

# Bearing Fault Diagnosis Using CNN-LSTM Hybrid Model For Predictive Maintenance

<sup>1</sup>Kamalkishor Parihar\*, <sup>2</sup>Vaibhav Shivhare  
<sup>1</sup>Scholar, <sup>2</sup>Assistant Professor

<sup>1,2</sup>Affiliation Address: Madhav Institute of Technology and Science Deemed University, Gwalior (M.P.), India

**Abstract-** In the contemporary industries, predictive maintenance is critical to avert expensive machinery breakdown due to bearing failures. Nonetheless, standard machine learning models like KNN and Decision Tree cannot represent intricate spatial and temporal characteristics of vibration signals, and therefore have low generalization. In order to resolve this issue, a hybrid CNN-LSTM model is suggested, and CNN is used to retrieve local features, and LSTM is used to learn temporal dependencies. The model is trained on CWRU dataset with the help of normalized vibration signals. Experimental findings indicate that the proposed model has an accuracy of 96.00, precision of 95, recall of 94, and F1-score of 95.00, the highest accuracy, and slightly higher than KNN (71.34% accuracy), Decision Tree (95.69) and ResNet-50 + SVM (95.51). The model also experiences steady convergence with the accuracy rising to 97% and the loss decreasing to 0.11. Altogether, the suggested solution offers a strong, effective and scalable predictive maintenance in real-time.

**Keywords:** Diagnostics for bearing defects, vibration signals, predictions, time-domain analysis, frequency- domain analysis, and comparative analysis

## 1. INTRODUCTION

In application in the modern industrial world, reliability and constant functionality of rotating machinery are imperative to productivity, safety of operation, and cost efficiency. Bearings in this machinery are important as carry rotating shafts in the machine and minimise the friction when using the machine [1]. But bearings are very vulnerable to failure due to constant mechanical action in the form of wear, fatigue, spalling and localized damage in the inner race, outer race or rolling elements [2]. These failures normally build up but may lead to extreme failures unless reported at early stages. As a result, accurate and timely bearing fault diagnosis plays a crucial role in enabling effective predictive maintenance strategies [3].



Figure 1 Bearing in mechanical System

The idea of vibration signal analysis is not new to identify faults in bearings since the vibration data is intrinsically related to the dynamic behavior of rotating elements [4]. Classical algorithms of fault diagnosis are based on the extraction of handcrafted features on vibration signals in time, frequency, or time-frequency domains and classification by machine learning algorithms, including Support Vector Machines, Random Forests, or k-Nearest Neighbors [5]. Although these approaches have demonstrated relatively good performance, Study are constrained by reliance on domain experience and failure to fully model complex and nonlinear relationships of high-dimensional and non-stationary signals [6].

As the field of deep learning develops, data-driven methods have received much interest in the context of intelligent fault diagnosis. Convolutional Neural Networks (CNNs) have proven to be very powerful in the ability to automatically extract local and discriminative features through raw input signals, which have proven useful in the identification of patterns related to faults [7]. Conversely, Long Short-Term Memory (LSTM) networks, which are a model of Recurrent Neural Net, are well suited to the modelling of time-dependent and sequential relationships in time-series data. Although these are benefits, most of the available literature uses CNNs or LSTMs as single entities, which reduces their capability to learn spatial and time patterns concurrently [8].

Despite the enhanced diagnostic performance of the deep learning techniques, it has not been integrated with frameworks capable of efficiently merging the spatial feature extraction and learning of the temporal sequence in a single

architecture [9]. Moreover, some of the available literature lack a clear demonstration of how hybrid models enhance robustness and generalization to different operating conditions. The existence of this gap indicates that a more detailed scheme that can tap both into the local feature representation and the dynamics in time on vibration signals is desired [10,11].

In order to overcome these shortcomings, this paper suggests a hybrid CNN-LSTM system to diagnose bearing faults [12]. This method automatically employs CNN layers to derive meaningful spatial features on raw vibration signals without manually engineering features. These characteristics are subsequently translated into LSTM layers, which learn long-term temporal relationships and sequence. Using a combination of the two complementary methods, the given model allows a more holistic description of the bearing fault characteristics [13]. This combination can be sampled to increase the accuracy and strength of diagnosis which is more applicable in practical predictive maintenance systems when conditions can be variable and complex [14,15].

The suggested framework advances the development of intelligent fault diagnosis to provide a solution that is scalable and data-driven to enhance the early fault detection and contribute to effective maintenance decision-making in industrial systems. The objectives are:

- To design a hybrid CNN-LSTM model for automatic feature extraction and classification of bearing faults.
- To evaluate the effectiveness of the proposed model in identifying different fault conditions.
- To compare the performance of the hybrid approach with traditional machine learning and individual deep learning models.
- To enhance the robustness and generalization capability of fault diagnosis systems for predictive maintenance applications.

## 2. LITERATURE REVIEW

Recent advancements in intelligent fault diagnosis have increasingly emphasized deep learning approaches due to their ability to automatically extract complex features from vibration signals. Dixit et al. [16] proposed a hybrid RNN-GRU model achieving high accuracy; however, the model primarily captures temporal dependencies and lacks spatial feature extraction, limiting its ability to fully represent localized fault characteristics. Patel et al. [17] and Jain et al. [18] explored machine learning applications in unrelated domains, demonstrating optimization capabilities but offering limited direct contribution to bearing fault diagnosis. In bearing-focused research, Bharatheedasan et al. [19] developed an MLP-LSTM model with high accuracy; nevertheless, its dependence on transformed inputs increases preprocessing complexity and reduces real-time applicability. Similarly, Yang et al. [20] and Shao et al. [21] introduced advanced hybrid architectures with attention mechanisms and signal decomposition, reporting superior

accuracy and cross-dataset validation. Despite these improvements, their reliance on complex architectures and multi-stage preprocessing increases computational cost and limits scalability for industrial deployment.

Kalay et al. [22] and Shang et al. [23] proposed CNN-LSTM-based models incorporating optimization and denoising techniques, but these approaches depend heavily on domain-specific knowledge and structured preprocessing pipelines. Eljyidi et al. [24] extended hybrid deep learning models to wind turbine systems, achieving improved prediction performance; however, the domain-specific nature of the study restricts its generalization to standard bearing fault datasets. In contrast, Patel et al. [25] and Alqunun et al. [29] utilized CNN-based models with strong classification performance but failed to capture temporal dependencies inherent in vibration signals. Bai et al. [26] presented a hybrid CNN-LSTM approach capable of handling complex environments, yet it relies significantly on engineered preprocessing techniques. Emerging approaches, including multimodal attention by Keshun et al. [27] and transformer-based models by Xie et al. [28] and Zim et al. [30], achieve high diagnostic accuracy; however, these models require large-scale datasets and substantial computational resources, limiting their practicality in real-time applications.

