Smarter Maintenance Machine Learning for Reliable Air Compressor Health Monitoring

Student Name: Khalid Alyahya | Student ID: g201572390
Student Name: Abdullah Alsagoor | Student ID: g201354050
Supervisor: Dr. Muzammil Behzad [muzammil.behzad@kfupm.edu.sa]
King Fahd University of Petroleum and Minerals, Saudi Arabia

Abstract

This paper re-examines the air compressor fault detection benchmark with a hybrid, end-to-end deep sequence-learning system that is learned on the raw vibration signals. This study is novel since it independently and rigorously assesses the performance of deep learning based on an all-encompassing, unseen, blind test set, thus setting a new reproducible baseline. The signals were divided into 4,340 development windows and 430 blind-test windows, the fixed length was 150 samples with stride of 10. The neural network models used were BiGRU (Model 1), Conv1D+GRU (Model 2), regularized Conv1D+BiGRU (Model 3), and non-regularized Conv1D+BiGRU (Model 4) deployed in the framework of PyTorch. Model 4 was the highest achiever in this experiment with a 93.26% accuracy in the blind test. This was better than the other models on the same blind test: Model 1 scored 74.42%, Model 2 scored 91.40%, and Model 3 (regularized Conv1D-BiGRU) scored 91.86%. The success of Model 4 proves the usefulness of particular hybrid architectures in this dataset. It validates the feasibility of effective end-to-end fault detection with the help of raw signals provided proper data pre-processing is implemented.

1. Introduction

Reciprocating compressors are important assets in process industries such as petrochemical and energy industries where it is important to have continuity in its operations. Malfunctions of such machines may cause unpredicted downtime, expensive maintenance procedures, and potentially safety incidents. Consequently, coming up with effective and sound fault diagnosis (FD) systems has emerged as a research focus. Vibration analysis is the most informative method of data-driven techniques because it is sensitive to degradation of mechanical components of the machines.

Previous researchers on an air-compressor dataset publicly available have found high, and in some cases perfect classification performance. These experiments however rely on large amounts of manual feature engineering and use simple random splits which can blur the generalization capacity of their model. As an example, Nambiar et al. (2024) obtained 100% accuracy through feature fusion and kNN, and Bhattacharyya et al. (2025) obtained

98.66% through a Weightless Neural Network. The reproducibility and scalability of such methods is limited by the fact that they are based on handcrafted features.

In this case, the study employs an end-to-end deep learning model, learning to utilize standardized raw vibration signal directly. Each of the four architectures is evaluated and a rigorous evaluation scheme; consisting of an independent blind test set that approximates real-life cases; is used. The findings reveal that development-set Cross-Validation (CV) performance and blind-set generalization are highly consistent, which shows that the non-regularized Conv1D-BiGRU architecture generalizes best, and it can achieve high accuracy without the intricate regularization. The findings provide a corrected and methodologically sound base of automated air-compressor fault diagnosis.

1.1 Problem Statement

The primary issue that was overcome in the current research is the use of manual feature engineering and simplified evaluation models. Earlier literature which reports near-perfect accuracy derives large collections of handcrafted features; statistical parameters, histogram descriptors, ARMA coefficients, converting signals data to 2D images; and expert tuning in order to generate strong performance. Although strong on this particular data, such pipelines are hard to scale and can fail to transfer to other operating conditions and equipment.

Moreover, the foregoing means that, past studies heavily rely on random train-test splits and accuracy-only reporting, which hides the actual generalization potential of the model. Many models are able to memorize temporal patterns rather than be able to learn transferable diagnostic features, even without chronological separation or blind assessment. Thus, a new standard is required; the new standard that excludes manual feature engineering and measures the performance based on strict and leakage-free blind testing. The paper bridges this methodological shortcoming by constructing automated deep sequence-learning architectures and only testing them on a completely unseen blind set.

1.2 Objectives

The following are the main objectives of this study:

- To create an end to end deep learning system that can learn discriminative representations using raw vibration windows without manual interaction features.
- To strictly test the model generalization with a development set that is completely independent of a blind test set that is also completely independent and separated in time and label.

