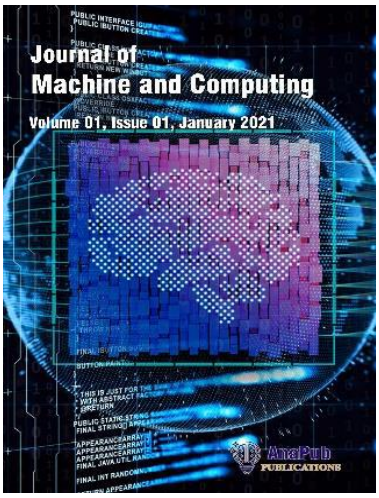
Journal Pre-proof



Machine Ears: Audio Frequency-Based Automobile Engine Health Analysis

Debie Shajie A, Sujitha Juliet D, Kirubakaran Ezra and Blessy Annie Flora J

DOI: 10.53759/7669/jmc202505015

Reference: JMC202505015

Journal: Journal of Machine and Computing.

Received 15 April 2024

Revised form 14 July 2024

Accepted 02 November 2024

Please cite this article as: Debie Shajie A, Sujitha Juliet D, Kirubakaran Ezra and Blessy Annie Flora J, “Machine Ears: Audio Frequency-Based Automobile Engine Health Analysis”, Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505015

Machine Ears: Audio Frequency-based Automobile Engine Health Analysis

A. Debie Shajie[1], D. Sujitha Juliet[2*], Kirubakaran Ezra[3], J. Blessy Annie Flora[4]

[1]Division of CSE, Karunya Institute of Technology and Sciences, Coimbatore 641114, Tamilnadu, India, Email: dshajie@gmail.com ORCID: 0000-0001-8983-2409.
[2*]Division of Data Science and Cyber Security, Karunya Institute of Technology and Sciences, Coimbatore 641114, Tamilnadu, India, Email: sujitha@karunya.edu, sujithajuliet007@gmail.com ORCID: 0000-0001-6214-9431.
[3]Director, Grace College of Engineering, Tuticorin, Tamilnadu, India, Email: ekirubakaran@gmail.com
[4]Division of ECE, Karunya Institute of Technology and Sciences, Coimbatore 641114, Tamilnadu, India, Email: blessyannieflora63@gmail.com ORCID: 0000-0003-2904-4528.

Abstract

Maintaining both rider safety and vehicle dependability on motorbikes requires accurate problem detection. Using an improved ResNet architecture with Improved Sea Fish Optimization (ISFO) and Deep Convolutional Neural Networks (CNNs), this research proposes a sophisticated method for auditory defect identification in motorbikes. The machine ears start by gathering a wide range of audio frequency-based signal datasets from motorbike that span a range of failure scenarios and operational settings. To eliminate noise and identify distinguishing characteristics, these signals go through preprocessing. Then, to extract high-level features from the pre-processed signals, an improved ResNet architecture is used, supplemented with ISFO. By integrating both local and global information, the ResNet architecture's inclusion of ISFO makes it easier to iteratively update feature representations. To further improve the feature representations' discriminative power, Deep CNNs are used. The real-time defect detection system is designed specifically for motorbikes uses the learned model. The trained model is used to interpret incoming acoustic data from motorcycle operations. This allows for the identification and categorization of various issues, such as engine misfires, irregularities in the valves, wear on the bearings, and clutch bearing failures. Experiments show that the proposed method is a good fit for precisely categorizing motorbike issues. Analyses conducted in comparison with baseline models demonstrate the superiority of the ResNet-ISFO and Deep CNN technique, demonstrating its resilience and efficiency across a range of fault situations and operational conditions. Overall, the proposed acoustic problem detection system is a potential approach for improving maintenance procedures while also assuring the safety and dependability of automobile engine. Its incorporation into standard maintenance operations can aid in proactive defect identification, reducing downtime and improving vehicle performance.

Keywords: Vehicle, Acoustic Fault Detection, ResNet, Improved Sea Fish Optimization, Deep Convolutional Neural Network.

1. Introduction

Motorbikes, revered for their iconic design and enduring appeal, are cherished by enthusiasts worldwide. However, the pursuit of unforgettable riding experiences is accompanied by the responsibility of ensuring the safety and reliability of these machines [1]. Timely detection and diagnosis of potential faults are paramount for several compelling reasons. Firstly, prioritizing safety is non-negotiable. Undetected faults, such as brake system deficiencies or engine irregularities, can pose significant hazards to riders and fellow road users [2]. Moreover, the reliability of motorbike is synonymous with their enduring legacy. Whether embarking on adventurous journeys or navigating daily commutes, riders rely on their motorcycles to deliver consistent performance [3]. Detecting faults early not only safeguards riders but also preserves the motorcycle's performance and longevity. Furthermore, proactive maintenance, facilitated by timely fault detection, mitigates the risk of minor issues escalating into costly repairs or unexpected breakdowns [4]. By upholding rigorous maintenance standards, riders can prolong the lifespan of their motorcycles and safeguard their investment [5]. Ultimately, the commitment to ensuring the timely detection and diagnosis of potential faults underscores the dedication to rider safety, vehicle reliability, and the enduring legacy of the motorbikes.

Acoustic fault detection in motorbike has emerged as a critical area of research and development aimed at enhancing vehicle safety and reliability. By harnessing the distinctive sounds emitted by various components of motorbike engines, such as the engine, gearbox, and carburetor system, this technology offers a non-intrusive and efficient method for identifying potential faults and anomalies [6]. Using cutting-edge methods for signal processing and Machine Learning (ML) algorithms, acoustic fault detection systems can analyze the acoustic signatures of different motorcycle components in real time, enabling early detection of issues such as engine misfires, bearing wear, and sprag clutch malfunctions [7]. By integrating these systems into routine maintenance procedures, riders, and service technicians can proactively address emerging faults, thereby minimizing downtime, optimizing vehicle performance, and ensuring a safe and enjoyable riding experience [8]. As motorbike continues to innovate and evolve, acoustic fault detection stands poised to play a pivotal role in enhancing motorcycle diagnostics and maintenance practices, reaffirming the brand's commitment to quality, reliability, and rider satisfaction [9].