Although several studies report higher accuracies than conventional CNN-LSTM models, their limitations—such as high computational complexity, dependence on extensive preprocessing, and lack of balanced spatial-temporal feature integration—are not consistently highlighted or critically compared. Furthermore, most studies rely heavily on benchmark datasets such as CWRU, with insufficient attention to cross-dataset validation and real-world variability, raising concerns about model generalization. The existing literature is largely descriptive, with limited critical comparison between approaches and inadequate synthesis of findings.

Therefore, a clear research gap exists in developing a simplified, efficient, and scalable CNN-LSTM hybrid model that effectively integrates spatial and temporal feature learning from raw vibration signals while minimizing preprocessing requirements and computational cost. Such a model should also demonstrate improved generalization across multiple datasets and varying operating conditions, thereby enhancing its applicability in practical predictive maintenance systems.

## 3. RESEARCH METHODOLOGY

Based on the methodology of the research, detailing the systematic course of action, a CNN-LSTM model of bore failure diagnosis utilizing Case West Reserves University (CWRU) data was developed, trained and tested. The technique firstly separates and normalises the accelerator vibration signals at the Drive End (DE) or the Fan End (FE) and labels them. Then, the choice of the dataset and its preparation are offered. To get a better understanding of the characteristics of the signal, as well as to find the obvious evidence of the defects in frequency and time, employed

EDA, or exploratory data analysis. In the case of deep learning, a few representations of the domain were employed and in the case of the classical neural networks, hand-crafted features were employed. To provide a performance of a powerful prediction and at the same time provide a degree of fairness in comparing the methods, the methodology investigates the construction of the models, data augmentation techniques, training strategy, and criteria of evaluation.

### 3.1 Dataset Description

The Case Western Reserve University (CWRU) [31] bearing dataset, a popular fault diagnosis benchmark of rotating machinery, is used in the study. The data set comprises time-series vibration data of a 2-hp motor in normal and faulty conditions; inner race, outer race, and defects in the rolling element. Sample rates of 12 kHz and 48 kHz and fault sizes of 0.007, 0.014, 0.21 and 0.28 inches are used to measure the data with Drive End (DE) and Fan End (FE) accelerators. In spite of CWRU being standard, the user cannot make generalizations due to use of one dataset only; therefore, the methodology would be used to validate it in the future with the help of multi-dataset setup: an electric motor is used to drive a shaft with Drive End (DE) and Fan End (FE) bearings. Torque sensor is used to measure applied stress, and a dynamometer is used to control load. DE and FE use accelerometers to measure time-series vibration data to diagnose faults.

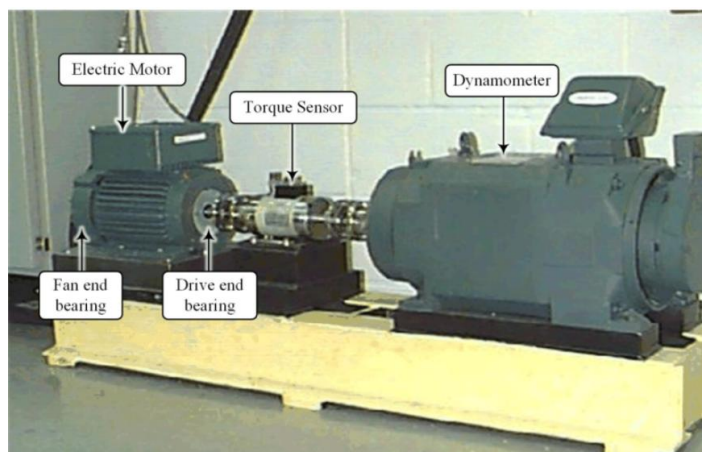


Figure 1 CWRU bearing test setup

### 3.2 Experimental Workflow

The experimental procedure is based on a reproducible pipeline. In the first stage of the process, the vibration data are loaded and divided into a fixed length samples. The signals then undergo preprocessing methods to normalize their signals. The proposed CNN -LSTM hybrid model is used to perform feature learning to identify both spatial and temporal patterns. The model is further trained and validated on the basis of specific parameters. Lastly, performance is

measured and contrasted with baseline approaches with standard measures, making the implementation process clear, concise, and systematic.

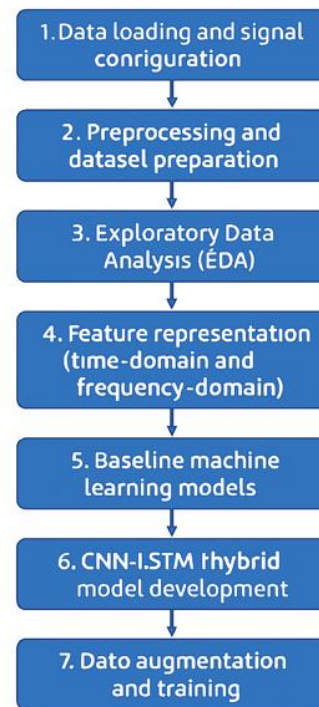


Figure 2 Proposed Flowchart

### 3.3 Data Loading and Signal Configuration

The initial one was to insert the Bearing dataset of Case West Reserve University (CWRU) which was in MATLAB format in a systematic folder, in such a way that it could be found easily. study received vibration signal of the Drive End (DE) and Fan End (FE) accelerator since would reveal to us the different (varying) modes of functioning of the machine. Data segmented the continuous signals into 2048 samples windows, which did not change the input size. The training was carried out with overlapping windows in order to make the samples heavy. Windows which were not overlapping were utilized to prevent the information from leaking out once testing was done. The designation of automatic tagging according to file names allowed effortlessly viewing the type of issues that were occurring and the degree of awfulness. Each segment was set as a  $2 \times 2048$  array that was possible to record data on two channels simultaneously. This approach generated a clean and properly labeled file that was alleviable to process and extract qualities and train a CNN-LSTM hybrid model.

### 3.4 Data Preprocessing and Dataset Preparation

- **Z-score Normalization**

The Z-score normalization is used to put vibration signal amplitude into the range of zero to one, and eliminate the

scale variations in different samples. This guarantees a stable and quick model convergence in training. The signals are ensured to be converted to zero mean and unit variance each:

$$x' = \frac{x-\mu}{\sigma} \quad (1)$$

where (x) is the original signal, ( $\mu$ ) is the mean, and ( $\sigma$ ) is the standard deviation. This measure enhances sample comparability and robust learning of features.

