# Intelligent Fault Diagnosis of Manufacturing Processes Using Extra Tree Classification Algorithm and Feature Selection Strategies

**SINA YOUSEFI** [ID], **SHEN YIN** [ID] **(Fellow, IEEE), AND MUHAMMAD GIBRAN ALFARIZI** [ID]

Department of Mechanical and Industrial Engineering, Faculty of Engineering, Norges Teknisk-Naturvitenskapelige Universitet, 7491 Trondheim, Norway

CORRESPONDING AUTHOR: MUHAMMAD GIBRAN ALFARIZI (e-mail: muhammad.g.alfarizi@ntnu.no).

**ABSTRACT** Fault diagnosis is integral to maintenance practices, ensuring optimal machinery functionality. While traditional methods relied on human expertise, intelligent fault diagnosis techniques, propelled by machine learning (ML) advancements, now offer automated fault identification. Despite their efficiency, a research gap exists, emphasizing the need for methods providing not just reliable fault identification but also in-depth causal factor analysis. This research introduces a novel approach using an extra tree classification algorithm and feature selection to identify fault importance in manufacturing processes. Compared with SVM, neural networks, and tree-based ML, the method enhances training and computational efficiency, achieving over 99% classification accuracy on prognostics and health management 2021 dataset. Importantly, the algorithm enables researchers to analyze individual fault causes, addressing a critical research gap. The study provides guidelines for further research, aiming to refine the proposed strategy. This work contributes to advancing fault diagnosis methodologies, combining automation with comprehensive causal analysis, crucial for both academic and industrial applications.

**INDEX TERMS** Deep learning, fault identification, intelligent fault diagnosis (IFD), machine learning (ML), machines.

## I. INTRODUCTION

The failure of machine parts can have severe consequences, from endangering lives to substantial financial casualties. Thus, maintaining industrial facilities is essential, and ensuring equipment availability, durability, and product quality depends heavily on maintenance [1], [2]. Identifying and assessing machine component conditions is crucial to enhancing machine safety and dependability while reducing operational and maintenance expenses [3]. However, timely and accurate fault detection in today's industrial facilities can be impractical, typically requiring one or more human professionals to assess machine performance in real-time [4]. Previous works in the field of fault diagnosis primarily relied on manual fault diagnosis and signal processing techniques [5]. These methods often required domain-specific expertise and were not efficient in handling the growing complexity of modern machinery. Therefore, there was a pressing need to develop automated fault diagnosis techniques that

could efficiently and accurately identify faults without extensive human intervention [6]. Alternative approaches, such as model-based diagnosis using artificial intelligence (AI) techniques, have gained prominence recently in scholarly and industrial contexts due to their ability to provide intuitive results with minimal expert intervention [7]. These methods use machine learning (ML) to adaptively create machinery diagnostic understanding using data rather than relying on the knowledge and skills of engineers. Intelligent fault diagnosis (IFD) intends to instantly create diagnostic models connecting gathered data and machine health [8]. Due to more effective paradigms and data availability, deep learning algorithms have also been successfully employed to identify errors intelligently [9].

The motivation for this research arises from the limitations of previous fault diagnosis methods. Manual diagnosis and traditional signal processing techniques were time-consuming, limited in scope, and dependent on expert

