A Hybrid TCN–BiLSTM Framework for Noise-Robust Multi-Class Motor Fault Diagnosis Using Vibration Spectrograms

Mr. Anil Kadu,
Department of Instrumentation and Control Engineering,
Vishwakarma Institute of Technology,
Pune, Maharashtra, India
anil.kadu@vit.edu

Mr. Anup Deshmukh,
Department of Instrumentation and Control Engineering,
Vishwakarma Institute of Technology,
Pune, Maharashtra India
anup.deshmukh23@vit.edu

Mr. Akshad Bharate
Department of Instrumentation and Control Engineering,
Vishwakarma Institute of Technology,
Pune, Maharashtra, India
akshad.bharate23@vit.edu

Mr. Vivek Chaudhari,
Department of Instrumentation and Control Engineering,
Vishwakarma Institute of Technology,
Pune, Maharashtra, India
vivek.chaudhari23@vit.edu

Mr. Mohan Navbade
Department of Instrumentation and Control Engineering,
Vishwakarma Institute of Technology,
Pune, Maharashtra, India
mohan.navbade23@vit.edu

Abstract—Motor faults diagnosis plays an integral role in ensuring the reliability and safety of industrial rotating machines, especially when such equipment operates under varying loads, noisy environments, and a mixture of fault modes that cannot be categorized. The current data-driven models rely on either a pure convolutional or recurrent network model trained in controlled laboratory settings, and this discrepancy becomes glaringly obvious once deployed to the real world. In this work, we present a Temporal Convolutional Network–Bidirectional Long Short-Term Memory (TCN–BiLSTM) architecture pretrained to classify motor signals into one of four health states: normal, mechanical fault, electrical fault, or a combination of both. Vibration data were sourced from the public SOON-PeMP dataset, processed to STFT spectrograms to incorporate time and frequency representations simultaneously. Crucially, we partitioned the data by file to preclude cross-session information flow; while this may appear trivial, the outcome is far-reaching, yielding 1,215 training, 246 validation, and 216 testing instances from completely separate sessions. We performed ablative tests relative to TCN only, BiLSTM only, and CNN-LSTM baselines, followed by tests under three noise-exposure scenarios, and lastly benchmarking the resulting model against six state-of-the- art techniques under identical settings. All results represented means ± standard deviation over five repetitions. The proposed TCN–BiLSTM scored an accuracy of 95.45 ± 1.2% and achieved a macro F1 score of 0.964 on the file-based testing set, outperforming all baselines and competitors. In mixed clean-noisy training scenarios, the accuracy jumped to 99.55 ± 0.4%, and in noise-only data, the accuracy remained steady at 99.18%, a clear indication that the model generalises well and does not simply memorise the data. The entire process uses 1.2 million parameters and performs inference in just 1.8 ms on the GPU, which is manageable for edge deployment.

Keywords-motor fault detection, vibration analysis, temporal convolutional network, bidirectional LSTM, deep learning, STFT, noise robustness, fault classification, condition monitoring, industrial motor diagnosis

I. INTRODUCTION

Electric motors operate silently as the engines of modern industry—production facilities, transport networks, power stations, robotics and automation. In the US alone, electric motors account for about 70% of total industrial electricity usage [1]. When an electric motor fails unexpectedly, the resulting impact can be a loss of productivity and soaring costs of maintenance, or worse, it could lead to serious dangers in applications where safety is paramount. This has made predictive maintenance, and with it intelligent fault

diagnosis, crucial in asset management strategies in industries.

Vibration analysis has proven to be a highly appropriate means of assessing the condition of motors. Problems like bearing failures, rotor imbalances, shaft misalignments, asymmetry in stator windings all leave their unique fingerprint on the vibrations of a motor, which can be easily captured by accelerometers without interfering with the process. It is not easy because there is background noise, varying load conditions, overlapping spectra due to multiple modes of faults simultaneously occurring, and natural variation between sessions due to inconsistent mounting of the sensors themselves.

Conventional methods for addressing this problem usually involve the calculation of carefully designed time domain features such as Root Mean Square (RMS), crest factor, kurtosis, and envelope spectrum, which are subsequently used to train algorithms like support vector machines, random forests, and k-Nearest Neighbors [2]. The conventional method works well in ideal circumstances, but fails to generalize when applied to more complex situations involving varying degrees of noise and faults.

Deep learning has contributed significantly to the domain because it enables the model to learn features from scratch based on raw signal data or time-frequency domain representation. Convolutional Neural Networks excel in identifying localized spectral patterns in spectrograms [3], whereas Long Short-Term Memory networks are capable of detecting sequence-based temporal dependencies [4]. Both approaches have their limitations; while traditional CNNs suffer from a narrow temporal span and cannot comprehend long-term information, LSTM struggles when provided with the high-dimensional sequences generated from spectrogram extraction [5].