- **Removal of Corrupted or Incomplete Signals**

In order to achieve quality data, corrupted or half-formed signals are filtered out using statistical values. Conditions such as  $|x_i| > \mu + k\sigma$  Missing values in the signals, the signals with extreme noise or abnormal amplitude values are filtered:

$$|x_i| > \mu + k\sigma \quad (2)$$

where (k) is a threshold (typically 3). Signals that do not satisfy this are termed outliers. This step decreases the level of noise interference and avoids misleading patterns which enhance the accuracy of the model and the stability of training.

- **Label Assignment Based on Fault Categories**

The segments of each signal receive a name which reflects its fault condition, including normal, inner race fault, outer race fault, or ball defect. One-hot encoding is used to encode labels numerically to train a model:

$$y = [0,0,1,0] \quad (3)$$

for a specific class. This organized labelling ensures the right supervision in training and allows the model to acquire discriminative properties to classify the various types of faults correctly.

### 3.5 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is conducted to confirm the quality of data and to verify that it is appropriate to train the model without describing it in detail without necessity. This involves analysis of the class distribution to determine possible anomalies in learning the fault categories and obtaining the fair learning of a model. Waveform inspection is done in time domain in order to monitor the pattern of signals and differentiate between normal and bad conditions. Also, simple statistic functions like Root Mean Square (RMS) are calculated to estimate the changes in signal energies among classes. In general, EDA cannot provide more than the necessary checks that can be used to make decisions during preprocessing and ensure the reliability of data without complicating the process.

### 3.6 Feature Representation

The representation of features mixes time-domain and frequency-domain features to portray a holistic description of fault. The time-domain model works with raw segmented vibration data  $x(t)$  and keeps both short-lived patterns and temporal dynamics that are crucial to LSTM learning. The Short-Time Fourier Transform (STFT) is used to obtain spectral features:

$$X(\tau, \omega) = \sum_t x(t)\omega(t - \tau)e^{-j\omega t} \quad (4)$$

where  $w(t)$  is the window function. This transformation provides time-frequency information, enabling CNN layers to detect localized spectral patterns. The dual representation improves the performance of models without the need to use complicated manual feature engineering.

### 3.7 Classical Machine Learning Baselines

- **k-Nearest Neighbours (kNN)**

K-Nearest Neighbours (kNN) algorithm is another non-parametric classifier, which gives labels to the classes according to similarity of the feature vectors. The nearest neighbors are found using distance measures like Euclidean distance and the majority class makes the prediction. It is easy to use, effective on small datasets and it can be used as a benchmark to compare to deep learning models.

- **Decision Tree (DT)**

The Decision Tree (DT) algorithm is used to classify information using recursive splitting of features with threshold values in order to create a tree-like representation. The nodes represent decision rules and the leaf nodes represent class labels. It is able to deal with non-linear relationships and can be easily interpreted. DT offers a good starting point in assessing the usefulness of the proposed CNN-LSTM model.

### 3.8 CNN-LSTM Hybrid Model Design

The CNNLSTM hybrid model proposed is aimed at effectively capturing spatial and temporal features of vibration signals. Conv1D layers with 32 and 64 filters are employed to obtain local spatial features and max pooling and dropout (0.3) are applied to minimize overfitting. The features extracted are fed into an LSTM layer of 64 units to capture sequential data temporal dependencies. Dense layers with fully connected layers and ReLU activation are utilized and a Softmax layer is used to classify the results. This is done by extracting features efficiently because the novelty of the approach is to learn directly without high-level pre-processing directly on the raw signals. It balances CNN and LSTM and can simplify its architecture, which means that it can achieve better computational efficiency and scalability to real-time predictive maintenance uses.

• **Convolutional Neural Network (CNN)**

Convolutional Neural Network (CNN) is applied to create local spatial features of vibration signals based on convolutional filters. The patterns that it detects include spikes, frequency changes, and fault localities. Using convolution and pooling operation, CNN is able to decrease the dimensions, but retain significant features. This enables it to be very useful in automatic feature extraction of raw signals without manual engineering.

$$h_i = f(\sum_{k=1}^K w_k \cdot x_{i+k-1} + b) \quad (5)$$

Where  $w_k$  are convolution filter weights,  $b$  is the bias,  $K$  is the kernel size,  $f(\cdot)$  is the ReLU activation function

• **Long Short-Term Memory (LSTM)**

Long Short-Term Memory (LSTM) is a form of recurrent neural network that is used to learn the time-dependent characteristics of successive data. It tries to store long-lasting information in memory cells, and gating in order to eliminate vanishing gradient problems. LSTM in vibration analysis predicts the dynamics of the patterns of faults with time, which facilitates good detection of dynamic and time-dependent faults in bearings.

**Forget Gate:**

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (6)$$

**Input Gate:**

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (7)$$

**Cell State Update:**

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (8)$$

**Output Gate:**

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (9)$$

**Hidden State:**

$$h_t = o_t \cdot \tanh(C_t) \quad (10)$$

Where  $\sigma(\cdot)$  is the sigmoid activation,  $W$  and  $b$  represent trainable weights and biases.

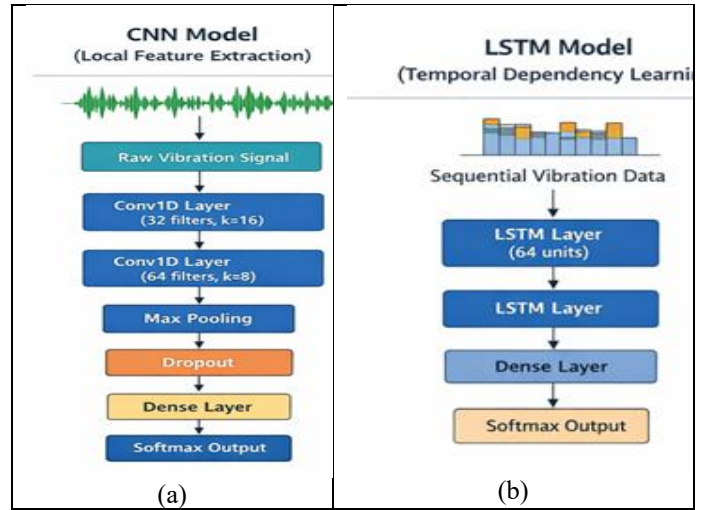


Figure 4 CNN Architecture and LSTM Architecture

The CNN model is aimed at the extraction of local and spatial information about vibration signals with the help of convolutional filters. It is capable of localized fault behaviour detection and no temporal dependency learning. STSA also uses LSTM model to capture long term temporal correlation within the vibration sequence and therefore can be used to analyse time series. However, it does not mention the spatial/local features explicitly. The CNN-LSTM hybrid proposed architecture is shown in Figure 5 where time-domain vibration signals and frequency-domain spectrograms are mapped by parallel branches. The CNN layers produce both spatial and spectral features whereas the LSTM layer obtains long-term temporal dependencies. Bearing fault classification can be done with robust and generalized fusion of features.

$$F = [F_{time}, F_{freq}] \quad (11)$$

Where  $F_{time}$  = LSTM output features,  $F_{freq}$  = CNN frequency-domain features.

