IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 June 2022, accepted 27 July 2022, date of publication 1 August 2022, date of current version 9 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3195939

RESEARCH ARTICLE

LightGBM-Based Fault Diagnosis of Rotating Machinery Under Changing Working Conditions Using Modified Recursive Feature Elimination

ALIREZA NEMAT SABERI[1,2], (Graduate Student Member, IEEE), ANOUAR BELAHCEN[1,3], (Senior Member, IEEE), JAN SOBRA[4], AND TOOMAS VAIMANN[3], (Senior Member, IEEE)

[1]Department of Electrical Engineering and Automation, Aalto University, 00076 Aalto, Finland
[2]ABB System Drives, 00380 Helsinki, Finland
[3]Department of Electrical Power Engineering and Mechatronics, Tallinn University of Technology, 19086 Tallinn, Estonia
[4]Faculty of Electrical Engineering, University of West Bohemia, 30100 Pilsen 3, Czech Republic

Corresponding author: Alireza Nemat Saberi (alireza.nematsaberi@aalto.fi)

This work was supported in part by the Academy of Finland Consortium under Grant 330747.

ABSTRACT This article presents an intelligent and accurate framework for fault diagnosis of induction motors using light gradient boosting machine (LightGBM). The proposed framework offers promising generalization ability when the testing data contains new unseen operating conditions unavailable during the training process. After the acquisition of vibration signals and feature extraction in multiple domains, we perform an iterative feature selection (FS) approach by utilizing a modified version of recursive feature elimination (RFE) and the features' importance scores obtained by LightGBM. To prevent overfitting and subsequent selection bias, an outer resampling loop encompasses the whole process of our RFE-LightGBM algorithm. Moreover, instead of the conventional resampling methods based on K-fold cross-validation (CV) or leave-one-out CV (LOOCV), we use a new scheme called leave-one-loading-out CV (*LOLO-CV*). Leveraging *LOLO-CV*, the proposed FS method identifies the optimal feature subset, making the fault diagnosis robust under changing operating conditions. Then, the final classification is performed with optimal feature subset by training a new LightGBM model with adjusted hyperparameters employing Bayesian optimization. Experimental results from two real case studies show that our proposed fault diagnosis framework achieves accuracies between 98.55% and 100% for various testing scenarios. For example, for the worst-case testing scenario in the bearing dataset of Case Western Reserve University where the no-load data (0hp) is absent during the training process and is only used for testing, the testing accuracy of LightGBM classifier before and after applying the proposed RFE-LightGBM-FS method is 88.04% to 97.23%, respectively. Using the Bayesian hyperparameter optimization further improves the accuracy to 98.55%.

INDEX TERMS Electrical machines, bearings, fault diagnosis, feature importance, gradient boosting, hyperparameter optimization, LightGBM, machine learning.

I. INTRODUCTION

Condition monitoring and fault diagnosis of induction motors (IMs) at incipient stages are critical to decrease maintenance costs and tremendous financial losses and avoid the long-term shutdown of industrial operations [1], [2]. The methods developed for the condition monitoring of IMs can be classified based on the various signals being recorded and analyzed, e.g., motor vibration, current, acoustic emission, temperature, and pressure [3]–[5]. Among them, vibration signal analysis is one of the most reliable, accurate, and standard methods widely utilized in machinery fault diagnosis [6]–[9].

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh.

Recently, intelligent data-driven fault diagnosis techniques based on either shallow machine learning (SML) or deep learning (DL) have become emerging in industry and academia [10], [11].

In the SML-based methods, after data acquisition, several handcrafted features must be extracted using signal-processing methods. Then the selected features are used in the final step (fault classification) to train various SML models such as support vector machine (SVM) [12], Naive Bayes [13], K-nearest neighbor (KNN) [14], ensemble boosted trees [15], etc.

The DL-based fault diagnosis employs deep architectures of neural networks (NNs) with many layers. They enable automatic extraction of abstract and informative features from the raw input signals and eliminate manual feature extraction requirements [16]. Although the classification accuracy of DL-based motor fault diagnosis methods is promising [17]–[19], using black-box models with few physical perspectives, they cannot provide enough interpretation about how the system operates well. Moreover, training deep NNs requires massive data, time, and computational resources, which is not practical in all industrial sectors [20], [21].

On the other hand, SML models such as SVM have simpler structures and can be trained with fewer computational requirements [22]–[24]. However, one of the biggest challenges in SML-based methods is dealing with the high dimensionality of the handcrafted features extracted from multiple domains, which deteriorates the classification accuracy due to overfitting [25]–[27]. Light Gradient Boosting Machine (LightGBM) is a recently-developed framework based on gradient boosting decision tree (GBDT) [28]. Besides its promising classification performance, LightGBM can rank and evaluate the feature scores. Feature scores define the total contribution of each feature to the splitting process of multiple decision tree learners. The calculation of each feature's score is based on the total gain of splits which use that feature in the structure of decision trees. Hence, using LightGBM as the classifier, it is also possible to implement an embedded feature selection (FS) and remove the redundant features that are not discriminative.

Another critical challenge in intelligent IM fault diagnosis methods is the lack of available data for all operational conditions. Few works in the literature have studied the more realistic and practical scenarios where the training and testing data originated from different working conditions, i.e., different loading levels or rotational speeds [29]. Highlighting this issue, Gangsar *et al.* [27] showed that when there is a difference between training and testing data working conditions, the average prediction accuracy of a multi-sensor SVM-based fault diagnosis system decreases significantly. It was shown in [27], [30] that the reason for this performance degradation is that variational working conditions lead to a distribution discrepancy between the training and testing feature sets, decreasing the generalization ability of the fault diagnosis methods. Stief *et al.* [31] utilized principal component analysis (PCA) to reduce the dependency of the extracted features on the loading levels and combined it with a two-stage Bayesian method to improve the performance generalization. Although the technique was reported to be effective in the accurate classification of different types of faults, it could not accurately discriminate the faults' severity. Another powerful approach for solving the distribution discrepancy of features is applying domain adaptation and transfer learning methods proposed in [30], [32]–[34]. However, the limitation of these methods is that they necessitate the availability of unlabeled data for the new working conditions. Satisfying this requirement is not practical in those industrial cases where neither labeled data nor unlabeled data are available for the motor's new working conditions.

Considering the above challenges of intelligent fault diagnosis, in this article, we propose an integrated SML-based framework using LightGBM to enhance the classification performance in terms of prediction accuracy and generalization ability. We used the vibration signals as the inputs of our fault diagnosis framework and extracted features in time domain, frequency domain, and time-frequency domain from the vibration signals obtained by the accelerometers. Thanks to the innovative techniques employed in LightGBM, such as gradient-based one-side sampling (GOSS), leaf-wise tree growth strategy, and histogram-based split finding [28], the proposed fault diagnosis method obtains high accuracy and efficient performance. We utilize the feature ranking capability of LightGBM to sort the features according to their importance. Then, we combine the feature scores with a modified recursive feature elimination (RFE) approach to select the best feature subset. In the proposed RFE-LightGBM-FS, an outer resampling loop encloses the whole process to avoid overfitting the training data and the selection bias. In addition, instead of the conventional resampling methods based on K-fold cross-validation (CV) or leave-one-out CV (LOOCV), we introduce a new resampling scheme called **leave-one-loading-out CV (*LOLO-CV*)** to increase the algorithm's generalization when test data contains new unseen loading conditions. The K-fold CV method split the original dataset randomly into K training and validation sets. LOOCV is also a special case of K-fold CV, where K is the number of observations [35], [36]. Therefore, in conventional resampling methods, the data from all the available IM operating conditions exist in all training sets, and there is high overlap between training and validation sets in terms of operating conditions. In other words, the conventional resampling methods select the training and validation sets without paying attention to the operating conditions under which the data is recorded. The reason is that domain knowledge is not used in the conventional resampling methods. On the other hand, *LOLO-CV* utilizes domain knowledge in dividing the data into training and validation sets. At each resampling iteration of the proposed FS strategy using *LOLO-CV*, all the data samples belonging to a specific loading level are excluded from the original training set and create a virtual validation set. Therefore, each resampling iteration contains

non-overlapping training and validation sets in terms of loading level. The proposed FS strategy improves the robustness of selected features against new unseen working conditions by efficiently utilizing the information in the different working conditions existing in the original training data. The main contributions of our work are described in the following points:

1) A new integrated fault diagnosis framework is presented, which first uses a basic default LightGBM model for obtaining the feature scores. In the next step, a modified RFE containing an outer resampling loop with a new *LOLO-CV* scheme is linked to the basic LightGBM. The *LOLO-CV* scheme efficiently leverages the information of multiple loading levels existing in the training set. It makes RFE identify an optimal feature subset that is robust to new unseen loading levels. Using RFE with *LOLO-CV* also leads to omitting the redundant and uninformative features. Finally, a new optimized LightGBM model is employed to perform the final classification task.
2) To improve the classification accuracy, we use Bayesian optimization to determine the most critical hyperparameters of the final LightGBM model.
3) To evaluate the model's robustness to limited data conditions, we build the testing sets using the data at specific loading conditions which is not included in the training data. The loading condition specified for the test data is neither used in FS nor hyperparameter optimization and is only for final model evaluation.
4) Two real case studies are used to validate the performance of the proposed algorithm. Experiments under different loading conditions prove the effectiveness and superiority of the proposed framework compared with the traditional fault diagnosis methods and related works.

The remainder of this paper begins with the theoretical knowledge of LightGBM in Section II. Section III presents the proposed fault diagnosis framework including the feature extraction in multiple domains and the proposed RFE-LightGBM FS approach. Section IV focuses on experimental results and evaluating the proposed method on two real case studies. The conclusion is given in Section V.

II. THEORETICAL FOUNDATION OF LIGHTGBM

In this section, the main principles and advantages of the LightGBM algorithm are clarified. GBDT is an iterative ensemble model achieving the final strong classification results by combining multiple base learners (i.e., weak decision trees) [37]. To improve the performance of the traditional GBDT, Chen *et al.* [38] presented XGBoost framework that supports parallel learning by CPU multi-threading, adds a regularization term to the loss function to deal with the overfitting, and applies the second-order Taylor approximation in optimizing the objective function.

Having the advantages of XGBoost, LightGBM [28] is a newer enhanced implementation of GBDT. One of the defects of XGBoost is using a **level-wise tree growth strategy** in which many nodes obtain low splitting gains and increase the computations without improving the accuracy. LightGBM solves this problem by adopting a **leaf-wise method**being faster and more accurate. The leaf-wise method detects the node with the highest gain at each layer and only splits that node, growing asymmetrical and deeper trees.

