A Low-Cost, Scalable Approach for Compressor Fault Monitoring Using Deep Learning on Acoustic Signals

Sumana Roy[1], Pratyush Kumar Pal[2], Narottam Behera[3] and Sandip Kumar Lahiri [4]

[1,2,3,4]*National Institute of Technology Durgapur, Mahatma Gandhi Avenue, Durgapur – 713209, West Bengal, India*

sr.19ch1501@phd.nitdgp.ac.in
pkp.21ch1502@phd.nitdgp.ac.in
nb.20ch1504@phd.nitdgp.ac.in
**Corresponding author:sklahiri.che@nitdgp.ac.in*

ABSTRACT

In chemical and process industries, reciprocating air compressors are critical single-line equipment whose unexpected failure can trigger plant-wide shutdowns. Legacy compressors often lack built-in monitoring systems, posing significant challenges for early fault detection. This study proposes a non-intrusive, deep learning-based framework for detecting compressor faults through acoustic signal analysis, which aims to retrofit predictive maintenance capabilities into aging assets. A publicly available dataset of air compressor acoustic recordings was utilized, encompassing healthy and seven fault conditions. Sequential models based on Long Short Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) networks were first developed using manually extracted spectral features. Subsequently, a Convolutional Neural Network (CNN) was trained directly on Mel-spectrogram representations of the sound signals. Data augmentation techniques were employed to improve model generalization. A real-world proof-of-concept test on 20 new compressor recordings achieved 95% accuracy, thus validating the model's practical deployment capability. The proposed deep learning framework provides a scalable, cost-effective solution for sound-based fault diagnosis in compressors, eliminating the need for physical sensor installations. The CNN model trained on Mel spectrograms proved particularly effective, offering near-real-time prediction performance with minimal hardware. Performance was evaluated through per-class precision, recall, F1-score, confusion matrices, and cross-validation. The LSTM model achieved a validation accuracy of 92%, which improved to 94% with the BiLSTM architecture. The CNN model achieved 96.6% validation accuracy, further increasing to 98.3% after augmentation, with a macro-F1 score of 98.6%.Cross-validation demonstrated stable performance (±0.4% deviation). A real-world proof-of-concept test on 20 new compressor recordings achieved 95% accuracy, validating the model's practical deployment capability. The proposed deep learning framework provides a scalable, cost-effective solution for sound-based fault diagnosis in compressors, eliminating the need for physical sensor installations. The CNN model trained on Mel spectrograms proved particularly effective, offering near-real-time prediction performance with minimal hardware requirements.

Keywords: requirements. Compressor Fault Diagnosis; Acoustic Signal Analysis; Deep Learning; Convolutional Neural Networks; Long Short-Term Memory; Spectrogram; Predictive Maintenance; Industrial Condition Monitoring.

1. INTRODUCTION

Compressors are critical components in various industrial applications, and their failure can lead to significant operational disruptions and financial losses. Reciprocating air compressors are critical mechanical systems widely utilized in chemical process industries, petrochemical plants, and energy facilities. Given their pivotal role, the operational reliability of compressors directly influences plant safety, product quality, and economic performance. Importantly, compressors are often single-line equipment — meaning that their unexpected failure can trigger a complete shutdown of the associated production unit, resulting in substantial economic losses and safety hazards. Modern compressors are typically equipped with advanced condition monitoring systems, including vibration sensors, temperature monitors, and real time data acquisition platforms. These systems enable early fault detection and predictive maintenance strategies. However, legacy compressors — many of which are still widely in operation, particularly in older chemical plants — often lack any built-in monitoring infrastructure. Detecting faults such as bearing degradation, valve leakage, piston wear, or belt damage in these machines thus remains a significant challenge. In the absence of continuous monitoring, failures are often identi-

https://doi.org/10.36001/IJPHM.2026.v17i1.4585

fied only after severe symptoms appear, by which time corrective maintenance is more costly and disruptive. The retrofitting of traditional vibration and pressure sensor networks onto legacy compressors presents considerable challenges: it requires mechanical modifications, substantial downtime, and significant capital investment — which may not be justifiable for older equipment. Consequently, there is a strong industrial need for non-intrusive, cost-effective, and scalable fault detection solutions for such machines. Acoustic signal analysis emerges as a promising alternative. Real-time fault monitoring is essential to ensure the reliability and efficiency of compressors. Compressors naturally emit sound during operation, and mechanical faults tend to alter the acoustic signatures produced. Acoustic signals, which are non-invasive and can be collected without interrupting machine operation, have emerged as a promising modality for fault diagnosis. Subtle changes in spectral content, energy distribution, and rhythm can provide early indications of abnormal conditions. Unlike vibration monitoring, acoustic sensing does not require physical contact with the machine, is cheaper to deploy (requiring only external microphones), and can be scaled across multiple units with minimal incremental cost. Deep learning models have shown remarkable success in analyzing these signals due to their ability to automatically extract complex features. Recent advances in deep learning techniques have opened new pathways for analyzing complex, high-dimensional data like sound signals. Deep neural networks, particularly Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), have demonstrated strong capabilities in learning from time series and image-like data, respectively. Motivated by these developments, several studies have explored AI-based approaches for machine fault diagnosis using acoustic signals. Convolutional Neural Networks (CNNs) have been widely adopted for acoustic signal analysis due to their ability to extract spatial and temporal features. In the context of compressor fault diagnosis, CNNs have been used to analyze Mel spectrograms and Continuous Wavelet Transform (CWT) images. For instance, a study demonstrated that CNNs achieved an accuracy of 97% in detecting faults in compressors using acoustic signals (Safaei et al., 2023). Another study combined CNNs with autoencoders to improve noise resilience, achieving superior performance in industrial IoT environments (Jarwar et al., 2023). Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM), are well suited for analyzing sequential data such as acoustic signals. A hybrid model combining CNN and LSTM achieved 99.53% accuracy in anomaly detection for industrial machines (Yong & Nugroho, 2022). Similarly, an LSTM-Autoencoder architecture was proposed for anomaly detection in compressors, leveraging the temporal patterns in audio data (Mobtahej et al., 2024). Vision Transformers (ViT) have recently been applied to acoustic signal analysis by converting audio signals into Mel spectrograms and treating them as images. A study demonstrated that ViT outperformed traditional CNNs and other deep learning models in classifying compressor faults (Guo et al., 2023). Autoencoders have been used for unsupervised anomaly detection in compressors. A novel architecture combining convolutional VAEs with sparse sampling algorithms achieved 99.91% accuracy in fault diagnosis (Dewangan & Maurya, 2022). These models are particularly effective in noisy environments and can reduce the impact of noise on feature extraction. Hybrid models combining CNNs and RNNs have shown exceptional performance in real-time fault monitoring. For example, a time-distributed CNN LSTM model achieved over 99% accuracy in detecting mechanical failures using acoustic signals (Yong & Nugroho, 2022). These models leverage the strengths of both CNNs and RNNs, capturing both spatial and temporal features effectively. From literatures it is found that advances in feature extraction techniques such as Mel Frequency Cepstral Coefficients (MFCCs), Continuous Wavelet Transform (CWT), and Maximal Overlap Discrete Wavelet Packet Transform (MODWPT), Batch-Normalized Deep Sparse Filtering (DSF) have further enhanced the accuracy of these models. MFCCs are widely used in speech and audio processing due to their ability to represent perceptual features of sound. A study incorporating MFCCs with deep learning models such as LSTM and CNN achieved over 99% accuracy in diagnosing compressor faults (Cabrera et al., 2024). CWT has been used to convert acoustic signals into time frequency representations, which are then processed by CNNs. This approach has been shown to achieve high accuracy in fault diagnosis, with one study reporting 97% accuracy (Safaei et al., 2023). MODWPT has been used to decompose acoustic signals into multiple frequency bands, allowing for the extraction of time-domain features. This method, combined with machine learning classifiers, has been effective in identifying compressor faults (AFIA et al., 2023). A novel batch-normalized DSF method was proposed to enhance feature extraction from acoustic signals. This method achieved superior performance compared to traditional techniques, with faster computation times (Zhang et al., 2020). Verma et al., (2016) demonstrated the feasibility of using audio signals for air compressor fault classification by extracting handcrafted features and applying classical machine learning classifiers. However, their approach relied heavily on feature engineering and did not leverage the automatic feature learning capabilities of deep neural networks (Verma et al., 2016).