knowledge. They were not equipped to handle the increasing complexity of machinery in various industries [10]. Moreover, in the pursuit of streamlined fault diagnosis, it is crucial to not merely categorize faults with high accuracy but to also equip researchers with the capability to conduct in-depth analyses, discerning the root causes behind these faults. This dual-pronged approach stands as an unaddressed research gap, one that offers the potential to greatly enhance the efficiency and efficacy of fault diagnosis methods, particularly in industrial contexts. This article proposes an ML-based model and a unique feature selection process to categorize errors based on input data from the prognostics and health management (PHM) mechanical defects dataset.[1] The PHM 2021 is a benchmark dataset for mechanical fault diagnosis designed for the PHM community. The dataset contains vibration signals and corresponding health status information from bearings with artificially induced faults, providing a realistic testbed for evaluating the performance of fault detection and diagnosis algorithms [11]. The proposed model uses the extra tree (ET) classifier to categorize mechanical faults based on the features in the dataset, contrasted with several other learning algorithms. The study's tasks include acquiring, preprocessing, and splitting the PHM dataset into two sets for training and testing, building the ET ML classification model, developing an adequate assessment procedure with metrics to evaluate the model's performance, fine-tuning model parameters for the ideal set of evaluation measures, contrasting outcomes with alternative approaches, and assessing the classification results for each class. With the motivation rooted in the shortcomings of previous fault diagnosis methods, this research contributes to the field by offering a more efficient and accurate approach to IFD. The proposed method not only enhances the reliability of machinery but also empowers researchers to analyze the causes of individual faults based on their essential characteristics. Our approach distinguishes itself by offering a multifaceted capacity. While excelling in categorizing mechanical failures with remarkable accuracy, consistently exceeding 99%, it further empowers researchers to conduct in-depth root cause analysis. This dual-pronged innovation enables fault identification and subsequent causality analysis, granting comprehensive insights into machinery health and fault scenarios.

The rest of this article is organized as follows. Section II investigates deep neural networks' significance and topologies for conducting IFD, the limitations, constraints, and possible practical approaches, including transfer learning. Section III examines aspects of the proposed model's workflow and implementation tools, followed by a detailed presentation of the suggested approach. Section IV presents and compares the experiment's results against several learning algorithms, evaluating the proposed method's results on the PHM dataset rigorously. Finally, Section V concludes this article.

---

[1]https://github.com/PHME-Datachallenge/Data-Challenge-2021

## II. RELATED WORKS

Science and technological advancements have led to the development of mechanical systems, such as those seen in wind turbines, aircraft, high-speed trains, and industrial gear. Engineers must develop strategies to guarantee the effectiveness of these systems. Machine fault identification is one of the key tactics for ongoing maintenance. This technique could help limit the escalation of abnormal events, minimize downtime, foresee residual life, and reduce productivity loss.

Incorporating advanced techniques and methodologies into the realm of fault diagnosis is crucial for enhancing the reliability and efficiency of machine health assessments. Notably, the field has witnessed significant innovations in optimization algorithms, such as the development of coevolutionary multiswarm adaptive differential evolution [12], which addresses issues like premature convergence and search stagnation, or the introduction of MS-RPNet, a novel hyperspectral image (HSI) classification network combining data-driven approaches with S3-PCA and 2D-SSA techniques [13], improving spectral accuracy in HSI classification. Furthermore, the advent of multistrategy competitive-cooperative coevolutionary algorithms, like MSCOEA, showcases a remarkable capability to balance uniformity and convergence in solving multiobjective optimization problems [14]. As technology evolves, issues related to data sharing and privacy protection become paramount, leading to innovative solutions such as the blockchain-based flight operation data sharing scheme [15], which ensures secure data sharing while preserving privacy and confidentiality. The integration of these advanced techniques marks a significant progression in the field of fault diagnosis, promising more accurate and efficient methodologies for maintaining industrial machinery. These advanced techniques complement the traditional methods of fault diagnosis that have relied on manual assessment and signal processing [16]. By introducing these innovations into the domain of machine health assessment, researchers aim to address challenges related to premature convergence, spectral uncertainty in image classification, multiobjective optimization, and secure data sharing. These advanced techniques offer more effective, accurate, and efficient approaches to fault diagnosis and maintenance practices in various industrial domains.

Furthermore, in actual conditions, various defect diagnosis techniques are used to gather helpful information from certain physical assets. Examples of situation monitoring for machine data contain environmental information, temperature, and pressure. It used to be possible to identify the types of equipment faults that occurred and where they originated using manual fault diagnosis and signal processing techniques. Nevertheless, in an engineering context, most maintainers must acquire the specific skills these solutions rely on. As a result, diagnostic defect schemes that are able to recognize appliance health issues automatically are preferred by industrial applications today [16]. Intelligent fault detection is anticipated to accomplish this goal via ML. To recognize machine problems in the past, IFD used well-known ML methods like

support vector machines. The diagnostic technique is split into three stages: data collection; synthetic attribute extraction; and health situation identification [17].

