COMPUTERS AND INFORMATICS
2757-8259 2026, Volume 6 Issue 2
DOI: 10.62189/ci.1974430 Research Article

A unified ensemble learning framework for electric motor fault diagnosis, thermal state prediction, and remaining useful life estimation

Yunus Kökver
Ankara University, Information Technologies Vocational School, Department of Computer Technologies, Ankara, Türkiye, ykokver@ankara.edu.tr
Neşet Koçkar
Ankara University, Faculty of Science, Department of Physics, Ankara, Türkiye, 22050082@ogrenci.ankara.edu.tr
Ahmet Yasin Selçuk
Ankara University, Faculty of Engineering, Department of Energy Systems Engineering, Ankara, Türkiye, 22290413@ogrenci.ankara.edu.tr
Fatih Ekinci
Ankara University, Institute of Artificial Intelligence, Ankara, Türkiye, fatihekinci@ankara.edu.tr
Abdullah Orman*
Ankara Yıldırım Beyazıt University, Vocational School of Technical Sciences, Department of Computer Technologies, Ankara, Türkiye, aorman@aybu.edu.tr
Mehmet Serdar Güzel
Ankara University, Faculty of Engineering, Department of Computer Engineering, Ankara, Türkiye, mguzel@ankara.edu.tr

Submitted: 27.06.2026
Accepted: 18.10.2026
Published: 31.12.2026

* *Corresponding Author*

Abstract:

Electric motors are critical components of modern industrial and transportation systems, making accurate fault diagnosis and predictive maintenance essential for operational reliability. This study proposes a unified machine learning framework for motor fault classification, temperature prediction, and remaining useful life (RUL) estimation using three publicly available electric motor datasets. The proposed pipeline integrates sliding-window segmentation, multi-domain feature extraction, consensus-based feature selection, and physically motivated descriptors derived from electrical and thermal measurements. Five machine learning algorithms, namely Random Forest, Support Vector Machine (SVM), XGBoost, LightGBM, and Long Short-Term Memory (LSTM), were systematically evaluated. For fault diagnosis on the Ottawa UOEMD-VAFCVS dataset, evaluated with a file-level train–test split in which all segments of a recording are kept in the same partition, LightGBM achieved the highest performance on the held-out recordings (Accuracy = 1.0000, F1-score = 1.0000, ROC-AUC = 1.0000). Sliding-window segmentation increased the number of trainable observations from 128 to 2,432 and improved the macro F1-score from 0.569 to above 0.994 across all models. For thermal prediction on the PMSM dataset, LightGBM obtained R^2 values of 0.993, 0.991, 0.989, and 0.987 for rotor, stator winding, stator tooth, and stator yoke temperatures, respectively, with mean absolute errors below 1.4°C. In an exploratory experiment on the synthetically generated EVIoT dataset, which contains no genuine run-to-failure trajectories, XGBoost and LightGBM reached $R^2 = 0.933$ on a 30-step smoothed RUL target; this value largely reflects the smoothing of the target and is reported only as a demonstration of the pipeline, not as validated prognostic capability. A Friedman test over N repeated runs indicated significant differences among the five classifiers ($\chi^2 = 11.556$, $p = 0.021$, Kendall's W = W). Within the evaluated configurations, feature-engineered gradient-boosting models matched or exceeded the LSTM baseline at low computational cost; a systematic comparison with end-to-end deep architecture is left for future work.

Keywords: *: Electric motor fault diagnosis, LightGBM, Machine learning, Predictive maintenance, Remaining useful life prediction.*

Published under the CC BY-NC 4.0 license by the peer-reviewed open access journal Computers and Informatics (C&I) at DergiPark (dergipark.org.tr/ci)

Cite this paper as: Kökver, Y., Koçkar, N., Selçuk, A.Y., Ekinci, F., Orman, A., Güzel, M.S., A unified ensemble learning framework for electric motor fault diagnosis, thermal state prediction, and remaining useful life estimation, *Computers and Informatics*, 2026; 6(2); 36-57, https://doi.org/10.62189/ci.1974430

1. INTRODUCTION

Electric motors are among the most critical components of modern industrial systems, electric vehicles, manufacturing facilities, and energy infrastructures [1,2]. Their operational reliability directly affects productivity, energy efficiency, maintenance costs, and system safety [3,4]. Mechanical and electrical faults such as bearing defects, broken rotor bars, stator winding failures, shaft imbalance, and thermal abnormalities can gradually degrade motor performance before leading to unexpected breakdowns [5-8]. Early identification of these faults is therefore essential for reducing downtime and extending equipment lifetime [9,10]. Recent advances in sensing technologies have enabled continuous acquisition of vibration, acoustic, electrical, and thermal measurements, creating large volumes of operational data suitable for intelligent condition monitoring [11-13].

Artificial intelligence and machine learning techniques have become increasingly important for extracting actionable information from industrial sensor data [14,15]. Numerous studies have investigated the use of Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, hybrid CNN–LSTM architectures, transfer learning approaches, and transformer-based models for fault diagnosis and predictive maintenance applications [16-18]. High classification accuracies have been reported on several benchmark motor datasets [22]. The growing availability of industrial data has further accelerated the development of data-driven maintenance strategies capable of identifying abnormal operating conditions before severe failures occur [25,26].

Feature engineering continues to play a central role in industrial predictive analytics [27,28]. Statistical descriptors derived from time-domain signals provide information about signal energy, amplitude distribution, and impulsive behavior [29,30]. Frequency-domain representations obtained with the fast Fourier transform [91] reveal spectral characteristics associated with specific fault mechanisms [31,32]. Power spectral density analysis, commonly estimated with Welch's averaged periodogram method [60], enables the investigation of energy distribution across different frequency bands [33,34]. In addition, physically motivated variables such as copper losses, mechanical power, current magnitude, and thermal gradients can incorporate domain knowledge directly into the feature space [51,52]. Such engineered quantities should be distinguished from physics-informed neural networks (PINNs), in which governing equations are embedded as constraints in the loss function of the learning model [35,36,85]; no such constraint is imposed in the present work. The integration of these complementary descriptors with ensemble learning algorithms has recently emerged as an effective strategy for improving predictive accuracy while maintaining computational efficiency and model interpretability [37,38].

Most existing studies focus on a single predictive maintenance task, typically fault classification, temperature estimation, or remaining useful life prediction [17,19,83]. Comparative investigations evaluating multiple machine learning paradigms under a unified framework remain relatively limited. The present study addresses this gap by proposing an integrated predictive maintenance pipeline combining sliding-window segmentation [42,95], multi-domain feature extraction [29-34], consensus-based feature selection [92,93], and physically motivated descriptors [51,52]. Five machine learning algorithms, namely Random Forest [57], Support Vector Machine [58], XGBoost [20], LightGBM [21], and LSTM [59], were systematically evaluated on three publicly available electric motor datasets representing fault diagnosis, thermal behavior prediction, and remaining useful life estimation problems. The study aims to identify robust and computationally efficient learning strategies for industrial motor monitoring while providing a comprehensive comparison across multiple predictive maintenance scenarios.

Table 1 contrasts representative studies with the present work. Existing approaches typically address a single task on a single dataset, and several report near-perfect accuracy without specifying how segments originating from the same recording are allocated to training and test partitions, which can

lead to optimistic estimates [94,97]. The present study differs in three respects: (i) the same pipeline structure, feature toolkit, and selection procedure are applied to classification and regression tasks on three datasets; (ii) the fault diagnosis experiments use a file-level split that prevents overlapping segments of a recording from appearing in both partitions; and (iii) the limitations of the data, notably the synthetic nature of the RUL dataset, are stated explicitly and the corresponding results are treated as exploratory.

Table 1. Representative studies and their scope compared with the present work

Study	Dataset / task	Approach	Scope and limitation relative to this study
Wahid et al. [17]	Microsoft PdM telemetry / machine failure prediction	Hybrid CNN–LSTM (compared with CNN and LSTM)	Single task; deep models only
Kirchgässner et al. [19]	PMSM temperature estimation	Residual CNN (TCN) and RNN	Thermal task only
Hossain et al. [83]	PMSM temperature forecasting	Physics-informed hybrid ensemble	Thermal task only
Sehri and Dumond [86]	UOEMD-VAFCVS data article	Dataset description	No benchmark models or evaluation protocol
This study	Fault diagnosis, thermal regression, exploratory RUL (three datasets)	Shared feature pipeline with RF, SVM, XGBoost, LightGBM, LSTM	File-level split; statistical testing; RUL explicitly treated as exploratory

2. METHODOLOGY

This study proposes a unified machine learning framework for electric motor fault diagnosis, thermal state prediction, and remaining useful life estimation. The framework integrates data preprocessing, sliding-window segmentation, multi-domain feature extraction, physically motivated feature engineering, consensus-based feature selection, and supervised learning algorithms within a single experimental pipeline. To evaluate the robustness and generalizability of the proposed approach across different predictive maintenance scenarios, three publicly available electric motor datasets were employed. These datasets represent complementary industrial tasks including fault classification, temperature prediction, and prognostic analysis. The overall workflow was applied consistently across all datasets while allowing task-specific adaptations during feature engineering and model optimization stages. In this study, the term "unified" refers to a shared methodological pipeline rather than to a single model trained jointly on all tasks. The elements shared across the three tasks are (i) the ordered sequence of pipeline stages (Table 5), (ii) the feature extraction toolkit comprising time-domain, frequency-domain, power spectral density, physically motivated, and rolling-window descriptors, (iii) the consensus feature selection procedure, (iv) the preprocessing and 85%/15% hold-out protocol, and (v) the evaluation and statistical testing routines. The task-specific elements are the segmentation strategy, the subset of descriptors applicable to the available sensor modalities, the target variable and learning type (classification or regression), and the model hyperparameters. Figure 1 summarises the shared and task-specific stages of the framework.