Moreover, LightGBM employs other innovative strategies that distinguish its performance from XGBOOST, such as **GOSS and histogram-based algorithm** of finding the best split points that are fully described in [28], [39].

Considering the training dataset with M instances and p features $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M}, (\mathbf{x}_i \in \mathbb{R}^p, y_i \in \mathbb{R})$, the predicted output of LightGBM model for the i-th sample, $\hat{y}_i$ is the combination of multiple weak decision trees as follows:

$$\hat{y}_i = \sum_{n=1}^{N} f_n\left(\mathbf{x}_i; \{R_{ln}\}_1^L\right), \tag{1}$$

where N is the total number of trees (i.e., number of iterations), and f_n is an L-leaf node (terminal node) decision tree at the n-th iteration that splits the feature space into L non-overlapping regions $\{R_{ln}\}_{l=1}^{L}$. The region R_{ln} represents the subset of feature space corresponding to the leaf node l in the n-th tree. Equation (2) defines f_n as follows:

$$f_n\left(\mathbf{x}; \{R_{ln}\}_1^L\right) = \sum_{l=1}^{L} \beta_{ln} 1\{\mathbf{x} \in R_{ln}\}, \tag{2}$$

where β_{ln} is the predicted score associated with the l-th leaf node, and $1\{\cdot\}$ is the indicator function that outputs 1 if the condition is true and 0 otherwise.

LightGBM trains the trees in an additive process. Let $\hat{y}_i^{(k)}$ be the predicted output of the i-th sample at the k-th iteration. The objective function in the k-th iteration is defined as

$$J^{(k)} = \sum_{i=1}^{M} loss\left(y_i, \hat{y}_i^{(k-1)} + f_k\left(\mathbf{x}_i; \{R_{lk}\}_1^L\right)\right) + \Omega\left(f_k\right) \tag{3}$$

$$\Omega\left(f_k\right) = \alpha L + \frac{1}{2}\lambda \sum\nolimits_{l=1}^{L} \beta_{lk}^2. \tag{4}$$

The term "*loss*" in (3) is the multi-class logistic loss function for classification problems [37]. The second part is the regularization term Ω that prevents the number of leaf nodes L and the leaf node scores $\{\beta_{lk}\}_{l=1}^{L}$ from increasing. α and λ are the corresponding tuning parameters. By using the second-order Taylor expansion and omitting the constant terms, (3) can be approximated as

$$J^{(k)} \approx \sum_{i=1}^{M}\left(g_{ik} f_k\left(\mathbf{x}_i; \{R_{lk}\}_1^L\right) + \frac{1}{2} h_{ik} f_k^2\left(\mathbf{x}_i; \{R_{lk}\}_1^L\right)\right) + \Omega\left(f_k\right), \tag{5}$$

where g_{ik} and h_{ik} are the first-order (gradient) and second-order (hessian) derivatives of loss function. At each iteration of the training, the optimal L-leaf node tree f_k^* must be found that minimizes $J^{(k)}$. Therefore, for each tree, the non-overlapping regions $\{R_{lk}\}_{l=1}^{L}$ and optimal leaf node scores $\left\{\beta_{lk}^*\right\}_{l=1}^{L}$ must be obtained. LightGBM adopts a two-step procedure to carry out this purpose. It firstly fits a regression tree on the pseudo-residuals $\widetilde{res}_{ik}$ of the previous tree (i.e., the negative first derivatives of loss function) as follows:

$$\widetilde{res}_{ik} = -g_{ik} = \partial_{\hat{y}_i^{(k-1)}} loss\left(y_i, \hat{y}_i^{(k-1)}\right), i = 1 \ldots M \tag{6}$$

$$\{R_{lk}\}_{l=1}^{L} = L-\text{leaf node regression tree}\left(\{\widetilde{res}_{ik}, \mathbf{x}_i\}_{i=1}^{M}\right). \tag{7}$$

Therefore, $\{R_{lk}\}_{l=1}^{L}$ is established. By combining (2), (4), (5), and (7), the objective function can be rewritten as

$$J^{(k)} \approx \sum_{l=1}^{L}\left[\left(\sum_{i\in I_{lk}} g_{ik}\right)\beta_{lk} + \frac{1}{2}\left(\sum_{i\in I_{lk}} h_{ik} + \lambda\right)\beta_{lk}^2\right] + \alpha L \tag{8}$$

$$I_{lk} = \{i \mid \mathbf{x}_i \in R_{lk}\}, \tag{9}$$

where I_{lk} is the subset of data instances at the leaf node l of the k-th tree. By setting the derivative of (8) to zero, the optimal score of the l-th leaf node β_{lk}^* and the corresponding minimized objective function $\mathrm{J}^{(k)^*}$ can be achieved as follows:

$$\beta_{lk}^* = -\frac{\sum_{i\in I_{lk}} g_{ik}}{\sum_{i\in I_{lk}} h_{ik} + \lambda} \tag{10}$$

$$J^{(k)^*} = -\frac{1}{2}\sum_{l=1}^{L}\frac{\left(\sum_{i\in I_{lk}} g_{ik}\right)^2}{\sum_{i\in I_{lk}} h_{ik} + \lambda} + \alpha L. \tag{11}$$

Suppose I_R and I_L are the sample subsets of right and left leaf nodes after splitting and $I = I_R \cup I_L$ is the sample subset of the original node. The splitting gain (i.e., the reduction of objective function after the split) is computed as

$$Gain_{split} = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i + \lambda}\right] - \alpha. \tag{12}$$

Higher values of gain are preferable in growing the trees. When splitting a leaf node, the gains associated with the segmentation points of candidate features are evaluated by (12). LightGBM selects the feature showing the maximum gain for splitting. Finally, feature scores are calculated according to the total splitting gain of each feature or the number of times it participated in the splitting process.

III. PROPOSED FAULT DIAGNOSIS METHODOLOGY

This article presents a new fault diagnosis approach offering high accuracy and generalization. Fig. 1 demonstrates the proposed workflow. Initially, vibration signals are collected from the accelerometers under various loading conditions. Before feature extraction, vibration signals are processed and divided into successive equal-length segments resulting in non-overlapping samples. The details of the remaining steps are described as follows.

A. MULTIPLE DOMAIN FEATURE EXTRACTION AND STANDARDIZATION

The performance of the intelligent IM fault diagnosis systems is highly dependent on the information contained in the extracted features. In this study, we calculated a total of 33 statistical features from the time domain, frequency domain, and time-frequency domain analysis of the preprocessed vibration signal samples. The constructed features are standard and commonly used in the previous IM fault diagnosis works, and their detailed description can be found in [6], [7], [31], [40]. Table 1 shows the formulations of the 15 statistical time-domain feature parameters TD_1 -TD_{15} - which are mean value, standard deviation, median value, skewness, kurtosis, square root amplitude, mean absolute deviation, peak to peak value, L1 norm (mean norm), L2 norm (mean-square norm), infinity norm (max norm), crest factor, impulse factor, margin factor, and shape factor, respectively.

The frequency domain features may capture the information that cannot be detected in the time-domain features. After performing fast Fourier transform (FFT) on each vibration signal sample, 12 statistical frequency domain features FD_1-FD_{12} are extracted from each sample's frequency spectrum summarized in Table 1. In the frequency domain, the energy of vibration signals may be reflected in the feature FD_1. The power spectrum convergence can be represented by features FD_2-FD_4, FD_6, and FD_{10}-FD_{12}. The position shift of the main frequencies can be seen in features FD_5 and FD_7-FD_9 [40].

In the time-frequency domain, we implemented discrete wavelet transform (DWT) on the vibration signal samples. The specific Daubechies mother wavelet with four vanishing moments "db4" in five levels is chosen to decompose the signal samples into five detail levels plus approximation level. Six wavelet domain features are computed as the percentage of energies associated with the wavelet coefficients of each decomposed level (WE_1 -WE_6) [6], [41].

LightGBM and other tree-based algorithms are inherently insensitive to the features' scales. However, for the sake of comparison, we also use the case studies of this paper to train other classification algorithms that require feature scaling such as SVM, KNN, etc. We also apply PCA on the dataset for feature visualization that needs feature scaling as well. Therefore, after the feature extraction step, the features are standardized i.e., subtracted by their average value, and divided by their standard deviation, where the average

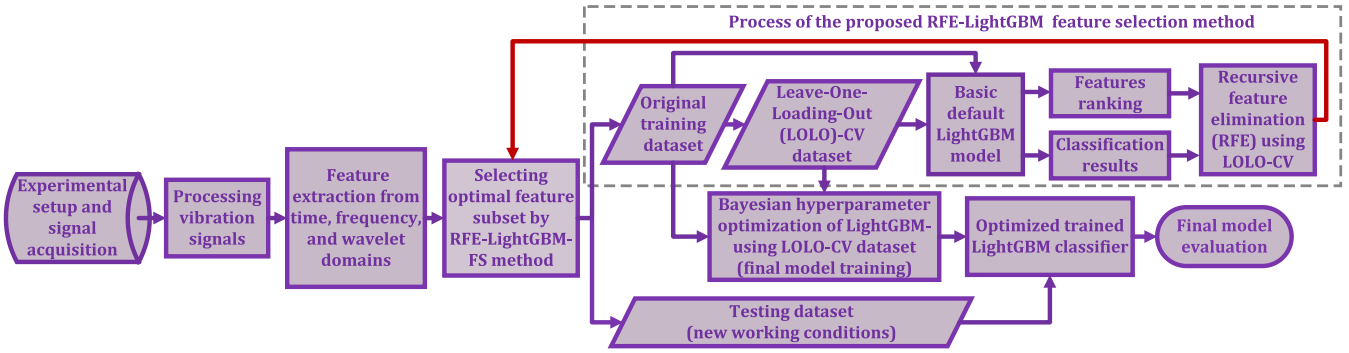

FIGURE 1. Proposed fault diagnosis architecture. (The workflow is represented from left to right from signal acquisition up to final model evaluation. The left part shows the data acquisition, processing, and feature extraction. The blocks located inside the dashed rectangle (upper right of the figure) represent the process of the proposed feature selection method in which only the original training dataset is used to form the LOLO-CV dataset. More details on the proposed RFE-LightGBM-FS are given in Algorithm 1. The red arrow shows the iterative process that connects the result of the RFE-LightGBM-FS to the extracted features to select the optimal feature subset. The lower right part shows the final model training using Bayesian hyperparameter optimization and also final model evaluation using the testing dataset that contains new unseen operating conditions.

TABLE 1. Time domain and frequency domain features.