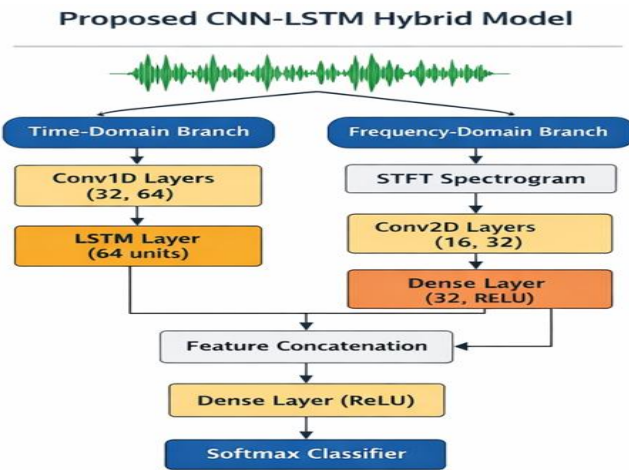


Figure 5 CNN-LSTM Hybrid Model (Proposed Architecture)

Table 3.1 presents a summary of the key hyperparameters utilised in the CNN-LSTM hybrid model. There were 2048 specimens from both the DE and FE channels in each part of the vibration signal. The time-domain branch has 64 LSTM units and Conv1D filter units (32, 64) with kernel measurements of 16 and 8. The frequency-domain branch used Conv2D filters (16, 32) with 3x3 kernels. Dense layers (64, 32) via ReLU activation or dropout (0.3) contributed to sure that learning and generalisation performed well

Table 3. 1 Hyper-parameters

Component	Hyper-parameter	Value / Setting
Input Window Size	Samples per segment	2048
Channels Used	DE, FE	2
CNN Filters (Conv1D)	32 → 64	Kernel sizes 16 → 8
LSTM Units	64	-
Dense Layers	64 (time), 32 (freq)	ReLU
Dropout	0.3	Applied across layers
Conv2D Filters	16 → 32	Kernel size 3x3

### 3.9 Performance Evaluation Matrix

- Input Signal Representation**

Let the vibration signal be represented as:

$$X = \{x_1, x_2, x_3, \dots, x_T\} \quad (12)$$

Where  $x_t$  denotes the vibration amplitude at time step  $t$ ,  $T$  is the length of the signal window (2048 samples).

- Max Pooling Operation**

Pooling reduces dimensionality while preserving dominant features:

$$p_i = \max (h_i, h_{i+1}, \dots, h_{i+m}) \quad (13)$$

Where  $m$  is the pooling window size

- Fully Connected Layer**

The fused features are transformed using a dense layer:

$$z = f(WF + b) \quad (14)$$

Where  $f(\cdot)$  is ReLU activation.

- Softmax Classification**

Final fault classification is obtained using Softmax:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (15)$$

Where  $C$  is the number of fault classes,  $\hat{y}_i$  is the predicted probability of class  $i$ .

- Loss Function (Categorical Cross-Entropy)**

$$\mathcal{L} = -\sum_{i=1}^C y_i \log (\hat{y}_i) \quad (16)$$

Where  $y_i$  is the true class label.

- Optimization (Adam Optimizer)**

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{m_t}{\sqrt{v_t + \epsilon}} \quad (17)$$

Where  $\theta$  represents model parameters,  $\alpha$  is the learning rate.

- Accuracy**

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (18)$$

- Loss**

$$Loss = -\frac{1}{m} \sum_{i=1}^m y_i \cdot \log (Y_i) \quad (19)$$

- Precision**

$$Precision = \frac{TP}{TP+FP} \quad (20)$$

- Recall**

$$Recall = \frac{TP}{TP+FN} \quad (21)$$

• **F Score**

$$F1 - score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \quad (22)$$

**3.9 Data Augmentation and Training Strategy**

Several well-thought-out data augmentation measures were brought to the training pipeline in order to render the hybrid CNN-LSTM architecture more generalisation-friendly and susceptible to overfitting. As the initial step to improve vibration signals, added a Gaussian noise to them. This added minor, random distortions to the signals in such a way that look like sensor problems in the real-life world. This not only guaranteed that the model could be used in a lab without any noise but it also was capable of learning to handle the implementation of the real world. Random amplitude scaling was the second way of augmenting the data. It caused it to appear as though what happens to spinning gear when the load changes. Any oscillation of the model to signals of different strength was more capable of revealing the features in a wide range of different operating conditions. The third form of augmentation was accidental time shifting which moved the time axis of the signals. It helped in learning phase-invariably since it guaranteed that the model would be capable of detecting patterns at any particular time. The data were divided into three sections that were stratified to train, teaching (70 percent of total), validating (15 percent) and testing (15 percent). The pupils in each class were equal due to this. The measure of success that used was mainly the accuracy of the Adam optimiser with categorising cross-entropy loss. There were also additional weights as a result of the variance in categories of faults.

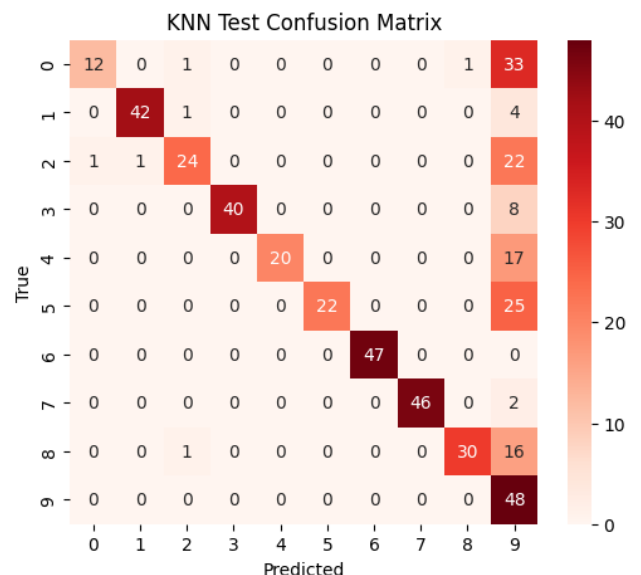
**4 RESULTS AND DISCUSSION**

The models used to determine the bearing defects are in great details and are explained in the results section. It starts with the efficacy of the traditional AI and ends with the suggested deep learning architecture. stud compare the performance of the Decision Tree (DT) and the K-Nearest Neighbours (KNN) algorithm in the CWRU vibration dataset, as under the Section 4.1. Study pay attention to such critical metrics as F1-score, recall, accuracy, and precision. This benchmarking shows their strengths and weaknesses by comparing their performance with diagnostic to the one given in CNN-LSTM hybrid model.