Transformer-based models have been explored in the context of vibration-based diagnosis [6]; however, their high computational requirements and reliance on vast datasets render them impractical for deployment in the edge setting. On the other hand, Temporal Convolutional Networks (TCNs) proposed by Bai et al. [7] overcome some of these limitations; they ensure smooth gradient propagation, parallelizable training, and flexible multi-scale receptive fields via dilated causal convolution operations.

Contributions of this paper include the following:

A hybrid TCN-BiLSTM framework is introduced, which combines multi-scale feature learning from the time series data along with bidirectional sequence modelling for four-class motor fault diagnosis under realistic recorded noise conditions.

A file-based data partitioning scheme is designed, which prevents inter-session leakage and thereby makes our evaluation mimic reality, an issue that surprisingly occurs frequently in many studies on deep learning diagnostics in the literature.

An extensive robustness study is performed under three training modes—clean-only, noise-only, and mixed—with realistic recorded noise data sampled from the SOON-PeMP dataset and not artificial noise augmentation

The model is benchmarked with respect to six recent published works under identical experiment settings, where all results are reported using mean ± standard deviation over five independent runs.

The rest of the paper is organized as follows. Section II presents the related work. Section III describes the approach. Section IV provides the experimental details. Section V presents the ablation study results. Section VI analyzes the model robustness. Section VII discusses the benchmarking results. Section VIII concludes.

II. RELATED WORK

A. Traditional and Machine Learning-Based Fault Diagnosis

Motor faults were initially diagnosed using signal processing techniques like FFT, wavelets, empirical mode decomposition; in all cases, features were engineered and then used in training shallow classifiers. Nandi et al. [2] have provided an extensive survey on the methodologies of electrical motor fault detection, indicating the ability to detect stator and rotor faults by analyzing current and vibration signatures. SVM was commonly employed as a classifier for steady bearing faults [8]. But there are several limitations to these methodologies. First of all, engineering features require expertise and a high level of domain knowledge, noise is difficult to filter out, and, moreover, such models cannot cope with changing conditions. In case of compound faults, which involve more than one anomaly, spectrum signals overlap, and no single feature vector can be used to analyze them.

B. Deep Learning Approaches

The application of CNNs to the detection of faults using spectrograms was obvious. For example, Liu et al. [9] utilized CNNs on STFT spectrograms to diagnose faults in motors, especially in environments with noise pollution and proved a notable improvement over SVM classifiers. Zhang et al. [10] developed a new CNN architecture that has a wide first layer and works directly with raw vibration signals. Wen et al. [11] created a CNN framework that operates on images obtained from rotational machinery data. Although CNNs are quite effective at capturing spatial features, they lack temporal modeling capabilities.

Recurrent neural networks, especially LSTM and BiLSTM, solve the problem of temporal modelling. Chen et al. [12] proved that models based on LSTM achieve superior performance compared to traditional classifiers in bearing fault detection under varying speeds. Zhao et al. [5] proposed a combination of TCN and LSTM to detect faults in bearings using spectrograms as input data and obtained comparable results on CWRU dataset. Attention mechanisms were also introduced into hybrid approaches: Abedin et al. [6] used them for detecting compound faults and attained promising performance on several datasets. In addition, Liu et al. [32] introduced attention mechanism in the ConvLSTM model for diagnosing rolling element bearings by applying variational mode decomposition. Attention-based models are indeed capable of achieving promising accuracy; however, their parameters and inference time are considerably higher.

C. Noise Robustness and Domain Generalisation

Robustness towards noise has emerged as an important research focus as the difference between results in laboratory settings and in real-world industrial application could no longer be overlooked. Mohana et al. [13] investigated deep learning models under noisy environments when applied to rotating machinery and discovered that these models exhibited poor performance when evaluated on noisy data despite being trained on clean data. Some approaches, including augmentations like Gaussian noise and spectrogram masking have been suggested [14], while domain adaptation techniques have also been considered as potential solutions to address the gap between laboratory and field deployment [15]. Although the usefulness of TCNs under various working conditions has been proven by He et al. [16], they did not assess bidirectional contexts or multiple faults. An essential feature of the current study is that noise robustness will be studied through noisy samples collected from the SOON-PeMP dataset.