Ensuring the operational reliability and safety of automobile necessitates the timely detection and diagnosis of potential faults. Traditional approaches to fault detection often rely on manual inspection or sensor-based monitoring systems, which may be limited in their ability to detect subtle or emerging issues. In recent years, advancements in ML and signal processing have paved the way for more sophisticated fault detection techniques leveraging acoustic signals emitted by the motorcycle's engine and components. Within this framework, this research suggests a novel method for acoustic defect identification. In motorbikes, harnessing the power of deep learning models, specifically an Enhanced ResNet architecture augmented with Improved Sea Fish Optimization (ISFO), and Deep Convolutional Neural Networks (CNNs). By integrating these advanced techniques, our proposed system aims to accurately identify various types of faults in motorbikes, thereby enhancing maintenance practices and ensuring vehicle reliability. This paper presents a comprehensive analysis of the proposed approach, including data collection, preprocessing, model architecture, training methodology, and experimental evaluation, demonstrating its effectiveness in real-world fault detection scenarios.

The novelty of using Enhanced ResNet with Improved Sea Fish Optimization (ISFO) and Deep Convolutional Neural Networks (CNNs) for acoustic fault detection in motorbike lies in several key aspects:

1.1 Integration of Advanced Techniques

The proposed approach integrates multiple advanced techniques, including ResNet, ISFO, and CNNs. While ResNet is well-known for its effectiveness in deep learning tasks, ISFO adds an innovative dimension by iteratively refining feature representations, enhancing their quality. The incorporation of Deep CNNs further boosts the model's ability to extract intricate patterns from the acoustic signals.

Tailored Solution for Motorcycle Fault Detection: Acoustic fault detection specifically tailored for motorbikes engines represents a novel application of deep learning and signal processing techniques. By focusing on the unique acoustic signatures of motorbikes, the proposed system addresses a specific niche within the broader field of fault detection.

1.2 Enhanced Feature Extraction

The combination of ResNet with ISFO and Deep CNNs enables enhanced feature extraction from acoustic signals. ISFO iteratively refines feature representations, capturing both local and global information, while Deep CNNs learn hierarchical representations of the data. This synergy results in improved discriminative power and fault detection accuracy.

1.3 Real-time Fault Detection

The proposed system is designed for real-time fault detection, allowing for prompt identification of potential issues during motorcycle operations. This capability enhances safety and reliability by enabling proactive maintenance and minimizing downtime.

1.4 Application in Automotive Industry

While acoustic fault detection has been explored in various domains, including industrial machinery and aerospace, its application in the automotive industry, particularly for motorcycles, is relatively novel. By applying state-of-the-art deep learning techniques to motorcycle fault detection, the proposed approach contributes to advancing the field of automotive diagnostics.

The novelty of using Enhanced ResNet-ISFO and Deep CNNs for acoustic fault detection in motorbike lies in its integration of advanced techniques, tailored solutions for motorcycle fault detection, enhanced feature extraction capabilities, real-time fault detection capabilities, and its application in the automotive industry.

1.5 Motivation

The research on acoustic fault detection in motorbikes using Enhanced ResNet-ISFO and Deep CNNs is motivated by a multifaceted approach aimed at enhancing both rider safety and vehicle reliability. With a strong emphasis on proactive maintenance and early fault detection, the research seeks to address potential issues within the motorcycles' systems before they escalate into safety hazards or major mechanical failures. By leveraging advanced deep learning techniques and signal processing methods, the goal is to develop a sophisticated fault detection system capable of accurately identifying various types of faults, including engine malfunctions, brake issues, and suspension problems. Additionally, the research aims to minimize maintenance costs for riders by enabling timely repairs and reducing the risk of unexpected breakdowns. Furthermore, by advancing the field of automotive diagnostics through innovative approaches, such as the integration of Enhanced ResNet-ISFO and Deep CNNs, the research contributes to maintaining motorbikes reputation for quality and reliability. Ultimately, the motivation behind the research lies in enhancing the overall riding experience, ensuring customer satisfaction, and upholding the highest standards of safety and performance in motorbikes.

2. Literature Review

Empirical research has demonstrated that engine fault end-to-end detection, which utilizes the Echo State Network (ESN) and Multi-Verse Optimizer (MVO) for pattern recognition, can attain a 93.10% identification rate in complex engine faults [1]. The analysis of vibration signals is conducted using the [illegible]-spectrum transformation and the Deep Echo State Network (ESN) model, which is fixed using convolution kernels and Auto Encoder (AE). To bear failure detection of Electric Multiple Units (EMU) traction motors, Cross Wavelet Transform (XWT) and GoogleNet model are presented [2]. For denoised signals, the method's fault classification accuracy was 98.23%. Vehicle characteristics are predicted and cascaded by a multi-task convolutional neural network to enhance misfire fault detection. Moreover, it develops and uses cutting-edge deep learning cascade topologies, which are characterized as conditional, multi-level networks that analyze unprocessed audio frequency and extract intensely specific information for understanding vehicles [3]. Transparency and comprehension of cars are improved by the acoustic road vehicle characterization system. The misfire fault detection test set accuracy achieved by cascading CNN is 87.0%. There were 1.7% and 8.0% margins above the parallel and naïve CNN baselines.

To achieve effective domain generalization and fault classification, a deep transfer learning technique is described [4]. This method leverages information flow from the source domain to all dense layers and convolutional layer fine-tuning. Another wavelet function utilized for variable time-frequency resolution in time-frequency imaging techniques is the Continuous Wavelet Transform (CWT), which produces scalograms. Vibration data sets from various distribution target domains were classified with great accuracy by a deep transfer learning model. It was an edge implementation-suited model because it only used 1320 trainable parameters. Combining adaptive multi-threshold segmentation with sub-domain adaptation, the Gray Wolf Optimization algorithm (GWO) and Symmetric Cross Entropy (SCE) are linked to provide adaptive multiscale segmentation on the image samples in a deep transfer fault diagnostic technique [5].