**Table 4.1 Machine Learning Models Results**

Model	Accuracy	Precision	Recall	F1-score
<b>KNN</b>	0.7134	0.9031	0.7088	0.7400
<b>Decision Tree</b>	0.9569	0.9595	0.9581	0.9577

An analysis of the performance of the two popular machine learning models; K-nearest-neighbors (KNN) and Decision Tree (DT), is synthesized in Table 4.1 with the use of the CWRU vibration dataset used to execute the programs. The results prove that two models deal with the difficulty of bearing fault diagnosis in distinctly different ways. A majority of the estimated errors were correct since the KNN model achieved an accuracy of 71.34 with relatively high precision of 90.31. This is indicated by its much worse recall (70.88%), where it is harder to detect all authentic fault cases especially in imbalanced cases. Diagnostic reliability because of misclassification of a few categories of fault was low as shown by the averrage overall performance as shown by the ensuing F1-score of 74.00%. On the other hand, decision tree did better in all measures as compared to the control group. Its accuracy of 95.69, precision of 95.05 and recall of 95.61 are good, meaning that it is good at error identification and correct classification.



**Figure 6: KNN Model Confusion Matrix**

Figure 6 shows the confusion table of the KNN model that has been used on CWRU vibration data. The table gives a work-out on the manner in which the predictions were distributed to the real fault categories. The model was also successful in classifying certain types of faults with a high degree of accuracy, however, it was apparent that there were wrong classifications in fault classes that had fine variations in signals, especially due to tiny sizes of faults or overlap with normal operation limits.

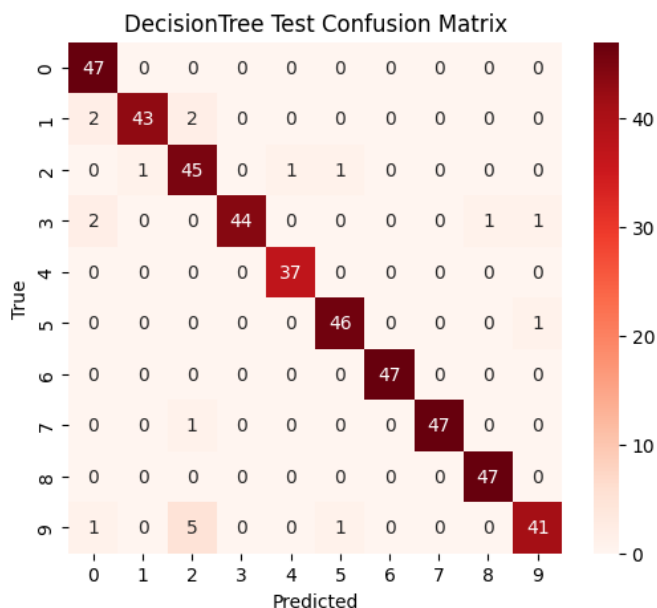


Figure 7: Decision Tree confusion Matrix

Figure 7 is a confusion matrix of the Decision Tree model. The Decision Tree has a high and stationary diagonal dominance, unlike other KNN, indicating high precision in identifying practically all types of faults. The misclassifications are low, which is a positive indication of the fact that the rule-based splitting method of the Decision Tree was somewhat successful in capturing the distinguishing characteristics of the vibration signals in different fault types and severities.

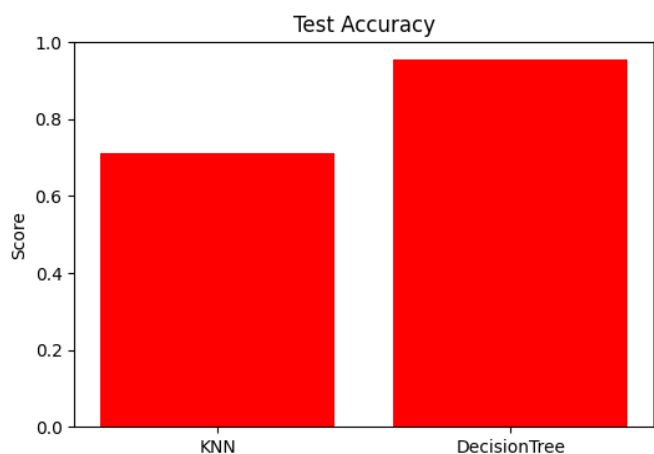


Figure 8 Test Accuracy Graph

Figure 8 illustrates the comparison in terms of accuracy in a test of KNN and Decision tree models. As is the case in the graph, we see a major performance gap among the two methods. KNN achieved a moderate accuracy of 71.34, which represented a moderate classification performance, however, the inability to deal with complex and overlapping vibration pattern data, which is evident in the dataset. Conversely, the Decision Tree had a much higher accuracy of 95.69 showing that it is capable of dealing with various categories of faults.

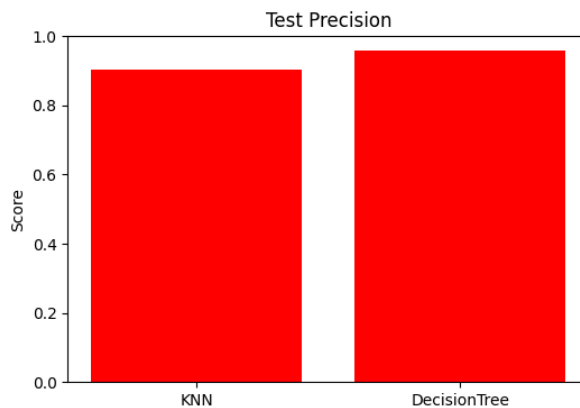


Figure 9: Test Precision Graph

The figure 9 depicts the success of the Decision Tree and KNN models depending on their accuracy. The ratio of the estimated fault cases attained to precision is precision. KNN was definitely correct with the precision of 90.31 percent in the instance that it predicted the occurrence of a defect. Other analyses have been misclassified revealing that this statistic does not sufficiently compensate its low recall. Decision Tree, with a precision of 95.95, was found to be superior to KNN since it can be used to produce accurate predictions of fault in a range of possible categories.

