features. Frequency analysis identified specific frequency bins most indicative of different fault types, providing valuable guidance for future diagnostic efforts. Additionally, by addressing the limitations of previous studies through the introduction of a large, diverse dataset encompassing various vehicle models and real-world operating conditions, we have significantly improved the robustness and generalizability of our fault diagnosis system. This framework provides a reliable and practical solution for accurate vehicle fault detection, paving the way for future advancements in automotive diagnostics.

References

1. Edmonds, E. One-in-Three U.S. Drivers Cannot Pay for an Unexpected Car Repair Bill. American Automobile Association, 2017. Available online: https://newsroom.aaa.com/2017/04/one-three-u-s-drivers-cannot-pay-unexpected-car-repair-bill (accessed on 28 April 2024).
2. Karaman, E.; Rende, H.; Akşahin, M.F. Recognition of Vehicles from Their Engine Sound. *Mühendis Ve Makina* **2019**, *60*, 148–164.
3. Xu, L.; Wang, T.; Xie, J.; Yang, J.; Gao, G. A Mechanism-Based Automatic Fault Diagnosis Method for Gearboxes. *Sensors* **2022**, *22*, 9150. [CrossRef] [PubMed]
4. Liu, Y.; Zhang, J.; Ma, L. A fault diagnosis approach for diesel engines based on self-adaptive WVD, improved FCBF and PECOC-RVM. *Neurocomputing* **2016**, *117*, 600–611. [CrossRef]
5. Wang, Y.; Ma, Q.; Zhu, Q.; Liu, X.; Zhao, L. An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine. *Appl. Acoust.* **2014**, *75*, 1–9. [CrossRef]
6. Feng, Z.; Zhang, D.; Zuo, M.J. Planetary Gearbox Fault diagnosis via Joint Amplitude and Frequency Demodulation Analysis Based on Variational Mode Decomposition. *Appl. Sci.* **2017**, *7*, 775. [CrossRef]
7. López-Torres, C.; Riba, J.-R.; Garcia, A.; Romeral, L. Detection of Eccentricity Faults in Five-Phase Ferrite-PM Assisted Synchronous Reluctance Machines. *Appl. Sci.* **2017**, *7*, 565. [CrossRef]
8. Qu, Y.; He, M.; Deutsch, J.; He, D. Detection of Pitting in Gears Using a Deep Sparse Autoencoder. *Appl. Sci.* **2017**, *7*, 515. [CrossRef]
9. Gao, C.; Xue, W.; Ren, Y.; Zhou, Y. Numerical Control Machine Tool Fault Diagnosis Using Hybrid Stationary Subspace Analysis and Least Squares Support Vector Machine with a Single Sensor. *Appl. Sci.* **2017**, *7*, 346. [CrossRef]
10. Lupea, I.; Lupea, M.; Coroian, A. Helical Gearbox Defect Detection with Machine Learning Using Regular Mesh Components and Sidebands. *Sensors* **2024**, *24*, 3337. [CrossRef]
11. Moshrefi, A.; Tawfik, H.H.; Elsayed, M.Y.; Nabki, F. Industrial Fault Detection Employing Meta Ensemble Model Based on Contact Sensor Ultrasonic Signal. *Sensors* **2024**, *24*, 2297. [CrossRef]
12. Morenas, J.d.L.; Moya-Fernández, F.; López-Gómez, J.A. The Edge Application of Machine Learning Techniques for Fault Diagnosis in Electrical Machines. *Sensors* **2023**, *23*, 2649. [CrossRef]
13. Qu, N.; Wei, W.; Hu, C. Series Arc Fault Detection Based on Multimodal Feature Fusion. *Sensors* **2023**, *23*, 7646. [CrossRef] [PubMed]

14. Yang, X.; Yang, J.; Jin, Y.; Liu, Z. A New Method for Bearing Fault Diagnosis across Machines Based on Envelope Spectrum and Conditional Metric Learning. *Sensors* **2024**, *24*, 2674. [CrossRef]

15. Abid, A.; Khan, M.T.; Iqbal, J. A review on fault detection and diagnosis techniques: Basics and beyond. *Artif. Intell. Rev.* **2021**, *54*, 3639–3664. [CrossRef]

16. Baydar, N.; Ball, A. Detection of Gear Failures via Vibration and Acoustic Signals Using Wavelet Transform. *Mech. Syst. Signal Process.* **2003**, *17*, 787–804. [CrossRef]

17. Jegadeeshwaran, R.; Sugumaran, V. Method and Apparatus for Fault Diagnosis of Automobile Brake System Using Vibration Signals. *Recent Patents Signal Process.* **2013**, *3*, 2–11. [CrossRef]

18. Barbieri, N.; Barbieri, G.d.S.V.; Martins, B.M.; Barbieri, L.d.S.V.; de Lima, K.F. Analysis of automotive gearbox faults using vibration signal. *Mech. Syst. Signal Process.* **2019**, *129*, 148–163. [CrossRef]

19. Jafarian, K.; Darjani, M.; Honarkar, Z. Vibration analysis for fault detection of automobile engine using PCA technique. In Proceedings of the 2016 4th International Conference on Control, Instrumentation, and Automation (ICCIA), Qazvin, Iran, 27–28 January 2016.