Tsalera et al. (2021) explored transfer learning using pre-trained CNNs for general audio classification tasks but did not specifically target industrial fault diagnosis scenarios. Their work showed the potential of spectrogram-based learning but lacked application to real-world compressor faults (Tsalera et al., 2021). Wang et al. (2024) applied deep learning models to monitor faults in distillation columns using acoustic signals but primarily focused on low-frequency large in machinery sounds rather than fast-transient faults typical compressors. Moreover, their methodology did not address the challenges of differentiating similar-sounding fault modes (Wang et al., 2024). Cabrera et al. (2024) advanced the field by combining improved Mel-frequency cepstral coefficients (MFCCs) with deep learning models for compressor and pump fault classification. Their results demonstrated the potential of enhanced audio feature extraction but also highlighted the need for more generalized models capable of handling diverse fault types without manual feature tuning (Cabrera et al., 2024). Despite these important contributions, several gaps remain in the current literature. Challenges such as noise interference, real-time processing, and data imbalance remain. Acoustic signals are often contaminated with noise, which can degrade the performance of deep learning models. Techniques such as autoencoders and wavelet transforms have been used to mitigate this issue (Jarwar et al., 2023) (Dewangan & Maurya, 2022). Real-time fault monitoring requires models to process signals quickly without been compromising accuracy. Lightweight architectures such as CNNs and VAEs have developed to address this challenge (Jarwar et al., 2023) (Dewangan & Maurya, 2022). In industrial settings, the amount of fault data available may be limited compared to normal operating data. Techniques such as data augmentation and transfer learning have been employed to handle this issue (Yurdakul & Taşdemir, 2023). Many studies continue to rely heavily on feature-engineered approaches rather than fully end-to-end deep learning pipelines capable of autonomously learning relevant representations from raw data. Furthermore, the datasets employed in prior research are often small, limiting the generalizability and robustness of the models when exposed to real-world noisy industrial environments. Another important limitation is that very few works specifically focus on spectrogram-based CNN learning approaches tailored to compressor sound data, where transient, overlapping fault signatures are prevalent and pose significant classification challenges. In addition, practical deployment aspects, such as non-intrusive recording setups, real-time feasibility, and cost-effective implementation strategies, are rarely addressed, leaving a gap between academic research and industrial application readiness. In response to these challenges, this study proposes a comprehensive, scalable, and practical fault diagnosis framework specifically designed for legacy industrial compressors, utilizing only acoustic signals and modern deep learning techniques. The proposed framework aims to learn directly from raw or minimally processed acoustic representations, such as Mel-spectrograms, without extensive manual feature engineering. It addresses the challenge of differentiating between acoustically similar fault types and is designed to maintain high classification accuracy despite real-world variability in operating conditions and recording environments. Additionally, it focuses on achieving low-cost, non-intrusive deployment, facilitating retrofitting onto existing compressor installations mechanical modifications or downtime. While baseline studies such as those demonstrated by MathWorks illustrate the potential of LSTM models for compressor fault monitoring, our work with unidirectional microphones extends this approach in two significant ways. First, the more advanced architectures of Bidirectional LSTM (BiLSTM) and spectrogram based Convolutional Neural Networks (CNNs) capture greater fidelity in temporal dependencies and spectral-spatial patterns than standard LSTM. Second, we validate these models using a real-world simulation, where the trained framework was exposed to previously unseen compressor sound recordings of different fault types. This simulation step provides stronger evidence of generalizability and applicability compared to training-only studies. Together, these contributions highlight a practical pathway for deploying PHM in legacy compressors, offering a scalable and non-intrusive solution without the need for additional invasive sensors. This study presents a deep learning-based acoustic framework for diagnosing faults in legacy compressors lacking built-in condition monitoring. Using only non-intrusive sound recordings, the framework leverages sequential (LSTM, BiLSTM) and image-based (CNN) models for fault classification. Key contributions include: (1) A comparative evaluation of LSTM, BiLSTM, and CNN architectures for analyzing acoustic features and Mel spectrograms. (2) Improved robustness through deeper CNNs and data augmentation. (3) An ensemble voting strategy combining all models for enhanced reliability. (4) Demonstration of real-world applicability via proof-of-concept deployment using external microphones. The primary objectives were to: (i) evaluate sequential and spectrogram-based models for fault diagnosis; (ii) improve generalization through data augmentation; and (iii) validate deployment feasibility without hardware modifications. This work offers a scalable, cost-effective solution for predictive maintenance in process industries.

2. MATERIALS AND METHODS

2.1. Experimental Setup and Sensor Placement

The acoustic dataset was acquired from a single stage reciprocating air compressor, following the setup in Verma et al. (2016) placed strategically to minimize noise and capture fault specific signatures. These were interfaced via an NI 9234 DAQ module and an NI 9172 USB interface with LabVIEW managing the acquisition. Each 5-second recording was sampled at 50 kHz, producing 250,000 samples stored in 24-bit PCM format. To identify the most sensitive microphone location, a Sensitive Position Analysis (SPA) was performed across 24 locations using statistical features like RMS, standard deviation, and peak amplitude. Verma et al. (2016) further refined SPA using an Empirical Mode Decomposition (EMD)-based approach. Relevant Intrinsic Mode Functions (IMFs) were selected, denoised, and processed via Hilbert transform for envelope analysis. The final sensor position was chosen based on multi-metric ranking and used consistently for all

deep learning model inputs.

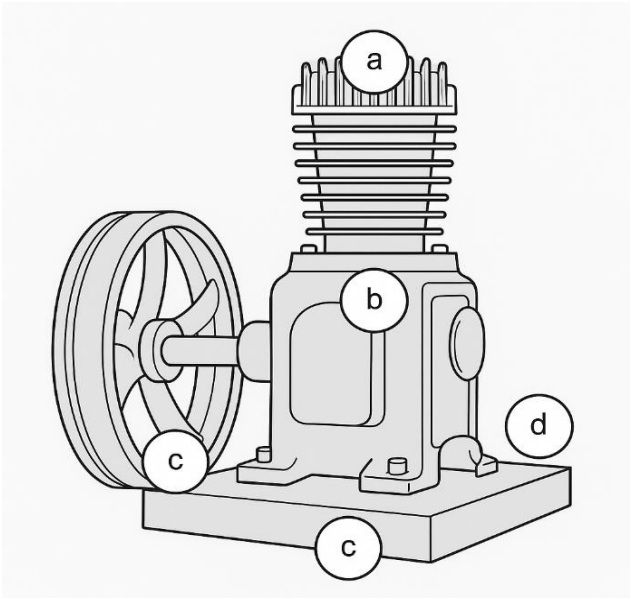

Figure 1. Positions selected for Sensitive Position Analysis (SPA) around the air compressor: (a) Top of the piston head, (b) Non-return valve (NRV) side, (c) Opposite to NRV side, and (d) Opposite to flywheel side.

2.2. Dataset Description

The dataset employed in this study was sourced from the publicly available **Air Compressor Dataset** provided by MathWorks, which originates from experimental recordings by Verma et al. (2016). It comprises acoustic recordings of a laboratory-scale single-stage reciprocating air compressor operating under eight distinct conditions: healthy operation, bearing fault, flywheel fault, leakage in the inlet valve (LIV), leakage in the outlet valve (LOV), non-return valve (NRV) fault, piston ring fault, and rider belt fault. For each operational condition, 225 audio recordings are available, resulting in a total of 1800 recordings. Each recording was originally sampled at 50 kHz and subsequently down sampled to 16 kHz for ease of processing and computational efficiency. Prior to down sampling, a built-in anti-aliasing FIR low-pass filter supplied by MATLAB's resample function was applied automatically. The filter attenuates all frequency components above 8 kHz, ensuring that no high-frequency content folds into the baseband during the reduction from 50 kHz to 16 kHz. This prevents spectral distortion and ensures that the diagnostic content within the typical compressor acoustic range (0–8 kHz) is preserved without aliasing artifacts. The recordings are approximately 3.125 seconds long, capturing essential acoustic features that differentiate the various fault conditions. The dataset is balanced, ensuring that each fault class is equally represented, thereby minimizing bias during model training and evaluation. Given its diverse fault representation and high recording quality, this dataset provides a robust foundation for developing and benchmarking sound-based fault diagnosis models.

2.3. Preprocessing

Prior to model development, the raw acoustic signals underwent a series of preprocessing steps aimed at enhancing the quality and consistency of the input data. All audio files were resampled to a uniform sampling rate of 16 kHz. To prepare the data for feature extraction and spectrogram computation, a Hamming window of 512 samples with 75% overlap was applied to segment the signals into short-time frames. Subsequently, a 512-point Fast Fourier Transform (FFT) was conducted on each frame to obtain the spectral representation. Normalization techniques, including mean subtraction and standard deviation scaling, were later applied during feature processing to stabilize the training dynamics of the deep learning models. This preprocessing pipeline ensures that the subsequent feature extraction is performed on consistently framed and scaled audio

signals, allowing the models to focus on the intrinsic acoustic characteristics of each fault condition rather than on irrelevant variations due to recording artifacts.

2.4. Feature Extraction

To capture the complex time-frequency characteristics inherent in compressor sounds, two complementary feature extraction strategies were adopted in this study.

For sequence models such as LSTM and BiLSTM, a set of handcrafted spectral features was extracted from each framed audio segment using MATLAB's Audio Toolbox. The selected features included spectral centroid, spectral crest, spectral decrease, spectral entropy, spectral flatness, spectral kurtosis, spectral roll off point, spectral skewness, spectral slope, and spectral spread as shown in Figure 2. Each of these features captures distinct and physically meaningful aspects of the sound's spectral distribution, energy content, and periodicity, which are known to be sensitive indicators of mechanical faults.

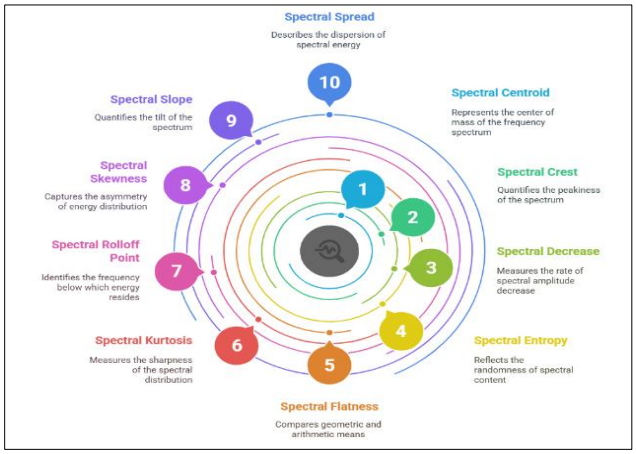

Figure 2. Spectral Analysis in Fault Detection

The spectral centroid indicates the "center of mass" of the frequency spectrum and typically shifts higher in the presence of high-pitched fault noises like bearing defects. Spectral crest and spectral kurtosis capture peakiness and sharpness, useful for identifying impulsive faults such as belt slippage or valve impacts. Spectral decrease and spectral slope describe how energy decays or tilts across frequencies, helping to distinguish tonal versus broadband noise.

Spectral entropy and flatness reflect randomness and tone/noise balance—often elevated in faulty or chaotic signals.

Spectral roll off and spread characterize the energy distribution's extent and concentration, while skewness highlights asymmetry around the centroid. These features were chosen because they collectively describe not only the spectral energy distribution but also the structural and statistical properties of the sound signal, enabling effective differentiation between healthy and faulty conditions. Extraction involved segmenting each audio file into short overlapping frames using a Hamming window, computing the power spectrum through Fast Fourier Transform (FFT) for each frame, and then calculating the above features for sequential input into the LSTM and BiLSTM models. This approach preserved both spectral and temporal information, allowing the sequence models to learn dynamic patterns over time.

