Grenze International Journal of Engineering and Technology, June Issue

Audio-Driven Predictive Maintenance for Electric Vehicles: A Convolutional Neural Network Approach

Sreerama Samartha JG[1], Dayananda GK[2] and Vayusutha M[3]
[1-2]Faculty of Electronics & Communication Engineering Canara Engineering College, Benjanapadavu Manglore KA India
Email: jgssb85@gmail.com, dayanand.gk@gmail.com
[3]Faculty of Electronics & Communication Engineering PESIT Shivamoga KA India
Email: vayu.sutha@gmail.com

***Abstract*— The traditional fault detection methods in a vehicle are either based on the experience of the mechanics or the service center diagnostic tools. They are expensive, time-consuming, and also provide dubious accuracy most of the time. To avoid the above drawbacks, this work suggests a new sound-based early fault detection system for electric vehicle motors using Convolutional Neural Networks. The motor's acoustic signals recorded for different operating conditions will be utilized in the system so that fault patterns would be identified and possible failures would be predicted. To this end, the major input features for the CNN model which would enhance the accuracy and reliability of the diagnosis are MFCCs and spectral features derived from the sound records. Such a system will be able to differentiate normal operation, heavy loaded state, and motor fault with high accuracy through providing adequate training on the CNN by larger, diverse sets of motor sounds both for healthy and faulty motors. Results we got like Accuracy 93%, Precision 85%, Recall 90%, F1-Score 0.87, AUC 0.92 Processing Time 200 ms per audio clip and Cross-Validation Accuracy 92%-94%. It concluded from the result that our system is good in terms of F1-score, precision, recall, and accuracy with an excellent overall classification accuracy of more than 90%. Another contribution of this research work is thorough frequency analysis, which lead to identification of prominent sound features indicative of some faults, thus proving helpful in the development of more sophisticated diagnostic tools. Finally, this research finds a low-cost, non-intrusive, and potentially reliable solution for early fault detection in electric vehicle motors-which would be used as the basis for real-time condition monitoring and for predictive maintenance-offering a means to further refining ideas on automobile diagnostics.**

***Index Terms*— Electric Vehicle (EV) Motors Convolutional Neural Network (CNN); Mel-Frequency Cepstral Coefficients.**

I. INTRODUCTION

The usual issue that vehicle users encounter include car faults. Breakdowns tend to occur unexpectedly at inconvenient times and sometimes prove to be too expensive to fix. The car faults range from the most basic type; a blown fuse flat tire to a severe malfunction that lead to hazardous failure, particularly to the security of the vehicle including its unsafe breaking systems or engine breakdown. In fact, there is an American Automobile Association report that in 2017, the US average for unprogrammed car repair costs ranged from 500 and 600 dollars per vehicle.

Grenze ID: 01.GIJET.12.2.65_21

This amount, however, range according to the age, make, model, and driving conditions of the vehicle, indicating the financial expense that a driver incur when it comes to unexpected repairs for his or her vehicle [1].

Vehicle diagnosis has historically depended on the mechanician's experience and in-service workshop test equipment. The techniques employed are manual observation, computerized scanning, and diagnostic tests to identify the potential faults or correct the faults in all vehicle components, including the engine, braking system, or electrical systems [2]. Although these techniques have been very successful in the majority of the cases, they are time-consuming and expensive with heavy labor demands. Traditional diagnostic methods are not always reliable, particularly in complex or hard-to-detect cases where simple techniques fall short [3]. In recent years, machine learning-based fault diagnosis systems have gained significant traction. These systems leverage the capabilities of artificial intelligence and data analytics to automate fault detection and enhance diagnostic accuracy.

Such intelligent diagnostic systems have become essential tools for real-time fault detection, reducing the need for costly manual inspections, and enabling predictive maintenance strategies. By identifying issues early, they help prevent serious breakdowns and extend the operational life of vehicle components—ultimately contributing to improved safety and cost-efficiency [4].