1.3 Scope of Study

This paper addresses fault diagnosis in the case of a single-acting, single-stage reciprocating air compressor using the publicly available vibration data with five fault classes, namely CVF, GOOD, IOVF, IVF and OVF. The experiment compares four neural networks on uniformly-timed windows of raw signals of fixed size and gives only leakage-free results on blind-test performance. They do not make any generalization to other types of compressors or actual industrial variability, and the paper does not discuss few-shot or zero-fault learning. The methodological focus is rather on the possibility of providing a proper baseline of automated, end-to-end learning of features under rigorous evaluation procedures.

2. Literature Review

The diagnostics and condition-based monitoring of rotating machinery; especially reciprocating compressors; have significantly changed during the last twenty years and nowadays the modality of analysis is based on vibration [6, 11]. The conventional data-driven fault diagnosis (DDFD) pipelines are based on multi-stage manual feature engineering, in which domain experts can extract statistical, spectral, or model-based features and then classify them [1, 3, 9]. These heuristic pipelines have a history of impressive performance on lab data and, however, their reliance on manual feature engineering and vulnerability to validation leakage have recently been pointed to as the primary drawbacks in systematic reviews [7]. Simultaneously, by providing end-to-end pattern learning on raw signals, deep learning has transformed the field, making it less dependent on hand-designed features and providing more scalable diagnostics models [5, 14, 16, 17, 18].

The classical DDFD process starts with the signal acquisition which is normally accomplished by vibration sensors due to their high sensitivity to mechanical degradation [1], though other modalities such as acoustic emission [13] and motor current signature analysis [2] are also investigated for specific fault types. A manually defined feature vector is then computed as a transformation of the raw wave-form to one of the statistical moments (RMS, kurtosis), histogram-based distributions, or the ARMA coefficients [1, 3, 9]. These are occasionally filtered using selection algorithms (e.g. J48 decision trees) to store only the most discriminative features and then classify with models such as SVMs, decision trees, weighted neural networks or WiSARD [1, 3, 10]. Although it works well in certain laboratory scenarios, this method demands heuristics developed by experts, and is not a sure thing when it comes to generalizing across machines, fault severity, and operating conditions [11].

When these limitations became more evident, researchers resorted to deep learning. 1D-CNNs were able to learn local, shift-invariant features on raw vibration signals, and had better performance than conventional ANNs even when given only basic preprocessing [5, 14]. The other methods transferred 1D signals to 2D representations and used ImageNet-trained convolutional networks (e.g., ResNet-50, VGG-16, GoogLeNet) on the radar plots, spectrograms, or other such encodings [12, 15, 17]. The same benchmark datasets recorded accuracies of above 97%

with these 2D-CNN-based models on a regular basis. The next architectures were hybrid architectures, which added CNN layers (to extract spatial features) with GRU or LSTM layers (to learn long-range temporal dependencies) [16, 17]. Such hybrid models were particularly applicable to mechanical signals where the sequences over time are very important.

More recent tendencies focus on the improvements of robustness like multimodal fusion, attention, and denoising. Fusion networks are time-frequency domain networks [8], whereas attention modules (e.g., CAM, CBAM) point to fault-relevant parts of the signal [4, 8]. The current denoising methods; such as low-rank matrix recovery [15] and integrated decomposition + GRU pipelines [18] are trying to fill the gap between laboratory-quality data and noisy industrial conditions. Although these are positive, the 2024 systematic review by Matania et al. [7] finds that the field remains roughly tested and not under realistic conditions; particularly, in terms of temporal leakage, overfitting, and reliance on human-crafted features. This research is consistent with the trend of complete automation and rigorous methods. It takes a raw-signal deep learning paradigm and uses a strict, chronological blind-test evaluation; it deals with the fundamental weaknesses that were repeatedly pointed in the literature over validation quality and reproducibility.

2.1 Related Work

Studies on this particular air-compressor data have traditionally been based on hand-crafted feature pipelines. Aravinth et al. (2016) set a baseline background by deriving statistical characteristics and training a J48 classifier, and the accuracy of the result was 98.33% [1]. Bhattacharyya et al. (2025) further developed this direction by adding statistical, histogram, and ARMA characteristics, and a Weightless Neural Network (WiSARD) was employed to achieve an accuracy of 98.66% [3]. These investigations revealed that there is a large amount of discriminative information present in the raw vibration waveform, but it needed a lot of manual processing.