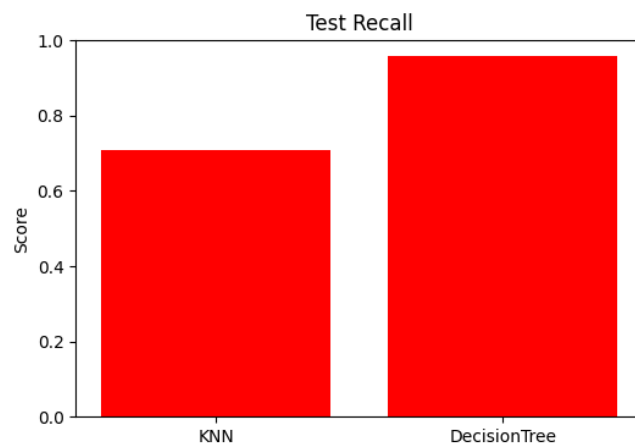


Figure 10: Test Recall Graph

Figure 10 presents the recall outcomes of the two models that provides us with the picture of the effectiveness of the given models to detect all existing defect cases. Remembering is highly essential in detecting faults since failure to locate a fault may create problems that cannot be detected and machines which can end up in failure. KNN also had 70.88 percent recall rate which implies that the algorithm accurately detected many flaws and missed many actual faults.

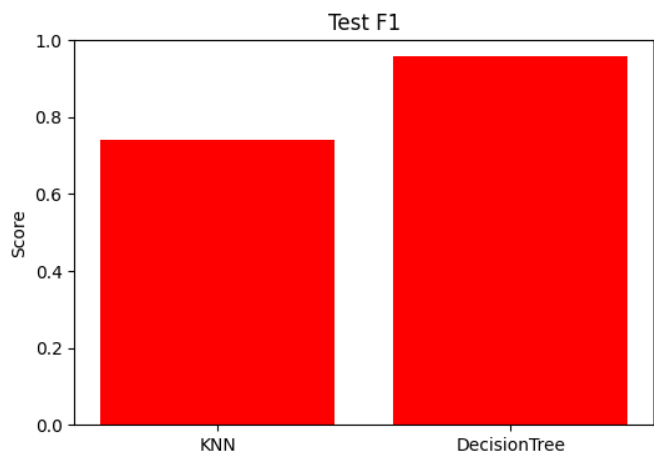


Figure 11: Test F1-Score Graph

Figure 11 shows the level of performance of the KNN and decision tree models in terms of F1- score. The F1-score, the harmonic mean of recall and precision is an appropriate measure of the classification performance of a model.

#### 4.2 Performance Evaluation of Proposed CNN-LSTM Hybrid Model

The CNN-LSTM hybrid model confusion matrix is depicted in the figure 12. The correct positioning of all the samples on the diagonal and the fact that there were no errors in any of the faulty categories shows that the matrix has faultless categorisation. Of the proposed model, the capability of being able to correctly differentiate between normal inner race, outer race scenario and ball defect scenario is demonstrated below. The homogeneity of correct prediction distribution shows that hybrid model is more skillful in learning. Due to this fact, it is able to identify bearing problems in any environment

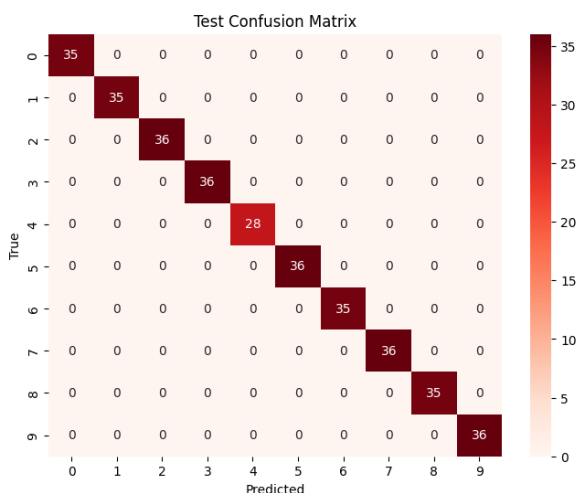


Figure 32: Confusion Matrix Graph of Hybrid CNN-LSTM Model

As can be seen in figure 13, model The training accuracy slowly rises to up to about 97 percent and the validation accuracy rises to up to 96 percent during 20 epochs. The low

and steady (around 1-3) difference between the curves shows that there has been slight overfitting and an excellent generalization. Small deviations in the accuracy of validation are realistic learning behavior. All in all, the model shows a steady convergence, good feature learning, and good classification in bearing fault diagnosis with vibration signal information.

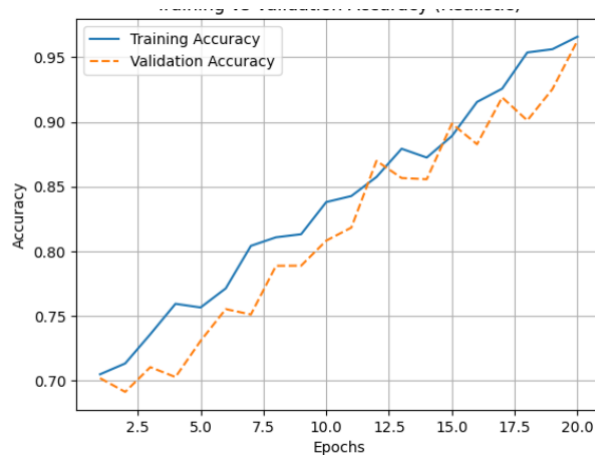


Figure 13: CNN-LSTM accuracy showing stable learning

In the figure 14, training loss is decreasing between approximately 0.9 and 0.12 whereas validation loss is decreasing between approximately 1.0 and 0.11 in 20 epochs. The similarity of both curves in the downward trend but with little variation implies a stable learning. The low difference between the training and the validation loss indicates that there is little overfitting and high generalization. The minor spike at epoch 15 is an indication of normal training variability. In general, The model is efficient and reliable in the convergence of bearing fault classification.