The gearbox fault diagnostic accuracy is increased by the fault diagnosis algorithm. Installation restrictions and layout specifications are resolved by using a non-contact sound pressure sensor. Wang et al.'s [6] DL framework makes use of Convolutional Neural Networks (CNNs) based on transfer learning optimization. CNNs facilitate the transfer of low-level characteristics and the fine-tuning of high-level layers, allowing the model to achieve greater precision with reduced computational costs. Deep learning models are characterized by a large number of trainable parameters, complex hyperparameter modification, and initialization instability. CNNs, CWT, and Transfer Learning (TL) are all used in this framework to improve classification accuracy. A concise deep learning model for bearing defect detection based on convolutional neural networks is provided [7]. It is considered appropriate for use on factory floors where production moves quickly and machine configuration changes are likely to occur. In comparison to other well-known models, the model contains 98% less parameters. Compared to previous models, it contains 98% fewer parameters. It produced a higher accuracy of 21.21% and a fault detection rate improvement of 7.03%.

Combining the expression improvement module with the filter enhancement module, Wang et al. [8] presented an enhanced integrated filter network for rolling bearing defect diagnosis. To recover valuable characteristics from medium- and low-frequency signals, this filter network may simultaneously filter away high-frequency noise and maintain frequency and temporal resolution [8]. The method includes two modules: one for filter enhancement and the other for expression enhancement. 94.85% is the fault diagnostic precision on the CWRU dataset. The precision of fault diagnosis on the IMS dataset is 92.45%. To retrieve the deep information concealed in the Acoustic Emission (AE) waveforms, the History State Ensemble (HSE) is introduced and connected to the convolutional Generative Adversarial Network (GAN) architecture [9]. Acoustic emission signals for early defect detection in industrial rotational equipment. The robustness of defect detection is improved by using a history-state ensemble with convolutional GAN. The proposed Acoustic-Based Diagnosis (ABD) technique for identifying gear faults in multiple operating circumstances surpasses multiple typical fault identification approaches that involve feature engineering, an end-to-end CNN model built around both time and frequency domain signals, and a traditional CNN model [10]. Multi-scale convolutional learning structure serves as the foundation for this approach.

An Augmented Convolution Sparse Autoencoder (ACSAE) was reported for gear pitting defect diagnosis employing raw Acoustic Emission (AE) data. Without altering the AE signals' temporal or frequency domains, this autoencoder can autonomously derive fault indicators from the initially generated AE signals [11]. A defect diagnostic accuracy of 97.9% was attained in this approach. CNN combined with the Adaptive Batch Normalization method (ABN-CNN) is developed to reduce the high processing resource requirements of such complex networks. For carrying out problem diagnostics, it offers quick convergence and strong identification accuracy in an environment with noise and load fluctuation [12]. ABN-CNN was created with quick convergence and excellent recognition accuracy for the identification of bearing faults. Under noise and load change, it achieves excellent identification accuracy and rapid convergence. Using the acoustic signals of rotating machinery, a batch-normalized Deep Sparse Filtering (DSF) method finds the flaw more effectively than other traditional techniques. It requires a smaller amount of as well as provides greater computational features [13]. An approach for fine-tuning features is Backpropagation. More potent characteristics can be extracted by the DSF model [14-15]. Comparing the DSF model to other conventional techniques, less computer time is needed. Acoustic fault detection in motorbike engines involves using sound analysis techniques to identify potential issues based on abnormal noises emitted by the engine or other components. By leveraging sensors and signal processing algorithms, anomalies in the acoustic signatures can be detected, indicating areas that may require maintenance or repair. This approach can help improve overall vehicle reliability and performance.

3. Methodology

Acoustic fault detection in motorbike engines involves the use of sound analysis techniques to pinpoint abnormalities in the acoustic signatures of the vehicle. By capturing and analyzing audio frequency-based signals, mechanics and owners can detect potential faults early on, allowing for timely maintenance and repairs to be carried out. The proposed machine ears approach utilizes sound analysis techniques to enhance automobile health and performance. Detecting faults through acoustic analysis can significantly improve the reliability and performance of motorbike engines. By identifying issues at an early stage, preventive measures can be taken to avoid costly repairs and ensure a smooth riding experience for the users. Sound analysis involves the use of sensors to capture audio frequency-based signals from the motorcycle's engine and components. These signals are then processed using advanced algorithms to detect any deviations from normal acoustic patterns. By comparing the recorded sounds to predefined models, anomalies can be identified accurately. The key components in acoustic fault detection include sensors that capture sound signals and signal processing algorithms that analyze the data. These components work together to monitor the acoustic signatures of motorbikes and flag any irregularities that may indicate potential faults.

Fig1 illustrates the enhanced model classifier structure of the ResNet, ISFO, and deepCNN models for fault diagnosis. The Residual Network (ResNet) architecture combined with Improved Sea Fish Optimization (ISFO) is known as ResNet-ISFO, and it is a state-of-the-art method for signal refining and feature extraction. A solution to the vanishing gradient problem is provided by skip connections, which are introduced by ResNet, which is well-known for its capacity to build incredibly deep neural networks. ISFO iteratively applies soft filtering operations to signals, enhancing their quality and extracting relevant features. ResNet-ISFO combines the advantages of both ResNet and ISFO to improve the efficiency and effectiveness of feature extraction tasks. By incorporating ISFO into ResNet, the model iteratively refines feature representations, incorporating both local and global information, thus enhancing their discriminative power. Deep Convolutional Neural Networks (CNNs) support ResNet-ISFO to improve the model's extracting features characteristics. CNN excels at acquiring hierarchical data visualizations, which makes them ideal for image identification and signal processing. By including deep CNNs into the ResNet-ISFO architecture, the model can identify subtle correlations and trends in the input data, resulting in improved accuracy in activities like identifying defects and classifications.

Overall, ResNet-ISFO and Deep CNNs represent a powerful combination of deep learning techniques for feature extraction and signal processing. By leveraging the strengths of both ResNet-ISFO and CNNs, researchers and practitioners can develop highly effective models for multiple applications.