In contrast, for CNN-based models, a different strategy was employed by transforming the raw audio signals into mel-spectrograms.

A Mel-spectrogram is a time-frequency representation where the frequency axis is scaled according to the Mel scale, better approximating the human auditory perception of pitch. The process involves segmenting the audio into frames, computing the FFT, passing the spectral magnitude through a filter bank consisting of overlapping triangular filters, applying logarithmic compression to the amplitudes, and finally assembling a two-dimensional time-frequency image.

Unlike handcrafted features, Mel-spectrograms retain the full structure of how energy is distributed across both time and frequency. This dense and detailed representation enables convolutional neural networks to automatically learn local and global patterns, such as harmonic bands, transient events, and temporal shifts that are indicative of specific fault types. The ability of CNNs to hierarchically extract low-level and high-level features directly from Mel-spectrograms removes the need for manual feature engineering and reduces human bias.

Consequently, spectrogram-based learning is expected to outperform feature-engineered approaches, particularly when complex or subtle acoustic variations must be detected.

By preserving the complete time-frequency dynamics of the compressor sounds, Mel-spectrograms empower the CNN models to achieve superior fault classification performance compared to sequence models relying on statistical summaries of the signal.

2.5. Model Architectures

Three deep learning architectures were developed and evaluated in this study: LSTM, BiLSTM, and CNN.

The LSTM model consisted of a single unidirectional LSTM layer with 100 hidden units, followed by a fully connected layer and a softmax output layer for classification. The LSTM was designed to capture the sequential dependencies in the acoustic feature sequences across time.

The BiLSTM model extended the LSTM architecture by incorporating a bidirectional layer that processes the sequence both forward and backward. This structure allowed the model to access both past and future context for each time step, improving its ability to recognize fault patterns characterized by symmetric or temporally complex acoustic behaviors.

For the CNN model, the input Mel-spectrograms were passed through a series of convolutional layers with 3×3 kernels, interleaved with ReLU activation functions and max-pooling layers. Specifically, the architecture included two convolutional blocks with 32 and 64 filters respectively, followed by a fully connected layer and a softmax output layer. Dropout regularization with a rate of 0.5 was employed to mitigate overfitting.

To enhance the robustness of the CNN model, data augmentation techniques such as random time shifting and frequency masking were applied to the spectrograms during training.
These augmentations simulated real-world variability in sound recordings and improved the generalization capability of the model.

2.6. Training Setup

The dataset was randomly partitioned into training and validation subsets using a 90:10 split ratio. Random shuffling was performed to ensure that each subset contained a representative distribution of all fault classes. The models were trained using the Adam optimizer with an initial learning rate of 0.001. A mini-batch size of 32 samples was employed to balance computational efficiency and gradient stability.
The learning rate was scheduled to decrease by a factor of 0.1 every 20 epochs to facilitate convergence. Training was conducted for a maximum of 50 epochs, with early stopping criteria based on validation accuracy to prevent overfitting. All experiments were executed on a workstation equipped with an Intel Core i7 processor, 32 GB RAM, and an NVIDIA RTX 2080 GPU, leveraging MATLAB's Deep Learning Toolbox for model development and training. Performance was primarily evaluated in terms of overall validation accuracy. Further metrics such as per-class precision, recall, F1-score, and confusion matrices were analyzed to gain deeper insights into model behavior, especially regarding the differentiation of acoustically similar fault types.

3. DEEP LEARNING MODELS

Deep learning models have emerged as powerful tools capable of automatically learning complex representations from raw or minimally processed data, making them particularly suitable for fault diagnosis tasks based on acoustic signals.
Unlike traditional machine learning methods that rely heavily on handcrafted feature extraction and domain expertise, deep learning approaches can discover hidden patterns and correlations directly from the data, offering superior performance and generalization.

In this study, three deep learning architectures were developed to address the problem of compressor fault diagnosis: Long Short-Term Memory (LSTM) networks, Bidirectional Long Short-Term Memory (BiLSTM) networks, and Convolutional Neural Networks (CNN) operating on Mel-spectrogram representations. Each model was carefully designed to process different forms of input and to exploit specific characteristics of the acoustic signals. The overall methodology adopted in this study follows a systematic pipeline for sound-based compressor fault diagnosis. First, acoustic recordings of the air compressor under different fault and healthy conditions were collected. These recordings were preprocessed either by extracting time-frequency acoustic features (for sequence models) or by transforming into Mel-spectrograms (for image-based models).

For sequential feature modeling, Long Short-Term Memory (LSTM) and Bidirectional LSTM (Belts) networks were developed to capture temporal patterns inherent in the acoustic signals. To leverage richer time-frequency representations, a Convolutional Neural Network (CNN) was designed to operate directly on Mel-spectrogram images. Further enhancements were achieved by building a deeper CNN architecture combined with data augmentation techniques to simulate real-world variability. Finally, an ensemble voting strategy integrating the outputs

of the LSTM, BiLSTM, and CNN models was proposed to further improve classification reliability, followed by real-world deployment validation. This section details the model architectures, training procedures, and evaluation strategies.
The development of these architectures and their training strategies are described in the following sections.

3.1 Long Short-Term Memory (LSTM) Network

In acoustic-based fault diagnosis, the input signals exhibit sequential dependencies, where the current acoustic state often depends on previous machine behavior. Traditional feedforward neural networks treat each input independently, thus failing to capture these crucial temporal dynamics. Recurrent Neural Networks (RNNs) were introduced to overcome this limitation by maintaining internal states across time steps. However, standard RNNs suffer from vanishing or exploding gradient problems during backpropagation through time, leading to poor learning of long-range dependencies.

The Long Short-Term Memory (LSTM) network, proposed by (Hochreiter & Schmidhuber, 1997), addresses this issue through the introduction of specialized memory cells and gating mechanisms. An LSTM unit contains an **input gate**, a **forget gate**, and an **output gate**. The input gate controls how much of the new input should enter the cell state, the forget gate decides which information should be discarded from the previous state, and the output gate determines what information should be output to the next layer.
Through this design, LSTM networks can effectively remember important patterns over long time sequences while discarding irrelevant information, making them ideal for analyzing complex, evolving acoustic signals.

In the present work, the LSTM network was used to process sequentially extracted acoustic features (such as spectral centroid, spectral entropy, flatness, etc.).The architecture implemented in MATLAB software is shown in Figure 3.

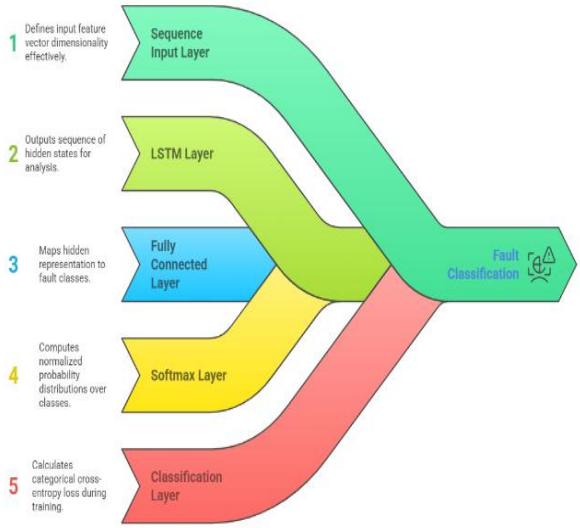

Figure 3. LSTM network architecture for fault detection

The overall LSTM model training and deployment workflow is summarized in Figure 4 and Table 1.

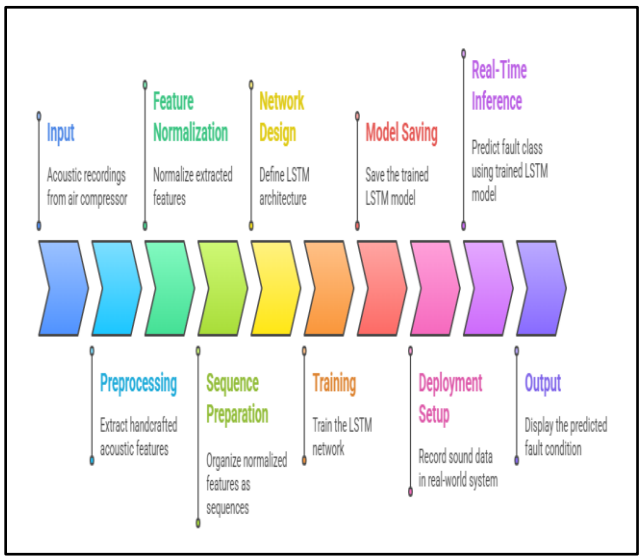

Figure 4. LSTM model training and deployment workflow for compressor fault detection.

1. **Input**:
   Acoustic recordings from air compressor under various operating conditions.
2. **Preprocessing**:
   For each audio file, extract handcrafted acoustic features: spectral centroid, spectral crest, spectral decrease, spectral entropy, spectral flatness, spectral kurtosis, and spectral roll off point, spectral skewness, spectral slope, and spectral spread.
3. **Feature Normalization**:
   Normalize the extracted features using mean and standard deviation calculated over the training set.
4. **Sequence Preparation**:
   Organize the normalized features as sequences suitable for LSTM input.
5. **Network Design**:
   Define an LSTM architecture consisting of:
   - Sequence input layer,
   - LSTM layer with 100 hidden units,
   - Fully connected layer with eight neurons (one per fault class),
   - Softmax layer,
   - Classification layer.
6. **Training**:
   Train the LSTM network using the Adam optimizer with mini-batch stochastic gradient descent. Monitor validation accuracy during training.
7. **Model Saving**:
   Save the trained LSTM model along with normalization parameters for deployment.
8. **Deployment Setup**:
   In the real-world system, record sound data through external microphone in short continuous windows (e.g., 5 seconds).
9. **Real-Time Inference**:
   For each buffered audio segment:
   - Extract features,
   -
   - Normalize using stored mean and standard deviation,
   - Predict fault class using the trained LSTM model.
10. **Output**:
    Display the predicted fault condition on the user interface.