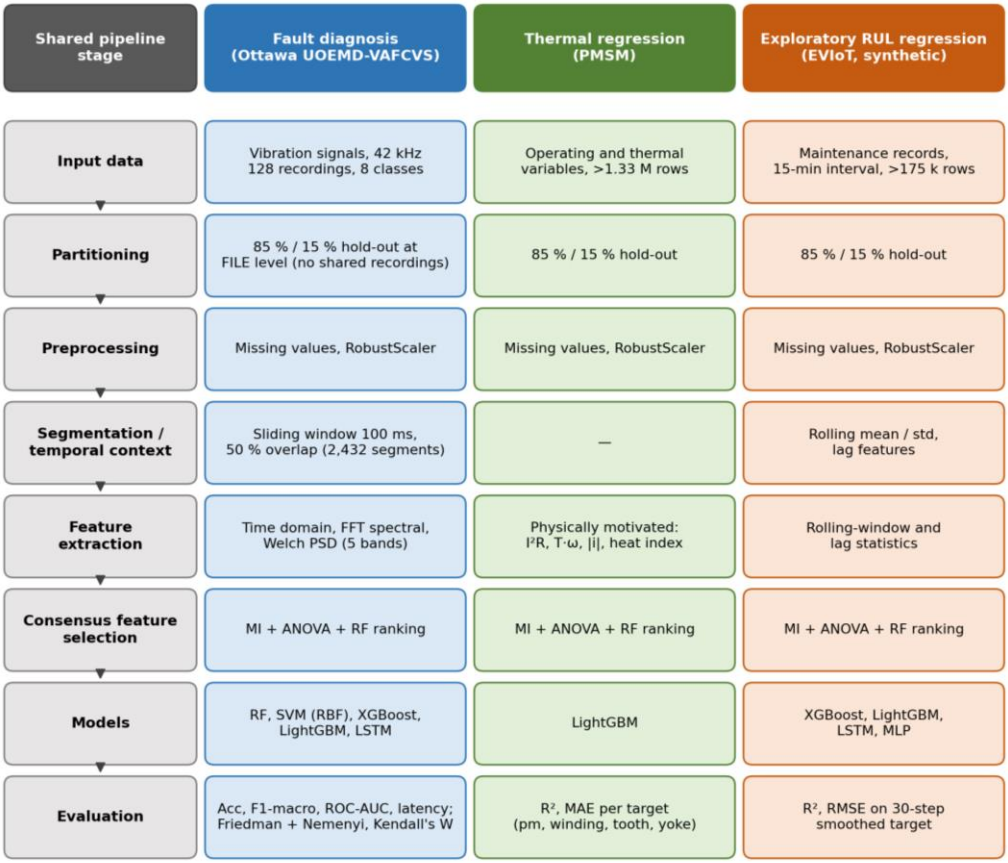


Figure 1. Overall workflow of the proposed framework. The left column lists the pipeline stages shared by all tasks; the coloured cells give the task-specific settings for fault diagnosis, thermal regression, and the exploratory RUL experiment.

2.1. Datasets

Three publicly available electric motor datasets were employed in the present study, each addressing a distinct predictive maintenance task. The datasets differ in their sensor modalities, sampling rates, and target variables, providing a heterogeneous experimental setting for the shared pipeline.

The University of Ottawa Electric Motor Dataset – Vibration and Acoustic Faults under Constant and Variable Speed Conditions (UOEMD-VAFCVS) [23,86] provides vibration, acoustic, and temperature recordings collected from eight D396 Marathon Electric three-phase induction motors at a sampling frequency of 42 kHz [86]. One healthy motor and seven motors, each carrying a single artificially introduced fault, were recorded under loaded and unloaded conditions at four constant and four variable speed profiles, yielding 16 recordings of 10 s for each health condition and 128 recordings in total [86]. The dataset is therefore balanced by the design of the original experiment; no resampling or class balancing was performed in this study. The eight health conditions are equally represented in the analysed data, each contributing 336,000 raw signal samples (2,688,000 samples in total). Table 2 lists the eight fault classes and their abbreviations [86].

Table 2. Eight health conditions in the Ottawa UOEMD-VAFCVS dataset (class codes and names as defined in the original data article [86])

Code	Fault Type	Code	Fault Type
B_R	Bowed Rotor	R_M	Rotor Misalignment
F_B	Faulty Bearings	R_U	Rotor Unbalance
H_H	Healthy	S_W	Stator Winding Fault
K_A	Broken Rotor Bars	V_U	Voltage Unbalance and Single Phasing

The Permanent Magnet Synchronous Motor (PMSM) Electric Motor Temperature dataset, obtained from Kaggle, comprises over 1.33 million observations (approximately 185 h of recordings) from 69 distinct measurement sessions (profile_id), sampled at 2 Hz [19]. The dataset provides eight input features describing the motor's operating state and four thermal target variables. Table 3 lists the principal attributes of the dataset [19].

Table 3. Principal attributes of the PMSM dataset [19]

Attribute	Value
Source	Kaggle — PMSM Motor Temperature Dataset
Total Observations	1,330,000+
Test Sessions	69 distinct sessions (profile_id)
Sampling	2 Hz (one row per 0.5 s)
Input Features	motor_speed, torque, i_d, i_q, u_d, u_q, coolant, ambient
Target Variables	pm (rotor T°), stator_yoke, stator_tooth, stator_winding

The EVIoT Predictive Maintenance dataset, also obtained from Kaggle, contains over 175,000 observations sampled at 15-minute intervals from electric vehicle motor systems [24]. The dataset is intended for remaining useful life (RUL) estimation. An important consideration regarding this dataset is that it is synthetically generated and does not contain genuine run-to-failure degradation trajectories, a property that is explicitly accounted for in the interpretation of the results. Because consecutive rows of this dataset are statistically independent, it is used here only as a surrogate dataset to demonstrate the regression branch of the pipeline, and the corresponding results are reported separately as an exploratory analysis (Section 3.6). Table 4 summarizes the dataset attributes [24].

Table 4. Principal attributes of the EVIoT Predictive Maintenance dataset [24]

Attribute	Value
Source	Kaggle — EVIoT-PredictiveMaint Dataset
Sampling Interval	15 minutes
Total Observations	175,000+
Objective	Remaining Useful Life (RUL) estimation
Important Note	Synthetically generated — no genuine degradation trend

2.2. Data Preprocessing and Partitioning Strategy

The proposed framework follows a structured pipeline that is applied consistently across the three datasets, with task-specific adjustments at the segmentation and feature engineering stages. The pipeline integrates data loading, preprocessing, signal segmentation, multi-domain feature extraction, consensus-based feature selection, model training across five learning algorithms, and statistical hypothesis testing [87,88]. Table 5 summarizes the stages of the pipeline together with the supporting libraries used at each stage.

Table 5. Stages of the proposed pipeline

Step	Operation	Tools
1	Data loading and merging	pandas, numpy
2	Preprocessing (missing values, normalization)	RobustScaler, sklearn
3	Signal segmentation (sliding window)	scipy.signal
4	Feature extraction (time + frequency + FFT)	scipy.fft, PyWavelets
5	Consensus feature selection (MI + ANOVA + RF)	sklearn, scipy
6	Model training (five algorithms)	XGBoost, LightGBM, sklearn, PyTorch
7	Statistical testing (Friedman)	scipy.stats

As summarised in Table 5, all pipeline stages were implemented in Python using pandas, NumPy, scikit-learn [39], SciPy [40], XGBoost [20], LightGBM [21], and PyTorch. The order of stages is preserved across all three datasets, with task-specific configurations applied as needed.

Raw vibration signals from the Ottawa UOEMD-VAFCVS dataset were segmented into temporally localized fragments using a sliding-window decomposition with a window length of 100 ms and 50% overlap between consecutive windows [41-43,95]. This procedure expanded the dataset from 128 file-level vectors to 2,432 segment-level examples (19 segments per recording), an approximately nineteen-fold increase in the number of trainable observations.

All experiments employed a fixed 85%/15% train–test split. For the Ottawa dataset, the split was performed at the file (recording) level: the 128 recordings were first assigned to the training or test partition, and the sliding-window segments were then generated separately within each partition. Consequently, all segments originating from the same recording belong to the same partition, and no overlapping or temporally adjacent segments of a recording appear in both the training and the test sets, which avoids the leakage that arises when overlapping segments are split at random [94,97]. The test partition contained n_test recordings (m_test segments) that were not used at any stage of training. The scaler and the consensus feature selection were fitted on the training partition only and then applied unchanged to the test partition. The same 85%/15% ratio was applied to all three datasets.

2.3. Descriptor Construction

A multi-domain feature engineering strategy was employed to construct a comprehensive and interpretable representation of the input data. The proposed approach integrates time-domain statistical descriptors, frequency-domain spectral characteristics, power spectral density measurements, physically motivated variables, and temporal rolling-window statistics [44,45]. This combination was designed to capture complementary aspects of motor behavior and provide informative features for both classification and regression tasks [44,45]. A consensus-based feature selection procedure was subsequently applied to identify the most relevant descriptors for downstream modelling [44,45].

Time-domain analysis included five widely used statistical features extracted from each signal segment: root mean square (RMS), peak value, crest factor, skewness, and kurtosis [46]. These descriptors characterize signal energy, amplitude distribution, impulsiveness, and waveform asymmetry [46]. Frequency-domain information was obtained using the Fast Fourier Transform (FFT) implemented through the scipy.fft library [40,47,91]. From each spectral representation, spectral centroid, spectral spread, spectral entropy, and dominant frequency were calculated to capture the distribution and complexity of spectral energy [48]. In addition, power spectral density (PSD) was estimated using the Welch method [60], and the resulting spectrum was partitioned into five frequency bands [49]. The integrated power within each band was retained as an additional feature describing energy distribution across different frequency regions [48,49].

To incorporate domain knowledge, several physically motivated variables were engineered from the underlying electromechanical measurements. These included copper loss (I^2R), mechanical power ($T \times \omega$), and current amplitude [51]. These quantities are standard electromechanical relations computed directly from the measured signals; unlike physics-informed neural networks [50,85], no physical law is imposed as a constraint during model training. For the PMSM dataset, an additional heat-index variable was calculated as the difference between rotor temperature and coolant temperature (pm − coolant), representing the thermal gradient within the cooling system [52]. Temporal information was incorporated through rolling-window statistics, including rolling means, rolling standard deviations, and lag-based features derived from previous observations [52,53]. These features were particularly important for the EVIoT remaining useful life estimation task, where temporal dependencies are not explicitly represented in the raw observations [54]. Finally, dimensionality reduction was performed using a consensus feature selection framework combining Mutual Information (MI), Analysis of Variance (ANOVA), and Random Forest feature importance scores [55,92,93]. Features consistently ranked highly across multiple criteria were retained, while low-ranking features were removed, resulting in a more robust and stable feature subset [54-56].

2.4. Model Architecture and Learning Framework

Five machine learning algorithms were evaluated within the unified pipeline: Random Forest, XGBoost, LightGBM, Support Vector Machine with an RBF kernel, and an LSTM network. The original formulations of these algorithms are given in [57], [20], [21], [58], and [59], respectively. Each algorithm was configured with task-specific hyperparameters as summarized in Table 6. The Random Forest model was deliberately constrained to limit its depth as a regularization measure, while the gradient-boosted methods (XGBoost and LightGBM) and the LSTM network were executed on a GPU.