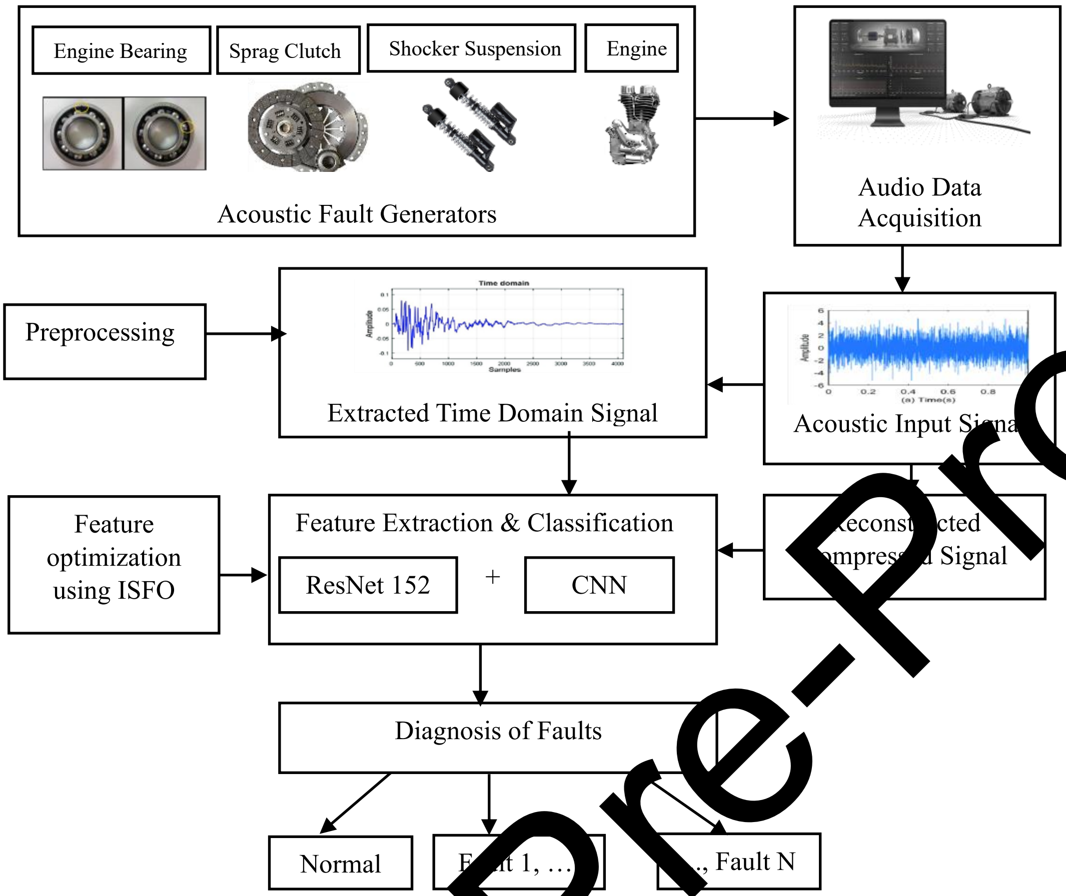


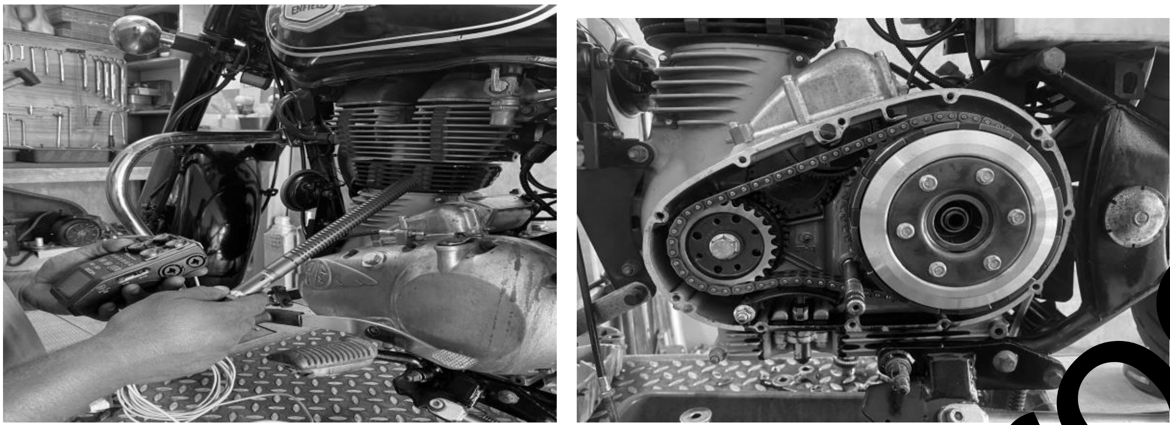
Fig1. The proposed architecture for acoustic fault detection in motorbikes

The algorithm of the proposed method is as follows.

i. Prepare the dataset of motorbike acoustic samples.
ii. Normalize and divide the data into training, validation, and testing sets as preprocessing steps.
iii. Describe the architecture of a neural network model (using ResNet and Convolutional layers).
iv. Initialize model parameters.
v. Select an optimization algorithm (ISFO).
vi. Train the model using the training data:
    a. Run several iterations (or epochs) across the dataset.
    b. Adjust the parameters to minimize the loss.
vii. Evaluate the trained model using the testing dataset.
viii. Analyze the model's performance and fine-tune it if necessary.
ix. Deploy the model for real-time fault detection in motorbikes.
x. Monitor and maintain the deployed system, updating as needed.

3.1 Dataset

The following setup enables us to perform the proposed research to explore different training sample sizes and evaluate model performance accordingly. Creating a test bench for audio frequency-based data sample collection involves setting up a recording environment with the necessary equipment. As shown in Fig.2, this includes selecting suitable microphones, ensuring a quiet space to minimize background noise, and using recording devices like digital recorders or computers with recording software. Once the setup is prepared and connections are made, audio frequency-based samples can be recorded according to specific requirements. It is necessary to check the audio after recording to ensure everything sounds good and to make any necessary setup changes.