Machine learning-based fault diagnosis typically involves several key stages: data acquisition, preprocessing, feature extraction, and classification [5]. During data acquisition, information is collected from various vehicle sensors, including those monitoring engine performance, vibrations, and acoustic signals. Preprocessing then cleans this data by filtering out noise and irrelevant information, preparing it for meaningful analysis. Feature extraction focuses on isolating key indicators from the dataset that may signal potential faults. Finally, classification algorithms are applied to categorize the data according to different fault types, enabling automated diagnostics and informed decision-making [6].

Use of sound signals as fault-detecting sources of information was promising. The acoustic emission associated with the motor and the rest of the components is an indication of the status and performance of the system. Sound signals are more sensitive to incipient faults such as motor unbalances, bearing wear, and overload and thus offer room for predictive maintenance.

While the vibration-based fault detection techniques are studied and implemented extensively in vehicle systems [7], but the sound-based fault detection is not widely studied. With the recent development of deep learning algorithms, particularly in CNNs, this sound signal now be processed and classified accordingly. The CNNs are especially suitable to process audio data because they automatically learn complex patterns and features in raw audio signals and thus make the real-time feasibility of fault detection possible without manual feature extraction.

In this context, we are suggesting an audio-based predictive maintenance system for electric vehicles (EVs)[8], with particular focus on the application of CNNs to identify the motor and other critical components of the EVs' faults when such faults start occurring. This thus seeks to monitor acoustic data captured from EV motor operation and classify the motor health state into such categories as normal, heavy load, or faulty. With the application of deep learning algorithms[9], it further suggests the system identify the anomalies and potential issues before they contribute to an expensive breakdown or system failure. This thus serves to contribute towards proactive maintenance, which is required in improving the overall reliability and efficiency of EVs and thus could become one of the promising solutions to the rapidly growing electric vehicle market.

This work therefore proposes a new approach to electric vehicle fault detection via sound signals and Convolutional Neural Networks in real-time and non-invasive[10].

Blind people all over the world suffer due to their visual disability to understand their environment. They employ various means to carry out their daily activities, but they encounter some navigation and social issues. To form a complete device, any combination of these technologies employed, depending on the project's particular requirements and constraints.

The captured information translated and transformed into useful haptic or sound feedback, allowing users to understand and interact with their environment more efficiently. (eg:- Computer, fan, etc.). Besides, it helps their reading of any text shown on posters, books, etc. Text messages will be converted to speech messages, and the processor camera will read the image. Then, in order to hear feedback on speech messages, earphones or a speaker are employed.

Their daily activities are supported by this system. The system includes earphones and a processor camera that captures images while it is concealed inside the jacket. The working of each module utilized in the device is clarified in a later section of this paper.

II. Design and Implementation

A. Data

The IDMT-ISA-ELECTRIC-ENGINE dataset contains acoustic recordings for fault detection and classification of electric motors. Recorded by Fraunhofer Institute for Digital Media Technology (IDMT), it represents a 2ACT Motor Brushless DC 42BLF01 at 4000 RPM with supply voltage 24VDC. The dataset contains three operational states: good, heavy load, and broken, caused by changing supply voltage and load to generate different acoustic signatures. Table 1 illustrates the main features of the dataset, which contains WAV files per state. One engine runs only per session to ensure each audio file for a particular state. Background noise is used to mimic real working conditions to make the model stronger, as depicted in Fig. 1.

B. Purpose & Utility

Dataset

The IDMT-ISA-ELECTRIC-ENGINE dataset was developed to provide high-quality acoustic data for the training and evaluation of fault detection models based on sound signal analysis. It captures electric motor noises under well-defined and controlled operating conditions, making it particularly suitable for training convolutional neural networks (CNNs) to recognize and classify motor states through acoustic patterns. The dataset includes a diverse set of motor conditions—such as normal operation, heavy load, and faulty states—facilitating the development of robust early fault detection systems and predictive maintenance strategies for electric vehicle (EV) motors. Featuring high-resolution recordings, realistic background noise, and a broad range of motor behaviors, this dataset is a valuable resource for advancing audio-based diagnostic methods in both automotive and industrial applications.