Table 6. Configuration of the five evaluated machine learning models

Model	Task	Key Parameters	Hardware
Random Forest	Fault diagnosis	100 trees, max_depth = 10 (constrained)	CPU
XGBoost (GPU)	Fault diagnosis + RUL	300 trees, lr = 0.05, max_depth = 7	RTX 5060
LightGBM (GPU)	Fault diagnosis + temperature	num_leaves = 15, max_depth = 6	RTX 5060
SVM (RBF)	Fault diagnosis	C = 10, gamma = scale	CPU
LSTM (PyTorch)	Fault diagnosis + RUL	2 layers, hidden = 128, 40 epochs	RTX 5060

The Random Forest classifier was implemented using scikit-learn with 100 decision trees and a maximum depth of 10 [61]. The depth constraint was applied deliberately to prevent the model from memorizing the training data and to provide a meaningful comparison against deeper ensemble methods [62]. The Random Forest model operated entirely on CPU [61,62]. XGBoost was configured with 300 estimators, a learning rate of 0.05, and a maximum tree depth of 7 [63]. The GPU-accelerated implementation was executed on an NVIDIA RTX 5060. XGBoost was applied to both the fault diagnosis task on the Ottawa dataset and the RUL estimation task on the EVIoT dataset [63,64]. LightGBM was configured with a maximum of 15 leaves per tree and a maximum depth of 6 [65]. The GPU-accelerated implementation was executed on an NVIDIA RTX 5060 [66]. The Support Vector Machine classifier was implemented with a radial basis function (RBF) kernel, a regularization parameter C = 10, and gamma set to the scikit-learn default 'scale' value [67]. The SVM was executed on CPU, providing a GPU-free baseline for the fault diagnosis task [67,68]. The LSTM network was implemented in PyTorch with two recurrent layers and a hidden state dimension of 128 [69]. The model was trained for 40 epochs on an NVIDIA RTX 5060 GPU [70]. The LSTM was applied to both the fault diagnosis task and the RUL estimation task [69,70].

In all cases, the test partition was not used for hyperparameter selection. The depth limits of Random Forest (max_depth = 10) and LightGBM (max_depth = 6, num_leaves = 15) were chosen deliberately as regularisation measures, given the small number of independent recordings in the Ottawa dataset.

Assignment of models to tasks. Not every algorithm was evaluated on every task (Table 6). The fault diagnosis task is the primary benchmark of the study, and all five algorithms were therefore compared on it. The kernel SVM was not applied to the PMSM and EVIoT datasets because its training cost grows between quadratically and cubically with the number of training samples, which is impractical for datasets with 10^5–10^6 observations [58]. For the PMSM thermal regression task, LightGBM, the best-performing model in the fault diagnosis benchmark and the most efficient model on large tabular data among those considered, was used. For the exploratory RUL experiment, the two gradient-boosting models were compared with a recurrent (LSTM) and a fully connected (MLP) neural baseline to examine whether sequence modelling offered any advantage on the surrogate data.

3. RESULTS

The results obtained from the three datasets are presented in this section. The overall fault diagnosis results on the Ottawa UOEMD-VAFCVS dataset are reported first, followed by the ROC and confusion-

matrix analysis (Section 3.1), performance stability and statistical testing (Section 3.2), computational cost (Section 3.3), and thermal behaviour across fault conditions (Section 3.4). Section 3.5 presents the temperature prediction results on the PMSM dataset, and Section 3.6 reports the exploratory RUL experiment on the EVIoT dataset. The comparison with existing literature and the limitations of the study are discussed in Section 4.

The proposed fault diagnosis pipeline was evaluated on the Ottawa UOEMD-VAFCVS dataset using five machine learning algorithms; Random Forest, XGBoost, LightGBM, Support Vector Machine with an RBF kernel, and an LSTM network. All experiments employed the file-level 85%/15% train–test split described in Section 2.2. Performance was quantified using accuracy, macro F1-score, ROC-AUC, training duration, and per-sample inference latency. Table 7 reports the aggregated results across all five models. Precision and recall are not tabulated separately because, owing to the balanced class distribution, they coincided with accuracy to within ±0.001 for every model.

Table 7. Fault diagnosis results on the held-out test recordings of the Ottawa UOEMD-VAFCVS dataset (file-level split)

Model	Accuracy	F1-macro	ROC-AUC	Train (s)	Infer (ms)
LightGBM	1.0000	1.0000	1.0000	5.8	0.008
XGBoost	0.9973	0.9973	1.0000	3.3	0.339
Random Forest	0.9945	0.9945	0.9999	0.1	0.046
SVM (RBF)	0.9918	0.9918	1.0000	0.4	0.066
LSTM	0.9753	0.9756	0.9989	3.9	0.004

Among all evaluated models, LightGBM achieved the highest classification performance, obtaining near-perfect scores under the evaluated test conditions across all primary evaluation metrics (Accuracy = F1-macro = ROC-AUC = 1.0000). Coupled with its sub-millisecond inference latency of 0.008 ms per sample, this combination of accuracy and speed suggests potential suitability for real-time deployment, subject to the limitations discussed in Section 4.1. XGBoost produced comparably strong predictive performance (F1-macro = 0.9973, ROC-AUC = 1.0000), although its inference latency remained higher than that of LightGBM (0.339 ms vs. 0.008 ms per sample). Despite the deliberately constrained hyperparameters (max_depth = 10), Random Forest attained a robust macro F1 of 0.9945 with only 0.1 s of training time, supporting the continued relevance of classical ensemble approaches for vibration-based fault diagnosis tasks.

The LSTM model obtained the lowest overall performance among the evaluated methods (F1-macro = 0.9756). In the present configuration, the LSTM baseline did not surpass the feature-engineered gradient-boosting models. Because only one recurrent configuration was evaluated and no end-to-end deep model operating on raw signals was included, this result should not be generalised to deep learning architectures as a whole; it is, however, consistent with benchmark studies reporting that tree-based ensembles remain highly competitive on tabular, feature-engineered inputs [96].

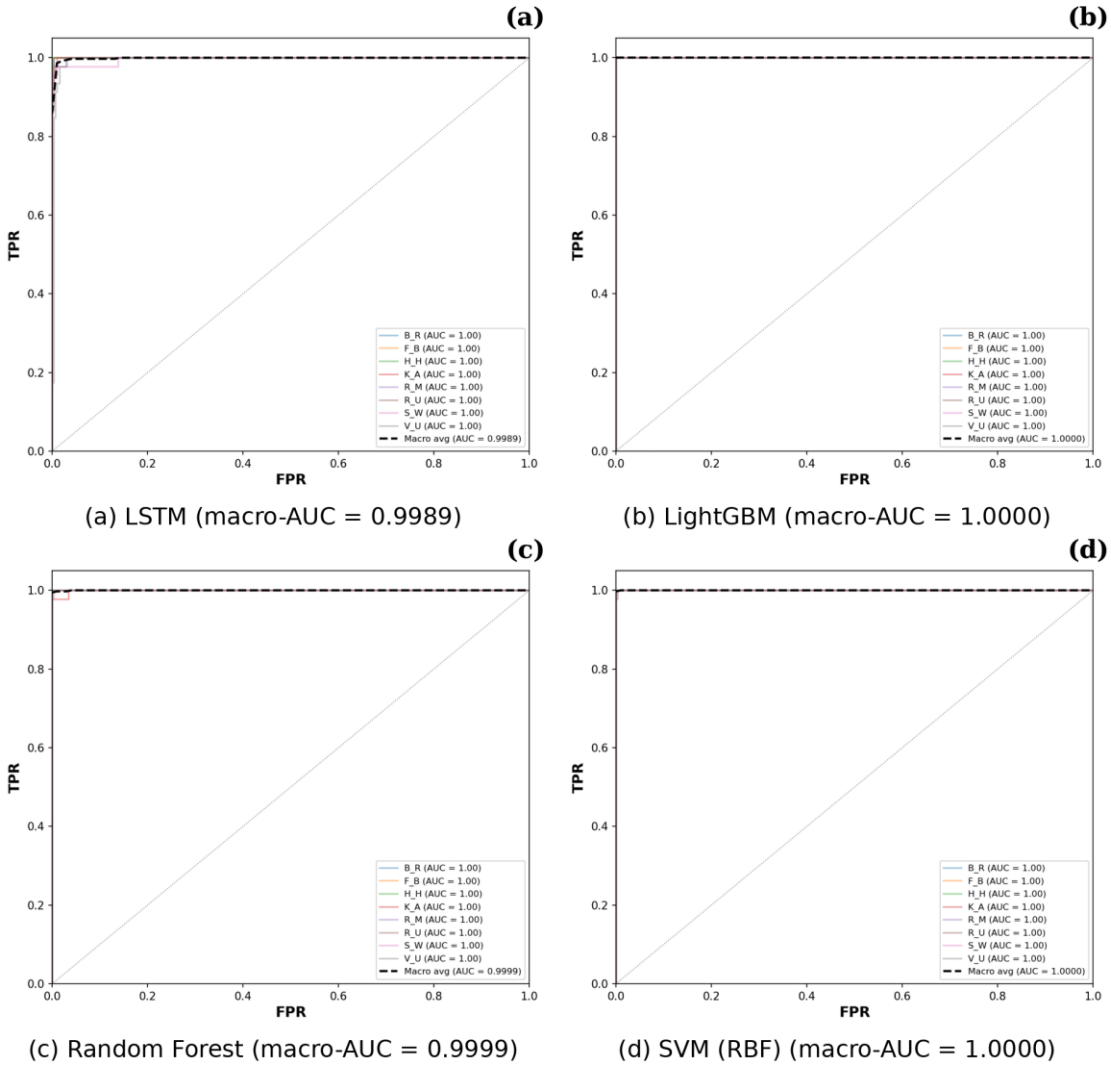
The results further illustrate the effectiveness of the proposed sliding-window segmentation strategy. As described in the methodology, the segmentation procedure expanded the dataset from 128 file-level vectors to 2,432 segment-level examples, an approximately nineteen-fold increase in the number of trainable observations. Under this segmentation protocol, and with the file-level split kept unchanged, the macro F1-score increased from 0.569 on the unsegmented baseline to above 0.994 for all evaluated models.

Several factors explain why all five models achieved uniformly high scores, and they should be considered when interpreting the results. First, the faults in UOEMD-VAFCVS were introduced artificially at a fixed severity and recorded in a laboratory under controlled conditions [86], so that the fault signatures are strong and stable. Second, the multi-domain descriptors summarise each 100 ms window into statistical and spectral quantities in which these signatures are close to linearly separable; once such

features are available, the choice of classifier has only a secondary effect, which is why the differences between models are below 0.025 in F1-macro. Third, and most importantly, each health condition was recorded on a single physical motor [86]. The file-level split prevents overlapping segments of the same recording from appearing in both partitions, but the training and test recordings of a given class still originate from the same motor, so a classifier may partly exploit motor-specific characteristics (for example, mounting or manufacturing tolerances) in addition to the fault signature itself. The reported scores should therefore be interpreted as within-machine performance on a controlled laboratory benchmark rather than as an estimate of generalisation to unseen machines. Finally, the test partition contains only n_test recordings, so that a single misclassified recording changes the accuracy by several percentage points and differences of less than 0.01 between models should not be over-interpreted. With respect to overfitting, the evidence available is that (i) the test recordings were never seen during training, (ii) scaling and feature selection were fitted on the training partition only, (iii) the depth of the tree-based models was deliberately constrained, and (iv) the ranking of the models was stable across repeated runs (Section 3.2).