Time-domain (TD) features

$TD_1 = \boldsymbol{x}_{mean} = \frac{\sum_{i=1}^{N} x_i}{N}$	$TD_7 = \boldsymbol{x}_{mad} = \frac{\sum_{i=1}^{N} \lvert x_i - \boldsymbol{x}_{mean} \rvert}{N}$	$TD_{13} = I_f = \frac{\lVert \boldsymbol{x} \rVert_\infty}{\lVert \boldsymbol{x} \rVert_1 / N}$
$TD_2 = \boldsymbol{x}_{std} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \boldsymbol{x}_{mean})^2}{N-1}}$	$TD_8 = \boldsymbol{x}_{p-p} = \boldsymbol{x}_{max} - \boldsymbol{x}_{min}$	$TD_{14} = m_f = \frac{\lVert \boldsymbol{x} \rVert_\infty}{\boldsymbol{x}_{sra}}$
$TD_3 = \boldsymbol{x}_{med} = \frac{x_{(\lvert (N+1)/2 \rvert)} + x_{(\lvert N/2 \rvert + 1)}}{2}$	$TD_9 = \lVert \boldsymbol{x} \rVert_1 = \sum_{i=1}^{N} \lvert x_i \rvert$	$TD_{15} = s_f = \frac{\lVert \boldsymbol{x} \rVert_2 (\sqrt{N})}{\lVert \boldsymbol{x} \rVert_1}$
$TD_4 = \boldsymbol{x}_{skew} = \frac{\sum_{i=1}^{N} (x_i - \boldsymbol{x}_{mean})^3}{(N-1)\boldsymbol{x}_{std}^3}$	$TD_{10} = \lVert \boldsymbol{x} \rVert_2 = \sqrt{\sum_{i=1}^{N} \lvert x_i \rvert^2}$	
$TD_5 = \boldsymbol{x}_{kurt} = \frac{\sum_{i=1}^{N} (x_i - \boldsymbol{x}_{mean})^4}{(N-1)\boldsymbol{x}_{std}^4}$	$TD_{11} = \lVert \boldsymbol{x} \rVert_\infty = max(\lvert x_i \rvert)$	
$TD_6 = \boldsymbol{x}_{sra} = (\frac{\sum_{i=1}^{N} (\sqrt{\lvert x_i \rvert})}{N})^2$	$TD_{12} = c_f = \frac{\lVert \boldsymbol{x} \rVert_\infty}{\lVert \boldsymbol{x} \rVert_2 / \sqrt{N}}$	

where x_i is the time-domain signal data point for i=1, 2,…,N, and N is the total number of data points.

Frequency-domain (FD) features

$FD_1 = \frac{\sum_{k=1}^{K} s(k)}{K}$	$FD_7 = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)}}$
$FD_2 = \frac{\sum_{k=1}^{K} (s(k) - FD_1)^2}{K-1}$	$FD_8 = \sqrt{\frac{\sum_{k=1}^{K} f_k^4 s(k)}{\sum_{k=1}^{K} f_k^2 s(k)}}$
$FD_3 = \frac{\sum_{k=1}^{K} (s(k) - FD_1)^3}{K(\sqrt{FD_2})^3}$	$FD_9 = \frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sqrt{\sum_{k=1}^{K} s(k) \sum_{k=1}^{K} f_k^4 s(k)}}$
$FD_4 = \frac{\sum_{k=1}^{K} (s(k) - FD_1)^4}{K FD_2^2}$	$FD_{10} = \frac{FD_6}{FD_5}$
$FD_5 = \frac{\sum_{k=1}^{K} f_k s(k)}{s(k)}$	$FD_{11} = \frac{\sum_{k=1}^{K} (f_k - FD_5)^3 s(k)}{K FD_6^3}$
$FD_6 = \frac{\sum_{k=1}^{K} (f_k - FD_5)^2 s(k)}{K}$	$FD_{12} = \frac{\sum_{k=1}^{K} (f_k - FD_5)^4 s(k)}{K FD_6^4}$

where f_k is the frequency value of the kth spectrum line, $S(k)$ is the kth spectrum, K is the number of spectrum lines.

values and standard deviations are computed from the train dataset.

B. RFE-LightGBM-FS USING LOLO-CV

The next step belongs to FS. As stated in section II, LightGBM (and other tree-based algorithms) can output the feature scores. However, selecting the optimal feature subset is another challenge affecting the fault diagnosis performance. This article proposes a straightforward method based on modified RFE to select a reduced feature subset having almost identical distributions under different loading conditions. The proposed FS also alleviates the classification algorithm's computational burden by removing the redundant features.

RFE is an iterative technique that uses the features ranking provided by training a model that can offer feature scores. At each step of RFE, the n least important features (n is user-defined) are iteratively removed from the feature set, and then the model (LightGBM) is retrained using the new reduced feature subset. An evaluation metric such as classification accuracy is iteratively estimated for each rebuilt LightGBM model. RFE finds the optimal feature subset achieving the maximum evaluation metric and selects it for the final model [35].

Algorithm 1 illustrates the process of the proposed RFE-LightGBM-FS. As shown in Fig. 1, the original test set containing new unseen working conditions does not take part in the whole process of our RFE-LightGBM-FS. Thus, Algorithm 1 only uses the original training set D as input, and it is possible to assess the robustness of the proposed method against varying working conditions by the original testing set.

To prevent the selected features from overfitting the training data and the subsequent selection bias in their iterative performance evaluation, an outer resampling at Line 1 encloses the whole FS loop as propounded by [35]. This contrasts with the traditional RFE methods in which the FS

Algorithm 1: Proposed RFE-LightGBM-FS

Inputs and definitions:
M: number of data samples in the original train set;
L: number of loading conditions in the original train set;
S: number of subset sizes being evaluated in the algorithm;
$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M}$ is the entire original train set;
G is an M-dimensional vector indicating each sample i ($i = 1 \ldots M$) belongs to which loading condition k ($k = 1 \ldots L$);

1: **for** $k = 1$ **to** L **do** *//k is the resampling (LOLO-CV) iteration*
2: $I_{train}^{k} \leftarrow \{\emptyset\}$, $I_{val}^{k} \leftarrow \{\emptyset\}$ *//* I_{train}^{k} is the nested train set, I_{val}^{k} is the nested validation set
3: **for** $i = 1$ **to** M **do** *//Partition D to* I_{train}^{k} *and* I_{val}^{k}
4: **if** $G[i] == k$ **then**
5: $I_{val}^{k} \leftarrow I_{val}^{k} \bigcup \{(\mathbf{x}_i, y_i)\}$
6: **else**
7: $I_{train}^{k} \leftarrow I_{train}^{k} \bigcup \{(\mathbf{x}_i, y_i)\}$
8: **end if**
9: **end for**
10: Train a *LightGBM* model on I_{train}^{k} using all features
11: Calculate and sort the feature scores
12: Evaluate the model performance on I_{val}^{k}
13: **for** *Each subset size* S_i $(i=1 \ldots S)$ **do**
14: Select and keep the S_i highest ranked features
15: Train a *LightGBM* model on I_{train}^{k} using S_i features
16: Evaluate the performance on I_{val}^{k} using S_i features
17: **end for**
18: **end for**
19: Obtain the performance profile with respect to different feature subset sizes S_i $(i=1 \ldots S)$ using I_{val}^{k} $(k = 1 \ldots L)$
20: Identify the optimal number of features S_i^*
21: Obtain final feature ranking using original train set D
22: Train the final LightGBM model on entire original train set D using optimal feature subset with top S_i^* features

Output: Optimal feature subset with top S_i^* features

loop is not included within the resampling process leading to information leakage into the selected model and overfitting the training data. Moreover, instead of the conventional K-fold CV or LOOCV resampling methods, we apply the *LOLO-CV* scheme in the proposed FS method. Generally, more than one operating condition is available in the training data for intelligent IM fault diagnosis practical cases. Considering this fact, *LOLO-CV* can fully utilize the information of the varying operating conditions in the training data in a simple way. At each iteration k of *LOLO-CV,* the data samples from the original train set D belonging to the k-th loading condition build a nested virtual validation split while the remaining Samples create a nested virtual train split (Algorithm 1, lines 2-9). Thus, making the non-overlapping nested train and validation sets resembles the testing scenarios containing new unseen working conditions and increases the robustness of the selected features. The number of resampling iterations equals the number of the loading levels L in the original training set. After finishing the L-th resampling iteration, Algorithm 1 averages the performance of each feature subset on L nested validation sets $\boldsymbol{I}_{\boldsymbol{val}}^{\boldsymbol{k}}$ $(k = 1 \ldots L)$ and then identifies the optimal number of features for the final model (Lines 19-20). Then, the complete original train set D is utilized to achieve the final feature ranking and train the final LightGBM model with top $\boldsymbol{S}_{\boldsymbol{i}}^{*}$ features (Lines 21-22).

As it is shown in Fig. 1, our proposed FS method employs a basic default LightGBM model for obtaining the feature scores and calculating evaluation metrics without performing any hyperparameters optimization at this stage. Using the basic default LightGBM model simplifies the iterative process of RFE and reduces the potentiality of overfitting to training data.

C. BAYESIAN HYPERPARAMETER OPTIMIZATION OF FINAL MODEL

After finding the optimal feature subset and performing FS, the proposed framework carries out the final classification task by training a new optimized LightGBM model. According to the principal theory of LightGBM discussed in section II, various hyperparameters affect the performance of LightGBM (and other tree-based algorithms), such as the total number of decision trees (iterations), maximum depth of each tree, the minimum value of gain for splitting the leaf nodes, the learning rate of each iteration, maximum number of leaves in each tree, etc. To improve the performance, we adjust the hyperparameters of the final LightGBM model using Bayesian optimization [42]. The basic idea of Bayesian optimization is to construct a surrogate probability model of the objective function and determine the configuration of adjusted hyperparameters that perform best on the surrogate model. It then evaluates the true objective function using the selected hyperparameters and updates the surrogate probability model by adding the new evaluation results. This process continues until a certain number of iterations or time limit is reached [42]. The surrogate function we used in this article for probability representation of the objective function is the tree-structured Parzen estimator [43], which is based on Bayesian reasoning.

Incorporating the previous evaluation results and not spending a considerable time finding non-optimal hyperparameters, Bayesian optimization is faster and more efficient than conventional hyperparameter optimization methods like grid search or random search. As shown in Fig. 1, a *LOLO-CV* scheme similar to the previous FS step evaluates the classification performance of different sets of hyperparameters to avoid overfitting.

IV. EXPERIMENTAL CASE STUDIES AND RESULT ANALYSIS

This section validates the efficacy of our proposed fault diagnosis framework and compares its performance with the existing intelligent methods using two case studies.