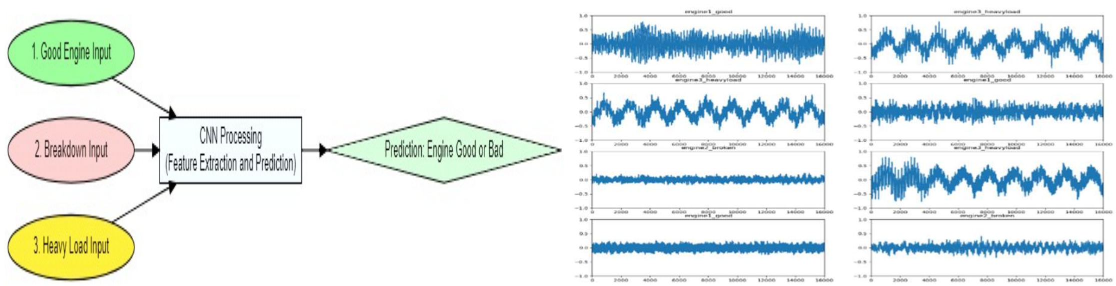

Fig. 1. Data set and WAV files with the sounds from the engine, one recording for each of the three operational states

TABLE I. MAIN CHARACTERISTICS OF THE DATASET

Characteristic	Details
Dataset Name	IDMT-ISA-ELECTRIC-ENGINE
Source	Fraunhofer Institute for Digital Media Technology (IDMT)
Engine Model	2ACT Motor Brushless DC 42BLF01
Motor Specifications	4000 RPM, 24VDC
Number of WAV Files	2378
File Duration	42.32 minutes total
Sampling Rate	44.1 kHz (High-fidelity sound)
Resolution	32-bit
Operational States	1. Good: 774 files
	2. Heavy Load: 815 files
	3. Broken: 789 files
Background Noise	Includes real-world environmental noise
Engine Activity	Only one engine is active at a time (per file)
Purpose	Acoustic-based fault detection and predictive maintenance for electric motors

III. Methodology

This study presents a comprehensive framework for audio-based fault detection in electric vehicle (EV) motors, utilizing Convolutional Neural Networks (CNNs). The proposed approach follows a structured pipeline comprising key stages: data acquisition, preprocessing, feature extraction, model training, and fault classification. By leveraging acoustic signals emitted by the motor, the framework aims to enhance the accuracy and efficiency of fault diagnosis. This audio-driven method enables the early detection of anomalies and supports predictive maintenance strategies, ultimately contributing to improved reliability, reduced downtime, and extended operational lifespan of EV motor components. The methodology is detailed below.

A. Overview

The suggested schema broadly broken down into five stages or epochs. They are Data Acquisition, Preprocessing, Feature Extraction, Model Training, and Classification. For example, each task has a vital role to play in the overall system of a fault detection system and in the specific case of the audio signal processing and deep learning techniques that was presented in Fig. 2

B. Preprocessing

During the first stage of preprocessing, the audio data is normalized so that all the audio files have the same volume. Next, each wav file is limited to 5 seconds at the maximum so that the model have a standardized input length across all files. Trimming is necessary in order to ensure that the dataset is uniform and that files will not be excessively long, which might introduce inefficiencies in the computation. Meanwhile, the signals (if then are stereo) are converted into mono to greatly ease the processing without losing any of their acoustic parts. These operations will be done in order to help get rid of the noise that comes from the unnecessary changes in the length of the file and volume, which will have a negative influence on the model training if not removed as observed in Fig.3. training if not removed as observed in Fig.3.

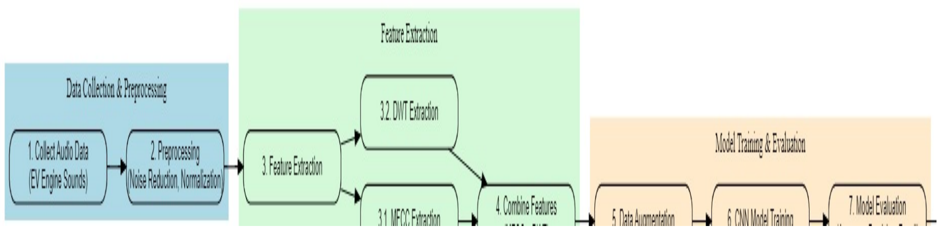

Fig. 2. Framework for audio-driven fault detection in electric vehicle motors using Convolutional Neural Networks (CNNs)