Nambiar et al. (2024) are the authors who combined several handcrafted feature groups and classified them using a Local k-NN model and demonstrated a 100% accuracy [9]. This finding highlights the strength of engineered characteristics as well as indicates the methodological dangers of random splits and non-temporal validation. Simultaneously, the alternatives of deep learning considered the automated feature extraction: Srivatsan et al. (2024) transformed the vibration signals into 2D radar images to be consumed by the pre-trained CNNs such as ResNet-50 and achieved an accuracy of 98.72% [12].

Combined, the literature demonstrates that both conventional and modern pipelines can perform well with random-split assessment, but neither of them can tackle time leakage or the necessity to strictly test blind-sets. Contrastingly, in the current study, performance is only evaluated on a blind test set that is separated in time. This

gives a more realistic evaluation of generalization and determines that models that are perfectly or nearly perfectly accurate on random splits can significantly depreciate on completely unseen data.

2.2 Limitations in Existing Approaches

Although the headline claims in the previous studies were quite accurate [1, 3, 9, 12], there are three critical weaknesses that limit their generalizability. To begin with, these techniques are fully based on handcrafted characteristics. This involves domain understanding, manual tuning, and a lot of preprocessing which might not be easily portable to other types of equipment, fault depths and even operating conditions [11]. Moreover, artifacts of the dataset might be inadvertently encoded by handcrafted features, resulting in pipelines which are fragile and only work in a laboratory.

Second, the current literature heavily relies on the use of simple random splits and accuracy-only reporting without paying attention to the time dependence of vibration data. Random splitting enables models to find future statistical trends during training because consecutive signal samples can be a part of the same operating segment, and this style of splitting inflates accuracy and masks weaknesses in generalization [7]. This is especially serious in the case of 1D vibration signals where two overlapping windows can give a high apparent performance even though untested generalization between different times or between fault transitions is possible.

Third, not many studies use strict and unseen blind-set assessment. In the absence of the isolation of a time-lagged test segment, one cannot tell whether a model is indeed learning discriminative mechanical features or just memorizing dataset-specific variations. This disjuncture has been brought up repeatedly in systematic reviews [7], particularly with regard to the challenge of the zero-fault shot whereby there are few examples of fault in the actual industrial context. The current paper aims to directly cover these limitations by applying raw-signal deep learning and testing only on a non-overlapping, chronologically separated blind test set of 430 windows.

3. Proposed Methodology

The proposed study will present an automated deep sequence-learning model that will remove manual feature engineering and learn the discriminative representations instead, based on the standardized raw vibration signals. This solution is driven by the fact that the conventional pipelines have disadvantages in that they are hand crafted, and that their performance has to be assessed in a random-split fashion, which overstates performance [1, 3, 9]. End-to-end architectures, on the other hand, have been shown to be highly able to learn hierarchical temporal patterns on mechanical signals when the evaluation is done with strict anti-leakage protocols [5, 14, 16, 17, 18].

The suggested methodology will compare four architectures, which are a baseline BiGRU (Model 1), a hybrid Conv1D+GRU (Model 2), a regularized Conv1D+BiGRU (Model 3), and a non-regularized Conv1D+BiGRU (Model 4). As expected, the non-regularized Conv1D +BiGRU (Model 4) exhibited the best generalization on the unseen blind test set, which was strictly speaking, at 0.9326 accuracy, closely followed by the regularized model (0.9186). This observation indicates the consistency of the PyTorch implementation: regularized and non-regularized hybrids converged equally, but the simpler non-regularized one had a slight advantage. The approach thus focuses on comparing empirically the models to establish optimum complexity.

The deep sequence-learning pipeline uses convolutional layers as automated local feature extractors, and a BiGRU layer to learn temporal relationships. The fully contained 150 -sample sequences are windowed over single-class regions to prevent label contamination. This design provides the integrity of time and the sequential character of the dataset. Testing is done only on a blind test set of 430 windows, separated by time, which is realistic because it gives a realistic idea of generalization and does not leakage due to overlapping windows or adjacent sections.

3.1 Existing Model and Challenges

Models that have been proposed previously to use this dataset have mostly been based on handcrafted features and classical classifiers, which have proven to work very well, but necessitate a lot of preprocessing and domain knowledge [3, 9]. The feature-fusion techniques, like that of Nambiar et al. [9], can reach a perfect accuracy by fusing statistical, histogram, and ARMA features and classifying the data using a local kNN model. Equally, WiSARD-based classifiers have achieved almost perfect accuracy with the same feature pipelines [3]. Despite the perceived impressiveness of these findings, they rely on non-temporal random divisions, and might not provide a generalization in the face of realistic chronological analysis.