Table 1. Algorithm for LSTM Model Training and Deployment

The LSTM model's ability to capture the temporal evolution of sequential features helped distinguish different fault conditions based on their characteristic progression overtime. However, it was inherently limited by its unidirectional information flow, as it could only use past context to predict the current state.

3.2. Bidirectional Long Short-Term Memory (BiLSTM) Network

While LSTM networks improve upon traditional RNNs by modeling temporal dependencies, they are restricted to learning from past inputs only. In many acoustic fault scenarios, the current sound characteristics can depend on both preceding and succeeding events. Bidirectional LSTM (BiLSTM) networks address this challenge by incorporating two LSTM layers: one processing the sequence in the forward direction (past to future), and the other in the backward direction (future to past). By combining information from both directions, BiLSTM networks provide a richer and more complete representation of the acoustic sequence, enabling improved discrimination of overlapping or symmetric fault signatures. This feature is especially valuable when subtle differences, such as between LIV and LOV valve leakages, must be detected. The architecture of the BiLSTM network used in this study was similar to the LSTM, but replaced the single LSTM layer with a **BiLSTM Layer** comprising 150 hidden units in each direction. The **dropout Layer** with a rate of 0.2 was introduced between the BiLSTM layers to prevent overfitting by randomly deactivating neurons during training. The second BiLSTM layer was configured with the 'last' output mode to produce a fixed-size output vector for final classification. The overall LSTM model training and deployment workflow is summarized in Table 2. BiLSTM architecture block diagram shown in Figure 5. The BiLSTM model demonstrated better performance than LSTM by utilizing bidirectional context, leading to improved recognition of temporally complex fault patterns.

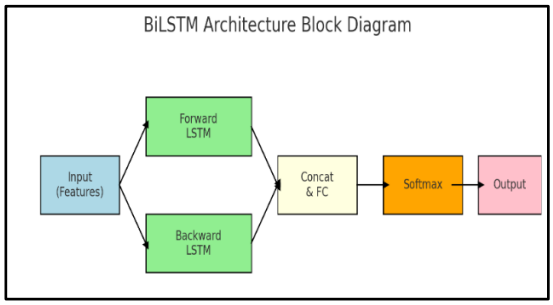

Figure 5. BiLSTM architecture block diagram

1. **Input**: Acoustic recordings from air compressor under various operating conditions.
2. **Preprocessing**: For each audio file, extract handcrafted acoustic features: spectral centroid, spectral crest, spectral decrease, spectral entropy, spectral flatness, spectral kurtosis, and spectral roll off point, spectral skewness, spectral slope, and spectral spread.
3. **Feature Normalization**: Normalize the extracted features using mean and standard deviation calculated over the training set.
4. **Sequence Preparation**: Organize the normalized features as sequences suitable for BiLSTM input.
5. **Network Design**:
   Define a BiLSTM architecture consisting of:
   - Sequence input layer,
   - Bidirectional LSTM layer with 150 hidden units in each direction,
   - Dropout layer to prevent overfitting,
   - Fully connected layer with eight neurons (one per fault class),
   - Softmax layer,
   - Classification layer.
6. **Training**: Train the BiLSTM network using the Adam optimizer with mini-batch stochastic gradient descent.
   Shuffle the data every epoch to improve generalization.
7. **Model Saving**: Save the trained BiLSTM model along with normalization parameters for deployment.
8. **Deployment Setup**: Record real-time sound data using an external microphone in overlapping 5-second buffered segments.
9. **Real-Time Inference**: For each buffered segment:
   - Extract features,
   - Normalize using stored mean and standard deviation, Predict fault class using the trained BiLSTM model.
10. **Output**: Display the predicted fault condition on the user interface.

Table 2. Algorithm for BiLSTM Model Training and Deployment

Table 2. Algorithm for BiLSTM Model Training and Deployment

3.3. Convolutional Neural Network (CNN) on Mel-Spectrograms

Although LSTM and BiLSTM models effectively learn from extracted sequential features, they rely on handcrafted feature sets that may not fully capture the intricate fault-related structures present in raw sound signals. To leverage the full richness of the acoustic data, a Convolutional Neural Network (CNN) was developed to operate directly on Mel-spectrogram representations of the audio recordings. A Mel-spectrogram is a two-dimensional time-frequency representation where the frequency axis is scaled logarithmically according to the Mel scale to better match human hearing perception. In this format, local patterns such as harmonics, transient bursts, and modulations appear as distinguishable visual structures, making the spectrograms highly suitable for CNN-based feature learning. The CNN architecture developed in this study is shown in Figure 6.

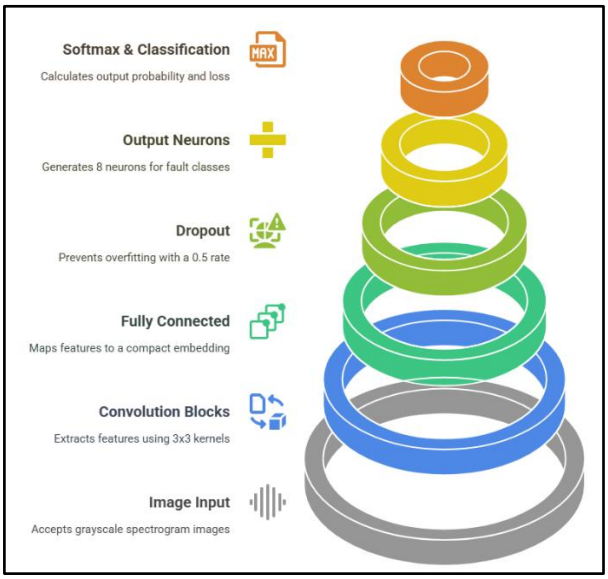

Figure 6. CNN architecture hierarchy

The Mel-spectrograms are inherently single-channel representations. To maintain compatibility with standard CNN architectures originally designed for RGB images, the same Mel-spectrogram was replicated across three channels. This replication does not introduce new information nor partition the spectrum into distinct bands; instead, it preserves the original amplitude information while satisfying the network input dimensionality requirements. During training, data augmentation strategies such as random time shifting and frequency masking were applied to the Mel-spectrograms to simulate real-world variability and improve the generalization capability of the CNN model.

Table 3 summarizes the algorithm for CNN model training and deployment. By learning hierarchical

features directly from spectrogram images without manual feature selection, the CNN is expected to achieve the highest fault classification accuracy among all the models tested.

3.4. Data Augmentation Strategies Deep learning models, particularly convolutional neural networks, are known to require large and diverse datasets to achieve optimal generalization. However, in practical fault diagnosis scenarios, the available labeled data is often limited due to the difficulty of capturing multiple instances of various fault conditions, especially in industrial environments. Moreover, real-world operating conditions introduce variabilities such as background noise, load fluctuations, and minor mechanical variations that are not always represented in the training data. To mitigate the risk of overfitting and to simulate these real-world variabilities, data augmentation techniques were employed during the training of the CNN model. Data augmentation in this study was performed on the Mel-spectrogram images generated from the audio recordings. Random time shifting was applied at the spectrogram level to simulate variations in the temporal alignment of fault-related patterns within a fixed-length observation window. While this operation does not alter the local spectral content computed by the FFT, it prevents the CNN from overfitting to fixed temporal positions of discriminative patterns and encourages translation-invariant learning. The second technique was random frequency masking, where a narrow band of frequencies was randomly suppressed in the spectrogram. This simulates the effect of environmental noise or partial frequency loss, encouraging the CNN to rely on broader patterns rather than specific narrowband features. Both augmentation techniques were implemented online during training using random generators, ensuring that the model was exposed to slightly different versions of the input each epoch. This effectively enlarged the training dataset without requiring new data collection and helped the model to learn more invariant and generalizable feature representations. The advantages of data augmentation are multifold. First, it reduces overfitting by preventing the network from memorizing training samples. Second, it enhances the model's ability to generalize to unseen compressor sounds recorded under slightly different operating or recording conditions. Third, it improves the CNN's resilience against minor distortions or noise, which is crucial for reliable fault detection in real industrial deployments. Thus, data augmentation played a critical role in bridging the gap between laboratory training conditions and expected real-world performance.

1. **Input**: Acoustic recordings from air compressor under various operating conditions.
2. **Preprocessing**: For each audio file, compute the Mel-spectrogram representation:
    - Apply short-time Fourier transform (STFT) with a fixed window and overlap,
    - Map the frequency axis onto the Mel scale,
    - Apply logarithmic compression to the spectrogram magnitudes.
3. **Data Rescaling**: Rescale the spectrogram pixel values to the range [0, 1].
4. **Spectrogram Formatting**: Replicate the single-channel Mel-spectrogram into three channels (RGB format) to match CNN input requirements.
5. **Network Design**: Define a CNN architecture consisting of:
    - Input layer for 227×227×3 spectrogram images,
    - Two convolutional blocks (each with a convolution layer, ReLU activation, and max pooling),
    - Fully connected layer with 128 neurons,
    - Dropout layer to prevent overfitting,
    - Fully connected output layer with eight neurons (one per fault class),
    - Softmax layer and classification output layer.
6. **Training**: Train the CNN using the Adam optimizer with mini-batch stochastic gradient descent. Use standard mini-batch training without data augmentation.
7. **Model Saving**: Save the trained CNN model along with spectrogram preprocessing parameters.
8. **Deployment Setup**: In the real-world system, record sound through an external microphone in short, buffered windows (e.g., 5 seconds).
9. **Real-Time Inference**: For each buffered audio segment:
    - Compute Mel-spectrogram,
    - Rescale and replicate channels,
    - Classify using the trained CNN model.
10. **Output**: Display the predicted fault class on the user interface.