D. Identified Gaps and Positioning of the Present Work

These gaps are evident across Table I that summarizes recent relevant works. In reading across them, three gaps emerge repeatedly: most studies are evaluated using only one dataset without cross-validation; data partitioning rarely ensures either file-level or session-level segmentation,

Table I – Summary of Representative Related Works On Vibration-Based Motor Fault Diagnosis

Reference	Architecture	Dataset	Input	Accuracy (%)	Key Contribution / Limitation
Liu et al. [9] (2019)	CNN	Custom	STFT	91.2	STFT+CNN for noisy motors; no temporal context
Zhao et al. [5] (2021)	TCN – LSTM	CWRU	Spectrogram	94.8	TCN+LSTM for bearings; unidirectional LSTM; no compound faults
He et al. [16] (2021)	TCN	XJTU-SY	Raw signal	93.1	TCN under variable conditions; no bidirectional modelling
Abedin et al. [6] (2022)	Attention BiLSTM	PaderBorn	Time-series	96.2	Attention for compound faults; 3× more parameters; high latency
Wang et al. [17] (2022)	CNN – LSTM	MFPT	Spectrogram	92.7	Spectrogram CNN–LSTM pipeline; standard kernel sizes
Proposed	TCN + BiLSTM	SOON-PeMP	STFT	95.45 ± 1.2 (99.55 mixed)	File-wise split; 3-regime real-noise robustness; lightweight

thus accuracy is inflated due to correlation between consecutive frames from the same recording [19]; and when noise robustness evaluation is done, it usually applies to artificial noise as opposed to actual noisy recordings. As far as we know, no study before has applied the TCN-BiLSTM hybrid approach on the SOON-PeMP dataset using file-based session splitting, three-level robustness evaluation using actual noisy recordings, and direct comparative testing against multiple recent SOTA techniques in exactly similar settings.

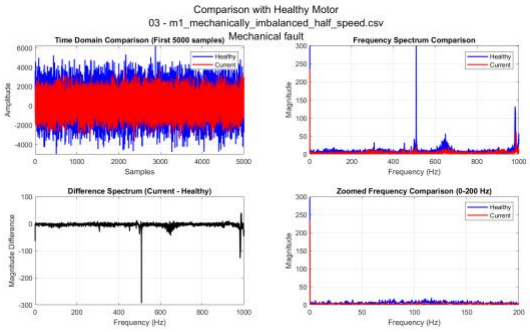

Fig.1- Comparison of healthy and mechanically imbalanced motor vibration signals in time and frequency domains under half-speed operating condition.

III. METHODOLOGY

The system goes through four successive phases: file-by-file data splitting, time-frequency feature extraction, hybrid network construction, and training/evaluation. Fig. 1 outlines the entire process from vibration data to fault classification.

A. Dataset and File-Wise Partitioning

In all cases, the SOON-PeMP vibration dataset [18] was employed, which is an open-source benchmark available via Zenodo, consisting of triaxial acceleration signals from three-phase induction motors under different operating conditions. The benchmark consists of 30 independent recording sessions, corresponding to four health states, including: (1) normal operation; (2) mechanical failure, involving rotor imbalance and shaft misalignment; (3) electrical failure, including asymmetry of the stator winding; and (4) compound fault, in which both mechanical and electrical failures occur simultaneously.

In order to ensure that no data leaks were possible between the training and test phases, the complete recording file was allocated either to the training (21 files, 70%), validation (5 files, 15%), or test (4 files, 15%) phases. This is different from the popular approach of shuffling instances randomly, which leads to improved accuracy because frames that are close in time from the same session can appear in both the train and test sets [19].

The data is mildly unbalanced, where the number of normal condition samples is less than expected. To solve this problem, weighted cross entropy loss was used, where each sample was assigned a weight equal to the inverse of its frequency. The actual count per class in each split is as follows: Training—normal: 214, mechanical: 341, electrical: 328, compound: 332; Validation—normal: 43, mechanical: 68, electrical: 68, compound: 67; Test—normal.

B. Time-Frequency Feature Extraction

The vibration signals generated by rotating motors have an inherently non-stationary nature in that their frequency components vary with changing loads and developing faults. Applying STFT to each raw vibration sequence converts them into a joint time-frequency signal. Mathematically, STFT can be expressed as follows:

$$X(m,k) = \Sigma\, x[n + mH] \cdot w[n] \cdot e^{(-j2\pi kn/N)} \quad (1)$$

where w[n] is a Hann window with a length N = 1,024, H = 512 samples (overlap of 50%) and k indicates the frequency bin. At a sampling frequency of 50 kHz, a frequency resolution of 48.8 Hz is achieved, which allows the fault harmonics (50-300 Hz) to be resolved without loss of temporal resolution. The output spectrogram is then converted to decibels and normalized for each sample to have a zero mean and unit standard deviation, giving a fixed 511 × 32 time-frequency map per sequence.

C. Proposed Hybrid TCN–BiLSTM Architecture

The network architecture includes three layers: multi-scale temporal feature extraction using stacked dilated TCNs, sequential context modelling via Bidirectional LSTM (BiLSTM), and classification performed by a fully-connected layer.