The main issue that this paper will deal with is the gap between the accuracy of the lab levels as seen in the previous literature and the actual performance that can be achieved when the models are evaluated in the absence of leakage. Randomized splits enable models to use overlapping temporal patterns to train and test, which artificially enhances their performance. Also, handcrafted pipelines are not scalable, and they are dependent on domain knowledge to decide on the features that are most important. It is these challenges that encourage the use of automated feature learning and rigid blind-test assessment in order to create a more honest baseline. The four-deep sequence-learning designs that were evaluated in this study explicitly face the following limitations: they avoid the design of features manually and impose temporal separation between development and blind test windows.

3.2 Proposed Enhancements

This work has a number of methodological improvements compared to the previous research. To start with, the suggested architecture does not require expert-domain feature engineering manually or otherwise, but instead, it uses a Conv1D-BiGRU pipeline that can extract spatial and temporal structure out of standardized windows [5, 14, 16, 17]. Second, to be robust; the study is tested on different architectures; with and without regularization; to make the empirical decision on which components actually contribute to generalization on purely unseen data.

Evidence of the PyTorch notebook demonstrates that the regularized Conv1D-BiGRU model (Model 3) with a blind accuracy of 91.86% was almost out-performed by the non-regularized (Model 4) architecture with a slight increase of accuracy of 93.26%. This confirms the fact that regularization is not harmful in this context but the less complex architecture was enough to learn the required discriminative features.

Third, in contrast to the previous literature that uses random splits, the study uses a leakage-free chronological evaluation pipeline. All the preprocessing elements (LabelEncoder, StandardScaler) are trained on the 4,430-window development set and applied to the blind set. The single-label regions are only extracted in their interior windows to ensure cross-boundary contamination is not obtained. Since overlapping windows nullify cross-validation reliability, the performance of the blind test is the only generalization measure. This makes the reported results a true reflection of temporal independent diagnostic ability.

3.3 Algorithm and Implementation

The algorithm starts by encoding the labels in the form of strings with the help of the LabelEncoder and normalizing the raw values of acceleration with the help of StandardScaler that is trained on the development set only. The standardized signal is divided into 150 samples long windows with a stride of 10 to come up with 4,430 windows in the development set and 430 windows in the blind test set. Extracting windows only in adjacent signal regions with the same fault label is done to prevent contamination of labels, and is enforced by group identifiers calculated based on transitions in the original label sequence.

PyTorch is used to implement all the models. The shape of input windows is (Batch, 1, 150) to permutate it to the shape required by Conv1D. Local structural information is captured by the Conv1D layers (in Models 2 - 4), whereas the GRU or BiGRU layers capture the bidirectional temporal context. Efficient training is done with a batch size of 128. The 5-fold stratified cross-validation was done using early stopping and learning-rate scheduling to find the most suitable training time. The training of final models was done with the number of

epochs being determined based on the average convergence of the CV: Model 1 (62 epochs), Model 2 (39 epochs), Model 3 (37 epochs), and Model 4 (63 epochs).

3.4 Loss Function and Optimization

The categorical cross-entropy loss function and Adam optimizer are used to train all the models, as is common in multi-class sequence classification tasks. Class weights are used to off-set small differences in class-frequency variations in development-set learning. EarlyStopping and ReduceLROnPlateau are some of the example callbacks that were used in exploratory CV to decide the number of epochs. In contrast to the earlier TF-based implementations where CV scores were not reliable, the PyTorch CV scores (Models 2, 3, and 4 with a CV accuracy of more than 98) demonstrated a high correlation with the blind test score. However, blind-test evaluation is the major measure of model performance to achieve zero temporal leakage.

The Conv1D-BiGRU non-regularized model (Model 4) incurred the smallest blind-set loss (0.2585) and highest accuracy (93.26%), and performed better than all other models; however, Models 2 and 3 were highly competitive. These results indicate that in this dataset, regularized and non-regularized hybrid designs show good generalization when tested on chronological blind-tests.