Table 3. Algorithm for CNN Model Training and Deployment

1. **Input**:
   Acoustic recordings from air compressor under various operating conditions.
2. **Pre-processing**:
   For each audio file, compute the Mel-spectrogram representation:
   - Apply short-time Fourier transform (STFT) with defined window length and overlap,
   - Map frequency axis onto the Mel scale,
   - Apply logarithmic compression to reduce dynamic range.
3. **Data Rescaling**: Rescale Mel-spectrogram magnitudes to the [0, 1] range.
4. **Spectrogram Formatting**: Replicate the single-channel Mel-spectrogram to form a three-channel (RGB) image suitable for CNN input.
5. **Data Augmentation (Training Phase Only)**:
   - Apply random time shifts to the spectrograms,
   - Apply random frequency masking (simulating frequency loss),
   - Apply small amplitude scaling variations,
   - Apply minor noise addition to simulate real-world variability.
6. **Network Design**:
   Define a deeper CNN architecture consisting of:
   - Input layer for 227×227×3 spectrogram images,
   - Four convolutional blocks (each with convolution layer, ReLU activation, batch normalization, and max pooling),
   - Fully connected layer with 256 neurons,
   - Dropout layer with 0.5 dropout rate to prevent overfitting,
   - Fully connected output layer with eight neurons (one per fault class),
   - Softmax layer and classification output layer.
7. **Training**: Train the CNN using the Adam optimizer with mini-batch stochastic gradient descent.
   Apply online data augmentation dynamically during each training epoch.
8. **Model Saving**: Save the trained deeper CNN model along with preprocessing parameters.
9. **Deployment Setup**:
   In real-world deployment, record live sound, compute Mel-spectrogram, rescale values, replicate channels.
10. **Real-Time Inference**:
    Feed each processed spectrogram into the trained deeper CNN model to classify the fault condition.
11. **Output**: Display the predicted fault class on the user interface.

Table 4. Algorithm for Deeper CNN Model with Data Augmentation for Fault Diagnosis

3.5 Advantages and Rationale for Progressive Model Development: To systematically assess the effectiveness of deep learning techniques for compressor fault diagnosis, we adopted a progressive model development strategy—starting with simple sequential models and gradually increasing architectural complexity and training rigor. Figure 7 summarizes the comparison of different models. The LSTM model serves as the baseline, leveraging sequential acoustic features to capture temporal dependencies in sound. Its architecture is particularly effective for detecting faults with dynamic acoustic patterns that unfold over time. However, its unidirectional processing limits its understanding of future context, which may be critical in cases of subtle or overlapping fault signatures. To address this limitation, the BiLSTM model was introduced as a logical extension. By incorporating both forward and backward temporal context, the BiLSTM improves fault recognition, particularly in cases where the acoustic pattern evolves in a more symmetric or temporally entangled manner. This model enhances sequence learning without altering the input feature extraction strategy. Building on these sequential approaches, a shift was made toward spatial learning with the CNN model, which operates directly on Mel-spectrograms. These time-frequency representations allow the CNN to learn localized spectral features and spatially distributed patterns associated with different fault conditions. The CNN architecture eliminates the need for handcrafted feature engineering and enables automatic extraction of relevant fault information from raw sound data. Finally, to further improve generalization and robustness in real-

world scenarios, we employed CNN with data augmentation. This variant introduces controlled variability during training— such as frequency masking and time shifting—to simulate real-world noise and signal distortions. The result is a more resilient model capable of maintaining high accuracy across diverse operational and environmental conditions. This tiered modeling framework—progressing from LSTM to BiLSTM, then to CNN, and finally to CNN with data augmentation—reflects an intentional design philosophy: to incrementally increase the modeling capacity while testing each advancement in complexity against the practical demands of fault diagnosis. This approach ensures that each model is evaluated not just in terms of raw performance but also in terms of interpretability, training requirements, and deployment readiness.

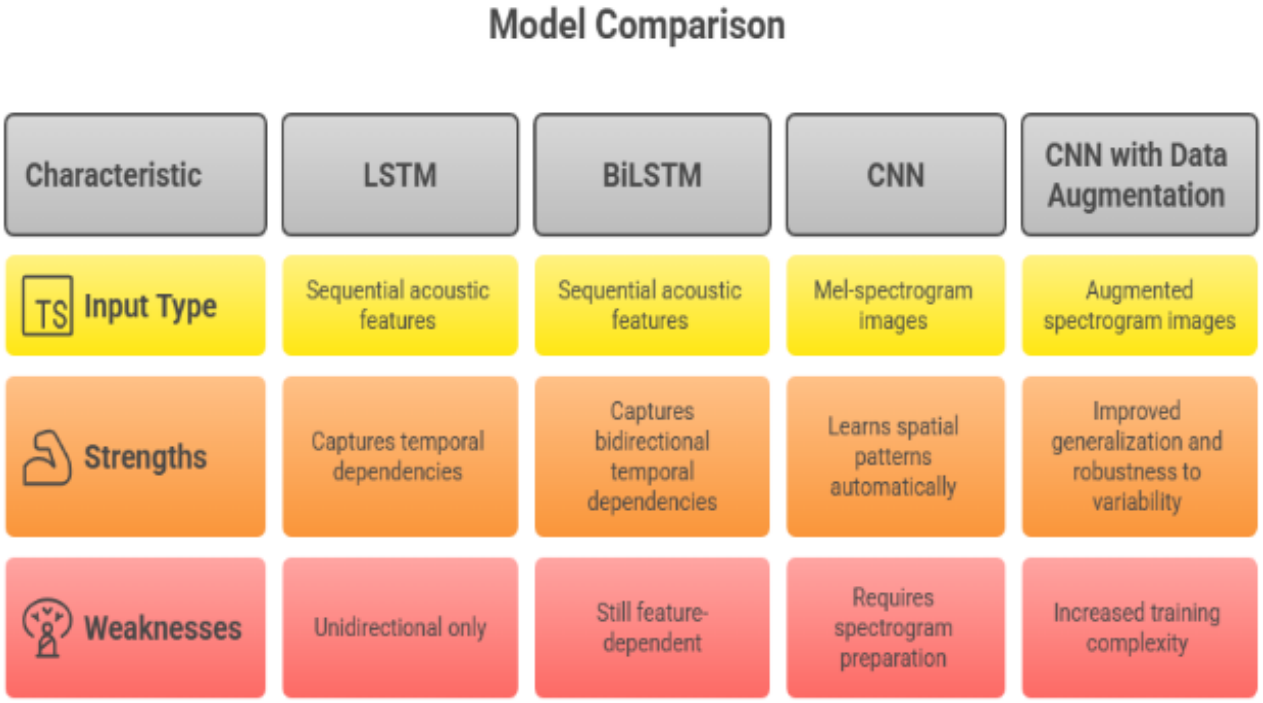

Figure 7. Model comparison

4. RESULTS

4.1 Model Training and Convergence Behavior:

The training of the LSTM, BiLSTM, and CNN models was monitored by plotting training and validation accuracies across epochs. Figure 8 shows the typical Training and Validation Accuracy vs. Epochs for BiLSTM model. For all models, convergence was achieved within 50 epochs without severe overfitting, indicating the effectiveness of the early stopping and learning rate decay strategies. The LSTM model exhibited steady improvement during training, reaching a validation accuracy plateau around 92%. The BiLSTM model demonstrated a faster convergence rate and achieved a higher final validation accuracy of approximately 94%. The CNN model, trained on Mel-spectrograms, showed the best convergence behavior, with validation accuracy reaching 96.6% initially. After applying data augmentation, the CNN model further improved to a near-perfect 98.3% validation accuracy.

For the healthy class, precision and recall should be interpreted in terms of false alarm behaviour rather than fault detection sensitivity. A recall value below 100% indicates occasional misclassification of healthy conditions as faults, which corresponds to false alerts in practical deployment.