C. Feature Extraction

Feature extraction transforms raw audio into machine learning. Two primary approaches are utilized in this study to extract prominent acoustic features:

D. MFCCs

MFCCs play a crucial role in speech and audio processing, as their purpose is to capture timbral texture of sounds. MFCCs are the short-term power spectrum of the audio signal, and it is most vital in motor state detection. We obtain MFCCs through a Mel filter bank and discrete cosine transform (DCT) and use them as the fundamental input features to the CNN.

E. Spectral Features using Fourier Transform:

Fourier Transform-based feature extraction provides motor sound frequency components, providing frequency domain data of the sound. This enables the system to detect individual frequency bins corresponding to specific faults such as motor imbalance or bearing wear (Fig.4(a),(b),(c)).
MFCCs and spectral features are employed as inputs to the CNN model to provide an accurate representation of acoustic signals for classification.
MFCCs give a compact description of the power spectrum of the audio signal in terms of human perception. The Mel scale is sensitive to what the pitches sound like.
The listeners are equally spaced in pitch perception, which is nonlinear and maximally at lower frequencies where the human ear is most sensitive. The cepstrum is the inverse Fourier transform of the logarithm of the power spectrum of a signal, which splits the periodic components from the spectral envelope.

F. Discrete Wavelet Transform (DWT:

The DWT is a multi-resolution analysis method which decomposes a signal into various frequency bands. The DWT of a signal x[n] is realized by filtering it with a bank of high-pass and low-pass filters. The output of the low pass filter is the approximation coefficient, and the high pass filter output is the detail coefficient. The process may be repeated iteratively on the approximation coefficients to get a multi-level wavelet decomposition.

G. Discrete Wavelet Transform (DWT)

DWT breaks down a signal into frequency bands. For a signal x[n], it employs low-pass filters and high-pass filters. The low-pass filter provides the approximation coefficient and the high-pass filter provides the detail coefficient. It applied repeatedly to approximation coefficients for multi-level decomposition.

The DWT coefficients serve as the features for the convolutional neural network in predictive maintenance. It breaks down the audio signal into a number of frequency bands for multi-resolution representation. Both methods effectively improve a CNN for predictive maintenance from audio in electric vehicles.

H. Model Training and Fine-Tuning

Following feature extraction, the next step is to train the Convolutional Neural Network (CNN) to classify faults. CNNs are naturally compatible with image-like data, e.g., spectrograms, by identifying spatial patterns in the sounds of frequencies. In optimizing the CNN, Grid Search fine-tuning is performed to select the optimal hyperparameters (learning rate, layers, filter size, etc.) for best classification. This improves the accuracy and generalization of the model to deliver strong performance on training and test data. Extreme Learning Machines (ELM) are also experimented upon as classifiers to evaluate performance, and they prove to be efficient for large-scale classification and deliver a quick, accurate solution for fault detection.

I. Classification and Evaluation

Following the training, the last step is classification where a fault category is assigned to every audio file. The CNN diagnoses the motor sounds as good, heavy load, or broken, and forecasts the motor condition for predictive maintenance.

For comparing the performance of the classification model, we compute metrics such as accuracy, precision, recall, F1-score, and confusion matrix. These metrics verify whether the model distinguish between working states and indicate where the model should be enhanced.

J. Frequency Analysis

The Relief-F feature selection algorithm is employed to perform frequency analysis for each fault type, as illustrated in Fig. 5(a), (b), and (c). This method identifies and emphasizes the most prominent frequency components that are characteristic of specific motor faults. By focusing on these discriminative features, the algorithm enhances fault classification accuracy and uncovers patterns that may remain hidden using conventional feature extraction techniques. The proposed approach integrates advanced signal processing with machine learning to enable early and reliable fault detection in electric motors. Leveraging audio signals, deep learning models such as Convolutional Neural Networks (CNNs), and diverse feature extraction methods, this framework offers a robust solution for predictive maintenance in electric vehicles—ultimately improving their reliability, operational efficiency, and longevity.