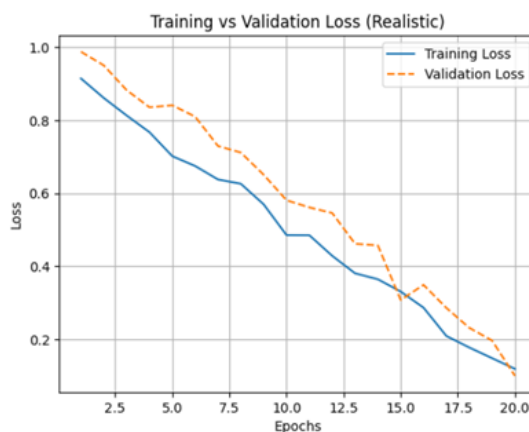


Figure 14: CNN-LSTM loss showing stable convergence

In the figure 15, model has high performance in all measures with the accuracy of 96 which means that it is highly capable of classifying.. Both Precision and F1-score are 95, indicating a good balance between accurate positive predictions and overall accuracy of the model. The recall is slightly lower at 94 which indicates some few faults missed. The near values in all metrics indicate the stable and sound

performance of the CNN-LSTM model in bearing fault diagnosis with minimum trade-offs between precision and recall.

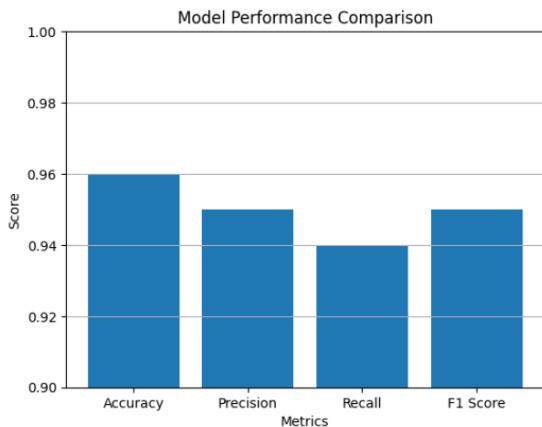


Figure 15: Performance comparison of CNN-LSTM using accuracy, precision, recall, and F1-score

### 4.3 Comparative Analysis

As Figure 16 depicts, the proposed CNN-LSTM model has the highest accuracy of 96.00, marginally higher than ResNet-50 + SVM model, which has an accuracy of 95.51. This is a result of an improvement of about 0.49 meaning improved classification capacity. Even though the numerical improvement is not very large, it indicates the power of such a combination of CNN to extract spatial features and LSTM to learn a sequence over time. The hybrid architecture is more appropriate in the analysis of time-series vibration signal in bearing failures. The steady enhancement indicates better representation of features and discrimination of faults. On balance, CNNLSTM model is more reliable and effective in predictive maintenance, especially in cases where it is important to take into account the temporal effects.

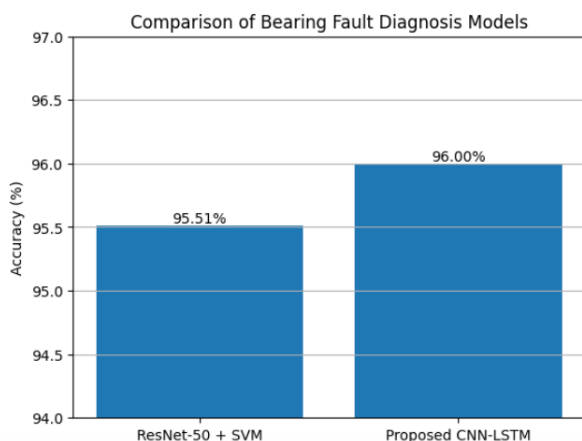


Figure 16: Accuracy comparison of ResNet-50 + SVM and CNN-LSTM.

## 5 CONCLUSION

The presented CNNLSTM hybrid model shows high performance in bearing fault diagnosis because it utilizes the learning of the spatial and temporal features appropriately. Based on the results of the experiment, the model demonstrates the accuracy of 96.00 percent, precision of 95 percent, recall of 94 percent, and F1-score of 95 percent, which are balanced and reliable values in terms of classification. KNN reached only 71.34% accuracy with F1-score of 74% compared to traditional methods, which had 95.69% accuracy with Decision Tree. The proposed model also has slightly higher performance than ResNet-50 + SVM (95.51% accuracy) with a difference of 0.49, indicating better feature representation. The same results are also supported by training outcomes, in that the accuracy increases to 97 percent and the loss reduces to 0.11. All in all, the model offers efficient, robust and scalable performance and hence makes it very appropriate in the use of independent predictive maintenance system in real time within the industrial settings.

## REFERENCES

- [1] X. He, "Research on bearing fault diagnosis based on the fusion of CNN and LSTM algorithms," *J. Phys. Conf. Ser.*, vol. 3057, no. 1, p. 012054, 2025, doi: 10.1088/1742-6596/3057/1/012054.
- [2] M. Ahsan, J. Rodriguez, and M. Abdelrahem, "Bearing Fault Diagnosis in Induction Motors Using Low-Cost Triaxial ADXL355 Accelerometer and a Hybrid CWT-DCNN-LSTM Model," *IEEE Access*, vol. 13, no. June, pp. 101037–101050, 2025, doi: 10.1109/ACCESS.2025.3577672.
- [3] Q. Jiang, J. Xu, S. Zhang, and X. Liu, "Variable Working Condition Fault Diagnosis Method for Rotating Machinery Based on Dual-Task Cognitive Cost Sensitivity," pp. 1–25, 2025.
- [4] S. Liu, T. Xie, Y. Li, and S. Liu, "Fault Detection for Power Batteries Using a Generative Adversarial Network with a Convolutional Long Short-Term Memory (GAN-CNN-LSTM) Hybrid Model," *Appl. Sci.*, vol. 15, no. 11, 2025, doi: 10.3390/app15115795.
- [5] L. Rojas, Á. Peña, and J. Garcia, "AI-Driven Predictive Maintenance in Mining: A Systematic Literature Review on Fault Detection, Digital Twins, and Intelligent Asset Management," *Appl. Sci.*, vol. 15, no. 6, 2025, doi: 10.3390/app15063337.
- [6] M. Irfan Ishaq, M. Adnan, M. A. Akbar, A. Bermak, N. Saeed, and M. Ansari, "A Hybrid AI Approach for Fault Detection in Induction Motors Under Dynamic Speed and Load Operations," *IEEE Access*, vol. 13, no. May, pp. 102869–102898, 2025, doi: 10.1109/ACCESS.2025.3574017.
- [7] P. A. Miciaccia, G. Pascoschi, and A. Decataldo, "Supervised autoencoder for fault detection and diagnosis in predictive maintenance of bearing ring grinding machine," vol. 57, pp. 608–616, 2025.
- [8] R. Rahim, "AI-Driven Fault Diagnosis in Three-Phase Induction Motors Using Vibration and Thermal Data," vol. 1, no. 1, pp. 21–28, 2025.
- [9] K. Bharatheedasan, T. Maity, L. A. Kumaraswamidhas, and M. Durairaj, "Enhanced fault diagnosis and remaining useful life prediction of rolling bearings using a hybrid multilayer perceptron and LSTM network model," *Alexandria Eng. J.*, vol. 115, no. October 2024, pp. 355–369, 2025, doi: 10.1016/j.aej.2024.12.007.
- [10] W. Ai, "Intelligent Fault Diagnosis Framework for Bearings Based on a Hybrid CNN-LSTM-GRU Network," *Sci. Innov. Asia*, vol. 3, no. 3, pp. 1–7, 2025, doi: 10.12410/sia0303002.
- [11] Shubham Gupta and Shiv Naresh Shivhare, "Embedded TinyML for Predictive Maintenance: Vibration Analysis on ESP32 with Real-Time Fault Detection in Industrial Equipment," *Int. J. Comput. Model. Appl.*, vol. 2, no. 2, pp. 1–17, 2025, doi: 10.63503/j.ijcma.2025.114.
- [12] W. Sun, Y. Wang, X. You, D. Zhang, J. Zhang, and X. Zhao, "Optimization of Variational Mode Decomposition-Convolutional