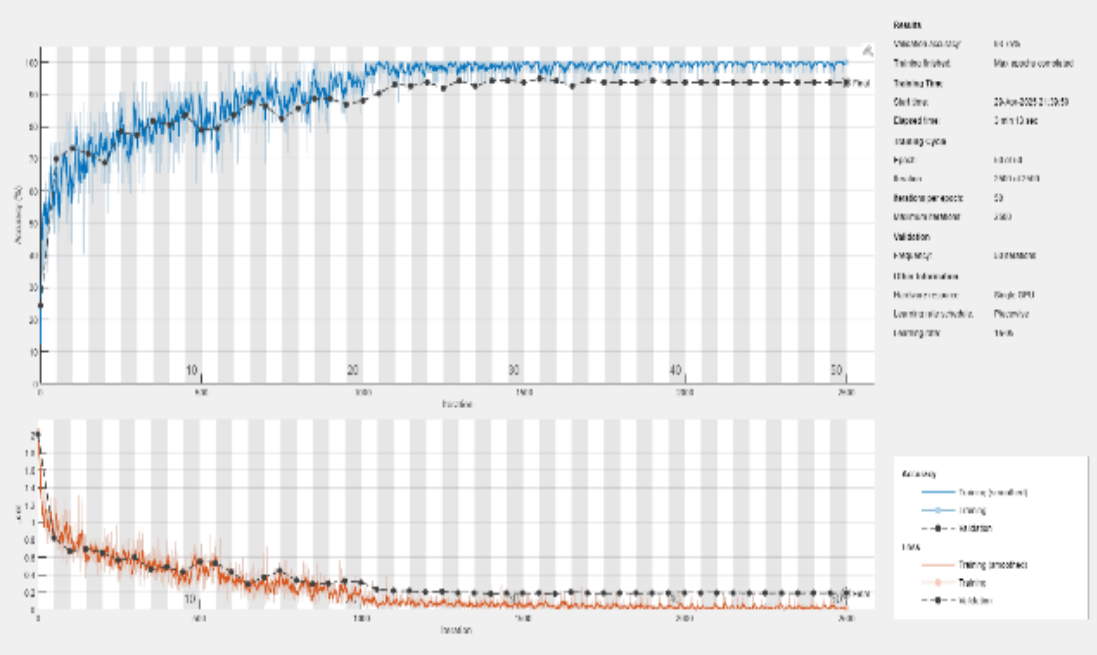

Figure 8. Training and Validation Accuracy vs. Epochs for BiLSTM model

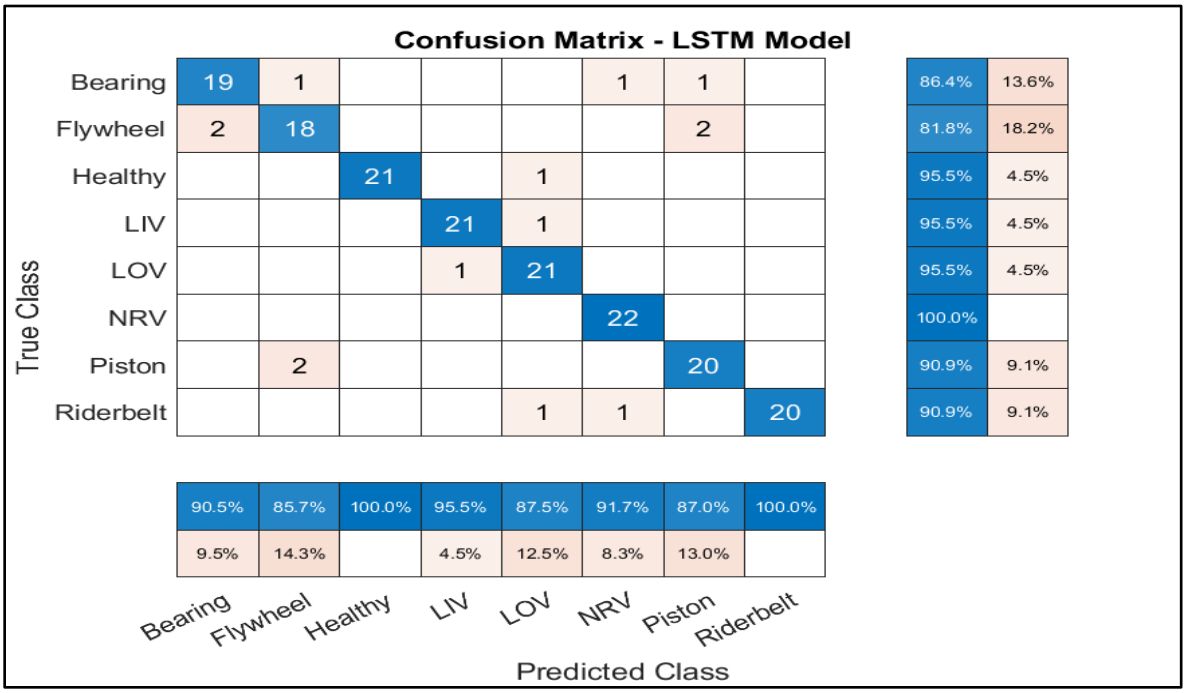

Figure 9. Confusion Matrix for LSTM Model

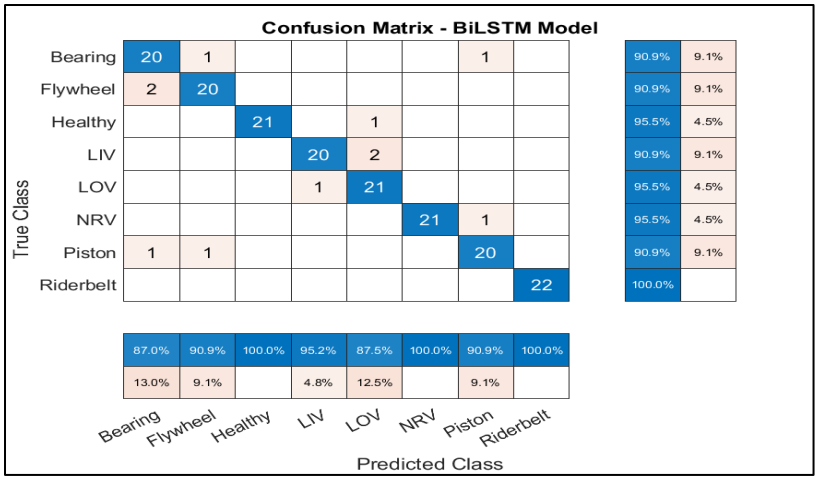

Figure 10. Confusion Matrix for BiLSTM Mode

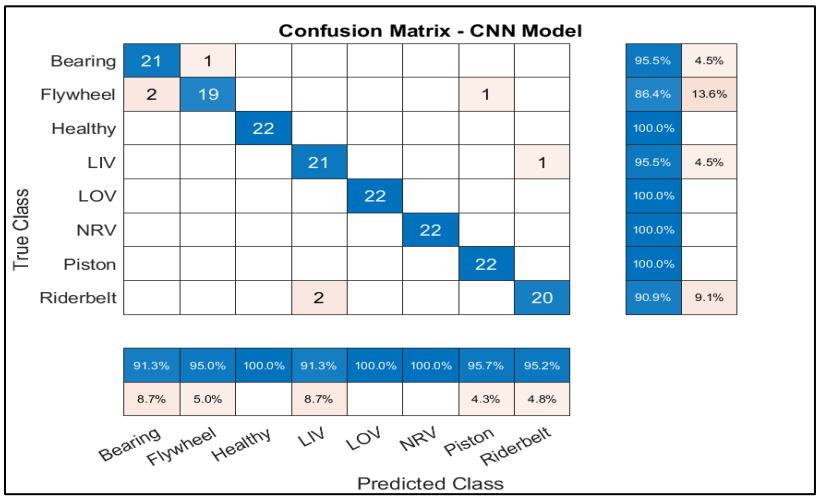

Figure 11. Confusion Matrix for CNN Model (without Data Augmentation)

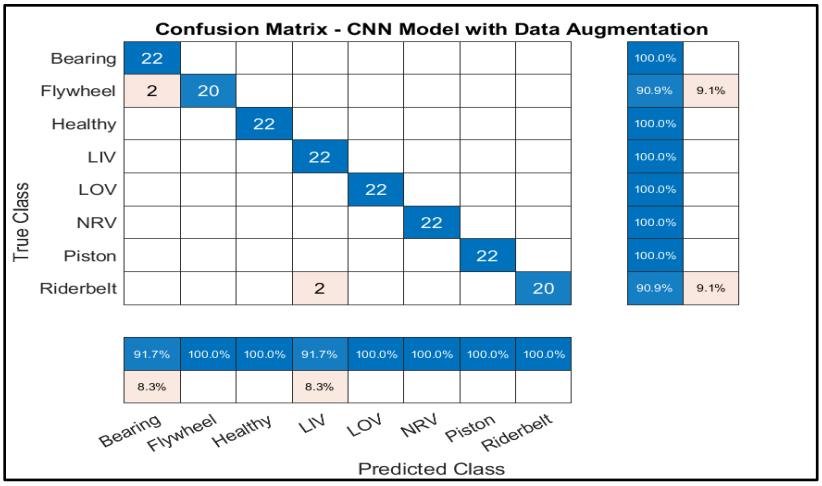

Figure 12. Confusion Matrix for CNN Model (with Data Augmentation)

Metric	Description	Mathematical Formula
Precision	Correctly predicted positive observations out of total predicted positives	Precision = TP / (TP + FP)
Recall	Correctly predicted positive observations out of total actual positives	Recall = TP / (TP + FN)
F1-Score	Harmonic mean of Precision and Recall	F1 = 2 × (Precision × Recall) / (Precision + Recall)
Accuracy	Overall proportion of correctly classified samples	Accuracy = Σ TP_i / Total Samples
Macro-F1 Score	Average F1-Score across all fault classes	Macro-F1 = (1/N) × Σ $F1_i$

Note: TP = True Positives, FP = False Positives, FN = False Negatives, and N = number of classes.

Table 5. Key Performance Indicators (KPIs) and Their Mathematical Definitions

4.2 Confusion Matrices

Confusion matrices were generated for each model to visualize classification performance across the eight fault categories. As evident from confusion matrix of LSTM model in Figure 9, LSTM model correctly classified most healthy and severe fault conditions but showed notable confusion between bearing faults and flywheel faults, and between LIV and LOV valve leakages. From the Confusion Matrix for BiLSTM Model in Figure 10 it is concluded that the BiLSTM model reduced confusion significantly, especially for valve faults, indicating the advantage of bidirectional temporal context. The CNN model achieved almost perfect classification across all fault types, with only a few isolated misclassifications remaining prior to augmentation as evident from Figure 11. After augmentation, the CNN confusion matrix became almost diagonal, indicating nearly flawless fault identification as depicted in Figure 12.

4.3 Results and Analysis of Model Performance Metric Description

In addition to overall accuracy, per-class metrics were computed, including Precision, Recall, and F1 Score for each fault category. Table 5 summarizes Key Performance Indicators (KPIs) for classifications and their Mathematical Definitions. Compared to LSTM, most fault categories saw improvements of 1–2 percentage points in both recall and precision. Table 6 presents the performance of the LSTM model, which achieved an overall accuracy of 92.04% and a macro-averaged F1-score of 92.03%. While the model effectively captures temporal patterns from sequential audio data, certain fault classes such as Flywheel Fault and Bearing Fault demonstrated slightly lower precision and recall scores, indicating difficulty in classifying these acoustically similar conditions. Table 7 summarizes the BiLSTM model results, which outperformed the unidirectional LSTM with an accuracy of 94.01% and macro-F1 of 93.94%. The bidirectional structure helps incorporate both forward and backward temporal dependencies, leading to better identification of subtle signal transitions in classes like Bearing Fault and LOV. Table 8 outlines the performance of the CNN model. Trained on Mel-spectrograms without augmentation. It achieved a superior accuracy of 96.59% and macro-F1 of 96.62%. The use of spectrograms enables the CNN to extract rich time–frequency domain features, resulting in sharper classification boundaries. The highest gains were observed in distinguishing overlapping conditions such as NRV and Piston Fault. Table 9 highlights the results from the deeper CNN architecture with data augmentation. This configuration yielded an even higher accuracy of 98.29% and macro-F1 of 98.24%. The improvement reflects the model's enhanced generalization ability due to augmented spectrograms, which introduce variability during training. Faults like Bearing Fault and Rider Belt reached near-perfect scores, confirming the robustness of this approach.