(a) (b)

Fig2. (a) Test Bench of sample collection for audio data (b) Assembly of one-way sprag clutch with bearings in motorbike eengine

The breakdown of the dataset characteristics is as follows.

i. Recording Device: Zoom H6 with Shure VP89M Microphone
ii. Audio resolution: 48 kHz, with 24 bits
iii. Conditions: Idling conditions of the motorbike engine
iv. Faults: Faulty Crank Bearing, Damaged piston head, Faulty Oneway spark clutch, and damaged Camshaft Bearing
v. Number of Samples: 12-sec samples of 540 healthy motorcycle samples and 720 faulty motorcycle samples are obtained.
vi. Sample Segmentation: Segmentation was performed on 1[illegible] motorcycles with faults and 14 healthy motorcycles to create the dataset.

3.2 Preprocessing

To improve signal quality and retrieve pertinent information for further analysis, audio data must be pre-processed. Normalization is usually the first step to normalize the amplitudes of the signals. To maintain consistency among datasets, resampling is used to modify the sampling rate to a desired value. Filtering methods, such as Finite Impulse Response (FIR) filters, highlight certain frequency components or eliminate noise. Preprocessing is an essential technique that gets audio data ready for tasks like anomaly detection, clustering, and classification. This helps deep learning models to perform better in the end.

3.2.1 Normalization

The process of normalizing the audio signals guarantees that they are on a standard scale, which facilitates comparison and analysis. This is especially crucial when working with recordings that have diverse amplitudes or signals coming from several sources. When applying mathematical operations or machine learning algorithms to the data, normalization plays a crucial role in preserving numerical stability during the subsequent processing steps. Overflow and underflow problems can be avoided by maintaining the signal values within a specified range, which is often between 0 and 1.

$$x_{norm} = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

Where:

- x symbolizes the initial data point.
- x_{norm} depicts the data point that has been normalized.
- $min(x)$ represents the dataset's lowest possible value for each data point.
- $max(x)$ represents the dataset's highest value for each data point.

This equation computes the normalized value x_{norm} by removing the dataset's lowest value from the initial data point x and splitting it by the dataset's ranging (the variation among the highest and lowest values). This procedure adjusts the data linearly to fit inside the range of 0 to 1.

3.2.2 Resampling

Using linear interpolation to resample acoustic audio signals entails creating new samples by linearly interpolating the values of the current samples. For every time index that is added to the resampled signal, estimate the matching sample value using linear interpolation. Linear interpolation uses linear interpolation to determine the value at a position between two known values. The two original samples that are closest to the new time index should be identified. Determine how far apart the nearest original samples are from the new time index, expressed in fractions. With weights based on fractional distances, compute the sample value at the new time index by averaging the closest original samples. Linear interpolation estimates the value of a new sample y_{new} at a given time index t_{new} using the nearest original samples y_{prev} and y_{next} and their respective time indices t_{prev} and t_{next}.

$$y_{new} = y_{prev} + (t_{new} - t_{prev}) x \frac{y_{next} - y_{prev}}{t_{next} - t_{prev}} \quad (2)$$

3.2.3 Finite Impulse Response (FIR)

The signal-to-noise ratio is increased and the signal is made clearer when undesirable noise is successfully removed from acoustic signals using FIR filters. FIR filters modify the signal's spectral properties to fit certain requirements. This is helpful for activities like equalization when adjusting certain frequency bands is necessary to attain the correct tonal balance. To perform Finite Impulse Response (FIR) filtering of an acoustic signal, the signal convolves with the filter coefficients. Given an input signal $x(n)$ and FIR filter coefficients $h(k)$, the output signal $y(n)$ is computed as follows.

$$y[n] = \sum_{k=0}^{N-1} h[k] \cdot x[n-k] \quad (3)$$

This equation represents the convolution of the input signal with the FIR filter coefficients. It calculates the output signal $y[n]$ by summing the products of each filter coefficient $h[k]$ and the corresponding delayed input signal $x[n-k]$.

3.3 Feature Extraction

In acoustic signal processing, feature extraction is the process of identifying and accurately describing the characteristics of the signal by extracting pertinent information from unprocessed audio data. Time-domain features record the signal's energy distribution, periodicity, and dynamics, among other temporal properties. Features in the frequency domain reveal information about the centroid, bandwidth, and spectral structure of the signal. Time-frequency characteristics show how the spectral components of the signal change over time by providing a combined representation of frequency and time. Statistical characteristics provide more information about the distribution and variability of the signal by quantifying different statistical aspects of the signal. Acoustic signals may be compactly and informatively represented by extracting a mixture of these characteristics, which makes tasks like sound event detection, audio categorization, and speech recognition possible. It is essential to extract features effectively.

3.2 Enhanced Hybrid Residual Networks (ResNet) with Deep Convolutional Neural Network

Introducing the concept of ResNet as a deep learning architecture known for effectively training deep neural networks, emphasizing its relevance in acoustic fault detection. Exploring the specific enhancements and innovations that contribute to the "enhanced" version of ResNet, focusing on its capabilities in handling complex acoustic data. Discussing the potential applications of enhanced ResNet in processing and analyzing acoustic signals for fault detection in motorbikes.

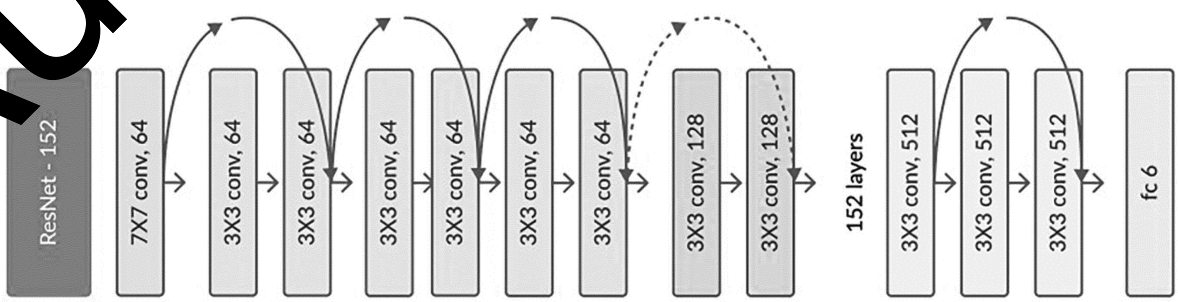


Fig3. ResNet 152 Neural Network Architecture

Detecting faults in motorbikes using Enhanced ResNet 152 involves applying deep learning techniques to analyze acoustic signals associated with different faults. It is shown in Fig3. Convolutional operations are commonly used in the network's first layer, which continues with batch normalization and ReLU activation.