3.1. ROC Curve and Confusion Matrix Analysis

Receiver Operating Characteristic (ROC) analysis was conducted to evaluate the class-level discrimination capability of each model. Figure 2 presents the one-vs-rest ROC curves for all five algorithms.



(a) LSTM (macro-AUC = 0.9989)

(b) LightGBM (macro-AUC = 1.0000)

(c) Random Forest (macro-AUC = 0.9999)

(d) SVM (RBF) (macro-AUC = 1.0000)

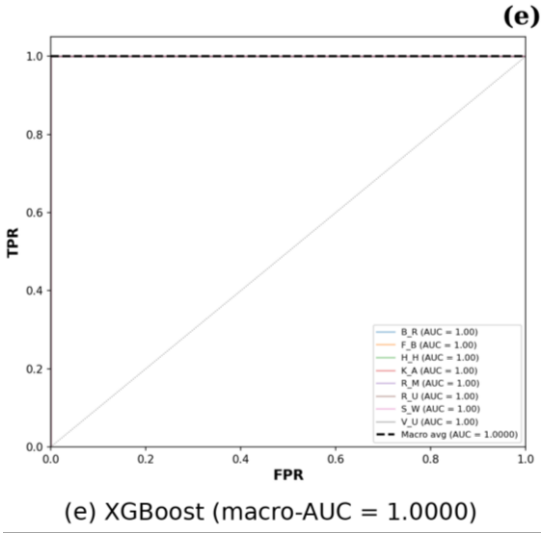
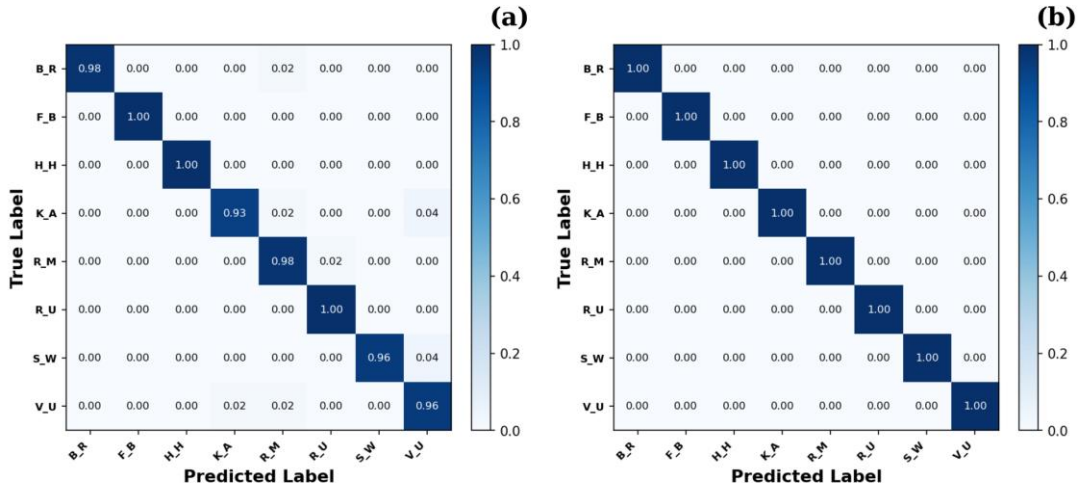


Figure 2. ROC curves (one-vs-rest) for all five models on the eight-class fault diagnosis task.

The LightGBM, XGBoost, and SVM models all achieved macro-averaged AUC values of 1.0000, indicating near-perfect class separability across the eight fault categories. Random Forest yielded a comparably high macro-AUC of 0.9999, with the marginal deviation concentrated in the K_A class. The LSTM model attained a macro-AUC of 0.9989; minor early-region deviations are visible in the K_A and V_U class curves (Figure 2a), reflecting probability-score calibration differences relative to the tree-based ensembles rather than a fundamental loss of discriminability. These findings indicate that the handcrafted statistical and spectral features contain highly discriminative information for fault separation, enabling tree-based ensemble methods to achieve reliable class separation.

To investigate class-wise classification behaviour, normalized confusion matrices were analysed for all five models and presented in Figure 3.



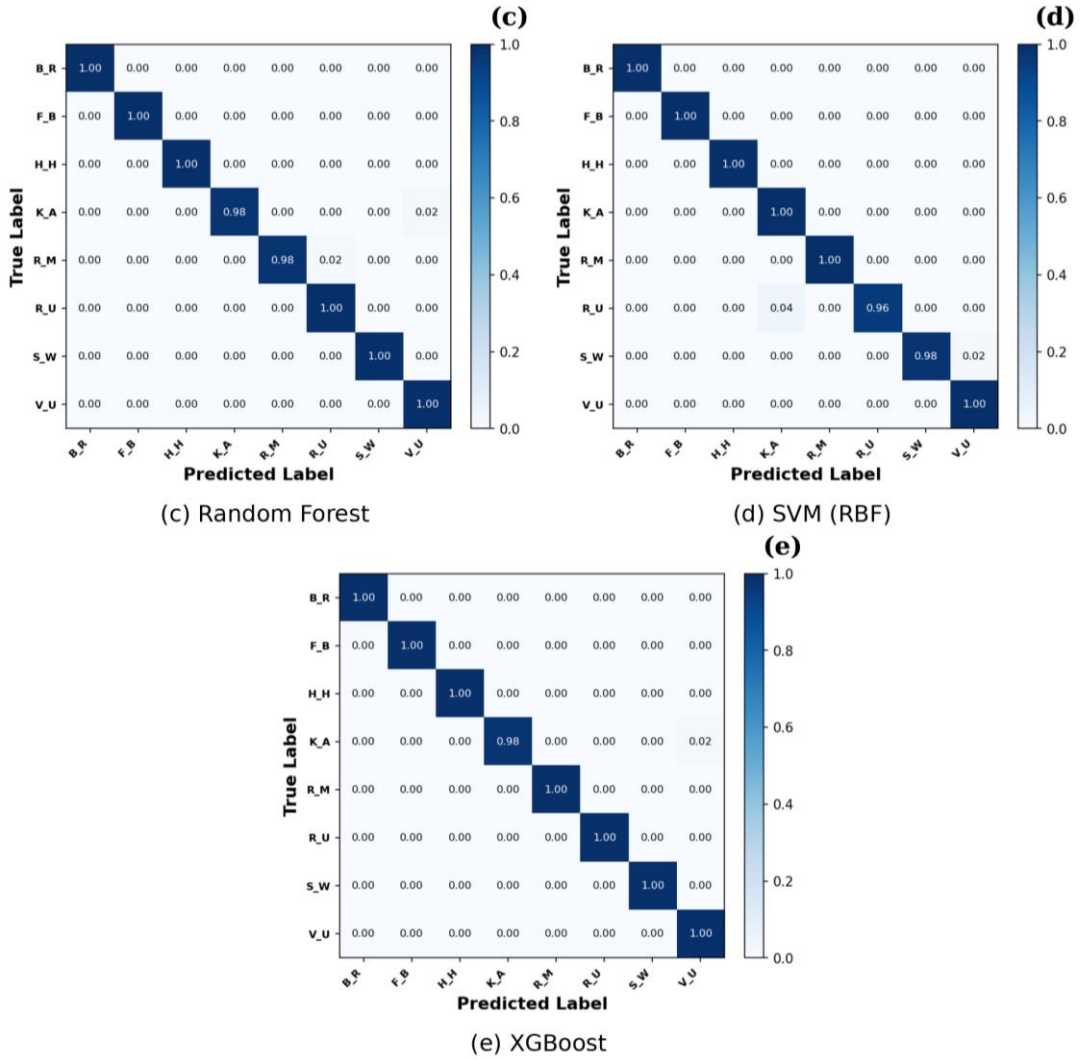


Figure 3. Normalized confusion matrices for all five models on the eight health conditions (B_R: bowed rotor; F_B: faulty bearings; H_H: healthy; K_A: broken rotor bars; R_M: rotor misalignment; R_U: rotor unbalance; S_W: stator winding fault; V_U: voltage unbalance and single phasing).

The confusion matrices reveal that LightGBM (Figure 3b) attained perfect classification across all eight fault classes, with no misclassifications observed in the evaluated test partition. XGBoost (Figure 3e) and Random Forest (Figure 3c) yielded near-diagonal matrices. Both models misclassified 2% of K_A samples as V_U, while Random Forest additionally misclassified 2% of R_M samples as R_U, which XGBoost classified without error. SVM (Figure 3d) misclassified 4% of R_U samples as K_A.

The most challenging class across all models was identified as K_A (broken rotor bars), which exhibited cross-class confusion with both V_U and R_M. The LSTM model (Figure 3a) showed the largest deviation from the diagonal, with 4% K_A → V_U, 2% K_A → R_M, 4% S_W → V_U, and 2% B_R → R_M misclassifications. Overall, the confusion matrix analysis indicates that the engineered features separate all health conditions well, with LightGBM yielding the highest class-level separability among the evaluated models.



Figure 4. Per-class F1 heatmap for all five models.

The per-class F1 heatmap in Figure 4 consolidates these observations: F_B, H_H, and S_W reach F1 = 1.000 for all tree-based models, whereas K_A is correctly classified across the entire test partition only by LightGBM (F1 = 1.000) - Random Forest and XGBoost both report 0.989, SVM 0.979, and LSTM 0.956. The V_U class shows a similar pattern, dropping to 0.936 for LSTM while remaining at 1.000 for LightGBM. This confirms that the marginal advantage of LightGBM is concentrated in the two most challenging fault categories rather than uniformly distributed across all classes.

3.2. Performance Stability and Statistical Analysis

Model robustness was assessed by examining the distribution of F1-macro scores across N independent training runs with different random seeds on the fixed file-level split, and statistical significance among the models was evaluated using the Friedman non-parametric test. Figure 5 presents the resulting F1-macro distributions.

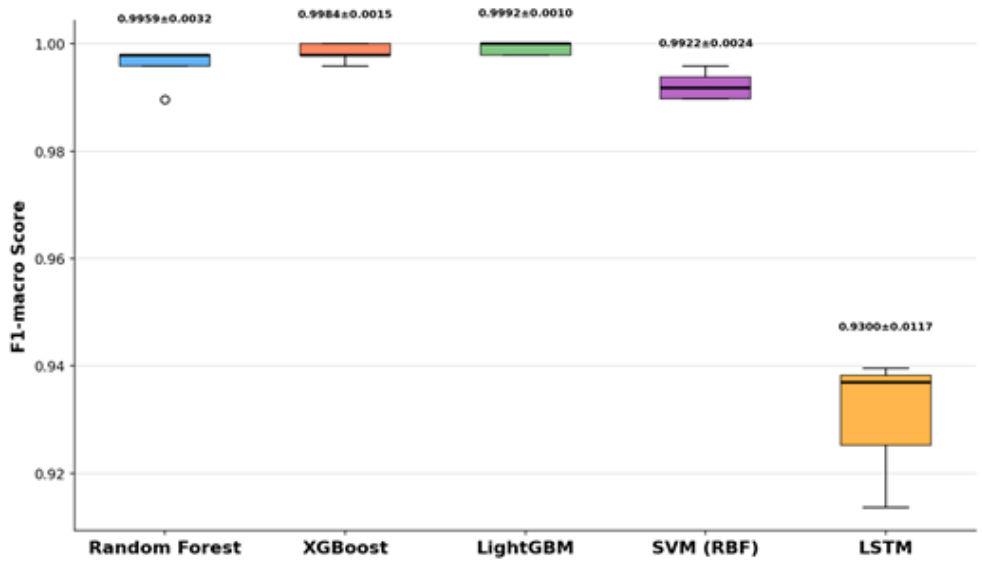


Figure 5. Box plots of the F1-macro distributions for all evaluated models.