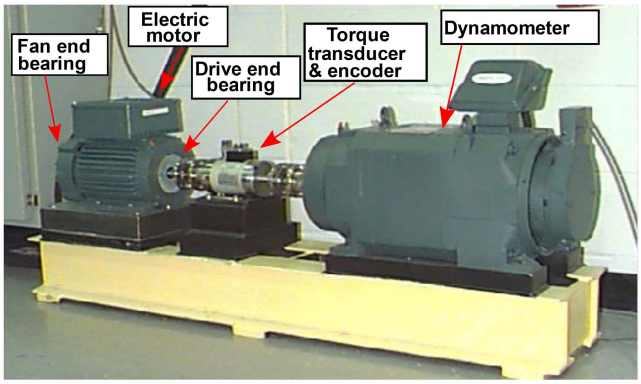

FIGURE 2. Experimental setup of CWRU bearing test rig [45].

TABLE 2. Considered class labels and their corresponding conditions for CWRU bearing dataset.

Fault type	Fault diameter (mm)	Class label
Normal	0	C1:NO
Inner race fault	0.18	C2:IF18
Inner race fault	0.36	C3:IF36
Inner race fault	0.54	C4:IF54
Ball fault	0.18	C5:BF18
Ball fault	0.36	C6:BF36
Ball fault	0.54	C7:BF54
Outer race fault	0.18	C8:OF18
Outer race fault	0.36	C9:OF36
Outer race fault	0.54	C10:OF54

TABLE 3. Various Testing Scenarios built for CWRU Bearing Dataset.

Scenario	Loads (hp) used for training	Loads (hp) used for testing
α	1-2-3	0
β	0-2-3	1
γ	0-1-3	2
δ	0-1-2	3
ε	0-1-2-3	0-1-2-3

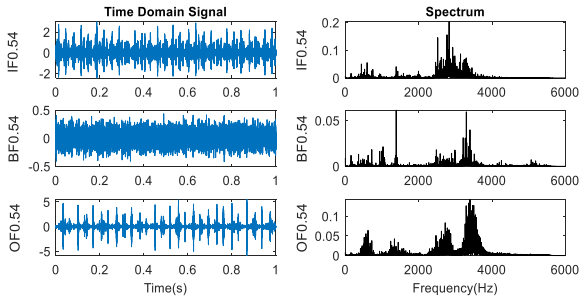

FIGURE 3. Vibration signals and their corresponding spectrums for CWRU dataset.

Rolling element bearings are among rotating machinery's most vulnerable and crucial components [44]. Therefore, we focus on the bearing faults in the first case study and evaluate our proposed algorithm's capabilities in diagnosing the bearing faults. The first dataset is the public Case Western Reserve University (CWRU) bearing fault dataset which is also studied in many research articles on fault diagnosis in the literature [44].

The second studied dataset belongs to our laboratory's induction machine (IM) setup. In addition to the bearing fault, which is the most common fault type, there are more fault types, including broken rotor bars and eccentricity faults [1]. A powerful fault diagnosis method must be general enough to present excellent performance for various fault types. Therefore, we also tested our algorithm on the IM dataset in which six types of non-bearing faults have been implemented.

We used python 3.8.5 to train the machine learning models. The libraries NumPy 1.19.2 and pandas 1.1.3 are utilized for preparing and preprocessing the data structures. Moreover, we imported Scikit-learn 0.23.2, XGboost 1.4.2, and LightGBM 3.2.1 to support the considered models. The following experiments are run on Windows 10, using a PC with Intel Core i7-9750H processor and 16 GB memory.

A. CASE 1: BEARING DATASET OF CASE WESTERN RESERVE UNIVERSITY (CWRU)

1) DATA PREPARATION

CWRU Bearing Data Center has prepared this public dataset as a benchmark [45]. As shown in Fig. 2, the experimental setup includes a 2hp IM, a torque transducer/encoder, and a dynamometer. The dataset contains vibration data collected by the accelerometers located at the drive-end and fan-end of an IM under four loading level conditions (0, 1, 2, and 3 hp). Three different fault types have been implemented on the drive-end bearings, including inner race fault (IF), outer race fault (OF), and ball fault (BF). Each fault type consists of three severity levels (diameters 0.18, 0.36, and 0.54 mm). Therefore, considering one normal condition and three fault types, each with three severity levels, there are a total of ten class labels for bearing health states, i.e., C1-C10. Table 2 illustrates the ten considered class labels and their corresponding fault types and diameters. The sampling frequency is 12kHz. The data acquired at each loading level creates a subset containing 500 samples where each sample consists of 2400 data points. We randomly selected 50% of the samples of each subset for training. Besides, to assess the performance of the fault diagnosis methods under varying operating conditions, we constructed five testing scenarios (α, β, γ, δ, and ε) shown in Table 3. In the first four scenarios, the training and testing sets are created from different loading levels, and the loading levels used for testing sets are removed from the corresponding training sets. Therefore, they are more critical than the last scenario because they can assess the generalization ability of the diagnosis methods.

Fig. 3 depicts three sample time-domain waveforms of raw vibration signals and their corresponding FFT spectrums for health states IF, BF, and OF at the severity level of 0.54mm. It is challenging to observe the hidden features discriminating the bearing states from the raw vibration signals. Hence, it is essential to extract features in multiple domains according to Section III.A.

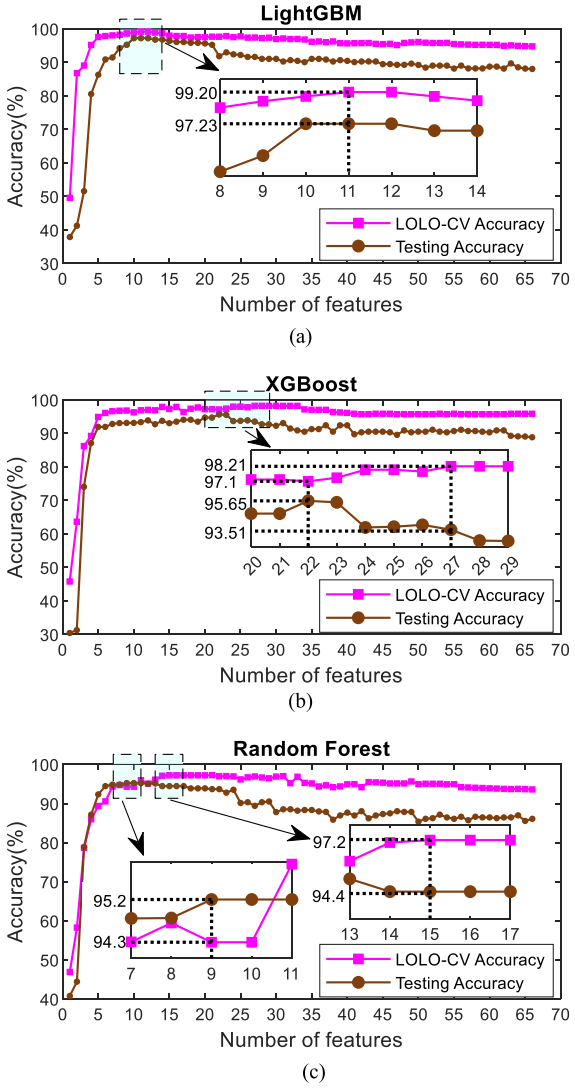

FIGURE 4. Profile of LOLO-CV and testing accuracy for CWRU dataset scenario α for different subsets of features. (a) RFE-LightGBM. (b) RFE-XGBoost. (c) RFE-RF.

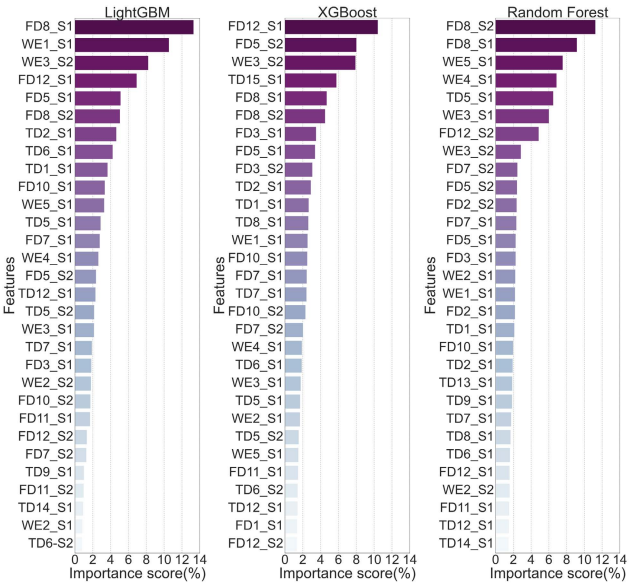

FIGURE 5. Top 30 features rankings obtained by LightGBM, XGBoost, and RF.

2) RESULTS, ANALYSIS, AND COMPARISON

Based on the methodology elucidated in section III and Fig.1, the proposed fault diagnosis framework is implemented on the CWRU bearing dataset. Initially, the data obtained from two vibration sensors at the drive-end and fan-end of the IM are processed, and 15 time domain features, 12 frequency domain features, and six wavelet energy features are extracted from each sensor's data leading to 66 features in total. Then, by training basic LightGBM models, the proposed RFE-LightGBM-FS is applied according to Algorithm 1 for each scenario. The total splitting gain in LightGBM is chosen as the criteria for calculating and evaluating the feature scores.

Among the five scenarios shown in Table 3, scenario α using the 0hp load as the original test data is taken as an example and is highlighted here because the no-load and light- load levels are potentially the most challenging conditions in fault diagnosis [27]. The resampling process in the RFE-LightGBM-FS of scenario α has three iterations ($L = 3$). Each loading (1hp, 2hp, and 3hp) is once excluded from the training set and included in the virtual nested validation set for *LOLO-CV* evaluation. Fig. 4a illustrates the RFE-LightGBM curve for scenario α concerning the *LOLO-CV* accuracy. For the sake of comparison, the testing accuracy for 0hp load is also shown in Fig. 4a for different subsets of features. It is observed that LightGBM achieves the maximum *LOLO-CV* accuracy (99.2%) when the top 11 most important features are selected. Correspondingly, the maximum testing accuracy (97.23%) is also achieved with the top 11 features. Therefore, applying the optimal feature subset for *LOLO-CV* accuracy determined by our proposed RFE-LightGBM-FS, we can reach the optimal testing performance, too, even when the test data contains new unseen loading levels.

In Algorithm 1, we applied XGBoost and Random Forest (RF) instead of LightGBM to obtain feature rankings (Lines 10-11 and 21) and perform classifications (lines 15 and 22) to compare their performance with RFE-LightGBM-FS. Figs. 4(b) and 4(c) depict the performance of RFE-XGBoost-FS and RFE-RF-FS for scenario α, respectively. In both figures, the optimal feature subsets achieving the highest testing accuracy differ from those achieving the highest *LOLO-CV* accuracy. In fact, using the proposed FS strategy of Algorithm 1, RFE-XGBoost-FS gives only 93.51% testing accuracy with the top 27 features, and RFE-RF-FS gives 94.4% testing accuracy with the top 15 features. Thus, we cannot achieve the best testing Performance by replacing LightGBM with XGBoost and RF in Algorithm 1.