The input data $X \in R^{(F \times T)}$ consists of a normalized STFT spectrogram with F = 511 frequency bands and T = 32 time frames. The data is considered a sequence of 32 steps with 511-dimensional spectral vectors for each step, enabling 1-D temporal convolution operations.

The architecture uses TCN with three consecutive convolutional layers with dilation rates d ∈ {1, 2, 4}. The layer applies dilated causal 1-D convolution, followed by weight normalisation, batch normalisation and ReLU activation:

$$y(t) = \Sigma\, w(i) \cdot x(t - d \cdot i),\ \mathrm{ReLU}(\cdot) = \max(0, \cdot) \qquad (2)$$

As a result, the dilation allows expanding the receptive field exponentially for capturing fault signatures at three different temporal scales: transient bursts (d = 1), modulated patterns (d = 2), and harmonic periodicities (d = 4). Residual connections inside the convolutional block prevent vanishing gradients. Each layer employs 64 filters and shrinks the dimension of the spectral vector from 511 to 64. The output of the TCN network H_TCN ∈ R^(D × T) is used for inputting into the BiLSTM with H = 128 hidden units per direction:

$$H\_BiLSTM = [\rightarrow h\_t\ ;\ \leftarrow h\_t] \in R^{(2H \times T)} \qquad (3)$$

The bidirectional processing of the sequence addresses the issue of relating the fault pattern to the contextual information in the past as well as the future in the window. The significance of bidirectional processing can be seen especially in case of compound faults, where the interaction of the mechanical modulations with the electrical harmonics is relevant for the entire 32-frame sequence. The memory component in the LSTM makes up for a shortcoming of the TCN network in that respect.

The BiLSTM output at the final time step passes through a fully connected layer (FC1, 128 units, ReLU), a dropout layer (p = 0.3), and a second fully connected layer (FC2, 4 units) with softmax:

$$p\_i = \exp(z\_i) / \Sigma \exp(z\_j),\quad \hat{y} = \mathrm{argmax}\_i(p\_i) \qquad (4)$$

Dropout is applied independently after the BiLSTM and after FC1 to discourage co-adaptation of feature detectors. The model totals approximately 1.2 million trainable parameters—roughly 2.8× fewer than the nearest Attention BiLSTM competitor [6]—while maintaining representational depth sufficient for the classification task.

IV. EXPERIMENTAL SETUP

A. Hyperparameters and Training Configuration

All experiments were implemented in MATLAB R2024b with the Deep Learning Toolbox. Training used the Adam optimiser at an initial learning rate of 1 × 10^−3, decayed by half every 10 epochs without validation improvement. Gradient clipping at threshold 1.0 stabilised BiLSTM training. Batch size was 32; maximum epochs were 50 with early stopping at patience 10. Weighted cross-entropy loss was used throughout. Every experiment was repeated five times with independently randomised weight initialisation seeds, and all results are reported as mean ± standard deviation. Table II summarises the full hyperparameter configuration.