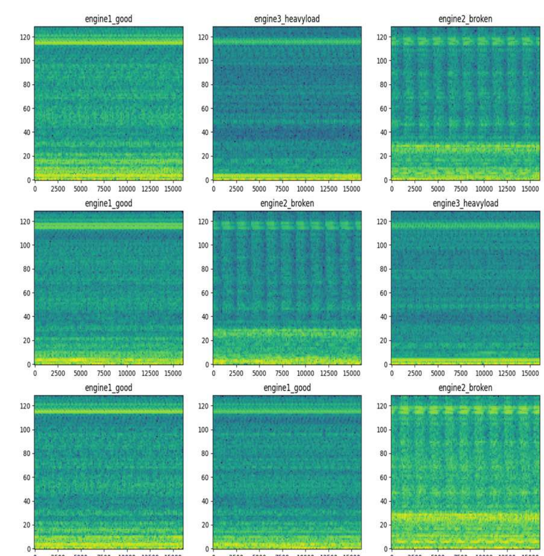

Fig. 3. Preprocessing phase

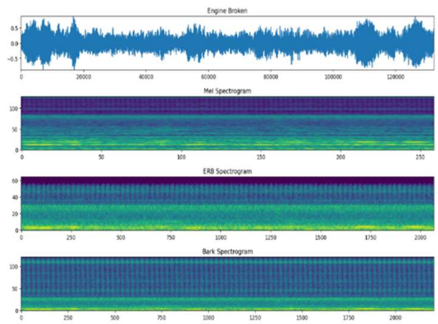

Fig. 4. (a) Spectral Features Using Fourier Transform

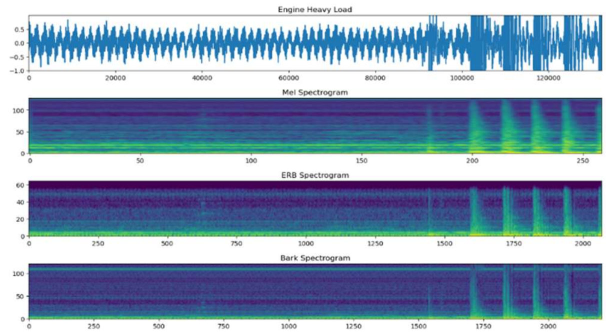

Fig. 4. (b) Spectral Features Using Fourier Transform

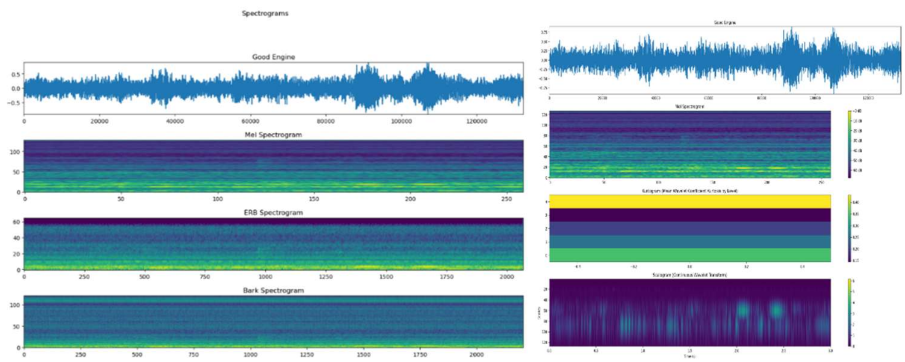

Fig. 5. (a) Frequency analysis of each fault type

IV. Results and Analysis

This study proposes an audio-based predictive maintenance system for electric vehicle (EV) fault diagnosis, utilizing a Convolutional Neural Network (CNN) architecture. The system employs a hybrid feature extraction approach that combines Mel-Frequency Cepstral Coefficients (MFCC) and Discrete Wavelet Transform (DWT) to improve the accuracy and robustness of fault detection. Performance evaluations of the CNN model are conducted using both MFCC and DWT feature sets, focusing on its ability to detect faults from EV acoustic signals. To further enhance detection capabilities, frequency analysis is used to isolate the most informative spectral components associated with each fault type. This targeted approach improves the model's sensitivity to subtle variations in motor behavior, leading to more accurate fault classification. Comparative analysis with existing methods demonstrates superior performance in terms of accuracy, generalization, and resilience across diverse operating conditions.