The reason behind the superiority of the RFE-LightGBM-FS method can be attributed to the different feature scores

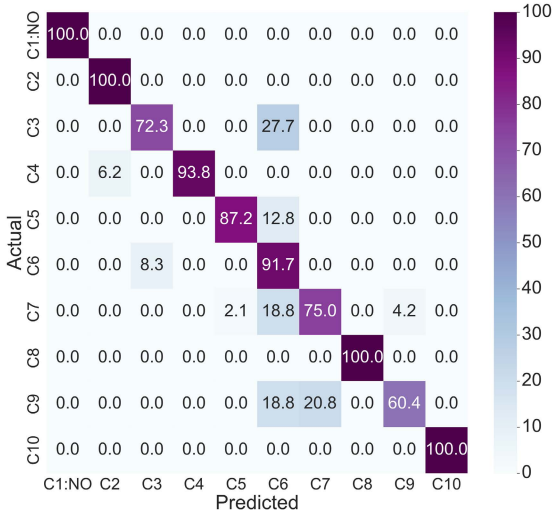

FIGURE 6. Testing confusion matrix of CWRU dataset scenario α for the basic LightGBM classifer without applying FS and hyperparameter optimization.

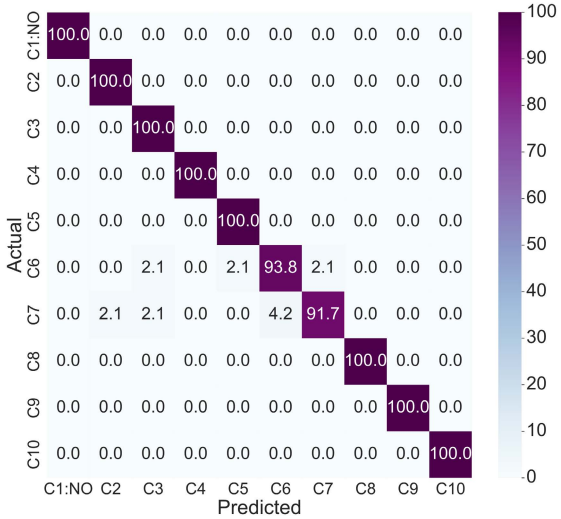

FIGURE 7. Testing confusion matrix of CWRU dataset scenario α for the LightGBM classifier after applying FS and hyperparameter optimization.

TABLE 4. Optimal LightGBM hyperparameters for CWRU dataset scenario α.

Hyperparameter	Description	Value
max_depth	Maximum depth of each trained tree	9
learning_rate	Multiplication applied per iteration	0.0623
num_iterations	Number of boosting iterations	725
num_leaves	Maximum leaves for each trained tree	63
feature_fraction	Proportion of features for each tree	0.45
min_sum_hessian_in_leaf	Minimum sum of hessians in one leaf	1.8
min_gain_to_split	Minimum gain for splitting the nodes	0.27
lambda_l1	L1 regularization	0.364

provided by LightGBM due to its leaf-wise tree growth strategy, explained in section II. Fig. 5 compares the top 30 most important features and their scores obtained by LightGBM, XGBoost, and RF models. For instance, FD8_S1 in Fig. 5 represents the eighth frequency domain feature from the first vibration sensor located in the drive-end of the IM, and WE3_S2 represents the third Wavelet energy feature from the second vibration sensor located in the fan-end of the IM. Although most of the top 30 features are shared between the three models, the feature rankings differ remarkably between the three cases in Fig. 5. According to the excellent performance of RFE-LightGBM-FS in selecting the optimal feature subset in Fig. 4a and utilizing the leaf-wise tree growth strategy, we can conclude that the feature ranking provided by LightGBM in Fig. 5 is more valid than XGBoost and RF.

Moreover, it can be seen that most of the features extracted from the drive-end vibration sensor (S1) have higher importance scores than the features obtained from the fan-end vibration sensor (S2) because the faults have been implemented on the drive-end bearing.

According to the confusion matrix shown in Fig. 6, the testing accuracy of a basic LightGBM classifier without applying FS is only 88.04% for scenario α. From Fig. 4a, we saw that adopting the proposed FS method improves the testing accuracy to 97.23%. The testing accuracy can further increase by applying Bayesian hyperparameter optimization for the final LightGBM classifier. Table 4 shows the adjusted LightGBM hyperparameters after performing Bayesian optimization. Fig. 7 illustrates the testing confusion matrix of the final optimized LightGBM classifier. After implementing the proposed FS and Bayesian optimization, the average testing accuracy becomes 98.55%.

To further explore the effectiveness of the proposed FS method, we applied PCA to reduce the features' dimensions and visualize their distribution in Fig. 8 for scenario α. Using the first three principal components (PCs), Fig. 8a depicts the distribution of features without implementing the FS method. It can be seen that there is a distribution discrepancy between the training and testing samples having identical class labels due to changing loading conditions, leading to misclassified samples in the confusion matrix of Fig. 6, particularly for classes C3, C5, C7, and C9. Moreover, because some features are redundant and uninformative, there is a significant overlap between the samples of class labels C3, C5, C6, C7, and C9, leading to poor classification performance. In contrast, after selecting the optimal subset with top 11 features in Fig. 8b, the distributions of the same class label samples are no longer sensitive to loading levels, and they are clustered together. In addition, the proposed FS has also increased the interclass separability between the samples of different fault types (C1-C10).

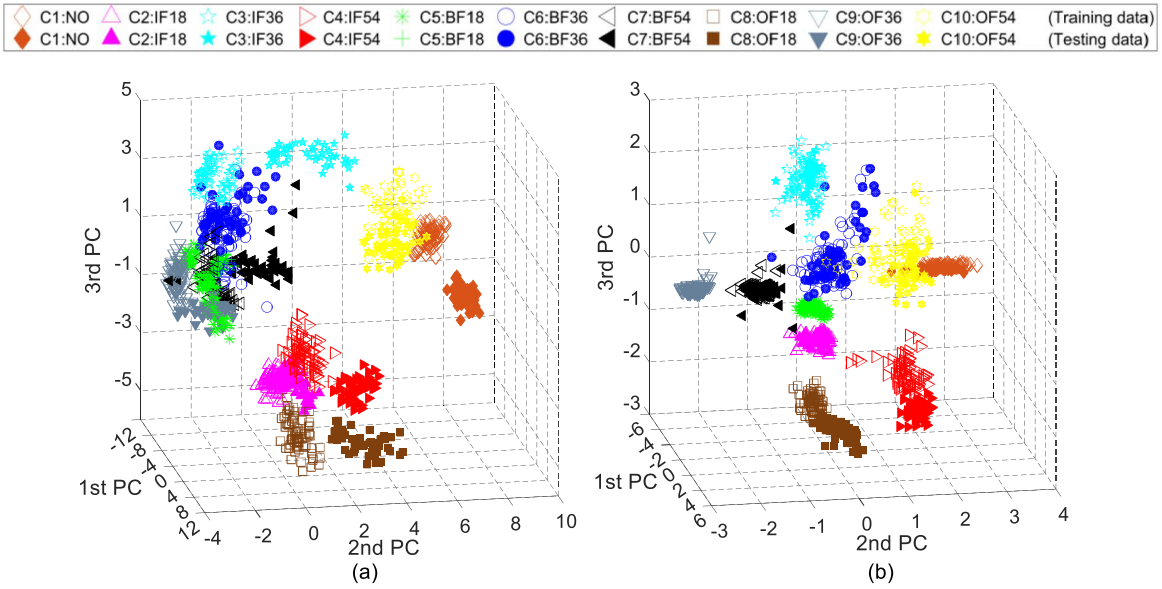

FIGURE 8. Visualization of feature distributions using the first three principal components (PCs) for CWRU dataset scenario α. (a) Without applying FS. (b) After applying the proposed RFE-LigthGBM-FS. (The train data samples from the loadings 1hp, 2hp, and 3hp are shown by hollow shapes, while filled shapes show the test data samples from 0hp load. Each of the ten class labels from C1 to C10 has its own shape and color.)

TABLE 5. Comparison of classification accuracies using CWRU dataset (%).

Fault diagnosis Method	Testing Scenarios under varying loading conditions				
	α	β	γ	δ	ε
1- Proposed RFE-LightGBM-FS with KNN classifier	94.86	97.17	96.27	96.54	97.74
2- Proposed RFE-LightGBM-FS with SVM classifier	95.93	98.57	98.34	96.27	98.93
3- Proposed RFE-LightGBM-FS with RF classifier	95.71	98.77	98.34	96.89	99.12
4- Proposed RFE-LightGBM-FS with XGBoost classifier	97.17	98.34	99.6	96.7	98.97
5- Proposed RFE-LightGBM-FS with ANN classifier	96.64	99.44	99.23	97.34	98.23
6- Proposed RFE-LightGBM-FS with LightGBM as classifier	**98.73**	**99.34**	**100**	**99.12**	**99.88**

Previously, Fig. 4 proved the superiority of LightGBM over XGBoost and RF in terms of providing the feature ranking. Table 5 compares the performance of six different settings of the proposed fault diagnosis framework for all scenarios. In methods (1-6) examined in Table 5, LightGBM is used for obtaining the feature rankings in the proposed FS (Algorithm1, lines 10-11 and 21). However, for evaluating different feature subsets (Lines 15-16) and obtaining the final classification results (Line 22), the methods (1-5) in Table 5 replace LightGBM with other existing optimized SML-based methods, i.e., KNN, SVM, RF, XGBoost, and artificial NN (ANN). For each experiment, ten trials are performed, and the average testing accuracies are reported. It can be seen that the proposed framework using LightGBM in all steps (method6 in Table 5) offers the highest average testing accuracies compared with the other methods for all the five scenarios.

Table 6 summarizes a number of fault diagnosis methods reported in the literature using the CWRU bearing dataset and compares their performance with the proposed LightGBM- based fault diagnosis framework. It can be seen that, unlike our proposed method that uses a shallow model, most of the recently developed methods in the literature are based on deep convolutional neural network (CNN). The methods based on deep transfer learning (DTL) [21], [46], [47] overcome the problems of high computational time and overfitting of conventional deep CNNs which are trained from scratch. According to Table 6, the DTL-based methods presented in [21] and [47] using VGG-16 and ResNet-50 models present the highest accuracies in the literature. Table 6 verifies the effectiveness of the proposed method in terms of accuracy compared with the state-of-the-art DL-based approaches reported in the literature. However, another key factor that should be considered and directly affects the efficiency of the fault diagnosis methods is the computational training time. As mentioned previously, the shallow models are more efficient than DL models in terms of computational requirements. The average training time of the proposed shallow LightGBM-based fault diagnosis method for the studied scenarios is 26.08 seconds; this is way less than the training times of the DTL-based methods in [21] and [47], which are reported to be in the range of 150-318 seconds

TABLE 6. Comparison of classification accuracies on CWRU bearing dataset for some published works (%).