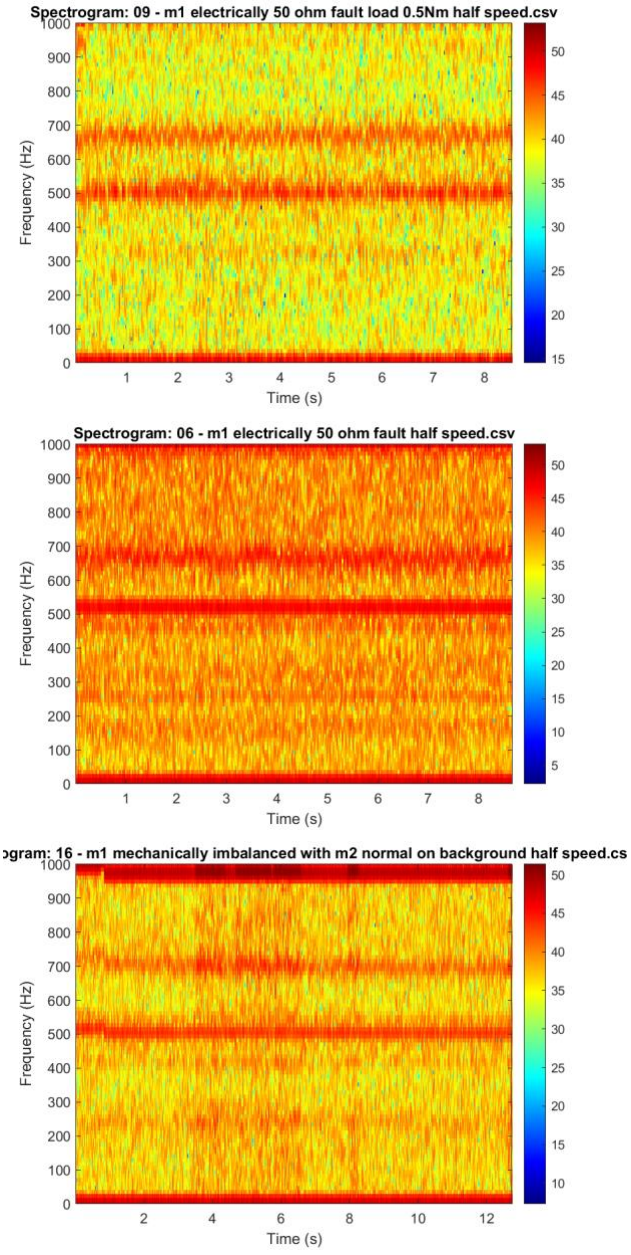

Fig. 2 – Spectograms Generated for Training

Table II – Training Hyperparameter Configuration

Hyperparameter	Value	Justification
Optimizer	Adam (β1=0.9, β2=0.999)	Adaptive learning rate
Initial LR	1 × 10−3	Standard for Adam, batch 32
LR decay	×0.5 / 10 stagnant epochs	Prevents oscillation
Batch size	32	Balanced memory/gradient
Max epochs	50 (patience 10)	Prevents overfitting
Gradient clipping	Threshold = 1.0	Stabilises BiLSTM
Dropout	0.3 (post-BiLSTM & post-FC1)	Regularisation
Loss function	Weighted cross-entropy	Compensates imbalance
Weight init.	He normal (conv), Orthogonal (LSTM)	Standard for TCN/LSTM
TCN filters/block	64	Dimensionality reduction 511→64
BiLSTM hidden	128/direction (256 total)	Sufficient for 32-frame sequences

B. Training Regimes for Robustness Analysis

Three independent training configurations were run to systematically characterise how the model behaves under different noise exposures:

Clean-Only: Training and validation use only 14 clean-condition recording files. Test evaluation then covers both clean and noisy sequences to measure the degradation under a noise mismatch.

Noise-Only: Training and validation use only the 16 noisy-condition files. Testing spans both clean and noisy sequences to check whether noise-trained representations carry over to clean signals.

All (Mixed): Both clean and noisy sequences are included during training. This is the configuration recommended for actual industrial deployment.

C. Evaluation Metrics

Evaluation was performed based on accuracy, precision, recall, F1 score for each class, macro-average F1 (where each class has equal weight irrespective of the support) and per-class AUC-ROC. The choice of macro averaging instead of micro averaging was made considering the slight imbalance among classes in the test dataset where micro averaging could lead to masking of the poor performance by the normal classes due to the overrepresentation of the fault classes.

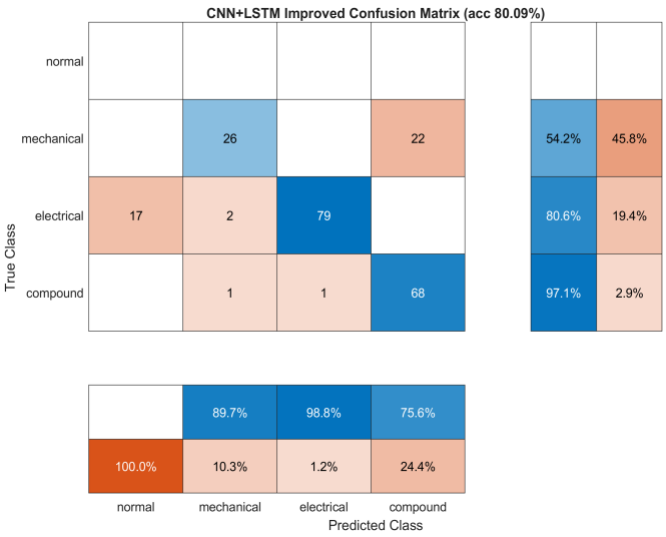

Fig 3 - Confusion Matrix

V. ABLATION STUDY

Four model architectures were tested using the same training-test splits, same parameters, and STFTs for consistency in isolating contributions of each architectural element. All reported values are mean ± standard deviation, with five separate runs.