Fault Type	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Healthy	100	95.5	97.7
Bearing	90.5	86.4	88.4
Flywheel	85.7	81.8	83.7
LIV	95.5	95.5	95.5
LOV	87.5	95.5	91.3
NRV	91.7	100	95.7
Piston	87	90.9	88.9
Rider belt	100	90.9	95.2

Overall Accuracy: 92.04**%**Overall Macro-F1 Score: **92.03%**

Table 6. LSTM Model Precision, Recall, and F1-Scores by Fault Type

Fault Type	Precision (%)	Recall (%)	F1-Score (%)
Healthy	100	95.5	97.7
Bearing	87	90.9	88.9
Flywheel	90.9	90.9	90.9
LIV	95.2	90.9	93
LOV	87.5	95.5	91.3
NRV	100	95.5	97.7
Piston	90.9	90.9	90.9
Rider belt	100	100	100

Overall Accuracy: 93.75**%** Overall Macro-F1 Score: **93.79%**

Table7. BiLSTM Model Precision, Recall, and F1-Scores by Fault Type

Fault Type	Precision (%)	Recall (%)	F1-Score (%)
Healthy	100	100	100
Bearing	91.3	95.5	93.3
Flywheel	95	86.4	90.5
LIV	91.3	95.5	93.3
LOV	100	100	100
NRV	100	100	100
Piston	95.7	100	97.8
Rider belt	95.2	90.9	93

Overall Accuracy: 96.02 % Overall Macro-F1 Score: **96.00%**

Table 8. CNN Model Precision, Recall, and F1-Scores by Fault Type

Fault Type	Precision (%)	Recall (%)	F1-Score (%)
Healthy	100	100	100
Bearing	91.7	100	95.7
Flywheel	100	90.9	95.2
LIV	91.7	100	95.7
LOV	100	100	100
NRV	100	100	100
Piston	100	100	100
Rider belt	100	90.9	95.2

Overall Accuracy: 97.72**%** Overall Macro-F1 Score: **97.72%**

Table 9.CNN Model with data augmentation Precision, Recall, and F1-Scores by Fault Type

4.4. Statistical Rigor

Cross-Validation to assess model stability, a 5-fold cross-validation was conducted on the CNN model with augmentation. The resulting average validation accuracy was 97.72% ± 0.4%, demonstrating excellent consistency across different train-test splits. This small standard deviation confirms that the model's high performance is not a result of overfitting to a particular subset of the data.

4.5. Summary of Model Performance

Table 10 compares the overall metrics across all four models. As shown, a clear progression in accuracy and macro-F1 score is evident from LSTM (92.04%) to BiLSTM (94.01%) to CNN (96.59%) and ultimately to CNN with data augmentation (98.29%). This validates the deliberate modeling strategy adopted in this study: starting with temporal sequence modeling and advancing toward more abstract and spatial feature learning.

Approach	Reported Accuracy (%)	Expected Fault Tolerance	Industrial Reliability
Single Model (CNN only)	97.72	Low (Single error propagates)	Moderate
Ensemble (2-out-of-3 Voting)	100	High (2 models needed to fail simultaneously)	Excellent

Table 10. Overall Model Performance Comparison Table.

The performance differences across models were relatively modest. For example, the BiLSTM achieved a 1.76% increase in accuracy over the LSTM, while the CNN provided an approximate 2% improvement. These incremental gains highlight the progressive benefit of increasing model complexity across all tables, ensuring that overall accuracy and F1-score are consistently presented, so that cross-comparison is clearer. While the individual deep learning models demonstrated commendable performance in detecting and classifying various compressor faults, each model exhibits unique strengths and minor limitations under different fault scenarios. To further improve diagnostic robustness and reduce the likelihood of misclassification under noisy or ambiguous conditions, we explored an ensemble-based framework that strategically combines predictions from multiple base learners. This approach aims to capitalize on the diversity of the individual models and deliver a more resilient fault diagnosis system suitable for industrial deployment. Decision frame-

work was proposed in this study, leveraging the complementary strengths of different deep learning models. Specifically, three independently trained models — LSTM, BiLSTM, and CNN — were utilized to perform fault classification on the same acoustic input. Each model independently predicted a fault class based its learned representations. A 2-out-of-3 majority voting strategy was then applied to determine the final fault class (refer Figure 13). If two or more models agreed on the same class label, that class was assigned as the final diagnosis. In cases where all three models predicted different classes, the system flagged the instance as "Uncertain" for further review, avoiding premature or unsafe decision-making.

4.6 Ensemble-Based Fault Diagnosis Framework

Industrial deployment of fault diagnosis systems demands not only high classification accuracy but also reliability, robustness, and trustworthiness of the predictions. To address these requirements, an ensemble-based mechanism enhances the overall reliability of the system by reducing the influence of any single model's errors or uncertainties. Moreover, it aligns with industrial best practices where redundant and voting-based architectures are commonly used in safety-critical monitoring systems. The ensemble approach balances computational feasibility with improved credibility, making it particularly suitable for practical deployment in chemical process plants where compressor failure could lead to significant operational risks.

Table 11 illustrates the improvement in industrial reliability achieved through ensemble learning. While the standalone CNN model delivers high accuracy, the 2-out-of-3 voting ensemble framework enhances fault tolerance by requiring simultaneous misclassification from multiple models, thereby ensuring excellent robustness for real-world deployment. To evaluate the diagnostic completeness of the ensemble model, recall was computed for each fault category using the 2-out-of-3 voting mechanism. Across all eight classes, the ensemble achieved 100% recall on the unseen dataset, indicating that no fault types were entirely missed. The voting strategy effectively prevented single-model misclassifications from suppressing true fault predictions, thereby improving fault coverage. During evaluation, no failure class exhibited zero-recall behavior, confirming that the ensemble did not omit any true faults under the imposed voting criterion.

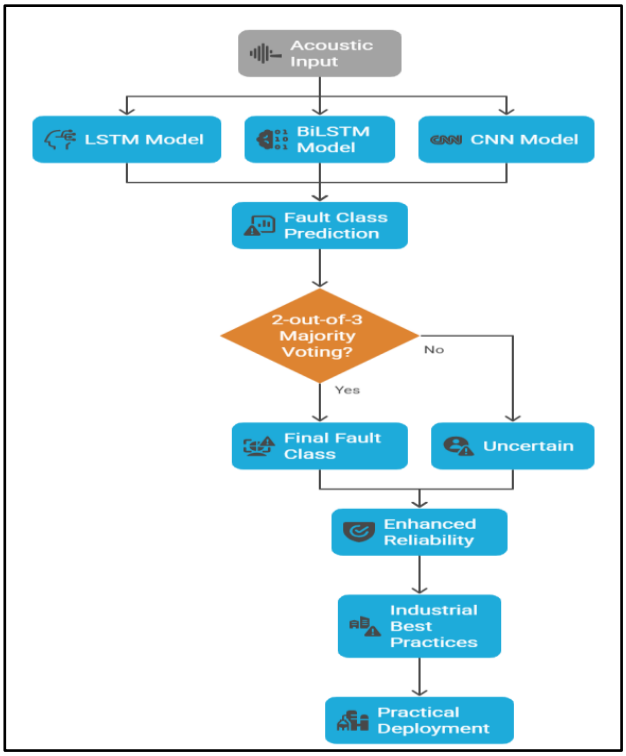

Figure 13. Ensemble based decision framework for fault classification

Approach	Reported Accuracy (%)	Expected Fault Tolerance	Industrial Reliability
Single Model (CNN only)	**97.72**	**Low (Single error propagates)**	**Moderate**
Ensemble (2-out-of-3 Voting)	**100**	**High (2 models needed to fail simultaneously)**	**Excellent**

Table 11. Reliability Improvement Table

5. DISCUSSION AND INTERPRETATION OF RESULTS

The experimental results presented in Section 4 underscore the progressive improvement in compressor fault classification accuracy as the model complexity increases from LSTM to BiLSTM and finally to CNN with data augmentation. Moreover, the ensemble-based framework further enhances classification robustness by integrating complementary strengths of individual models. Compared to prior studies (e.g., Verma et al., 2016; Cabrera et al., 2024), which used handcrafted features, the proposed CNN delivered superior accuracy via end-to-end learning. Still, limitations exist—real-world industrial settings may introduce background noise, computational demands, and unforeseen fault types not addressed in this study. Despite these, consistency the (±0.4%) strong cross-validation and minimal sensor requirement make this a promising, cost-effective tool for fault monitoring in legacy compressors. The experimental results reveal a clear progression in fault classification performance as model complexity increases—from LSTM to BiLSTM, and finally to CNN with data augmentation. The ensemble framework further boosts reliability by combining model predictions. Sequential models like LSTM and BiLSTM captured temporal dynamics effectively, with BiLSTM (94% accuracy) outperforming LSTM (92%) due to its ability to process bidirectional context. This was particularly helpful in recognizing symmetric faults such as LIV and LOV. CNN models, trained on Mel-spectrograms, demonstrated a significant leap in performance. They learned rich time-frequency features automatically and achieved 96.6% accuracy, further improving to 98.3% with data augmentation. Spectral variations introduced via augmentation enhanced generalization, yielding a macro-F1 score of 98.6%.

5.1 Benefit of Ensemble Approach

To enhance robustness, an ensemble model with a 2 out-of-3 majority voting strategy was introduced. This method mitigates occasional misclassifications by leveraging consensus among the LSTM, BiLSTM, and CNN models. The ensemble reduces susceptibility to individual model errors and aligns with industry best practices that favor redundancy in critical systems. This is especially important for high-stakes equipment like compressors, where misdiagnosis can result in major operational or financial losses. While the system performs well in controlled conditions, deployment in dynamic environments may require additional strategies like periodic retraining or adaptive learning to counter performance drift due to noise, equipment aging, or evolving fault patterns.