20. Ahmed, R.; El Sayed, M.; Gadsden, S.A.; Tjong, J.; Habibi, S. Automotive Internal-Combustion-Engine Fault Detection and Classification Using Artificial Neural Network Techniques. *IEEE Trans. Veh. Technol.* **2015**, *64*, 21–33. [CrossRef]

21. Taghizadeh-Alisaraei, A.; Mahdavian, A. Fault detection of injectors in diesel engines using vibration time-frequency analysis. *Appl. Acoust.* **2019**, *143*, 48–58. [CrossRef]

22. Wang, C.; Zhang, Y.; Zhong, Z. Fault diagnosis for diesel valve trains based on time–frequency images. *Mech. Syst. Signal Process.* **2008**, *22*, 1981–1993. [CrossRef]

23. Madain, M.; Al-Mosaiden, A.; Al-khassaweneh, M. Fault diagnosis in vehicle engines using sound recognition techniques. In Proceedings of the 2010 IEEE International Conference on Electro/Information Technology, Normal, IL, USA, 20–22 May 2010.

24. Wu, J.-D.; Liu, C.-H. Investigation of engine fault diagnosis using discrete wavelet transform and neural network. *Expert Syst. Appl.* **2008**, *35*, 1200–1213. [CrossRef]

25. Figlus, T.; Liščák, Š.; Wilk, A.; Łazarz, B. Condition monitoring of engine timing system by using wavelet packet decomposition of a acoustic signal. *J. Mech. Sci. Technol.* **2014**, *28*, 1663–1671. [CrossRef]

26. Mofleh, A.; Shmroukh, A.; Ghazaly, N. Fault Detection and Classification of Spark Ignition Engine Based on Acoustic Signals and Artificial Neural Network. *Int. J. Mech. Prod. Eng. Res. Dev.* **2020**, *10*, 5571–5578.

27. Navea, R.; Sybingco, E. Design and Implementation of an Acoustic-Based Car Engine Fault Diagnostic System in the Android Platform. In Proceedings of the International Research Conference in Higher Education, Manila, Philippines, 26–28 October 2013.

28. Siegel, J.; Kumar, S.; Ehrenberg, I.; Sarma, S. Engine Misfire Detection with Pervasive Mobile Audio. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Turin, Italy, 19–23 September 2016.

29. Yılmaz, G.; Mete, N.F.; Umugabekazi, U.; Aydemir, Ç. Dalgacık Dönüşümü ve Özbağlanım Model Parametreleri Öznitelikleri ile Otomobil Motor Seslerinden Arıza Tespiti. *J. Investig. Eng. Technol.* **2020**, *3*, 48–54.

30. Abdul, Z.K.; Al-Talabani, A.K. Mel Frequency Cepstral Coefficient and its Applications: A Review. *IEEE Access* **2022**, *10*, 122136–122158. [CrossRef]

31. Weeks, M.; Bayoumi, M. Discrete Wavelet Transform: Architectures, Design and Performance Issues. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **2003**, *35*, 155–178. [CrossRef]

32. Huang, G.; Huang, G.-B.; Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *64*, 32–48. [CrossRef] [PubMed]

33. Chegni, A.M.; Ghavami, B.; Eftekhari, M. A GPU-based accelerated ELM and deep-ELM training algorithms for traditional and deep neural networks classifiers. *Intell. Syst. Appl.* **2022**, *15*, 200098. [CrossRef]

34. Qureshi, S.A.; Hussain, L.; Alshahrani, H.M.; Abbas, S.R.; Nour, M.K.; Fatima, N.; Khalid, M.I.; Sohail, H.; Mohamed, A.; Hilal, A.M. Gunshots Localization and Classification Model Based on Wind Noise Sensitivity Analysis Using Extreme Learning Machine. *IEEE Access* **2022**, *10*, 87302–87321. [CrossRef]

35. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]

36. Zhang, J.; Li, Y.; Xiao, W.; Zhang, Z. Robust extreme learning machine for modeling with unknown noise. *J. Frankl. Inst.* **2020**, *357*, 9885–9908. [CrossRef]

37. Bashir, M.B.; Latiff, M.S.B.A.; Coulibaly, Y.; Yousif, A. A survey of grid-based searching techniques for large scale distributed data. *J. Netw. Comput. Appl.* **2016**, *60*, 170–179. [CrossRef]

38. Górny, K.; Kuwałek, P.; Pietrowski, W. Increasing Electric Vehicles Reliability by Non-Invasive Diagnosis of Motor Winding Faults. *Energies* **2021**, *14*, 2510. [CrossRef]

39. Kira, K.; Rendell, L.A. A Practical Approach to Feature Selection. In *Machine Learning Proceedings 1992*; Sleeman, D., Edwards, P., Eds.; Morgan Kaufmann: San Francisco, CA, USA, 1992; pp. 249–256.

40. Azadi, S.; Soltani, A. Fault detection of vehicle suspension system using wavelet analysis. *Veh. Syst. Dyn.* **2009**, *47*, 403–418. [CrossRef]

41. Kemalkar, A.K.; Bairagi, V.K. Engine fault diagnosis using sound analysis. In Proceedings of the 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India, 9–10 September 2016.