The performance distributions demonstrate that LightGBM and XGBoost not only achieved superior mean F1 scores but also exhibited the lowest performance variance (standard deviations of 0.0010 and 0.0015, respectively). Random Forest and SVM produced similarly narrow distributions (standard deviations of 0.0032 and 0.0024). In contrast, the LSTM model showed noticeably larger variability — its interquartile range spans approximately 0.92 to 0.94, indicating reduced stability under different training conditions.

The Friedman test [88] was applied with the five models as treatments and the N runs as blocks ($k = 5$, $N = N$). It yielded $\chi^2 = 11.556$ ($df = 4$, $p = 0.021$), indicating that the models are not equivalent at the $\alpha = 0.05$ level. The effect size, expressed as Kendall's coefficient of concordance [90], $W = \chi^2/[N(k - 1)] = W$, indicates weak/moderate/strong agreement of the model rankings across runs. Pairwise differences were examined with the Nemenyi post-hoc test [87,89]; with a critical difference of CD = CD at $\alpha = 0.05$. Because all runs share the same train–test partition, they are not fully independent, and the test captures variability due to model initialisation rather than data sampling. Its outcome is therefore interpreted as supporting evidence for the observed ranking on this dataset, namely that the gradient-boosting models achieved higher and more stable F1-macro scores than the LSTM baseline, rather than as a general statement about ensemble and deep learning methods.

3.3. Computational Cost Analysis

Beyond predictive performance, computational efficiency was evaluated in terms of training duration and per-sample inference latency. Figure 6 summarizes both measurements. Training time was measured as the wall-clock duration of the fitting call. Per-sample inference latency was obtained by predicting the entire test partition in a single batch of m_test samples, dividing the elapsed wall-clock time by the number of samples, and averaging over R repetitions after w warm-up calls. Random Forest and SVM were executed on CPU (CPU model), whereas XGBoost, LightGBM, and LSTM were executed on an NVIDIA RTX 5060 GPU. Because this convention measures batched throughput, it favours GPU-executed models and is not equivalent to the latency of a single-sample (batch size = 1) prediction; the values should therefore be compared primarily within the same hardware class.

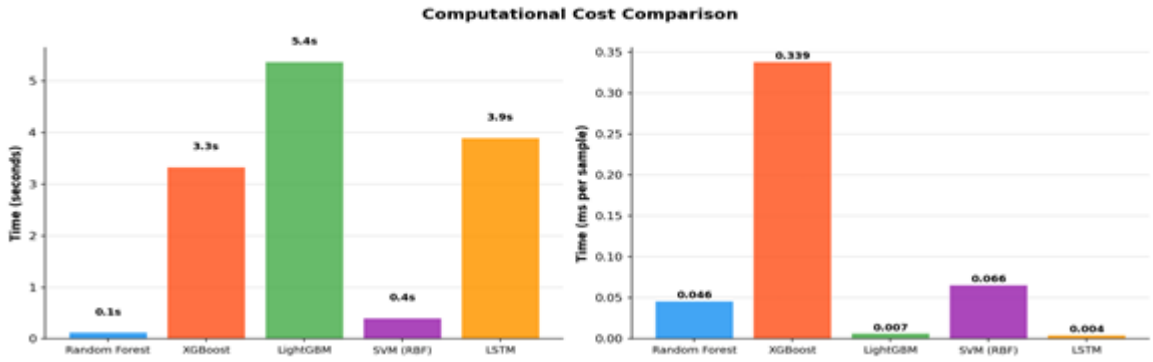


Figure 6. Computational cost comparison: (left) training duration in seconds; (right) per-sample inference latency in milliseconds.

Random Forest achieved the shortest training duration of 0.1 s on CPU, whereas LightGBM required the longest training time among the evaluated ensemble methods (5.8 s). Despite this longer training stage, LightGBM provided a per-sample inference cost of 0.008 ms, suggesting suitability for high-throughput industrial monitoring scenarios and edge-based deployment. XGBoost displayed a higher inference cost (0.339 ms per sample) that may limit its applicability in high-throughput real-time scenarios.

LSTM exhibited a 3.9 s training time and a per-sample cost of 0.004 ms. This value is the GPU inference time of the full test batch divided by the batch size and therefore reflects parallel batch throughput on

the GPU; it does not imply that a single LSTM prediction is faster than a single prediction of the CPU-based models. In addition, the larger performance variance reported in Section 3.2 may reduce the reliability of the LSTM for practical deployment. SVM, operating entirely on CPU, provided a competitive inference time of 0.066 ms without requiring GPU resources, making it an attractive option for resource-constrained edge deployments. In general, these findings indicate that classical and gradient-boosted machine learning approaches offer a more favourable trade-off between predictive performance and deployment efficiency than the evaluated LSTM configuration for this fault diagnosis task.

3.4. Thermal Behaviour Across Fault Conditions

To further investigate the relationship between fault conditions and thermal behaviour, an ANOVA was performed on the temperature sensor measurements recorded in the Ottawa UOEMD-VAFCVS dataset, grouped by fault class. The analysis revealed that motors operating under a stator winding fault (S_W) exhibited the highest mean temperature (33.8 °C), whereas healthy motors (H_H) operated at the lowest mean temperature (27.7 °C). The ANOVA test confirmed that this difference was statistically significant ($F = 84.864$, $p < 0.05$).

This finding establishes a quantitative link between fault conditions and thermal behaviour: motor fault states are not only distinguishable through vibration signatures, as demonstrated in Sections 3.1–3.3, but also exhibit measurable thermal signatures that may serve as complementary diagnostic indicators. Motivated by this observation, the following section evaluates a dedicated regression framework for rotor and stator temperature estimation in permanent magnet synchronous motors using the PMSM Electric Motor Temperature dataset.

3.5. Temperature Prediction Results on the PMSM Dataset

The PMSM Electric Motor Temperature dataset was used to evaluate the proposed regression pipeline for rotor and stator temperature prediction. Four thermal targets were estimated simultaneously: rotor permanent magnet temperature (pm), stator winding temperature, stator tooth temperature, and stator yoke temperature. The feature space incorporated physically motivated variables, including the heat-index (pm − coolant), copper loss (I^2R), mechanical power ($T \times \omega$), and current amplitude, in addition to raw sensor measurements. Table 8 reports the regression performance.

Table 8. Temperature prediction results on the PMSM dataset using LightGBM (GPU) (Green row: primary safety-critical target. All four targets achieve $R^2 > 0.987$ and MAE < 1.4°C.)

Target Variable	R^2	MAE (°C)	Most Important Feature
pm (Rotor Temperature)	0.993	1.09	heat_index (pm−coolant)
stator_winding	0.991	1.24	coolant
stator_tooth	0.989	1.31	motor_speed
stator_yoke	0.987	1.38	i_q (q-axis current)

The LightGBM GPU model achieved the highest overall predictive performance, obtaining an R^2 of 0.993 and a mean absolute error (MAE) of 1.09°C for rotor temperature (pm) estimation, indicating high predictive accuracy relevant to thermal protection systems in PMSM drives. Comparable performance levels were obtained for the three stator-related thermal variables (R^2 of 0.991, 0.989, and 0.987 for stator_winding, stator_tooth, and stator_yoke, respectively), with mean absolute errors ranging from 1.24°C to 1.38°C.

Among all extracted features, the engineered heat-index variable — representing the temperature differential between the rotor and the coolant circuit — emerged as the most influential predictor for rotor temperature estimation. For the stator-related thermal variables, the most influential features were

the coolant temperature for stator_winding, motor_speed for stator_tooth, and the q-axis current (i_q) for stator_yoke.

3.6. Exploratory Analysis: RUL Regression on a Synthetic Surrogate Dataset

The experiment reported in this section is exploratory. As noted in Section 2.1, the EVIoT dataset is synthetically generated, its consecutive rows are statistically independent, and it contains no genuine run-to-failure degradation trajectories. The purpose of this section is therefore limited to demonstrating that the regression branch of the pipeline can be executed end to end on a time-indexed maintenance dataset; the results must not be interpreted as evidence of remaining-useful-life estimation capability. The dataset contains more than 175,000 rows sampled at 15-minute intervals. Sliding-window statistics, lag-based temporal features, and rolling-mean representations were incorporated into the feature space to provide temporal context. Table 9 reports the regression performance on the 30-step smoothed RUL target.

Table 9. Exploratory RUL regression results on the synthetic EVIoT Predictive Maintenance dataset (30-step smoothed target; not indicative of prognostic capability)

Model	R^2 (Smooth-30)	RMSE	Method
XGBoost (GPU)	0.933	3.98	Sliding window + lag features
LightGBM (GPU)	0.933	3.98	Sliding window + lag features
LSTM (PyTorch)	0.883	5.25	40 epochs, RTX 5060 (20 s)
MLP (sklearn)	0.826	6.42	Fully connected layers
Raw baseline	~0.03	—	No feature engineering

On this surrogate dataset, the raw-data baseline produced an R^2 of approximately 0.03, whereas the engineered pipeline reached $R^2 = 0.933$ for both XGBoost and LightGBM. LSTM achieved comparatively lower performance ($R^2 = 0.883$, RMSE = 5.25), while MLP-based regression produced the weakest results among the evaluated learning methods ($R^2 = 0.826$).

These values must be interpreted in light of how the target was constructed. The raw RUL values of consecutive rows are statistically independent, which explains the near-zero R^2 of the raw baseline. Applying a 30-step rolling mean to the target introduces strong autocorrelation between adjacent target values, and lag and rolling features computed from the same smoothed series then make the target highly predictable even though no physical degradation process is present in the data. The increase from $R^2 \approx 0.03$ to $R^2 = 0.933$ is therefore an artifact of the target smoothing and does not measure the ability of the models to estimate remaining useful life. Validation of the RUL branch on genuine run-to-failure data, such as the PRONOSTIA/FEMTO-ST bearing or NASA C-MAPSS datasets, is required before any prognostic claim can be made.