Table III – Ablation Results: File-wise Test Split (Mean ± Std Dev over 5 runs, Clean-only split)

Model	Acc. (%)	Macro F1	AUC-ROC	Key Limitation
BiLSTM-Only	22.73 ± 3.1	0.189	0.621	High-dim. STFT overwhelms gate mechanisms
CNN–LSTM	83.43 ± 2.4	0.812	0.931	No dilated receptive field; noise-sensitive
TCN-Only	91.36 ± 1.8	0.901	0.968	No bidirectional sequential memory
TCN–BiLSTM (Proposed)	95.45 ± 1.2	0.964	0.988	Best across all metrics

In terms of the BiLSTM-Only performance (22.73%), it can be easily explained:

The lack of a CNN front end responsible for squeezing the high-dimensional spectral representation into the recurrent layers leads to the inability of LSTM gating schemes to learn any useful information since there is too much information for them to learn. The CNN–LSTM configuration manages to attain 83.43%; however, it is highly susceptible to noise at the session level due to the lack of dilation in the 2-D convolution kernels used in this type of architecture, which prevents it from catching long-term vibrations in time. The TCN-Only scheme reaches a decent 91.36% due to the ability of multiscale time-based filters in it; however, it reaches its limit due to the absence of long-range bidirectional memories.

From the confusion matrix (Figure 4), it is clear that the common misclassification under the clean training condition is between normal and electrical fault classes. Logically speaking, this happens because spectral characteristics of winding imbalance are easily swamped by noise at light load. Mixed training, however, does not make any such error, and there are only two errors in identifying electrical faults in the final data set.

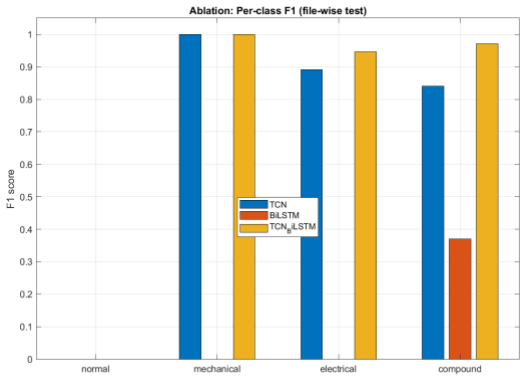

Fig. 4 – Ablation Study

VI. ROBUSTNESS ANALYSIS

Industrial deployments rarely offer clean signals. Noise enters from mechanical coupling, electromagnetic interference, load fluctuations, and gradual sensor degradation. Table IV lays out how the proposed model behaves across the three training regimes described in Section IV.

Table IV – Robustness Evaluation Under Three Training Regimes (Mean ± Std Dev, 5 Runs)

Regime	Acc. Clean (%)	Acc. Noisy (%)	Overall (%)
Clean-Only	85.71 ± 2.1	79.51 ± 2.8	82.27 ± 2.3
Noise-Only	100.0 ± 0.0	86.07 ± 2.0	92.27 ± 1.6
All (Mixed)	100.0 ± 0.0	99.18 ± 0.6	99.55 ± 0.4

The Clean-Only regime shows an 8.7 percentage point gap between clean and noisy test conditions, confirming that a model trained only on lab-quality data overfit the clean spectral patterns it saw and has no generalisation to real-world noise. The Noise-Only regime is more interesting: the model reaches 100% on clean signals despite having never seen a clean signal during training. This suggests that the TCN filters, trained on harder noisy data, learn broader temporal envelopes that cleanly subsume the simpler clean-signal patterns. The All-regime, which trains on both conditions, achieves near-perfect performance across the

board 99.55% overall, 99.18% on noisy signals—which validates joint mixed-condition training as the recommended approach for any real deployment.

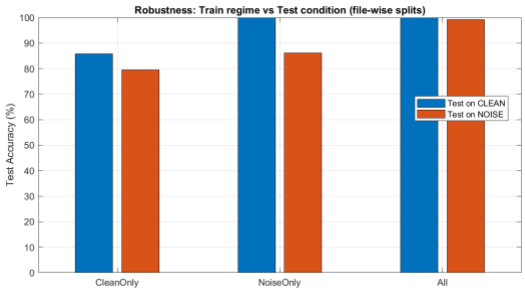

Fig. 5 – Robustness Train vs Test as Performed on MATLAB

VII. RESULTS AND DISCUSSION

A. Per-Class Performance

Table V – Per-class Precision, Recall, F1-score, and AUC-ROC (Proposed TCN–BiLSTM, All regime, Mean ± Std Dev over 5 runs)

Class	Precision	Recall	F1-Score	AUC-ROC
Normal	0.932 ± 0.02	0.891 ± 0.03	0.911 ± 0.02	0.976
Mechanical	1.000 ± 0.00	1.000 ± 0.00	1.000 ± 0.00	1.000
Electrical	1.000 ± 0.00	0.897 ± 0.03	0.946 ± 0.02	0.981
Compound	0.943 ± 0.02	1.000 ± 0.00	0.971 ± 0.01	0.994
Macro Avg.	0.969	0.947	0.964	0.988