4. Experimental Design and Evaluation

This paper considers the performance of four Deep Learning (DL) architectures in two phases. The models were first trained with a 5-fold Stratified Cross-Validation (CV) system on the 4,430 windows of the development set (90% of the data). In this step, we enabled shuffling to make sure that the models were able to learn the strong features that were acquired throughout the entire timeline instead of remembering a particular order. This aided in establishing the stability of the model and the number of epochs to converge.

A LabelEncoder was used to pre-process the fault classes of the development set and a StandardScaler was used to pre-process the vibration signals. The scaling parameters were only fitted on the development set to avoid data leakage and the blind test was then applied. Strict boundaries of each class were adhered to in window extraction to make sure that every window did not have mixed fault signals. We have reported performance measures other than accuracy such as Precision, Recall, F1-Score, ROC-AUC and PR-AUC to have a comprehensive view of diagnostic ability.

4.1 Datasets and Preprocessing

The data samples are continuous blocks of vibrations observed on a reciprocating air compressor in five different conditions namely CVF (Control Valve Fault), GOOD (Healthy), IOVF (Inlet/Outlet Valve Fault), IVF (Inlet Valve Fault) and OVF (Outlet Valve Fault). The raw data (50,000 readings) was divided in time: 45,000 to be

developed and 5,000 to be in the blind test. To normalize the raw vibration signals, we normalized them to the mean of zero and a standard deviation of one. In the case of windowing, we took a sequence length of 150 samples and a stride of 10. This produced 4, 430 training and 430 test windows.

4.2 Performance Metrics

In this study, more performance measures were considered than accuracy to measure classification performance of the blind test:

- **Accuracy**: The general percentage of the correctly classified windows.
- **Classification Report**: The strengths and weaknesses of discriminating among different types of faults are shown in per-class precision, recall, and F1-scores.
- **Macro-averaged ROC-AUC**: his measures the separability of each class against all the others but equally across classes. A value of 0.5 denotes random performance.
- **Macro-averaged PR-AUC**: Evaluates the precision recall tradeoff of each of the classes separately and averages the scores, which gives information about the behavior of the model in the case of imbalance.
- **Confusion Matrix**: This is an informative tool that shows the misclassifications and the errors that are class specific.

These measures provide a better understanding of the model performance than the accuracy in itself, which is particularly significant in the light of the fact that certain types of faults (e.g., GOOD and CVF) have subtle variations and therefore demand high discriminative sensitivity.

4.3 Experiment Setup

The models in this experiment [section 3] were implemented with PyTorch and NumPy and preprocessing and metrics reporting with scikit-learn. Data batches of 128 were loaded into the models to optimize the use of the GPUs. The loss objective of this classification problem was a categorical cross-entropy loss which was minimized through Adam. In the 5-fold Cross Validation, EarlyStopping and ReduceLROnPlateau were used to train the four models to approximate stable epoch numbers.

According to the PyTorch CV results, we set the last training epochs of the final blind test as follows:

- **Model 1 (Baseline BiGRU)**: 62 epochs
- **Model 2 (Conv1D + GRU)**: 39 epochs
- **Model 3 (Reg. Conv1D–BiGRU)**: 37 epochs
- **Model 4 (Non-Reg. Conv1D–BiGRU)**: 63 epochs (best performing)

These are the precise values of the CV setup and the counts of these epochs are associated with the performance characteristics of each model. Every conclusive judgment mentioned in the below sections is a pure product of the blind test.

4.4 Results Comparative Analysis

Table 1 of the blind test results indicated the four models' accuracy, Loss, ROC-AUC (Macro) and PR-AUC (Macro) of the four models:

Table 1: Blind Test Set Performance (Exact Metrics)

Model	CV Accuracy	Blind Accuracy	F1 Score	ROC-AUC	PR-AUC	Loss
Baseline BiGRU (Model 1)	83.25%	74.42%	0.7361	0.9333	0.8148	0.6524
Conv1D + GRU (Model 2)	99.53%	91.40%	0.9130	0.9846	0.9475	0.3177
Reg. Conv1D–BiGRU (Model 3)	99.82%	91.86%	0.9178	0.9916	0.9688	0.2275
Non-Reg. Conv1D–BiGRU (Model 4)	98.85%	93.26%	0.9325	0.9842	0.9389	0.2585

Model 4 (Non-Regularized Conv1D-BiGRU) was the most effective as it scored 93.26% on the blind test. It was able to learn the hierarchy of features: convolution layers learned the local vibration spikes, whereas bidirectional GRU learned time dependency. Structurally, Model 3 is the same as Model 4, except that it has Batch Normalization and Dropout. It achieved a very high accuracy (91.86%), which showed that regularization in the PyTorch model did not cause convergence problems as predicted in the TensorFlow experiments. Nevertheless, Model 4 had a very small advantage and it may indicate that very little regularization is required to work with this data.