4. DISCUSSION

In the evaluated configurations, the gradient-boosting models achieved higher scores than the LSTM baseline on the fault diagnosis task. LightGBM achieved the highest fault diagnosis performance on the Ottawa UOEMD-VAFCVS dataset [86], reaching perfect classification results on the held-out test recordings (Accuracy = 1.0000, F1-score = 1.0000, ROC-AUC = 1.0000) [71-73], while also providing one of the lowest inference latencies. XGBoost produced comparable performance with a macro F1-score of 0.9973 [74], whereas Random Forest maintained robust classification capability despite its intentionally constrained architecture [75]. SVM achieved competitive results without requiring GPU resources [76]. In contrast, LSTM obtained the lowest classification performance (F1-score = 0.9756) and exhibited greater variability across repeated experiments [77]. The same ranking was observed in the exploratory

RUL experiment [79-81]; however, because those results are dominated by the smoothing of a synthetic target (Section 3.6), they provide no independent evidence of prognostic performance. Taken together, the findings suggest that, when informative engineered features are available, gradient-boosting models are strong and computationally inexpensive baselines; whether they also surpass end-to-end deep architectures operating on raw signals was not tested in this study. A similar observation has been reported for PMSM temperature estimation, where comparatively simple supervised models were shown to be competitive with parameter-heavy deep networks in terms of accuracy, model size, and data demand [98].

The proposed framework also highlights the importance of feature engineering and data representation in predictive maintenance applications. Sliding-window segmentation increased the number of trainable observations from 128 to 2,432 and improved the macro F1-score from 0.569 to above 0.994 across all evaluated classifiers, demonstrating that data preparation can be as influential as model selection [82]. The PMSM temperature prediction experiments further indicated the usefulness of physically motivated descriptors [83], with LightGBM achieving R^2 values between 0.987 and 0.993 and mean absolute errors below 1.4°C for all thermal targets [78,84]. The Friedman test (χ^2 = 11.556, p = 0.021, W = W) supports the observed ranking of the classifiers on this dataset [87]. Collectively, these findings suggest that the combination of multi-domain feature extraction, consensus-based feature selection, and gradient-boosting algorithms provides an accurate and computationally efficient solution for laboratory-scale electric motor fault diagnosis and thermal estimation [82-84,86], within the limitations discussed below.

4.1. Limitations of the Study

The findings of this study are subject to several limitations related to the data, the risk of overfitting, and the generalisability of the models.

Datasets. In the Ottawa UOEMD-VAFCVS dataset, each health condition was recorded on a single motor, with faults introduced artificially at a fixed severity in a laboratory environment [86]. Although the file-level split prevents segment-level leakage, training and test recordings of a class come from the same machine, and the models may partly learn machine-specific characteristics. The class-wise temperature differences reported in Section 3.4 may likewise reflect differences between individual motors as well as the faults themselves. The PMSM dataset was recorded on a single laboratory test bench [19], and the EVIoT dataset is synthetic and lacks genuine degradation trajectories, so that the RUL results are exploratory only (Section 3.6).

Overfitting and evaluation protocol. The evaluation relies on a single fixed 85%/15% hold-out split. With 128 recordings in total, the test partition is small, and a single misclassified recording changes the accuracy by several percentage points. Repeated runs vary only the model initialisation on the same split, so the Friedman test does not capture variability due to data partitioning. Nested or grouped cross-validation over recordings would provide tighter estimates of generalisation error.

Generalisability. The models were not evaluated across machines, across operating conditions (for example, training on some speed profiles and testing on others), under different fault severities, or under additional noise. Only one LSTM configuration was considered, and no end-to-end deep model operating on raw signals was included; the comparison with deep learning is therefore limited to the evaluated baseline. Inference latencies were measured as batched throughput on heterogeneous hardware (Section 3.3) and should not be interpreted as single-sample latencies on embedded devices. Addressing these points, in particular cross-machine validation and RUL validation on real run-to-failure data, is the main direction for future work.

5. CONCLUSION

This study presented a unified machine learning framework for electric motor fault diagnosis, temperature prediction, and remaining useful life estimation using three publicly available datasets. The proposed pipeline combined sliding-window segmentation, multi-domain feature extraction, physically motivated descriptors, and consensus-based feature selection within a common predictive maintenance framework. LightGBM achieved the highest fault diagnosis performance on the Ottawa UOEMD-VAFCVS dataset with perfect classification metrics on the held-out test recordings of a file-level split (Accuracy = 1.0000, F1-score = 1.0000, ROC-AUC = 1.0000) and a low per-sample inference cost. For PMSM thermal prediction, LightGBM achieved R^2 values ranging from 0.987 to 0.993 with mean absolute errors below 1.4°C. An exploratory experiment on the synthetic EVIoT dataset showed that the pipeline can be applied to RUL-type regression; however, the R^2 of 0.933 obtained there reflects the smoothing of a synthetic target and does not constitute validated prognostic capability. A Friedman test indicated significant differences among the evaluated classifiers. The results indicate that gradient-boosting models combined with domain-informed feature engineering can provide accurate and computationally efficient solutions for laboratory-scale motor fault diagnosis and thermal estimation, while validation across different machines and on genuine run-to-failure data remains necessary.

Future studies may evaluate additional machine learning and deep learning architectures, including CatBoost, Extra Trees, AdaBoost, HistGradient Boosting, TabNet, Temporal Fusion Transformer, Transformer Encoder Networks, Informer, AutoGluon, N-BEATS, GRU, and Graph Neural Networks. The proposed framework may also be validated using other publicly available predictive maintenance datasets such as Paderborn University Bearing Dataset, Case Western Reserve University Bearing Dataset, MFPT Bearing Dataset, IMS Bearing Dataset, PRONOSTIA Dataset, SECOM Manufacturing Dataset, NASA C-MAPSS Turbofan Dataset, FEMTO-ST Bearing Dataset, Tennessee Eastman Process Dataset, and MIMII Industrial Sound Dataset. In addition, explainable artificial intelligence techniques including SHAP, LIME, Integrated Gradients, DeepSHAP, Permutation Importance, Partial Dependence Plots, Accumulated Local Effects, Counterfactual Explanations, Anchors, and Grad-CAM may be incorporated to improve model transparency and support decision-making in real-world industrial environments.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Author Contributions (CRediT): Conceptualization, A.Y.S., N.K., F.E., and Y.K.; Data curation, F.E., M.S.G. and A.O.; Formal analysis, F.E., A.Y.S., N.K. and M.S.G.; Funding acquisition, M.S.G., A.O and F.E.; Investigation, M.S.G, Y.K, N.K. and A.Y.S.; Methodology, F.E., Y.K. and M.S.G.; Project administration, .Y.K., N.K., F.E., and A.O.; Resources, A.O., M.S.G, A.Y.S., N.K. and F.E.; Software, M.S.G., Y.K. and F.E.; Supervision, F.E., and M.S.G.; Validation, M.S.G. and F.E. Visualization, M.S.G, A.O., Y.K. and F.E ; Writing – original draft, M.S.G, A.O., Y.K. and F.E; Writing – review & editing, M.S.G, A.O. and F.E. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The datasets used to train and evaluate the fault detection, temperature prediction, and remaining useful life (RUL) estimation models in this study were obtained from three publicly accessible sources. Vibration, acoustic, and temperature data used for the 8-class fault detection task were retrieved from the University of Ottawa Electric Motor Dataset (https://www.kaggle.com/datasets/guilhermeparcerao/university-of-ottawa-electric-motor-dataset), which provides open access to signals sampled at 42 kHz from D396 Marathon Electric three-phase motors under various load and speed conditions. Continuous sensor measurements used for real-time rotor and stator thermal modeling were sourced from the PMSM Electric Motor Temperature repository (https://www.kaggle.com/datasets/wkirgsn/electric-motor-temperature). This dataset contains over 1.3 million continuous samples of motor speed, torque, currents, voltages, and temperatures. The predictive

maintenance records used as RUL forecasting samples were sourced from the EVIoT-PredictiveMaint dataset (https://share.google/ZPF6sBAt5tuH5zysj). This dataset contains records sampled at 15-minute intervals. The full feature set employed across all five classifiers, including time-domain metrics (RMS, peak value, crest factor, skewness, kurtosis), frequency-domain descriptors (FFT spectral center, spread, entropy, dominant frequency, Welch PSD), physics-derived quantities (copper loss, mechanical power, current amplitude), and sliding window statistics (rolling mean, rolling standard deviation, lag features), was derived entirely from these open-access sources. No proprietary data were used in this study. The code, trained model outputs, and prediction results generated in this work will be made available upon reasonable request to the corresponding author.

REFERENCES

[1] Tiwari D, Miscandlon J, Tiwari A, Jewell GW. A review of circular economy research for electric motors and the role of Industry 4.0 technologies. *Sustainability*. 2021;13(17):9668. doi: https://doi.org/10.3390/su13179668

[2] Cai W, Wu X, Zhou M, Liang Y, Wang Y. Review and development of electric motor systems and electric powertrains for new energy vehicles. *Automotive Innovation*. 2021;4(1):3-22. doi: https://doi.org/10.1007/s42154-021-00139-z

[3] Ghemari Z, Belkhiri S. Improving the maintenance of electric motors: enhancing reliability and reducing downtime. *Safety and Reliability*. 2025:1-30. doi: https://doi.org/10.1080/09617353.2025.2594229

[4] Wróblewski A, Krot P, Zimroz R, Mayer T, Peltola J. Review of linear electric motor hammers—an energy-saving and eco-friendly solution in industry. *Energies*. 2023;16(2):959. doi: https://doi.org/10.3390/en16020959

[5] Sheikh MA, Bakhsh ST, Irfan M, Nor NBM, Nowakowski G. A review to diagnose faults related to three-phase industrial induction motors. *Journal of Failure Analysis and Prevention*. 2022;22(4):1546-1557. doi: https://doi.org/10.1007/s11668-022-01445-2

[6] Lokhande R, Mishra SK, Ronanki D, Shakya P, Edachery V, Koottaparambil L. Review on tribological and vibration aspects in mechanical bearings of electric vehicles: effect of bearing current, shaft voltage, and electric discharge material spalling current. *Lubricants*. 2025;13(8):349. doi: https://doi.org/10.3390/lubricants13080349

[7] Garcia-Calva T, Morinigo-Sotelo D, Fernandez-Cavero V, Romero-Troncoso R. Early detection of faults in induction motors—a review. *Energies*. 2022;15(21):7855. doi: https://doi.org/10.3390/en15217855

[8] Habyarimana M, Adebiyi AA. A review of artificial intelligence applications in predicting faults in electrical machines. *Energies*. 2025;18(7):1616. doi: https://doi.org/10.3390/en18071616

[9] Akbar S, Vaimann T, Asad B, Kallaste A, Sardar MU, Kudelina K. State-of-the-art techniques for fault diagnosis in electrical machines: advancements and future directions. *Energies*. 2023;16(17):6345. doi: https://doi.org/10.3390/en16176345