	Fault diagnosis Method Description	Testing scenario	
		δ	ε
1	Semi-supervised distance-preserving self-organizing map trained with time domain statistical features, frequency domain features related to the characteristic frequencies of the bearing components , and wavelet energy features [48].	-	95.8
2	Empirical Mode Decomposition (EMD) is used to decompose the signals into a group of intrinsic mode functions (IMFs). Then, an SVM classifier optimized by inter-cluster distance (ICD) is trained with permutation entropy features of first five IMFs [12].	-	97.91
3	Two-stage NN. In the first stage, sparse filtering learns representative features from vibration signals in an unsupervised manner. Then, in the second stage, a softmax regression is used to perform the classification in a supervised manner [8].	-	99.66
4	Energy-fluctuated multiscale feature mining approach based on wavelet packet energy (WPE) image and deep convolutional NN (CNN) [49].	96.8	99.8
5	Deep CNN (VGG-16 architecture) trained from scratch based on the time-frequency distributions of the original sensor data which is converted to images using continuous wavelet transform (CWT) [21].	96.47	98.85
6	Transfer learning: A deep CNN (VGG-16 architecture) is pre-trained via large datasets of natural images and then is fine-tuned using the time-frequency images of original sensor data obtained by CWT [21].	98.8	99.95
7	A combined CNN-SVM framework is used in which CNN (ResNet-18 architecture) extracts representative features from the time-frequency images of original sensor data obtained by CWT. Then, SVM performs the classification task [46].	-	98.75
8	DL-based method using shunt-wound restricted Boltzmann machines (RBMs) that learns the features in a layer-wise manner from raw vibration signals directly [50].	-	98.4
9	Deep transfer convolutional neural network (ResNet-50 architecture) in which the time-frequency representations of the raw data obtained by CWT are used as inputs [47].	99.72	99.95
10	A combined deep CNN-SVM framework that uses particle swarm optimization for tuning the SVM hyperparameters [51].	-	98.2
11	Deep belief network (DBN) with Genetic Algorithm-based hyperparameter optimization using parallel computing [52].	-	98.44
12	Switchable normalization-based CNN (SN-CNN) model trained with the scalogram images obtained by CWT [53].	-	99.88
13	An improved PixelHop framework based on successive subspace learning [54].	-	98.91
14	**Proposed RFE-LightGBM-FS with LightGBM classifier**	**99.12**	**99.88**

TABLE 7. Considered class labels and their corresponding conditions for IM dataset.

Fault type	Class label
Normal	C1: NO
Dynamic Eccentricity	C2: Ecc
Two consecutive broken rotor bars (BRBs)	C3: 2 BRBs
Three consecutive BRBs	C4: 3 BRBs
Two non-consecutive BRBs	C5: 2 NC-BRBs
Two non-consecutive BRBs and Eccentricity	C6: 2 NC-BRBs& Ecc

for different scenarios in CWRU bearing dataset. Therefore, the performance of the proposed method is proven to be promising in terms of both accuracy and efficiency.

B. CASE 2: INDUCTION MACHINE DATASET OF AALTO UNIVERSITY

In this section, we further examine the performance of the proposed fault diagnosis system by conducting experiments on the IM dataset at our laboratory.

1) DATA PREPARATION

This setup consists of two 18kW IMs that are connected back-to-back via their shaft. The IM measurement setup and its configuration are depicted in Figs. 9(a) and 9(b), respectively. The vibration signals were measured from the first IM (tested machine) fed from a sinusoidal voltage supply at 50 Hz. The second IM operating as the loading machine was connected to a frequency converter to provide various loading levels. Three Kistler 8763B050AB accelerometers were evenly arranged at positions 120 degrees apart on the

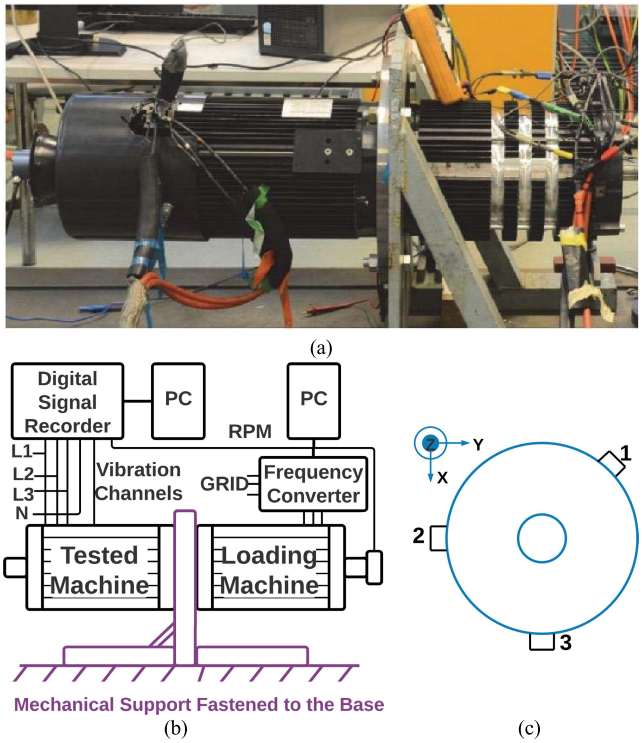

FIGURE 9. (a) IM measurement setup. (b) Setup configuration. (c) Position of accelerometers.

TABLE 8. Various Testing Scenarios created for IM Dataset.

Scenario	Loads used for training	Loads used for testing
α	FL-HL	NL
β	FL-NL	HL
γ	HL-NL	FL
ε	FL-HL-NL	FL-HL-NL

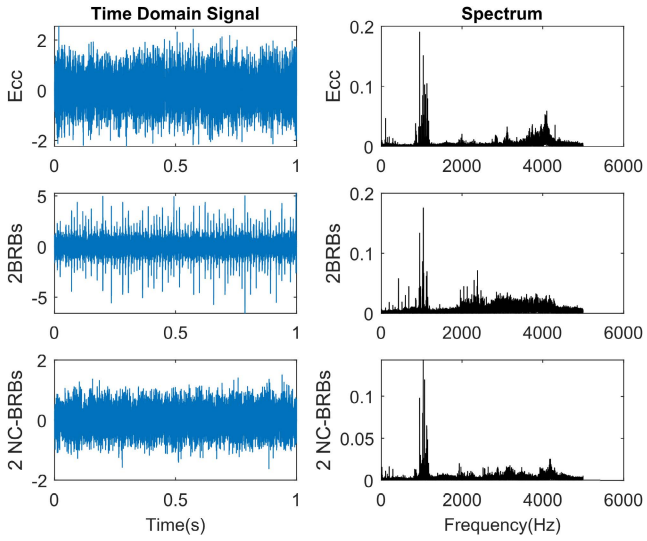

FIGURE 10. Three sample raw vibration signals and their spectrums for IM setup dataset measured in FL condition.

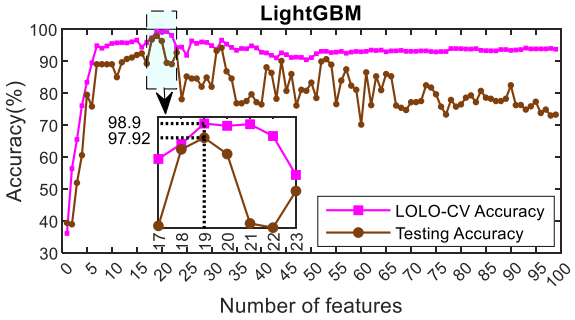

FIGURE 11. Profile of LOLO-CV and testing accuracy of IM dataset scenario α for different subsets of features.

circumference of the tested IM. Fig. 9c shows the position of the accelerometers. The vibration signals were collected under full-load (FL), half-load (HL), and no-load (NL) conditions where the currents were 40A, 30A, and 18A, respectively. The following defects were implemented on the tested IM: dynamic eccentricity with 28.5% severity (Ecc), two consecutive broken rotor bars (2 BRBs), 3 BRBs, two non-consecutive (NC) BRBs (2 NC-BRBs), and simultaneous fault of 2 NC-BRBs and Ecc. Hence, with the normal condition, there are a total of six class labels, i.e., C1-C6. Table 7 depicts the six considered class labels and their corresponding fault conditions. The sampling frequency is 10kHz. The data acquired at each loading

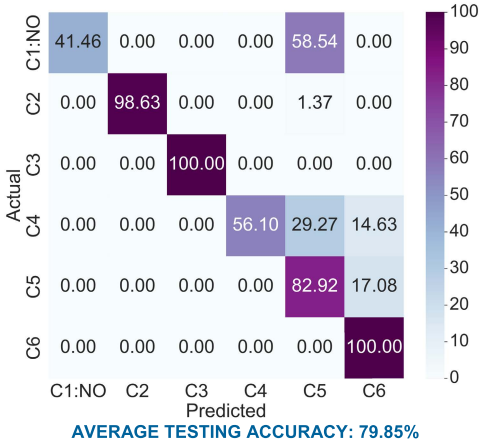

FIGURE 12. Testing confusion matrix of IM datset scenario α for the basic LightGBM classifer without applying FS and hyperparameter optimization.

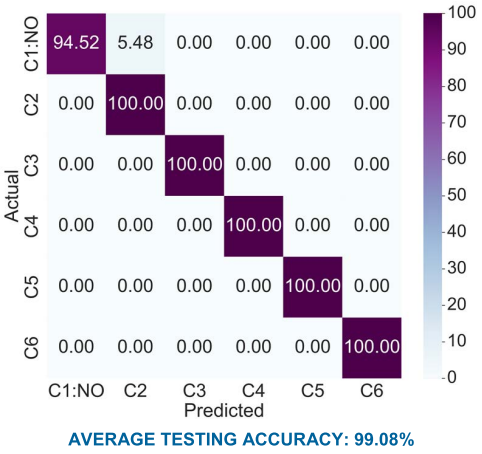

FIGURE 13. Testing confusion matrix of IM dataset scenario α for the LightGBM classifier after applying FS and hyperparameter optimization.

level creates a subset containing 400 samples where each sample consists of 2400 data points. We randomly selected 50% of the samples of each subset for training. Table 8 shows the four testing scenarios (α, β, γ, and ε) that are built like the previous CWRU case study. In the first three scenarios, the loading level used for testing is entirely unseen during the training process. The raw vibration signals for three sample health states under FL level and their spectrums are illustrated in Fig. 10.