The BiGRU (Model 1) alone did not perform as well as the hybrid models stalling at 74.42% accuracy. This proves that the signal needs to be preprocessed by convolutional feature extraction to be fed in the RNN layers. Model 2 (Unidirectional) was similar to Model 3, which suggests that the popular performance of the convolutional front-end is the key factor, and bidirectionality only contributes to success.

Confusion Matrix Analysis.

Looking at the confusion matrix of the best model (Model 4), we observe that the model detection of IVF, OVF and IOVF is almost perfect with the F1 scores of 0.96, 1.00 and 0.96 respectively. The major source of mistake is the mix up of GOOD (Healthy) and CVF (Control Valve Fault). The model at times confused GOOD with CVF such that the vibration signature of a control valve fault has a high overlap with the healthy state than other faults.

4.5 Ablation Study

In order to know why Model 4 was successful we examine what they have eliminated in the other models. The standalone BiGRU (Model 1) did not perform as well as hybrid models, as without the Conv1D layers to scan the vibrations signals and prepare them to RNN, the accuracy was around 74%, compared to the over 91% of hybrid models. The removal of Bidirectionality (Model 2) resulted in 91.40%, which is competitive yet slightly lower than the BiGRU hybrids, which means that the bidirectional context brings marginal gains.

An important point was identified in the course of the Regularization analysis. However, unlike the previous TF-based results, where regularization obliterated the performance, the PyTorch version demonstrated that regularization (Model 3) achieved great results (91.86%). But when it was eliminated (Model 4), it increased the performance to 93.26%. This confirms that although regularization is not detrimental the unfiltered signal intensity is the most explicit in drawing the decision boundaries that are possible in this particular data set.

5. Extended Contributions

One important conclusion of this research is that the consistency between Cross-Validation (CV) scores and final Blind Test scores is high when the PyTorch framework is used. This theorizes the emphasis of the research on the instability of the framework, rather than the stability of the hybrid architecture.

The most essential finding is that the hybrid models (Models 2, 3 and 4) all had almost perfect CV accuracy (> 98%) and high accuracy (> 91%) on the pure unseen blind test set. The non-regularized hybrid model (Model 4) had an average CV and Blind Test accuracy of 98.85% and 93.26%, respectively. The correlation of the CV with the Test scores is high, which confirms the design of the experiment, which indicates that the high performance does not come as a consequence of overfitting or data leakage.

The large CV scores show that the model was able to learn generalized features that are able to manage the variance in the development set. Applying the model to continuous and sequential blind test set, the model was able to transfer this learning to unknown future data. This consistency serves as confirmation of the applicability of the architecture in raw vibration analysis to the fact that the model has actually learned the physics of the faults and not has learned noise or a batch-specific artifact.

6. Conclusion and Future Work

The paper was able to re-establish a new, reproducible, air compressor fault detection benchmark based on a fully automated, end-to-end Deep Learning framework, and reduced the use of human-generated, handcrafted features by a large margin. The main findings are that deep learning is effective on raw signals and provides the highest accuracy of 93.26% with the Model 4 (Conv1D-BiGRU). This shows that DL has the ability to automatically extract features and do so in a solid framework such as PyTorch.

Additionally, this kind of vibration signal data does not need a lot of signal data to be regularized, although regularization (Model 3) worked well, the non-regularized model (Model 4) was more accurate, indicating that aggressive dropout or batch normalization is not necessarily needed in this sequence length and data volume. Architecturally, the local feature extraction (Conv1D) and temporal context (BiGRU) were needed as the performance dramatically decreased (~19 %) when the convolutional front-end (Model 1) was eliminated. For further improvement of addressing fault classification performance, it is worth investigating deployment of deeper architectures such the ones that have attention mechanisms (Transformers) or Bi-directional LSTM models, pooling vibrations signals from different populations to expand the confidence of the model's generalizability across rotating machinery.