[10] Yousuf M, Alsuwian T, Amin AA, Fareed S, Hamza M. IoT-based health monitoring and fault detection of industrial AC induction motor for efficient predictive maintenance. *Measurement and Control*. 2024;57(8):1146-1160. doi: https://doi.org/10.1177/0020294024123147

[11] Kumar RR, Andriollo M, Cirrincione G, Cirrincione M, Tortella A. A comprehensive review of conventional and intelligence-based approaches for the fault diagnosis and condition monitoring of induction motors. *Energies*. 2022;15(23):8938. doi: https://doi.org/10.3390/en15238938

[12] Shah R, Mittal V, Lotwin M. Recent advances in vibration analysis for predictive maintenance of modern automotive powertrains. *Vibration*. 2025;8(4):68. doi: https://doi.org/10.3390/vibration8040068

[13] Hamani K, Kuchar M, Kubatko M, Kirschner S. Advancements in induction motor fault diagnosis and condition monitoring: a comprehensive review. *Sensors*. 2025;25(19):5942. doi: https://doi.org/10.3390/s25195942

[14] Zong Z, Guan Y. AI-driven intelligent data analytics and predictive analysis in Industry 4.0: transforming knowledge, innovation, and efficiency. *Journal of the Knowledge Economy*. 2025;16(1):864-903. doi: https://doi.org/10.1007/s13132-024-02001-z

[15] Yuan S-M, Hong Z-W, Cheng W-K. Artificial intelligence and deep learning in sensors and applications. *Sensors*. 2024;24(10):3258. doi: https://doi.org/10.3390/s24103258

[16] Villegas-Ch W, Gaibor-Naranjo W, Sanchez-Viteri S. Application of deep learning techniques for the optimization of industrial processes through the fusion of sensory data. *International Journal of Computational Intelligence Systems*. 2024;17(1):187. doi: https://doi.org/10.1007/s44196-024-00596-4

[17] Wahid A, Breslin JG, Intizar MA. Prediction of machine failure in Industry 4.0: a hybrid CNN-LSTM framework. *Applied Sciences*. 2022;12(9):4221. doi: https://doi.org/10.3390/app12094221

[18] Vignesh RG, Fancy C. Predictive intelligence: a hybrid CNN-LSTM architecture with explainable AI for industrial machine failure prediction. In: 2026 4th International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA); 2026. p. 1-7. Institute of Electrical and Electronics Engineers. doi: https://doi.org/10.1109/AIMLA67915.2026.11522343
[19] Kirchgässner W, Wallscheid O, Böcker J. Estimating electric motor temperatures with deep residual machine learning. *Institute of Electrical and Electronics Engineers Transactions on Power Electronics*. 2021;36(7):7480-7488. doi: https://doi.org/10.1109/TPEL.2020.3045596
[20] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13–17; San Francisco, CA, USA. New York (NY): Association for Computing Machinery; 2016. p. 785-794. doi: https://doi.org/10.1145/2939672.2939785
[21] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems. 2017;30. Available from: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
[22] Lei Y, et al. Applications of machine learning to machine fault diagnosis: a review and roadmap. *Mechanical Systems and Signal Processing*. 2020;138:106587.
[23] Sehri M, Dumond P. University of Ottawa electric motor dataset – vibration and acoustic faults under constant and variable speed conditions (UOEMD-VAFCVS) [dataset]. Mendeley Data. 2026;2. doi: https://doi.org/10.17632/msxs4vj48g.2
[24] EVIoT-PredictiveMaint Dataset [dataset on the Internet]. Kaggle; 2026 [cited 2026 Apr 10]. Available from: https://www.kaggle.com/datasets/datasetengineer/eviot-predictivemaint-dataset/data
[25] Bousdekis A, Lepenioti K, Apostolou D, Mentzas G. A review of data-driven decision-making methods for Industry 4.0 maintenance applications. *Electronics*. 2021;10(7):828. doi: https://doi.org/10.3390/electronics10070828
[26] Garcia J, Rios-Colque L, Peña A, Rojas L. Condition monitoring and predictive maintenance in industrial equipment: an NLP-assisted review of signal processing, hybrid models, and implementation challenges. *Applied Sciences*. 2025;15(10):5465. doi: https://doi.org/10.3390/app15105465
[27] Haque BT, Rahman MA. Data preparation and feature engineering in high-dimensional predictive modeling. *Journal of Sustainable Development and Policy*. 2024;3(04):205-244. doi: https://doi.org/10.63125/5txtr530
[28] Miah MR, Islam MM. Data preprocessing and feature engineering strategies for large-scale predictive modeling applications. *Review of Applied Science and Technology*. 2024;3(01):263-302. doi: https://doi.org/10.63125/tqqqed47
[29] Umar M, Siddique MF, Kim J, Kim JM. A distribution-level statistical framework for reliable pipeline leak detection using multi-domain signal analysis. *Scientific Reports*. 2026. doi: https://doi.org/10.1038/s41598-026-56335-6
[30] Zimmermann M, Dostert K. Analysis and modeling of impulsive noise in broad-band powerline communications. *Institute of Electrical and Electronics Engineers Transactions on Electromagnetic Compatibility*. 2002;44(1):249-258. doi: https://doi.org/10.1109/15.990732
[31] Wang Y, Jiang H, Tong B, Song S. Rolling bearing fault diagnosis via Meta-BOHB optimized CNN–Transformer model and time-frequency domain analysis. *Sensors*. 2025;25(22):6920. doi: https://doi.org/10.3390/s25226920
[32] Hu B, Zhu J, Liang S, Ting L. Multi-domain representation learning for bearing fault diagnosis with phase and transient preservation. *Applied Sciences*. 2026;16(4):1846. doi: https://doi.org/10.3390/app16041846
[33] Yang Z. Systematic analysis on power spectral density estimation and its practical application. In: American Institute of Physics Conference Proceedings. Vol. 3435, No. 1; 2026. p. 020010. American Institute of Physics Publishing LLC. doi: https://doi.org/10.1063/5.0331080
[34] Duff EP, Johnston LA, Xiong J, Fox PT, Mareels I, Egan GF. The power of spectral density analysis for mapping endogenous BOLD signal fluctuations. *Human Brain Mapping*. 2008;29(7):778-790. doi: https://doi.org/10.1002/hbm.20601
[35] Nyangon J. Engineering hybrid physics-informed neural networks for next-generation electricity systems: a state-of-the-art review. arXiv preprint arXiv:2605.21903. 2026. doi: https://doi.org/10.48550/arXiv.2605.21903
[36] Du Z, Lu R. Physics-informed neural networks for advanced thermal management in electronics and battery systems: a review of recent developments and future prospects. *Batteries*. 2025;11(6):204. doi: https://doi.org/10.3390/batteries11060204
[37] Svetnik V, Wang T, Tong C, Liaw A, Sheridan RP, Song Q. Boosting: an ensemble learning tool for compound classification and QSAR modeling. *Journal of Chemical Information and Modeling*. 2005;45(3):786-799. doi: https://doi.org/10.1021/ci0500379

[38] Abimannan S, El-Alfy ESM, Chang YS, Hussain S, Shukla S, Satheesh D. Ensemble multifeatured deep learning models and applications: a survey. *Institute of Electrical and Electronics Engineers Access*. 2023;11:107194-107217. doi: https://doi.org/10.1109/ACCESS.2023.3320042
[39] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.
[40] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020;17(3):261-272. doi: https://doi.org/10.1038/s41592-019-0686-2
[41] Huang J, Tong J. Rolling bearing acoustic-vibration fusion fault diagnosis based on heterogeneous modal perception and knowledge distillation. *Electronics*. 2026;15(8):1631. doi: https://doi.org/10.3390/electronics15081631
[42] Garg S, et al. An overlapping sliding window and combined features based emotion recognition system for EEG signals. *Applied Computing and Informatics*. 2025;21(1-2):114-130. doi: https://doi.org/10.1108/ACI-05-2021-0130
[43] Sehri M, et al. Temporal domain vibration fault diagnosis of drone blades via selective embedding. *Machines*. 2026;14(2):241. doi: https://doi.org/10.3390/machines14020241
[44] Zeng M, et al. Research and application of a model selection forecasting system for wind speed and theoretical power generation. *Future Internet*. 2025;18(1):7. doi: https://doi.org/10.3390/fi18010007
[45] Xia Y, et al. Simulation-guided interpretable fault diagnosis of hydraulic directional control valves under limited fault data conditions. *Sensors*. 2026;26(7):2052. doi: https://doi.org/10.3390/s26072052
[46] Yadav OP, Pahuja GL. Bearing health assessment using time domain analysis of vibration signal. *International Journal of Image, Graphics and Signal Processing*. 2020;11(3):27. doi: https://doi.org/10.5815/ijigsp.2020.03.04
[47] Dua X, et al. Research on Bode data processing method based on Python. In: 2024 Institute of Electrical and Electronics Engineers 2nd International Conference on Control, Electronics and Computer Technology (ICCECT); 2024. Institute of Electrical and Electronics Engineers. doi: https://doi.org/10.1109/ICCECT60629.2024
[48] Giri MK, Majumder S. On eigenvalue-based cooperative spectrum sensing using feature extraction and maximum entropy fuzzy clustering. *Journal of Ambient Intelligence and Humanized Computing*. 2023;14(8):10053-10067. doi: https://doi.org/10.1007/s12652-021-03670-3
[49] Ameera A, Saidatul A, Ibrahim Z. Analysis of EEG spectrum bands using power spectral density for pleasure and displeasure state. *Institute of Physics Conference Series: Materials Science and Engineering*. 2019;557(1):012030. doi: https://doi.org/10.1088/1757-899X/557/1/012030
[50] Deng Y, et al. Physics-informed learning in artificial electromagnetic materials. *Applied Physics Reviews*. 2025;12(1). doi: https://doi.org/10.1063/5.0232675
[51] Yu W, et al. Comparative analysis of AC copper loss with round copper wire and flat copper wire of high-speed stator-PM flux-switching machine. *Institute of Electrical and Electronics Engineers Transactions on Industry Applications*. 2022;58(6):7131-7142. doi: https://doi.org/10.1109/TIA.2022.3194042
[52] Zhou A, et al. Rotor temperature safety prediction method of PMSM for electric vehicle on real-time energy equivalence. *Mathematical Problems in Engineering*. 2020;2020(1):3213052. doi: https://doi.org/10.1155/2020/3213052
[53] Salsabila A, Sari AP, Junaidi A. Predicting CO levels using LSTM and rolling-features. *Journal of Artificial Intelligence and Engineering Applications*. 2025;4(3):2178-2183. doi: https://doi.org/10.59934/jaiea.v4i3.1126
[54] Hou Y, et al. Evidential domain adaptation for remaining useful life prediction with incomplete degradation. *Institute of Electrical and Electronics Engineers Transactions on Instrumentation and Measurement*. 2025. doi: https://doi.org/10.1109/TIM.2025.3551977
[55] Robindro K, et al. Hybrid distributed feature selection using particle swarm optimization-mutual information. *Data Science and Management*. 2024;7(1):64-73. doi: https://doi.org/10.1016/j.dsm.2023.10.003
[56] Theng D, Bhoyar KK. Feature selection techniques for machine learning: a survey of more than two decades of research. *Knowledge and Information Systems*. 2024;66(3):1575-1637. doi: https://doi.org/10.1007/s10115-023-02010-5
[57] Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32. doi: https://doi.org/10.1023/A:1010933404324
[58] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20(3):273-297. doi: https://doi.org/10.1007/BF00994018
[59] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*. 1997;9(8):1735-1780. doi: https://doi.org/10.1162/neco.1997.9.8.1735
[60] Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *Institute of Electrical and Electronics Engineers Transactions on Audio and Electroacoustics*. 1967;15(2):70-73. doi: https://doi.org/10.1109/TAU.1967.1161901