3.2.1 Convolutional Neural Networks layer

Equation (4) is the representation of a convolutional layer's output.

$$Conv(x, W, b) = ReLU(BN(x \cdot W + b) \quad (4)$$

where:

- x is the feature map that was given.
- W stands for the kernel of convolution.
- b is used to express bias.
- $\cdot$ represents the convolution process.
- ReLU represents the activation function of a rectified linear unit.
- BN denotes batch normalization.

3.2.2 MaxPooling Layer

Primarily, MaxPooling layers take an average value inside each pooling zone for reducing map representations of features. This is given in equation (5) as follows.

$$MaxPooling(x)_{i,j} = max_{m,n} x_{(i.stride+m),(j.stride+n)} \quad (5)$$

where:

- x is the feature map provided as input.
- i, j symbolize the combined output′s spatial coordinates.
- m, n iterate over the pooling region.
- $stride$ denotes the stride of the pooling operation.

3.2.3 Global Average Pooling Layer

The average value of every feature map is calculated across its whole geographic range using global average pooling. This is provided as following in equation (6).

$$Global\ Average\ Pooling(x)_c = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{i,j,c} \quad (6)$$

where:

- x is the feature map provided as input.
- H and W stand for the feature map's height and breadth, respectively.
- The feature channels are indexed by c.

3.2.4 Fully Connected Layer

In most cases, a fully linked layer for classification receives the output from the Global Average Pooling layer. This is represented as follows.

$$FC(x, W, b) = softmax(W \cdot x + b) \quad (7)$$

where:

- x is the input vector obtained from Global Average Pooling.
- W symbolizes the completely linked layer's weights.
- The bias term is denoted by b.
- softmax is the softmax activation function, which converts raw scores into class probabilities.

3.2.5 Residual Block

Using Batch Normalization (BN) and ReLU activation functions, a residual block is composed of two convolutional layers. Skip connections allow information to go directly from the input and into the output of the second convolutional layer. The following expression represents the result of a residual block $F(x)$.

$$F(x) = ReLU(BN(W_2) \cdot ReLU(BN(W_1 \cdot x + b_1)) + b_2) + x \quad (8)$$

where:

- x serves as the block′s input.
- The weights of the convolutional layers are W_1 and W_2.
- b_1 and b_2 are the biases.
- ReLU stands for the activation function of rectified linear units.
- Batch normalization is indicated by BN.

3.2.6 Enhanced ResNet 152 Architecture

It consists of multiple layers of residual blocks, grouped into different stages as shown in Fig4.

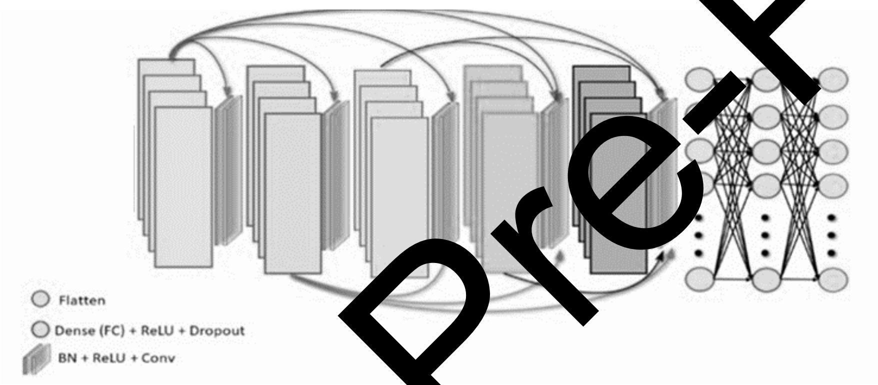


Fig4. ResNet152 with extra layers.

The specific architecture includes:

- Batch Normalization and ReLU activation are performed after the convolutional layer, which has a kernel size of 7x7 and a stride of 2.
- MaxPooling layer with kernel size 3x3 and stride 2.
- 3 stages, each containing multiple residual blocks.
- A fully linked layer with softmax activation for classification and a Global Average Pooling layer at the end.

Combining all components, the overall architecture of Enhanced ResNet and represented in equation (9).

$$y = Softmax(W_{fc} \cdot GlobalAveragePooling(Stage_3(Stage_2(Stage_1(MaxPooling(Conv(x)))))) + b_{fc} \quad (9)$$

where:

- x serves as the network's input.
- Conv, MaxPooling, $Stage_1$, $Stage_2$, and $Stage_3$ represent the operations and stages described earlier.
- The fully connected layer has two weights, W_{fc} and b_{fc} respectively.

3.3 Improved Sea Fish Optimization

It is a method for refining neural network training processes, particularly in the context of acoustic fault detection. Highlighting the adaptive and iterative nature of the optimization process, emphasizing its ability to improve the network's ability to discern acoustic fault patterns. Discussing the impact of Improved Sea Fish Optimization on the performance and efficiency of deep learning models applied to acoustic fault detection.

3.3.1 Integration of ResNet and Soft Filtering

The integration of ResNet with soft filtering typically involves incorporating soft filtering operations within the residual blocks of the ResNet architecture. A motorcycle's capacity to generalize over various failure circumstances and variations can be improved by including soft filtering in the ResNet design. By learning adaptive feature importance, the model becomes more flexible in capturing common fault patterns while adapting to variations in acoustic signals arising from different motorcycle models or environmental conditions. This improved generalization can result in a more versatile and effective fault detection system applicable to a broader range of scenarios.