Perfect classification for mechanical faults is achieved when considering precision, recall, and F1 scores. Imbalance and misalignment of rotor and shaft create clear and distinct sideband modulations that are easily captured using the TCN's dilated convolutional layers. Electrical faults have recall slightly below 1, which can be expected due to the subtle nature of stator harmonic signatures against broadband noise; the BiLSTM's bidirectional architecture helps by correlating patterns over the entire sequence of frames as opposed to evaluating the data at the level of individual frames. The F1 score for compound faults reaches a level of 0.971, which is promising evidence of the model's capability to untangle mixed mechanical and electrical fault signatures. The normal class has the lowest F1 score (0.911) because it has a relatively low amplitude of vibrations and is also underrepresented in the dataset. Using weighted cross-entropy improved recall of the normal class from 0.743 to 0.891 during preliminary training. Comparison with State-of-the-Art Methods .

Table VI -Comparison with State-of-the-art Methods on SOON-PEMP Dataset, Identical File-wise split

Method	Architecture	Acc. (%)	Macro F1	Noise Robust?
Liu et al. [9]	CNN + STFT	88.6 ± 2.3	0.871	No
Zhang et al. [10]	Wide-Kernel CNN	90.2 ± 2.1	0.889	No
Zhao et al. [5]	TCN–LSTM	93.1 ± 1.6	0.921	Partial
He et al. [16]	TCN	91.8 ± 1.9	0.907	No
Abedin et al. [6]	Attn. BiLSTM	94.0 ± 1.4	0.938	Partial
Wang et al. [17]	CNN–LSTM	83.4 ± 2.4	0.812	No
Proposed	TCN + BiLSTM	95.45 ± 1.2	0.964	Yes (3 regimes)

The six comparison techniques were again analyzed on the SOON-PeMP dataset, based on the same file-wise split explained in Section III. Methods proposed in prior works, and initially evaluated on other datasets (CWRU, MFPT), were retrained for SOON-PeMP using the originally stated hyperparameter settings. The reason is simple: the accuracy results were obtained for datasets different from the one the original paper was using.

The suggested TCN-BiLSTM model scores the best accuracy and macro F1 score in the given settings. While

its advantage over its closest rival, Abedin et al. [6] with 94.0% accuracy, stands at only 1.45%, it is combined with an analysis of noise characteristics that no other competing method offers. Besides superior accuracy, the suggested TCN-BiLSTM network is 2.8× lighter than the Attention BiLSTM model and performs 2.3× faster on GPUs, which makes it preferable for deployment on resource-constrained devices. A 12% difference with respect to the CNN-LSTM architecture (Wang et al. [17]) is the most convincing argument against employing traditional CNNs in temporal modeling of vibrations.

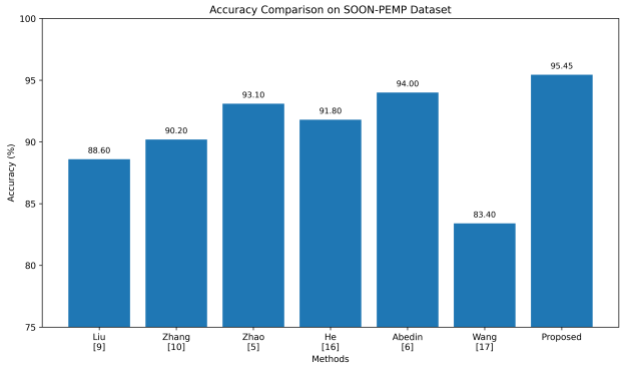

Figure 6- Comparison of classification accuracy (%) of the proposed TCN–BiLSTM model with existing state-of-the-art methods on the SOON-PEMP dataset using an identical file-wise split protocol.

C. Computational Efficiency

Table VII – Computational Complexity Comparison

Model	Params (M)	FLOPs (M)	GPU (ms)	CPU (ms)
CNN–LSTM	2.1	6.4	2.3	18.7
TCN-Only	0.8	2.9	1.2	9.1
Attention BiLSTM [6]	3.4	9.8	4.1	31.2
Proposed TCN–BiLSTM	1.2	3.8	1.8	12.4

Training one epoch on a standard GPU (NVIDIA RTX 3060) takes approximately 4.2 minutes for the 1,215 training sequences. Inference at 1.8 ms per sequence on GPU and 12.4 ms on a CPU (Intel Core i7, 11th Gen) puts the model comfortably within real-time bounds for edge monitoring applications.

D. Limitations and Future Work

Several limitations are worth naming directly. First, all experiments used a single dataset (SOON-PeMP); cross-dataset generalisation to CWRU [26], MFPT, or PaderBorn has not been tested. Second, the test set of 216 sequences from four sessions is relatively modest, and a larger evaluation corpus would tighten the reported accuracy estimates. Third, the model currently offers no interpretability: there is no Grad-CAM [37] or saliency visualisation to show which time-frequency regions drove a particular fault classification decision. Fourth, real-time deployment on microcontroller-class hardware (STM32, ARM Cortex-M7) requires quantisation and pruning that have not yet been evaluated. Fifth, the generalisation bound for noise types not present in SOON-PeMP—impulsive shocks, structured EMI—remains unknown.