7. References

1. Aravinth, S., Kanna, K. R., & Sugumaran, V. (2016). Air compressor fault diagnosis through vibration signals using statistical features and J48 algorithms. Indian Journal of Science and Technology, 9(47).
2. Ayankoso, S., Dutta, A., He, Y., Gu, F., Ball, A., & Pal, S. K. (2024). Performance of vibration and current signals in the fault diagnosis of induction motors using deep learning and machine learning techniques. Structural Health Monitoring, 1–17.
3. Bhattacharyya, A., Sridharan, N. V., Sivakumar, A., & Vaithiyanathan, S. (2025). Detection and diagnosis of air compressor faults using weightless neural networks. Advances in Mechanical Engineering, 17(5), 1–16.
4. Du, L. (2024). Fault diagnosis method of rotating machinery based on MSResNet feature fusion and CAM. Journal of Vibroengineering, 26(7), 1600–1607.
5. Guo, F., Zhang, Y., Wang, Y., Wang, P., Ren, P., Guo, R., & Wang, X. (2020). Fault detection of reciprocating compressor valve based on one-dimensional convolutional neural network. Mathematical Problems in Engineering, 2020, Article ID 8058723.
6. Lv, Q., Yu, X., Ma, H., Ye, J., Wu, W., & Wang, X. (2021). Applications of machine learning to reciprocating compressor fault diagnosis: A review. Processes, 9(6), 909.
7. Matania, O., Dattner, I., Bortman, J., Kenett, R. S., & Parmet, Y. (2024). A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery: Limitations and challenges. Journal of Sound and Vibration, 590, 118562.
8. Mo, C., Huang, K., Li, W., & Xu, K. (2024). A lightweight and efficient multimodal feature fusion network for bearing fault diagnosis in industrial applications. Sensors, 24(22), 7139.
9. Nambiar, A., Venkatesh, N. S., Aravinth, S., Sugumaran, V., Ramteke, S. M., & Marian, M. (2024). Prediction of air compressor faults with feature fusion and machine learning. Knowledge-Based Systems, 304, 112519.
10. Oliveira, J. C. M., Pontes, K. V., Sartori, I., & Embiruçu, M. (2017). Fault detection and diagnosis in dynamic systems using weightless neural networks. Expert Systems with Applications, 84, 200–219.
11. Sahu, A. R., Palei, S. K., & Mishra, A. (2024). Data-driven fault diagnosis approaches for industrial equipment: A review. Expert Systems, 41(1), e13360.
12. Srivatsan, B., Venkatesh, S. N., Aravinth, S., Sugumaran, V., Dhanraj, J. A., Solomon, J. M., & Vaidhyanathan, R. M. (2024). Fault diagnosis of air compressors using transfer learning: A comparative study of pre-trained networks and hyperparameter optimization. Journal of Low Frequency Noise, Vibration and Active Control, 43(4), 1877–1894.
13. Wang, Y., Xue, C., Jia, X., & Peng, X. (2015). Fault diagnosis of reciprocating compressor valve with the method integrating acoustic emission signal and simulated valve motion. Mechanical Systems and Signal Processing, 56-57, 197–212.

14. Xiao, S., Nie, A., Zhang, Z., Liu, S., Song, M., & Zhang, H. (2020). Fault diagnosis of a reciprocating compressor air valve based on deep learning. Applied Sciences, 10(18), 6596.
15. Yu, T., Long, C., & Feng, G. (2024). Research on fault diagnosis algorithm of air compressor based on low-rank matrix recovery. International Journal of Computer Science, 51(12), 2010–2016.
16. Ahsan, M., & Salah, M. M. (2023). Efficient DCNN-LSTM model for fault diagnosis of raw vibration signals: Applications to variable speed rotating machines and diverse fault depths datasets. Symmetry, 15(7), 1413.
17. Zhang, Y., Zhou, T., Huang, X., Cao, L., & Zhou, Q. (2021). Fault diagnosis of rotating machinery based on recurrent neural networks. Measurement, 171, 108774.
18. Zhou, H., Chen, W., Liu, J., Cheng, L., & Xia, M. (2024). Trustworthy and intelligent fault diagnosis with effective denoising and evidential stacked GRU neural network. Journal of Intelligent Manufacturing, 35, 3523–3542.