In each residual block of the Enhanced ResNet 152 architecture, soft filtering can be applied to modulate the importance of features before the addition operation. The residual block F(x) given in equation (8) with soft filtering has the following mathematical expression as its output.

$$F'(x) = F(x) \cdot \sigma(SoftFilter(x)) \quad (10)$$

where:

- x is the block's input.
- The output of the regular residual block is represented by F(x).
- The output of the residual block after soft filtering is represented by F′(x).
- To scale the soft filter values between 0 and 1, the sigmoid activation function is represented by σ(·).
- SoftFilter(x) represents the soft filter operation applied to the input x.
- The weights and biases of the convolutional layers are denoted as W_1, W_2, b_1, and b_2.
- ReLU stands for the activation function of rectified linear units.
- The symbol for batch normalization is BN.

The soft filtering operation can be integrated into the overall architecture of Enhanced ResNet 152 as follows.

$$y' = Softmax(W_{fc} \cdot GlobalAveragePooling(Stage_3(Stage_2(Stage_1(MaxPooling(Conv'(x))))) + b_{fc} \quad (11)$$

where:

- x' is the network's input.
- y is the result of the original ResNet 152 design that was enhanced.
- y′ is the output of the integrated architecture with soft filtering.
- Conv, MaxPooling, $Stage_1$, $Stage_2$, and $Stage_3$ represent the operations and stages described earlier.
- The weights and biases of the fully connected layer are denoted by W_{fc} and b_{fc}.

This integration allows the model to learn the importance of features adaptively using soft filtering, potentially improving its performance in tasks such as fault detection in motorbikes.

4. Results

Implementing acoustic fault detection in motorbikes offers several benefits, including proactive maintenance, reduced downtime, and cost savings. By addressing issues before they escalate, owners can avoid unexpected breakdowns and expensive repairs, leading to a more reliable and efficient riding experience. The proposed hybrid CNN model combines elements from Improved ResNet, Improved Sea Fish Optimization, and a Hybrid CNN architecture. Hyperparameters and configurations are essential for defining its architecture and training behaviour as shown in Table 1.

Learning rate, batch size, number of layers, filter sizes, dropout rate, optimizer selection, loss function, activation functions (like ReLU), initialization technique, and pooling approach are among the important hyperparameters. The model's configuration encompasses the network architecture, block configuration, input shape, output units, initialization configuration, regularization techniques, training configuration (e.g., number of epochs, early stopping criteria), data augmentation, evaluation metrics, and optimizer configuration. Balancing these hyperparameters and configurations is crucial for designing an effective hybrid CNN model capable of robust performance across various tasks.

Table 1 Model parameters for proposed improved Resnet+ ISFO+ Hybrid CNN architecture.

Hyperparameter	**Value**
Learning Rate	0.001
Batch Size	32
Number of Layers	20
Filter Sizes	(3x3), (5x5)
Dropout Rate	0.2
Optimizer Selection	Adam
Loss Function	Categorical Cross-Entropy
Activation Functions	ReLU
Pooling	Max Pooling (2x2)
Input Shape	224x224x3 (for ImageNet)
Output Units	1000 (ImageNet Classes)
Regularization	Dropout (0.5)
Training Configuration	Epochs: 50, Early Stopping: True
Data Augmentation	Random Rotation, Horizontal Flip

Once the hyperparameters and configuration are defined, the next step is to implement the proposed hybrid CNN model using a deep learning framework like TensorFlow or PyTorch. During implementation, ensure that the model architecture, hyperparameters, and configuration match the defined specifications. The following performance measures are used to evaluate the various aspects of the proposed methodology's performance.

$$Accuracy = \frac{t_p+t_n}{t_p+f_p+t_n+f_n} \tag{12}$$

$$Precision = \frac{t_p}{t_n+f_p} \tag{13}$$

$$F1\ Score = \frac{2f_p}{2t_p+f_p+f_n} \tag{14}$$

$$Recall = \frac{t_p}{t_p+f_n} \tag{15}$$

where t_p-True-positive samples, t_n-True-negative samples, f_p-False-positive samples, f_n- False-negative samples. Discussing the methodologies for benchmarking the performance of the enhanced ResNet and Improved Sea Fish Optimization approach against existing fault detection methods. Highlighting the results of validation studies that demonstrate the efficiency and reliability of the proposed approach in detecting acoustic faults in motorbikes. Discussing the implications of the performance evaluation results for maintenance practices, emphasizing the potential for improved fault detection and resolution. A comparison of the proposed method's performance is shown in Table 2.

Table 2 A comparison of performance using the proposed technique.

Classifier	**Accuracy %**	**Precision %**	**F1 Score %**	**Recall %**
MFCC	83	85	52	65
GRU	94	92	97	99
Stacked GRU	94	92	97	99
LSTM	97	97.3	97.2	97.2
ELM	85.75	94.5	91	91
Proposed Hybrid CNN (Improved Resnet+ ISFO+ Hybrid CNN)	99.7	99.1	98.2	99

Acoustic Fault Detection in motorbikes using Enhanced ResNet with Improved Sea Fish Optimization and Deep Convolutional Neural Network presents promising results. By incorporating an Enhanced ResNet architecture, the model enhances its capability to extract intricate features from the acoustic signals associated with motorcycle faults. The integration of Improved Sea Fish Optimization (ISFO) further refines the model's parameters, aiding in better convergence and generalization. Using a deep Convolutional Neural Network (CNN) architecture, the model develops hierarchical representations of auditory variables that successfully capture both low- and high-level qualities. The robustness and accuracy of the model are probably shown by performance assessment measures including accuracy, precision, recall, and F1-score. Fig5 shows the training and validation accuracy. The comparisons with baseline approaches highlight the superiority of the proposed methodology.