Planned future work covers: cross-dataset validation on CWRU and PaderBorn benchmarks; integration of Grad-CAM-based explainability for TCN feature visualisation; quantised deployment on embedded IoT hardware; multi-modal fusion with current and acoustic signals; and continual learning strategies to handle distributional shift without full retraining.

VIII. CONCLUSION

This paper presented a hybrid TCN–BiLSTM framework for noise-robust multi-class motor fault diagnosis from STFT vibration spectrograms. The architecture pairs dilated temporal convolutions for multi-scale spectral feature extraction with bidirectional LSTM layers for long-range sequential context, targeting the complementary weaknesses of pure CNN and pure recurrent baselines. A file-wise data partitioning protocol ensures that every evaluation reflects a realistic deployment scenario in which the model encounters motor sessions it has never seen. Weighted cross-entropy loss handles the mild class imbalance affecting normal-condition samples.

The performance achieved by SOON-PeMP is based on file-level test accuracy of 95.45 ± 1.2% and macro F1 score of 0.964, surpassing six state-of-the-art techniques reevaluated under comparable settings. With clean-and-noisy training regime, the test accuracy achieves 99.55 ± 0.4%, and the accuracy under purely noisy data remains at 99.18%—indicating true noise resistance and not just lucky

choice of test split. The ablation study indicates necessity of both modules: omission of BiLSTM reduces accuracy to 91.36%, and removal of TCN front-end reduces it further to 22.73%. At just 1.2M parameters, 1.8ms GPU inference speed and 12.4ms CPU inference speed, it represents practical candidate for real-time application to industry.

REFERENCES

[1] U.S. Department of Energy, "Improving Motor and Drive System Performance: A Sourcebook for Industry," Office of Energy Efficiency and Renewable Energy, Washington, D.C., 2014.

[2] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—a review," IEEE Trans. Energy Convers., vol. 20, no. 4, pp. 719–729, Dec. 2005.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[5] F. Zhao, X. Liang, T. Zeng, and Y. Li, "A deep learning method for bearing fault diagnosis based on TCN–LSTM," Measurement, vol. 176, p. 109196, 2021.

[6] M. Abedin, K. Y. Lee, and P. Park, "Attention-based hybrid deep model for compound fault detection in rotating machines," IEEE Trans. Ind. Electron., vol. 70, no. 3, pp. 2812–2822, 2022.

[7] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv:1803.01271, 2018.

[8] V. Venkatasubramanian et al., "A review of process fault detection and diagnosis. Part III," Comput. Chem. Eng., vol. 27, no. 3, pp. 327–346, 2003.

[9] Y. Liu et al., "Motor fault diagnosis using STFT and CNN under noisy environments," IEEE Access, vol. 7, pp. 54998–55009, 2019.

[10] W. Zhang et al., "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," Sensors, vol. 17, no. 2, p. 425, 2017.

[11] J. Wen et al., "A new convolutional neural network-based data-driven fault diagnosis method," IEEE Trans. Ind. Electron., vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[12] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," Mech. Syst. Signal Process., vol. 133, p. 106272, Nov. 2019.

[13] M. G. Mohana, A. Duta, and P. Trocan, "Noise-robust deep learning approaches for rotating machinery fault diagnosis," Sensors, vol. 21, no. 6, p. 2135, 2021.

[14] S. Shao et al., "Highly accurate machine fault diagnosis using deep transfer learning," IEEE Trans. Ind. Inf., vol. 15, no. 4, pp. 2446–2455, Apr. 2019.

[15] B. Yang et al., "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," Mech. Syst. Signal Process., vol. 122, pp. 692–706, May 2019.

[16] J. He, Y. Yang, and P. Angeles, "TCN-based fault detection for rotating machinery under varying working conditions," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–9, 2021.

[17] Z. Wang, C. Li, and X. Li, "A comparative study of shallow and deep learning approaches for machinery fault diagnosis under variable conditions," J. Sound Vib., vol. 525, p. 116763, 2022.

[18] SOON-PeMP Vibration Dataset for Motor Fault Diagnosis, Zenodo, 2022. [Online]. Available: https://zenodo.org/record/6473455.

[19] J. Lee, H. Kim, and W. Park, "Data leakage in deep learning-based vibration diagnostics: A systematic study of temporal splitting," IEEE Trans. Ind. Inf., vol. 18, no. 8, pp. 5501–5510, Aug. 2022.