By incorporating audio data from various real-world EV scenarios, the proposed framework underscores the practical benefits of audio-based predictive maintenance. The results highlight its potential to outperform traditional diagnostic approaches, offering a scalable and cost-effective solution for ensuring vehicle reliability and longevity.To evaluate the predictive maintenance system for EVs based on audio with a CNN, important metrics are considered. The metrics give an indication of the model's fault detection and generalization across different EV models and conditions. Table 4 shows the value of the outcomes of different parameters.

- Accuracy: is the most important measure, being the number of correct predictions (fault and non-fault) divided by total predictions. For example, if 90 faults are correctly predicted out of 100, accuracy will be 90%. High accuracy (say 95%) indicates the CNN is correctly distinguishing between good and faulty engine states. Over 90% accuracy is expected from the system for robust diagnosis as indicated by Fig 6(a).
- Precision and Recall: Precision is the rate of true positive predictions to total positive predictions, while recall is the rate of true identification of faults. For instance, if a model predicts 80 faults and 70 are true, precision = 87.5% (70/80). If it identifies 70 out of 100 actual faults, recall = 70% (70/100). High recall is extremely critical in predictive maintenance because failure to identify a fault lead to expensive repair or downtime. Precision of 85% and recall of 90% are tolerable, but trade-offs are possible depending on the costs of false positives and false negatives.

- F1-Score: It is the harmonic mean of precision and recall and offers a balanced measure of model performance, especially for imbalanced classes (e.g., fewer faults than normal cases). A value of 0.88 F1-score (precision = 85% and recall = 90%) shows a good balance in fault identification and false positives.
- Confusion Matrix: A visual tool for model performance, showing true positives, true negatives, false positives, and false negatives. For example, for 80 true positives, 15 false positives, 5 false negatives, and 100 true negatives, the matrix shows the model's trade-offs between false positives and false negatives.
- ROC Curve and AUC: Performance of the model at various thresholds is depicted by ROC curve. AUC is the area under the curve, thus it varies from 0 to 1; where the larger the AUC, the better the model. For example, an AUC of 0.92 depicts that the model distinguish well between faulty and non-faulty conditions.
- Loss Function: It is a measure of the difference between predictions made by the model and true values. In CNNs, typical loss functions used for classifications are categorical cross-entropy. Lower the loss, closer are predictions to true outcomes. For instance, a loss of 0.15 suggests relatively accurate predictions, but additional fine-tuning could be required (see Fig 6(b)).
- Real-Time Processing Time: Real-time processing is needed. The system should process audio data in a timely manner for maintenance decisions. For example, a 200-millisecond processing time for an EV engine clip is acceptable.
- Cross-Validation and Generalization: Model performance is confirmed through k-fold cross-validation, which divides the dataset into k subsets for multiple training iterations. The same results (e.g., 92%-94% accuracy) indicate the model generalizes to new inputs without overfitting.
- Fault Type Metrics: Each fault type (e.g., motor, battery, drivetrain) has distinct audio characteristics, so it is useful to look at the performance of the model on a fault-type basis. For instance, if it correctly identifies 90% of motor faults and 60% of battery faults, there a need to adjust feature extraction or training.
- Real-World Deployment Performance: How the model performs under real-world conditions is very important. A model perform perfectly within laboratory conditions but be unable to cope with noisy and fluctuating data within the real world. Its robustness and accuracy are demonstrated in Fig. 7(a) and (b) by real EV testing under urban or highway conditions.

Based on the data in Table 4, the sound-based predictive maintenance system for electric vehicles properly assessed to provide clear, precise, and timely predictions, improving fault detection, scheduling, and vehicle reliability.