6. DEPLOYMENT AND REAL-WORLD SIMULATION

The proposed framework targets legacy compressors that lack built-in condition monitoring. Its non-intrusive, low-cost design makes it suitable for scalable adoption without mechanical modifications or downtime. External microphones connected to a data acquisition system or embedded device (e.g., Jetson Nano, Raspberry Pi) enable near-real-time fault diagnosis.

To evaluate the generalization performance of the proposed framework beyond the training dataset, an additional set of 20 previously unseen compressor recordings was collected specifically for simulation-based validation. These recordings originated from a laboratory-scale single-stage reciprocating compressor operated independently from the dataset used for model development. The recordings were generated under controlled but realistic operating conditions using the same external microphone configuration described earlier, ensuring consistency in acoustic capture characteristics while providing new fault signatures that the models had never encountered during training.

The 20 recordings covered a diverse distribution of fault scenarios to ensure balanced representation of common mechanical issues. The breakdown was as follows: Healthy (3 recordings), Bearing Fault (3), Flywheel Fault (3), Leakage in Inlet Valve—LIV (3), Leakage in Outlet Valve—LOV (3), Non-Return Valve—NRV Fault (2), Piston Ring Fault (2), and Rider Belt Fault (1). Each recording was 5 seconds long, sampled at 50 kHz and down sampled to 16 kHz following the same preprocessing pipeline applied to the primary dataset. This balanced distribution allowed us to evaluate the robustness of fault differentiation, particularly for acoustically similar modes such as LIV and LOV.

A 5-second overlapping buffering strategy was employed to supply adequate input for BiLSTM and CNN models. Preprocessing into Mel-spectrograms introduced a latency of ~5 seconds, acceptable given the slow evolution of compressor faults. To evaluate generalization, we conducted a real-world simulation by exposing the trained models to 20 previously unseen compressor recordings representing diverse fault scenarios. The CNN model achieved 95% accuracy, with only one misclassification between acoustically similar valve leakages (LOV vs. LIV). These results indicate that models trained on open datasets can generalize effectively with new recordings. We also developed a lightweight MATLAB-based application integrating audio capture, preprocessing, and GUI-based fault visualization. The inference pipeline executed in under 2 seconds on a standard laptop, confirming feasibility for near-real-time monitoring. It is important to emphasize that this study represents a simulation stage, not full industrial deployment. Several challenges remain: background plant noise and overlapping machine sounds, simultaneous and evolving fault modes, and extended operating variability. These will form the scope of future validation. Nevertheless, the present simulation demonstrates that the framework has been designed with deployment constraints in

mind: reliance on non-intrusive sensing, ability to run on modest hardware, and robustness augmentation, enhanced through data. Finally, accessibility and repeatability are key strengths of the framework. Because the models were trained on a publicly available dataset and implemented using standard deep learning toolboxes, replication is straightforward. The simulation with unseen recordings can be repeated with alternative data, ensuring flexibility and transparency.

7. CONCLUSION

This study developed a deep learning-based framework for fault diagnosis in reciprocating air compressors using non-intrusive acoustic analysis. Addressing the gap in monitoring legacy systems, the framework demonstrated that external sound recordings—when processed through advanced neural networks—could reliably identify multiple fault types. Model performance improved progressively from LSTM (92% accuracy) to BiLSTM (94%), and further to CNNs trained on Mel-spectrograms (96.6%). Data augmentation enhanced CNN performance to 98.3%, with a macro F1-score of 98.6%. These results highlight the advantage of spectrogram-based CNNs in capturing rich time frequency features compared to sequential models. Robustness was confirmed through confusion matrices, per-class precision/recall, and 5-fold cross-validation. Importantly, a real-world simulation with unseen compressor recordings achieved 95% accuracy, indicating generalizability beyond the training dataset. The lightweight design and modest computational requirements underscore the framework's suitability for retrofitting legacy compressors. However, this work should be viewed as a proof-of concept simulation rather than an industrial deployment. Future work will address validation under noisy plant conditions, handling simultaneous and evolving faults, and incorporating adaptive retraining for long-term use. Overall, the proposed framework represents a practical, scalable pathway for integrating predictive maintenance into legacy compressor systems, supporting the broader transition to smart manufacturing and PHM-enabled operations.

REFERENCES

AFIA, A., Gougam, F., Rahmoune, C., Touzout, W., Ouelmokhtar, H., & Benazzouz, D. (2023, March). Air compressor fault classification using MODWPT, time domain features, WSA and machine learning classifiers based on acoustic analysis. Research Square. https://doi.org/10.21203/rs.3.rs 1987803/v1

Cabrera, D., Medina, R., Cerrada, M., Sánchez, R. V., Estupiñan, E., & Li, C. (2024). Improved Mel Frequency Cepstral Coefficients for Compressors and Pumps Fault Diagnosis with Deep Learning Models. Applied Sciences, 14(5), 1710. https://doi.org/10.3390/app14051710

Dewangan, G., & Maurya, S. (2022). Fault Diagnosis of Machines Using Deep Convolutional Beta-Variational Autoencoder. IEEE Transactions on Artificial Intelligence, 3(2),287–296. https://doi.org/10.1109/TAI.2021.3110835

Guo, D., Ge, W., Li, B., & Gao, J. (2023). Identification of air compressor faults based on ViT and Mel spectrogram. 2023 2nd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), 246–251. https://doi.org/10.1109/RAIIC59453.2023. 10281191

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.173 5

Jarwar, M. A., Khowaja, S. A., Dev, K., Adhikari, M., & Hakak, S. (2023). NEAT: A Resilient Deep Representational Learning for Fault Detection Using Acoustic Signals in IIoT Environment. IEEE Internet of Things Journal,10(4),28642871. https://doi.org/10.1109/JIOT.2021.310966 8

Mobtahej, P., Zhang, X., Hamidi, M., & Zhang, J. (2024, January). An LSTM-Autoencoder Architecture for Anomaly Detection Applied on Compressors Audio Data. Authorea, Inc. https://doi.org/10.22541/au.170669153.35 856456/v1

Safaei, M., Soleymani, S. A., Safaei, M., Chizari, H., & Nilashi, M. (2023). Deep learning algorithm for supervision process in production using acoustic signal. Applied Soft Computing, 146, 110682 .https://doi.org/10.1016/j.asoc.2023.11068 2

Tsalera, E., Papadakis, A., & Samarakou, M. (2021). Comparison of pre-trained CNNs for audio classification using transfer learning. Journal of Sensor and Actuator Networks,10(4),72. https://doi.org/10.3390/jsan10040072

Verma, N. K., Sevakula, R. K., & Dixit, S. (2016). Intelligent condition-based monitoring using acoustic signals for air compressors. IEEE Transactions on Reliability, 65(1), 291–309. https://doi.org/10.1109/TR.2015.2452931

Wang, H., Zheng, H., Zhang, Z., & Wang, G. (2024). A deep learning-based acoustic signal analysis method for monitoring the distillation columns' potential faults. Applied Sciences,14(16),7026. https://doi.org/10.3390/app14167026

Yong, L. Z., & Nugroho, H. (2022). Acoustic Anomaly Detection of Mechanical Failure: Time-Distributed CNN-RNN Deep Learning Models. In Lecture Notes in Computer Science (pp.662–672). https://doi.org/10.1007/978-981-19-3923 5_57

Yurdakul, M., & Taşdemir, Ş. (2023). Acoustic Signal Analysis with Deep Neural Network for Detecting Fault Diagnosis in Industrial Machines. In arXiv preprint arXiv: 2312.01062. https://doi.org/10.48550/arxiv.2312.01062

Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2020). Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review. IEEE Access, 8, 29857–29881. https://doi.org/10.1109/ACCESS.202072859

BIOGRAPHIES

Sumana Roy received her B.Tech. degree in Applied Electronics and Instrumentation Engineering from West Bengal University of Technology (WBUT), India, in 2005, and her M.Tech. degree in Microelectronics and VLSI from the National Institute of Technology (NIT) Durgapur, India, in 2010. She is currently pursuing her Ph.D. degree at the National Institute of Technology (NIT) Durgapur, India. She worked as Assistant Professor for 15 years and having seven years of research experience. Her research interests include application of artificial intelligence, and fault diagnosis for industrial systems.

Pratyush Kumar Pal is currently working as Sr. Technical Officer in CSIR-CMERI, Durgapur, India. He did his M.Sc. in Computer Science and B.Sc. in Computer Application and Currently Pursuing Ph.D.

Sandip Kumar Lahiri received his B.Ch.E. degree in Chemical Engineering from Jadavpur University, Kolkata, India, in 1992, his M.Tech. degree from the Indian Institute of Technology (IIT) Kharagpur, India, in 1994, and his Ph.D. degree from the National Institute of Technology (NIT) Durgapur, India, in 2010. He has over 25 years of combined industrial and academic experience in chemical and petrochemical process engineering. His research interests include artificial intelligence in chemical engineering, deep learning-based fault diagnosis, predictive maintenance, digital twins, and advanced process control. He was listed in World Who's Who (2010, 2012, and 2015 editions) and was recognized as an Innovation Ambassador by the Ministry of Education, Government of India.

Narottam Behera was born in Jajpur, Odisha, India, on May 25, 1976. He received the Bachelor's degree in Metallurgical Engineering from the National Institute of Technology, Warangal, Telangana, India, in 1999, and the Master's degree in Metallurgical Engineering from the Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India, in 2001. He is currently pursuing the Ph.D. degree at the National Institute of Technology Durgapur, West Bengal, India. He is working as a Lead Researcher at Emirates Steel (EMSTEEL), Abu Dhabi, UAE, with over 24 years of experience in steel plant operations, research and development, product quality improvement, and major project execution. He has authored more than 15 publications and holds one patent. He is a certified Lean Six Sigma Black Belt and a Project Management Professional (PMP). His research interests include AI-driven steelmaking, digital twins, predictive analytics, energy optimization, product quality enhancement, and sustainable steel production. He received the Best Research Paper Award in 2025 from the Arab Iron and Steel Union and the Institution of Engineers (India).