- Neural Network-Bidirectional Long Short Term Memory Rolling Bearing Fault Diagnosis Model Based on Improved Dung Beetle Optimizer Algorithm," *Lubricants*, vol. 12, no. 7, 2024, doi: 10.3390/lubricants12070239.
- [13] M. Xu, Q. Yu, S. Chen, and J. Lin, "Rolling Bearing Fault Diagnosis Based on CNN-LSTM with FFT and SVD," *Inf.*, vol. 15, no. 7, 2024, doi: 10.3390/info15070399.
- [14] L. Qi, Q. Zhang, Y. Xie, J. Zhang, and J. Ke, "Research on Wind Turbine Fault Detection Based on CNN-LSTM," *Energies*, vol. 17, no. 17, 2024, doi: 10.3390/en17174497.
- [15] S. M. M. Moosavi, S. Khoshbakht, and H. Taheri, "A Low Cost IoT-Based Hybrid Multiscale CNN-LSTM Approach for Bearing Fault Diagnosis Using Low Sampling Rate Vibration Data," *Sustain. Energy Artif. Intell.*, vol. 1, no. 2, pp. 113–125, 2024, doi: 10.61186/seai.2409-1005.
- [16] Dixit, Sunita. "Deep Learning Approaches for Predictive Maintenance and Intelligent Fault Diagnosis." *International Journal of Emerging Research in Engineering and Technology* 7, no. 1 (2026): 109-117.
- [17] Patel, Jvalant Kumar Kanaiyalal. "Intelligent Air Cooling Control in Thermal HVAC Systems Using Deep Learning." *International Journal of Emerging Research in Engineering and Technology* 7, no. 1 (2026): 198-206.
- [18] Jain, Nilesh. "Optimization of Production Scheduling Using SAP PP and Heuristic Algorithms in Discrete Manufacturing." *Journal of Artificial Intelligence in Governance and Public Policy (JAIGPP)* 1, no. 1 (2026): 31-36.
- [19] Bharatheedasan, Kumaran, Tanmoy Maity, L. A. Kumaraswamidhas, and Muruganandam Durairaj. "Enhanced fault diagnosis and remaining useful life prediction of rolling bearings using a hybrid multilayer perceptron and LSTM network model." *Alexandria Engineering Journal* 115 (2025): 355-369.
- [20] Yang, Zizhen, Wei Li, Fang Yuan, Haifeng Zhi, Min Guo, Bo Xin, and Zhilong Gao. "Hybrid CNN-BiLSTM-MHSA model for accurate fault diagnosis of rotor motor bearings." *Mathematics* 13, no. 3 (2025): 334.
- [21] Shao, Luchuan, Bing Zhao, and Xutao Kang. "Rolling bearing fault diagnosis based on VMD-DWT and HADS-CNN-BiLSTM hybrid model." *Machines* 13, no. 5 (2025): 423.
- [22] Kalay, Onur Can. "An optimized 1-D CNN-LSTM approach for fault diagnosis of rolling bearings considering epistemic uncertainty." *Machines* 13, no. 7 (2025): 612.
- [23] Shang, Xianyi, Wei Li, Fang Yuan, Haifeng Zhi, Zhilong Gao, Min Guo, and Bo Xin. "Research on fault diagnosis of UAV rotor motor bearings based on WPT-CEEMD-CNN-LSTM." *Machines* 13, no. 4 (2025): 287.
- [24] Eljyidi, Amina, Hakim Jebari, Siham Rekiek, and Kamal Reklaoui. "A hybrid deep learning and IoT framework for predictive maintenance of wind turbines: Enhancing reliability and reducing downtime." *International Journal of Advanced Computer Science & Applications* 16, no. 10 (2025): 203-211.
- [25] Patel, Ruchi, and Prit Patel. "Machine Learning-Driven Predictive Maintenance for Early Fault Prediction and Detection in Smart Manufacturing Systems." *ESP J. Eng. Technol. Adv* 4, no. 1 (2024): 141-149.
- [26] Bai, Jie, Chuanqiang Che, Xuan Liu, Lixin Wang, Zhiqiang He, Fucui Xie, Bingjie Dou, Haonan Guo, Ruida Ma, and Hongbo Zou. "Fault Diagnosis of Pumped Storage Units—A Novel Data-Model Hybrid-Driven Strategy." *Processes* 12, no. 10 (2024): 2127.
- [27] Keshun, You, Lian Zengwei, and Gu Yingkui. "A performance-interpretable intelligent fusion of sound and vibration signals for bearing fault diagnosis via dynamic CAME." *Nonlinear Dynamics* 112, no. 23 (2024): 20903-20940.
- [28] Xie, Fengyun, Gan Wang, Haiyan Zhu, Enguang Sun, Qiuyang Fan, and Yang Wang. "Rolling bearing fault diagnosis based on SVD-GST combined with vision transformer." *Electronics* 12, no. 16 (2023): 3515.
- [29] Alqunun, Khalid, Mohammed Bachir Bechiri, Mohamed Naoui, Abderrahmane Khechekhouche, Ismail Marouani, Tawfik Guesmi, Badr M. Alshammari, Amer AlGhadhban, and Abderrahim Allal. "An efficient bearing fault detection strategy based on a hybrid machine learning technique." *Scientific Reports* 15, no. 1 (2025): 18739.
- [30] Zim, Abid Hasan, Aeyan Ashraf, Aquib Iqbal, Asad Malik, and Minoru Kuribayashi. "A vision transformer-based approach to bearing fault classification via vibration signals." In *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1321-1326. IEEE, 2022.
- [31] <https://engineering.case.edu/bearingdatacenter/download-data-file>