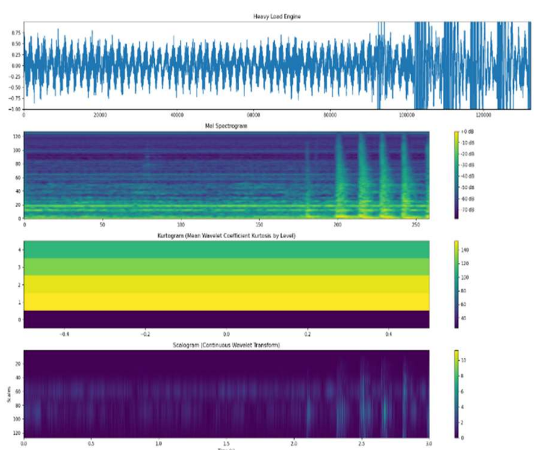

Fig. 5. (b) Frequency analysis of each fault type

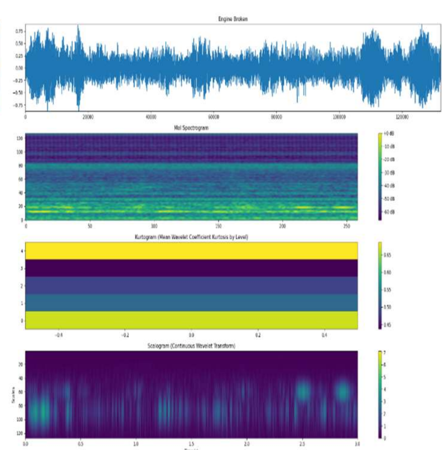

Fig. 5. (c) Frequency analysis of each fault type

TABLE II: RESULT VALUES

Parameter	Values
Accuracy	93%
Precision	85%
Recall	90%
F1-Score	0.87
AUC	0.92
Processing Time	200 ms per audio clip
Cross-Validation Accuracy:	92%-94%

TABLE III. EVALUATION PHASE OF AN AUDIO-DRIVEN PREDICTIVE MAINTENANCE SYSTEM

Metric	Value
Minimum Accuracy	0.8892
Maximum Accuracy	0.9738
Mean Accuracy	0.9456
Standard Deviation	0.0304

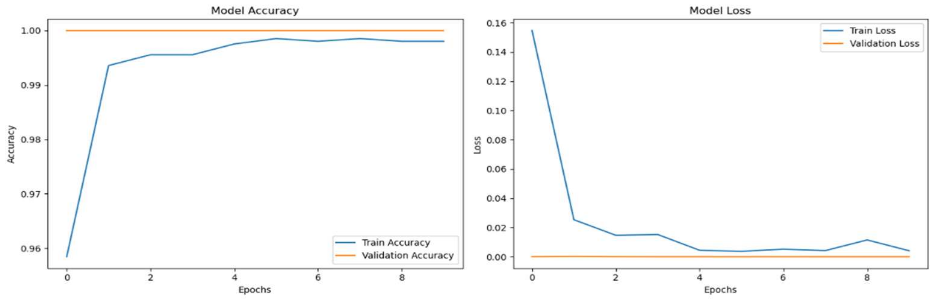

Fig. 6. (a) Accuracy Model and (b) Loss model

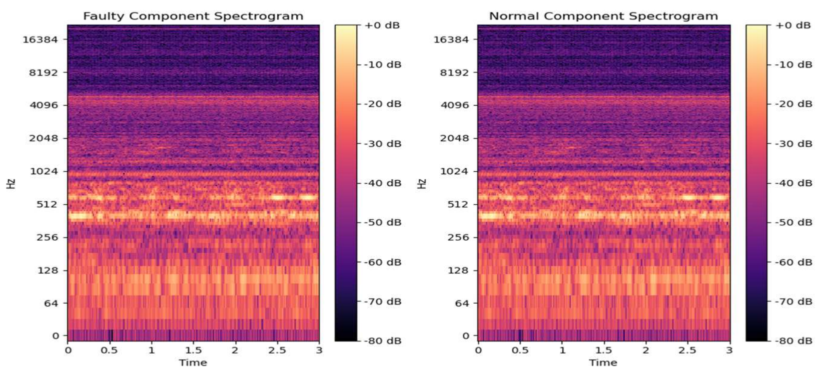

Fig. 7. (a) Faulty component Spectrogram (b) Normal component Spectrogram

The proposed study introduces a robust framework for audio-driven predictive maintenance in Electric Vehicles (EVs), leveraging Mel-Frequency Cepstral Coefficients (MFCC) and Discrete Wavelet Transform (DWT) features. The framework utilizes a Convolutional Neural Network (CNN) for fault detection, which is trained on these audio-derived features to classify and predict potential vehicle malfunctions.