2) RESULTS, ANALYSIS, AND COMPARISON

The same fault diagnosis methodology as the previous CWRU case is carried out on the IM dataset. In total, 99 features are calculated from the three vibration sensors data. Taking scenario α as an example, Fig. 11 displays the RFE-LightGBM-FS curve and the profile of testing accuracy. After adopting the proposed RFE-LightGBM-FS, the feature subset containing the top 19 features is determined to be the optimal one as it gives the maximum *LOLO-CV* accuracy (98.9%). According to Fig. 11, this

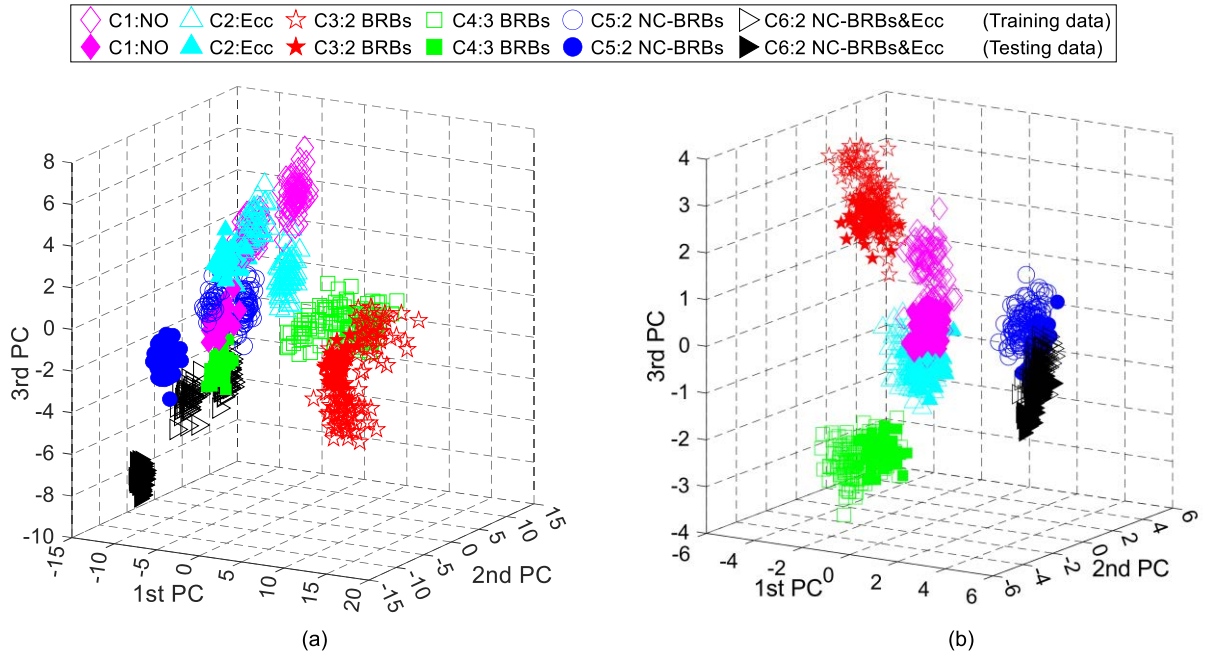

FIGURE 14. Visualization of feature distributions using the first three principal components (PCs) for IM dataset scenario α. (a) Without applying FS. (b) After applying the proposed RFE-LigthGBM-FS. (The train data samples from full-load and half-load conditions are shown by hollow shapes, while filled shapes show the test data samples from No-load condition. Each of the six class labels from C1 to C6 has its own shape and color.)

TABLE 9. Optimal LightGBM hyperparameters for IM dataset scenario α.

Hyperparameter	Description	Value
max_depth	Maximum depth of each trained tree	7
learning_rate	Multiplication applied per iteration	0.094
num_iterations	Number of boosting iterations	514
num_leaves	Maximum leaves for each trained tree	15
feature_fraction	Proportion of features for each tree	0.52
min_sum_hessian_in_leaf	Minimum sum of hessians in one leaf	4.3
min_gain_to_split	Minimum gain for splitting the nodes	0.33
lambda_l1	L1 regularization	0.82

feature subset also matches the highest testing accuracy (97.92%). Thus, applying the proposed FS technique leads to finding the optimal feature subset offering the highest testing accuracy.

Fig. 12 shows the testing confusion matrix of a basic LightGBM classifier for scenario α, indicating the average testing accuracy of 79.85%. Previously, we saw from Fig. 11 that applying the proposed FS method improves the testing accuracy to 97.92%. Fig. 13 illustrates the testing confusion matrix of scenario α after adopting the proposed RFE-LightGBM-FS and Bayesian hyperparameter optimization. According to Fig. 13, hyperparameter optimization further improves the testing accuracy to 99.08%. The optimized LightGBM hyperparameters are listed in Table 9.

The average computational training time of the proposed shallow LightGBM-based fault diagnosis framework for the studied scenarios of IM dataset is 43.12 seconds. It is higher than the computational training time of the same method trained on the CWRU dataset because the number of sensors and hence, the number of features is increased in IM case study.

Figs. 14(a) and 14(b) demonstrate the 3D PCA visualization of the features' distributions before and after performing RFE- LightGBM-FS, respectively. According to Fig. 14(a), the training and testing samples of similar class labels are separated because of varying loading levels. On the other hand, Fig. 14(b) shows that the proposed FS strategy can perfectly decrease the distance between the same class label features at various loading conditions and improve the classification accuracy.

Table 10 evaluates the classification results of the six variations of the proposed fault diagnosis framework. In all methods assessed in Table 10, the feature rankings are provided by the LightGBM model, but in methods (1-5), evaluation of different feature subsets and final classifications in Algorithm 1 are performed by other SML-based methods. The results indicate the superiority of method 6, in which LightGBM is used for all steps of Algorithm 1, including obtaining feature rankings, evaluating feature subsets, and final classification.

TABLE 10. Comparison of classification accuracies using IM dataset (%).

Fault diagnosis Method	Testing Scenarios under varying loading conditions			
	α	β	γ	δ
1- Proposed RFE-LightGBM-FS with KNN classifier	93.55	96.42	94.88	98.95
2- Proposed RFE-LightGBM-FS with SVM classifier	96.26	97.22	98.17	99.14
3- Proposed RFE-LightGBM-FS with RF classifier	96.26	96.55	95.49	98.38
4- Proposed RFE-LightGBM-FS with XGBoost classifier	97.87	98.38	97.22	99.26
5- Proposed RFE-LightGBM-FS with ANN classifier	97.22	97.64	96.55	98.92
6- Proposed RFE-LightGBM-FS with LightGBM as classifier	**99.08**	**99.79**	**99.46**	**99.92**

V. CONCLUSION AND FUTURE WORK

In this article, a theoretical and experimental study of a new fault diagnosis framework was presented that offers high accuracy and generalization ability for the testing scenarios in which data originates from new operating conditions being unavailable during training. Leveraging the LightGBM's ability to provide feature ranking, we proposed an efficient FS strategy combining LightGBM, RFE, and a *LOLO-CV*-based resampling process. Moreover, we performed Bayesian hyperparameter optimization to enhance the final classification result. Two experimental case studies, i.e., bearing dataset of CWRU and induction IM dataset obtained in our laboratory, were utilized to examine the proposed fault diagnosis system's effectiveness and accuracy.

Considering the changing operating conditions, we studied various testing scenarios in which a particular loading level is removed from the dataset and only used for testing. The results demonstrated that the proposed RFE-LightGBM-FS method could identify the optimal subset of features that are not sensitive to changing operating conditions and offer high separability between various class labels. The evaluation results proved that for various testing scenarios of CWRU and IM datasets, the proposed fault diagnosis framework achieved 98.55% to 100% accuracy. In both case studies, we highlighted the most challenging testing scenario (scenario α) where the data measured under the no-load condition did not participate in the training process and was only used for testing. According to the results, the classification accuracy of scenario α using a basic LightGBM classifier without any FS implementation or hyperparameter tuning was 88.04% and 79.85% for CWRU and IM datasets, respectively. Meanwhile, the proposed RFE-LightGBM-FS method increased the testing accuracies to 97.23% and 97.92%, respectively. The results also showed that applying the Bayesian hyperparameter optimization further increases accuracy to 98.55% and 99.08%.

Future research work includes:

- Extending the proposed framework to be able to locate the faults in addition to detecting and discriminating the fault types by implementing and studying the faults occurring in both drive-end and fan-end bearings.
- Considering and implementing the primary causes of bearing failure and more realistic faults such as lubrication degradation of bearings, overheating, excessive loads, and corrosion.
- Embedded system implementation of the proposed framework to assess its ability in real-time operation.

REFERENCES

[1] 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, doi: 10.1109/TEC.2005.847955.

[2] S. Kumar, D. Mukherjee, P. K. Guchhait, R. Banerjee, A. K. Srivastava, D. N. Vishwakarma, and R. K. Saket, "A comprehensive review of condition based prognostic maintenance (CBPM) for induction motor," *IEEE Access*, vol. 7, pp. 90690–90704, 2019, doi: 10.1109/ACCESS.2019.2926527.

[3] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018, doi: 10.1109/TMECH.2017.2728371.

[4] G. H. Bazan, P. R. Scalassara, W. Endo, and A. Goedtel, "Information theoretical measurements from induction motors under several load and voltage conditions for bearing faults classification," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3640–3650, Jun. 2020, doi: 10.1109/TII.2019.2939678.

[5] M. N. S. K. Shabbir, X. Liang, and S. Chakrabarti, "An ANOVA-based fault diagnosis approach for variable frequency drive-fed induction motors," *IEEE Trans. Energy Convers.*, vol. 36, no. 1, pp. 500–512, Mar. 2021, doi: 10.1109/TEC.2020.3003838.

[6] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2441–2451, May 2014, doi: 10.1109/TIE.2013.2273471.

[7] H. Helmi and A. Forouzantabar, "Rolling bearing fault detection of electric motor using time domain and frequency domain features extraction and ANFIS," *IET Electr. Power Appl.*, vol. 13, no. 5, pp. 662–669, May 2019, doi: 10.1049/iet-epa.2018.5274.

[8] Y. Lei, F. Jia, J. Lin, and S. Xing, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, Jan. 2016, doi: 10.1109/TIE.2016.2519325.

[9] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016, doi: 10.1016/j.measurement.2016.07.054.