[61] Testas A. Random forest classification with scikit-learn and PySpark. In: Distributed Machine Learning with PySpark: Migrating Effortlessly from Pandas and Scikit-Learn; 2023. p. 243-258. Berkeley, CA: Apress. doi: https://doi.org/10.1007/978-1-4842-9751-3_9

[62] Rane N, Choudhary SP, Rane J. Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions. *Studies in Medical and Health Sciences*. 2024;1(2):18-41. doi: https://doi.org/10.48185/smhs.v1i2.1225

[63] Zhou L, Zhu Q, Chen Q, Wang P, Huang H. Predicting hospital outpatient volume using XGBoost: a machine learning approach. *Scientific Reports*. 2025;15(1):17028. doi: https://doi.org/10.1038/s41598-025-01265-y

[64] Moniruzzaman M, Okilly AH, Choi S, Baek J, Mannan TI, Islam Z. A comprehensive study of machine learning algorithms for GPU based real-time monitoring and lifetime prediction of IGBTs. In: 2024 Institute of Electrical and Electronics Engineers Applied Power Electronics Conference and Exposition (APEC); 2024. p. 2678-2684. Institute of Electrical and Electronics Engineers. doi: https://doi.org/10.1109/APEC48139.2024.10509167

[65] Lu B, Huang H, Wu Z, Zhang T, Gu Y, Wang F, Shu Z. Utilizing LightGBM to explore the characterization of PM2.5 emission patterns from broadleaf tree combustion in Northeastern China. *Forests*. 2025;16(5):836. doi: https://doi.org/10.3390/f16050836

[66] Zhang S, Duan L, Jiang H. Accelerating Gaussian beam tracing method with dynamic parallelism on graphics processing units. *Computer Physics Communications*. 2025;315:109722. doi: https://doi.org/10.1016/j.cpc.2025.109722

[67] Sekar PC, Shnain AH, Rani MJ, Durgadevi G, Chethana R. A support vector machine with elastic net regularization and radial basis function based spectrum sensing for cognitive radio networks. In: 2024 First International Conference on Software, Systems and Information Technology (SSITCON); 2024. p. 1-5. Institute of Electrical and Electronics Engineers. doi: https://doi.org/10.1109/SSITCON62437.2024.10796356

[68] Zhang Y, Peng Y, Sun J, Gui G, Lin Y, Mao S. GPU-free specific emitter identification using signal feature embedded broad learning. *Institute of Electrical and Electronics Engineers Internet of Things Journal*. 2023;10(14):13028-13039. doi: https://doi.org/10.1109/JIOT.2023.3257479

[69] Wielgosz M, Skoczeń A, Mertik M. Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2017;867:40-50. doi: https://doi.org/10.1016/j.nima.2017.06.020

[70] Pakki KR, Repala ESC, Eluri DB. Performance review of deep learning models on different GPU systems. In: 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI); 2025. p. 1151-1158. Institute of Electrical and Electronics Engineers. doi: https://doi.org/10.1109/ICMSCI62561.2025.10894456

[71] Sinha BB, Ahsan M, Dhanalakshmi R. LightGBM empowered by whale optimization for thyroid disease detection. *International Journal of Information Technology*. 2023;15(4):2053-2062. doi: https://doi.org/10.1007/s41870-023-01261-3

[72] Yuniawan F, Mandala S. LightGBM-based ensemble boosting for detection and anatomical localization of myocardial infarction using 12-lead ECG morphological features. In: 2025 Institute of Electrical and Electronics Engineers International Symposium on Future Telecommunication Technologies (SOFTT); 2025. p. 105-111. Institute of Electrical and Electronics Engineers. doi: https://doi.org/10.1109/SOFTT67007.2025.11213071

[73] Donepudi S, Nakka R, Thota KK, Ajmeera M, Praveen SP, Sindhura S. Enhancing person-centric health care for diabetes prediction: a comparative study of LightGBM, XGBoost, and hybrid LIGB model. In: Enabling Person-Centric Healthcare Using Ambient Assistive Technology, Volume 2: Personalized and Patient-Centric Healthcare Services in AAT; 2025. p. 127-155. Cham: Springer Nature Switzerland. doi: https://doi.org/10.1007/978-3-031-82516-3_7

[74] Sow TH, Mehdi A. Enhancing IDS performance through a comparative analysis of random forest, XGBoost, and deep neural networks. *Machine Learning with Applications*. 2025;22:100738. doi: https://doi.org/10.1016/j.mlwa.2025.100738

[75] Sun Y, Wang H, Xue B, Jin Y, Yen GG, Zhang M. Surrogate-assisted evolutionary deep learning using an end-to-end random forest-based performance predictor. *Institute of Electrical and Electronics Engineers Transactions on Evolutionary Computation*. 2019;24(2):350-364. doi: https://doi.org/10.1109/TEVC.2019.2924461

[76] Wen Z, Shi J, He B, Chen J, Chen Y. Efficient multi-class probabilistic SVMs on GPUs. *Institute of Electrical and Electronics Engineers Transactions on Knowledge and Data Engineering*. 2018;31(9):1693-1706. doi: https://doi.org/10.1109/TKDE.2018.2866097

[77] Omer SZ, Hashim F, Sali A, Ahmad FA. Binary classification of low-rate DoS attacks using long short-term memory feed-forward (LSTM-FF) intrusion detection system (IDS). *Engineering Science and Technology, an International Journal*. 2025;66:102049. doi: https://doi.org/10.1016/j.jestch.2025.102049

[78] Dai Z, Huang W. Improving energy management practices through accurate building energy consumption prediction: analyzing the performance of LightGBM, RF, and XGBoost models with advanced optimization strategies. *Electrical Engineering*. 2025;107(9):12583-12605. doi: https://doi.org/10.1007/s00202-025-03167-8
[79] Shehadeh A, Alshboul O, Al Mamlook RE, Hamedat O. Machine learning models for predicting the residual value of heavy construction equipment: an evaluation of modified decision tree, LightGBM, and XGBoost regression. *Automation in Construction*. 2021;129:103827. doi: https://doi.org/10.1016/j.autcon.2021.103827
[80] Kilinc HC, Apak S, Ozkan F, Ergin ME, Yurtsever A. Multimodal fusion of optimized GRU–LSTM with self-attention layer for hydrological time series forecasting. *Water Resources Management*. 2024;38(15):6045-6062. doi: https://doi.org/10.1007/s11269-024-03943-4
[81] Kocer A. Prediction of the higher heating values of biomass using machine learning methods based on proximate and ultimate analysis. *Journal of Mechanical Science and Technology*. 2024;38(3):1569-1574. doi: https://doi.org/10.1007/s12206-024-0247-1
[82] Xie J, Hu K, Li G, Guo Y. CNN-based driving maneuver classification using multi-sliding window fusion. *Expert Systems with Applications*. 2021;169:114442. doi: https://doi.org/10.1016/j.eswa.2020.114442
[83] Hossain RB, Hasan MMA, Khan MI, Ahmed M, Lin Y, Pan X. A physics-informed hybrid ensemble for robust and high-fidelity temperature forecasting in PMSMs. *World Electric Vehicle Journal*. 2026;17(3):133. doi: https://doi.org/10.3390/wevj17030133
[84] Zhang H, Li W, Wang G, Song F, Wen Z, Zhang H, et al. Predicting stomatal conductance of chili peppers using TPE-optimized LightGBM and SHAP feature analysis based on UAVs' hyperspectral, thermal infrared imagery, and meteorological data. *Computers and Electronics in Agriculture*. 2025;231:110036. doi: https://doi.org/10.1016/j.compag.2025.110036
[85] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*. 2019;378:686-707. doi: https://doi.org/10.1016/j.jcp.2018.10.045
[86] Sehri M, Dumond P. University of Ottawa constant and variable speed electric motor vibration and acoustic fault signature dataset. *Data in Brief*. 2024;53:110144. doi: https://doi.org/10.1016/j.dib.2024.110144
[87] Demšar J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*. 2006;7:1-30.
[88] Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*. 1937;32(200):675-701. doi: https://doi.org/10.1080/01621459.1937.10503522
[89] Nemenyi PB. Distribution-free multiple comparisons [PhD thesis]. Princeton (NJ): Princeton University; 1963.
[90] Kendall MG, Babington Smith B. The problem of m rankings. *Annals of Mathematical Statistics*. 1939;10(3):275-287. doi: https://doi.org/10.1214/aoms/1177732186
[91] Cooley JW, Tukey JW. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*. 1965;19(90):297-301. doi: https://doi.org/10.1090/S0025-5718-1965-0178586-1
[92] Saeys Y, Abeel T, Van de Peer Y. Robust feature selection using ensemble feature selection techniques. In: Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2008). Lecture Notes in Computer Science, vol 5212. Berlin: Springer; 2008. p. 313-325. doi: https://doi.org/10.1007/978-3-540-87481-2_21
[93] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Physical Review E*. 2004;69(6):066138. doi: https://doi.org/10.1103/PhysRevE.69.066138
[94] Kaufman S, Rosset S, Perlich C, Stitelman O. Leakage in data mining: formulation, detection, and avoidance. *Association for Computing Machinery Transactions on Knowledge Discovery from Data*. 2012;6(4):15. doi: https://doi.org/10.1145/2382577.2382579
[95] Zhang W, Peng G, Li C, Chen Y, Zhang Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*. 2017;17(2):425. doi: https://doi.org/10.3390/s17020425
[96] Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? In: Advances in Neural Information Processing Systems 35 (NeurIPS 2022). 2022. p. 507-520.
[97] Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*. 2023;4(9):100804. doi: https://doi.org/10.1016/j.patter.2023.100804
[98] Kirchgässner W, Wallscheid O, Böcker J. Data-driven permanent magnet temperature estimation in synchronous motors with supervised machine learning: a benchmark. *Institute of Electrical and Electronics Engineers Transactions on Energy Conversion*. 2021;36(3):2059-2067. doi: https://doi.org/10.1109/TEC.2021.3052546