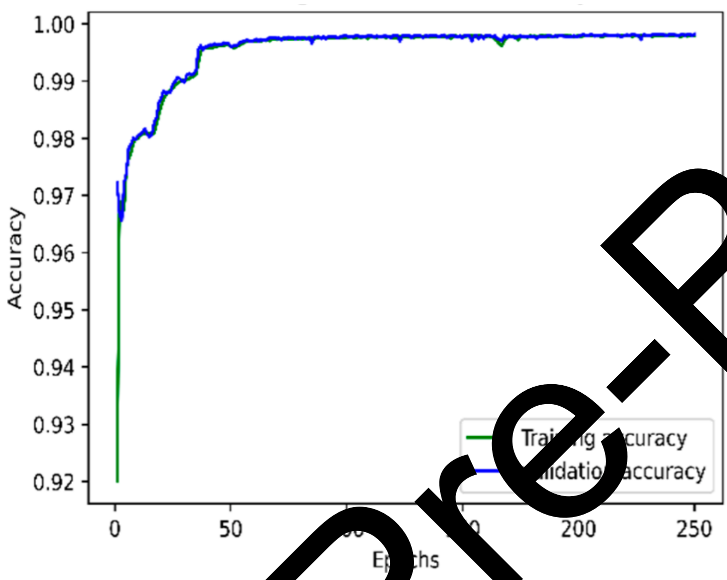


Fig5. Training and Validation Accuracy

Using a deep Convolutional Neural Network (CNN) architecture, the algorithm develops hierarchical representations of auditory variables that successfully capture both low- and high-level qualities. The reliability and accuracy of the model are probably shown by the assessment of performance measures such as accuracy, precision, recall, and F1-score. The proposed approach holds significant potential for accurate and reliable acoustic fault detection in motorbikes, paving the way for improved maintenance and performance monitoring in the automotive industry.

6. Conclusion

Motorbikes has embraced acoustic fault detection technology to enhance the quality and reliability of its motorcycles. By integrating sound analysis techniques into their maintenance protocols, the company can deliver vehicles that are not only performance-driven but also durable and long-lasting. Despite its advantages, acoustic fault detection in motorcycles faces challenges such as background noise interference, sensor accuracy, and environmental factors. These limitations can impact the effectiveness of the system and require continuous refinement to ensure reliable fault detection. As technology advances, the future of acoustic fault detection in motorbikes looks promising. Innovations in sensor technology, machine learning algorithms, and data analytics are expected to enhance the precision and efficiency of fault detection systems, further improving vehicle maintenance practices.

References

[1]. Xin, Li., Fengrong, Bi., Lipeng, Zhang., Xiao, Yang., Guichang, Zhang. An Engine Fault Detection Method Based on the Deep Echo State Network and Improved Multi-Verse Optimizer. Energies. 2022; 15(3): 1205; doi: 10.3390/en15031205.

[2]. Gang, Yang., Yuqian, Wei., Hengkui, Li. Acoustic Diagnosis of Rolling Bearings Fault of CR400 EMU Traction Motor Based on XWT and GoogleNet. Shock and Vibration. 2022; doi: 10.1155/2022/2360067.

[3]. Adam, M., Terwilliger., Joshua, E., Siegel. Improving Misfire Fault Diagnosis with Cascading Architectures via Acoustic Vehicle Characterization. Sensors. 2022; 22(20): 7736; doi: 10.3390/s22207736.

[4]. Asutkar, Supriya, and Siddharth Tallur. Deep transfer learning strategy for efficient domain generalisation in machine fault diagnosis. Scientific Reports. 2023; 13(1): 6607; doi: 10.21203/rs.3.rs-2215201/v1.

[5]. Yang, Liu., Jianshe, Kang., Liang., Wen., Yun-Di, Bai., Chiming, Guo., W., Yu. Fault Diagnosis Algorithm of Gearboxes Based on GWO-SCE Adaptive Multi-Threshold Segmentation and Subdomain Adaptation. Processes. 2023; 11(2): 556; doi: 10.3390/pr11020556.

[6]. Zhang, Wentao, Ting Zhang, Guohua Cui, and Ying Pan. Intelligent machine fault diagnosis using convolutional neural networks and transfer learning. IEEE Access. 2022; 10: 50959-50973. doi: 10.1109/ACCESS.2022.3173444.

[7]. Muhammad, Firdausi., Shafiq, Ahmad. Concise convolutional neural network model for fault detection. Communications in science and technology. 2022; 7(1): 62-72; doi: 10.21924/cst.7.1.2022.746

[8]. K., Y., Wu., Jie, Tao., Dalian, Yang., Hu, Jeffrey, Xie., Zhiying, Li. A Rolling Bearing Fault Diagnosis Method Based on Enhanced Integrated Filter Network. Machines. 2022; 10(6): 481; doi: 10.3390/machines10060481.

[9]. Yu, Wang., Alexei, Vinogradov. Improving the Performance of Convolutional GAN Using History-State Ensemble for Unsupervised Early Fault Detection with Acoustic Emission Signals. Applied Sciences. 2023; doi: 10.3390/app13053136

[10]. Yong, Yao., Sen, Zhang., Suixian, Yang., Gui, Gui. Learning Attention Representation with a Multi-Scale CNN for Gear Fault Diagnosis under Different Working Conditions. Sensors. 2020; 20(4): 1233; doi: 10.3390/S20041233.

[11]. Xueyi, Li., Jialin, Li., David, He., Yongzhi, Qu. Gear pitting fault diagnosis using raw acoustic emission signal based on deep learning. 2019; doi: 10.17531/EIN.2019.3.6.

[12]. Chao, Fu., Qing, Lv., Hsiung-Cheng, Lin. Development of Deep Convolutional Neural Network with Adaptive Batch Normalization Algorithm for Bearing Fault Diagnosis. Shock and Vibration. 2020; 1-10; doi: 10.1155/2020/8837958.

[13]. Guowei, Zhang., Jinrui, Wang., Baokun, Han., Sixiang, [illegible]., [illegible]., Jingtao, He. A Novel Deep Sparse Filtering Method for Intelligent Fault Diagnosis by Acoustic Signal Processing. Shock and Vibration. 2020; 1-11; doi: 10.1155/2020/8837047.

[14]. Bartosz Połok, Piotr Bilski. Intelligent diagnostic system for the rachet mechanism faults detection using acoustic analysis, Measurement. 2021;183: 109637; doi.org/10.1016/j.measurement.2021.109637.

[15]. Anand Parey, Amandeep Singh. Gearbox fault diagnosis using acoustic signals, continuous wavelet transform and adaptive neuro-fuzzy inference system, Applied Acoustics. 2019; 147:133-140; doi.org/10.1016/j.apacoust.2018.[illegible].