[10] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018, doi: 10.1016/j.ymssp.2018.02.016.

[11] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, and M. Saif, "A semi-supervised diagnostic framework based on the surface estimation of faulty distributions," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1277–1286, Mar. 2019, doi: 10.1109/TII.2018.2851961.

[12] X. Zhang, Y. Liang, and J. Zhou, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, Jun. 2015, doi: 10.1016/j.measurement.2015.03.017.

[13] R. H. C. Palácios, I. N. da Silva, A. Goedtel, and W. F. Godoy, "A comprehensive evaluation of intelligent classifiers for fault identification in three-phase induction motors," *Electr. Power Syst. Res.*, vol. 127, pp. 249–258, Oct. 2015, doi: 10.1016/j.epsr.2015.06.008.

[14] M. S. Safizadeh and S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Inf. Fusion*, vol. 18, pp. 1–8, Jul. 2014, doi: 10.1016/j.inffus.2013.10.002.

[15] R. Razavi-Far, M. Farajzadeh-Zanjani, and M. Saif, "An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2758–2769, Dec. 2017, doi: 10.1109/TII.2017.2755064.

[16] J. S. L. Senanayaka, H. Van Khang, and K. G. Robbersmyr, "Multiple classifiers and data fusion for robust diagnosis of gearbox mixed faults," *IEEE Trans. Ind. Informat.*, vol. 15, no. 8, pp. 4569–4579, Aug. 2019, doi: 10.1109/TII.2018.2883357.

[17] W. Sun, R. Zhao, R. Yan, S. Shao, and X. Chen, "Convolutional discriminative feature learning for induction motor fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1350–1359, Jun. 2017, doi: 10.1109/TII.2017.2672988.

[18] J. Long, S. Zhang, and C. Li, "Evolving deep echo state networks for intelligent fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4928–4937, Jul. 2020, doi: 10.1109/TII.2019.2938884.

[19] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020, doi: 10.1109/ACCESS.2020.2972859.

[20] I. Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez, and R. J. Romero-Troncoso, "An experimental comparative evaluation of machine learning techniques for motor fault diagnosis under various operating conditions," *IEEE Trans. Ind. Appl.*, vol. 54, no. 3, pp. 2215–2224, May/Jun. 2018, doi: 10.1109/TIA.2018.2801863.

[21] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019, doi: 10.1109/TII.2018.2864759.

[22] W. Lang, Y. Hu, C. Gong, X. Zhang, H. Xu, and J. Deng, "Artificial intelligence-based technique for fault detection and diagnosis of EV motors: A review," *IEEE Trans. Transport. Electrific.*, vol. 8, no. 1, pp. 384–406, Mar. 2022, doi: 10.1109/TTE.2021.3110318.

[23] P. Gangsar and R. Tiwari, "Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review," *Mech. Syst. Signal Process.*, vol. 144, Oct. 2020, Art. no. 106908, doi: 10.1016/j.ymssp.2020.106908.

[24] B. Playe and V. Stoven, "Evaluation of deep and shallow learning methods in chemogenomics for the prediction of drugs specificity," *J. Cheminformatics*, vol. 12, no. 1, p. 11, Dec. 2020, doi: 10.1186/s13321-020-0413-0.

[25] A. N. Saberi, S. Sandirasegaram, A. Belahcen, T. Vaimann, and J. Sobra, "Multi-sensor fault diagnosis of induction motors using random forests and support vector machine," in *Proc. Int. Conf. Electr. Mach. (ICEM)*, Aug. 2020, pp. 1404–1410, doi: 10.1109/ICEM49940.2020.9270689.

[26] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, 2018, doi: 10.1109/access.2018.2818678.

[27] P. Gangsar and R. Tiwari, "A support vector machine based fault diagnostics of induction motors for practical situation of multi-sensor limited data case," *Measurement*, vol. 135, pp. 694–711, Mar. 2019, doi: 10.1016/j.measurement.2018.12.011.

[28] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst*, 2017, pp. 3146–3154.

[29] P. Gangsar and R. Tiwari, "Multifault diagnosis of induction motor at intermediate operating conditions using wavelet packet transform and support vector machine," *J. Dyn. Syst., Meas., Control*, vol. 140, no. 8, pp. 1–11, Aug. 2018, doi: 10.1115/1.4039204.

[30] X. Li and W. Zhang, "Deep learning-based partial domain adaptation method on intelligent machinery fault diagnostics," *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4351–4361, May 2021, doi: 10.1109/TIE.2020.2984968.

[31] A. Stief, J. R. Ottewill, J. Baranowski, and M. Orkisz, "A PCA and two-stage Bayesian sensor fusion approach for diagnosing electrical and mechanical faults in induction motors," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9510–9520, Dec. 2019, doi: 10.1109/tie.2019.2891453.

[32] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019, doi: 10.1109/TIE.2018.2877090.

[33] W. Qian, S. Li, and J. Wang, "A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions," *IEEE Access*, vol. 6, pp. 69907–69917, 2018, doi: 10.1109/ACCESS.2018.2880770.

[34] H. Zheng, R. Wang, Y. Yang, J. Yin, Y. Li, Y. Li, and M. Xu, "Cross-domain fault diagnosis using knowledge transfer strategy: A review," *IEEE Access*, vol. 7, pp. 129260–129290, 2019, doi: 10.1109/ACCESS.2019.2939876.

[35] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning?: With Applications in R*. Cham, Switzerland: Springer, 2013.

[36] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Cham, Switzerland: Springer, 2013.

[37] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[38] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[39] S. Luo and T. Chen, "Two derivative algorithms of gradient boosting decision tree for silicon content in blast furnace system prediction," *IEEE Access*, vol. 8, pp. 196112–196122, 2020, doi: 10.1109/ACCESS.2020.3034566.

[40] J. Qu, Z. Zhang, and T. Gong, "A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion," *Neurocomputing*, vol. 171, pp. 837–853, Jan. 2016, doi: 10.1016/j.neucom.2015.07.020.

[41] R. Yan and R. X. Gao, "Energy-based feature extraction for defect diagnosis in rotary machines," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3130–3139, Sep. 2009, doi: 10.1109/TIM.2009.2016886.

[42] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated Machine Learning: Methods, Systems, Challenges*. Cham, Switzerland: Springer, 2019.

[43] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, 2011, pp. 1–9.

[44] D. Neupane and J. Seok, "Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review," *IEEE Access*, vol. 8, pp. 93155–93178, 2020, doi: 10.1109/ACCESS.2020.2990528.

[45] *Case Western Reserve University Bearing Data Center*. Accessed: Apr. 20, 2022. [Online]. Available: https://engineering.case.edu/bearingdatacenter/welcome

[46] L. Yuan, D. Lian, X. Kang, Y. Chen, and K. Zhai, "Rolling bearing fault diagnosis based on convolutional neural network and support vector machine," *IEEE Access*, vol. 8, pp. 137395–137406, 2020, doi: 10.1109/ACCESS.2020.3012053.

[47] Z. Chen, J. Cen, and J. Xiong, "Rolling bearing fault diagnosis using time-frequency analysis and deep transfer convolutional neural network," *IEEE Access*, vol. 8, pp. 150248–150261, 2020, doi: 10.1109/ACCESS.2020.3016888.

[48] W. Li, S. Zhang, and G. He, "Semisupervised distance-preserving self-organizing map for machine-defect detection and classification," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 869–879, May 2013, doi: 10.1109/TIM.2013.2245180.

[49] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 8, pp. 1926–1935, Aug. 2017, doi: 10.1109/TIM.2017.2674738.

[50] T. Pan, J. Chen, J. Pan, and Z. Zhou, "A deep learning network via shunt-wound restricted Boltzmann machines using raw data for fault detection," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4852–4862, Jul. 2020, doi: 10.1109/TIM.2019.2953436.

[51] X. Zhang, P. Han, L. Xu, F. Zhang, Y. Wang, and L. Gao, "Research on bearing fault diagnosis of wind turbine gearbox based on 1DCNN-PSO-SVM," *IEEE Access*, vol. 8, pp. 192248–192258, 2020, doi: 10.1109/ACCESS.2020.3032719.

[52] C. Guo, L. Li, Y. Hu, and J. Yan, "A deep learning based fault diagnosis method with hyperparameter optimization by using parallel computing," *IEEE Access*, vol. 8, pp. 131248–131256, 2020, doi: 10.1109/ACCESS.2020.3009644.

[53] D. Neupane, Y. Kim, and J. Seok, "Bearing fault detection using scalogram and switchable normalization-based CNN (SN-CNN)," *IEEE Access*, vol. 9, pp. 88151–88166, 2021, doi: 10.1109/ACCESS.2021.3089698.

[54] L. Wan, Z. Zhou, K. Gong, G. Zhang, Y. Li, and C. Li, "An improved PixelHop framework and its application in rolling bearing fault diagnosis," *IEEE Access*, vol. 9, pp. 139755–139770, 2021, doi: 10.1109/ACCESS.2021.3119752.

ALIREZA NEMAT SABERI (Graduate Student Member, IEEE) received the M.Sc. degree in electrical power engineering from the University of Tehran, Tehran, Iran, in 2017, and is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland. He is currently a Data Science Engineer at ABB System Drives, Helsinki, Finland. His research interests include intelligent fault diagnosis and condition monitoring of industrial apparatus and systems, applied machine learning and deep learning, and design and modeling of non-conventional electrical machines and drives.

ANOUAR BELAHCEN (Senior Member, IEEE) was born in Morocco, in 1963. He received the M.Sc. (Tech.) and Doctor (Tech.) degrees from the Helsinki University of Technology, Finland, in 1998 and 2004, respectively. He is currently a Professor in energy and power at Aalto University, Finland, and a Professor in electrical machines at the Tallinn University of Technology, Estonia. His research interests include numerical modeling of electrical machines, magnetic materials, coupled magneto-mechanical problems, magnetic forces, magnetostriction, and fault diagnostics of electrical machines.

JAN SOBRA was born in Tabor, Czech Republic, in 1987. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the University of West Bohemia, Pilsen, Czech Republic, in 2010, 2012, and 2018, respectively. He is holding an academic and research position at the Faculty of Electrical Engineering, University of West Bohemia, Pilsen, Czech Republic. His research interests include the modeling and diagnostics of electrical machines.

TOOMAS VAIMANN (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Tallinn University of Technology, Estonia, in 2007, 2009, and 2014, respectively. He is currently a Senior Researcher with the Department of Electrical Power Engineering and Mechatronics, Tallinn University of Technology. He has been working in several companies as an Electrical Engineer. His main research interest includes diagnostics of electrical machines. He is a member of the Estonian Society of Moritz Hermann Jacobi and the Estonian Society for Electrical Power Engineering.