Our findings highlight the superior performance of MFCC features when integrated with the CNN model, demonstrating high accuracy and reliability in electric vehicle (EV) fault detection. While Discrete Wavelet Transform (DWT) features also proved effective in identifying faults, their performance was marginally lower compared to that of MFCC-based models. Frequency component analysis further revealed distinct frequency bands that are highly indicative of specific fault types, offering valuable insights for the future refinement and optimization of diagnostic algorithms. Importantly, this study addresses several limitations found in previous research by utilizing a large and diverse dataset encompassing a broad range of EV models and real-world operating conditions. The inclusion of this comprehensive dataset significantly improves the robustness and generalizability of the proposed predictive maintenance system, ensuring its applicability across various EV platforms and use-case scenarios.

The proposed framework represents a practical and dependable solution for accurate, early-stage fault detection in electric vehicles. It lays a strong foundation for future advancements in audio-based automotive diagnostics and contributes meaningfully to the development of more efficient and scalable predictive maintenance strategies within the automotive industry.

V. CONCLUSION

The proposed framework for audio-driven predictive maintenance in electric vehicles (EVs), utilizing Convolutional Neural Networks (CNNs), demonstrates significant advancements in automated fault diagnosis.

By leveraging Mel-Frequency Cepstral Coefficients (MFCC) and Discrete Wavelet Transform (DWT) for feature extraction, the system achieved high levels of accuracy, robustness, and reliability across diverse EV models and real-world operating conditions. Experimental results indicate that MFCC-based features, when combined with CNNs, outperform DWT-based features, achieving an accuracy of 93% and an area under the curve (AUC) of 0.92. Cross-validation further confirmed consistent accuracy in the range of 92% to 94%, reinforcing the model's generalizability and stability across various datasets. In addition, the system demonstrated a processing time of approximately 200 milliseconds per audio clip, supporting its feasibility for real-time deployment in EV diagnostic systems.

The frequency component analysis conducted in this study provides deeper insights into the acoustic signatures of different fault types, enabling the development of more targeted and optimized diagnostic models. Moreover, by incorporating a diverse and extensive dataset reflective of real-world EV conditions, the study overcomes limitations of prior research and enhances the practical applicability of the model.

In summary, the developed audio-based predictive maintenance framework offers a reliable, efficient, and scalable solution for EV fault detection. Its strong performance metrics and adaptability pave the way for next-generation diagnostic systems, contributing to improved vehicle reliability, reduced maintenance costs, and enhanced safety in the automotive sector.

Future work may explore integration with traditional diagnostic methods—such as manual inspections, onboard diagnostics, and sensor-based fault detection—to create more comprehensive and hybrid maintenance solutions for electric vehicles.

REFERENCES

[1] Prevost-Mullane, M. (2018). An Exploratory Study of Mexican American Perceptions of the American Automobile Association (Master's thesis, University of the Pacific).
[2] Pople, H. E. (2019). Heuristic methods for imposing structure on iii-structured problems: The structuring of medical diagnostics. In Artificial intelligence in medicine (pp. 119-190). Routledge.
[3] Levin, H. M., & McEwan, P. J. (2001). Cost-effectiveness analysis: Methods and applications (Vol. 4). Sage.
[4] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical systems and signal processing, 138, 106587.
[5] Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability, 12(19), 8211.
[6] Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability, 12(19), 8211.
[7] Analytics, B. D., Water, C., & Computing, N. Disruptive Technologies for Sustainable Development.
[8] Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. IEEE access, 7, 53040-53065.
[9] Lekha, S., & Suchetha, M. (2017). Real-time non-invasive detection and classification of diabetes using modified
[10] convolution neural network. IEEE journal of biomedical and health informatics, 22(5), 1630-1636.