Sensors are mounted on the equipment during the data-collecting phase to gather data continuously. Numerous sensors, including currents, temperature, and vibration, have been employed for automatic status assessments as sensor technology has advanced [18], [19]. The subsequent step in building an automatic defect recognition scheme using conventional ML models is extracting critical characteristics from the information obtained during data collection stages. The feature extraction phase uses signal processing techniques, including time-domain and Fourier spectral analysis to extract representative features from the recorded signals [20]. Subsequently, these models learn the labeled data in the training stage to identify machinery health issues when presented with unlabeled input samples. To accomplish this, tagged samples are used during the initial training of the diagnostic models. The remainder of this part is separated into subsections based on the various sorts of these procedures and the methods IFD research utilizes them.

### A. ARTIFICIAL NEURAL NETWORKS (ANN)

ANNs are a subset of ML fundamental to deep learning methodologies. A primal input layer, one or more hidden layers, and an output layer help compensate for a neural network. These algorithms are potent tools in computer science and AI after they have been adjusted for better accuracy [21]. Shallow neural networks have occasionally been utilized to find defects based on input data, even though deep learning methods were employed in most researches on IFD involving ANNs. For instance, Bernieri et al.'s [22] work was among the pioneering studies to apply an ANN-based method for real-time issue detection. They showed that ANNs might aid identification of schemes and defect recognition in situations requiring quick response times. When this research was released in 2009, most ANN techniques were single-step time-series prediction methods, indicating that deep learning networks were not utilized. Lei et al. [20] proposed a two-step training method for ANN-based intelligent machine failure detection in another similar research. Furthermore, using ANNs, another research created an intelligent system that can recognize three recurring occurrences in a PV array, including healthy and short circuit failures and string disconnection [23].

### B. SUPPORT VECTOR MACHINES (SVMS)

The SVM algorithm seeks an N-dimensional space hyperplane that divides data points. Decision boundaries, known as hyperplanes, are used to classify data. In the late 1990s, the initial efforts were made to use SVMs for mechanical condition control and issue diagnostics. Multiple research papers later proposed SVM-based fault diagnostic techniques, displaying more outstanding fault diagnosis capabilities than traditional ML algorithms [24]. These approaches used a variety of kernel functions and cross validation. For instance,

Samanta [25] identified characteristics from the vibration signals of a spinning tool with both standard and unreliable gears. Sugumaran et al. [26] employed the proximal SVM in order to effectively identify defects using numerical information in other inquiries utilizing a DTree model to select the most important attributes of an example collection for a classification problem. Piliougine et al. [27] also introduce ML-based techniques to detect partial shading-induced mismatches in photovoltaic arrays, achieving accurate module fault detection with SVM and decision tree (DT) models. Zhang et al. [28] suggested employing collaborative empirical mode decay to break down vibrational signals into a cluster of intrinsic mode functions when bearings contain flaws. Datta and Sarkar [29] examined vibrational inspection, acoustic methods, and pipeline leakage discovery based on SVM. Stetco et al. [30] reported the most recent study on ML methods for monitoring wind turbine conditions. Most models utilize a dataset called SCADA or simulated data, with classification making up roughly two-thirds of the techniques and regression providing the remaining one-third.

### C. DECISION TREES

The goal of using a DT is to create an ML model that can predict the type or quantity of the target variable. In its simplest versions, DTs are straightforward algorithms that are easy to observe and understand. These models, however, could be excessively straightforward for problems with more intricate elements [31]. Numerous tree-based methods were created to increase accurateness and maintain dispensation competence. Many people use ensemble techniques, which aggregate DTs to improve prediction accuracy. These methods include the XGBoost and random forest (RF) algorithms [32], [33].

The DT's precision in fault recognition can be proved by means of test data and expert information because it is easy to understand and grasp [34]. For instance, Zhao et al. [35] provide a decision-tree-based system for categorizing and detecting defects. Widely accessible PV system data, such as PV array voltage, operational temperature, and irradiance, are utilized as characteristics in the training and test sets. The trained DT models demonstrated good flaw detection and classification accuracy in experiments. Yan et al. [34] used the categorization and regression tree approach for DT induction as a data-driven diagnostic tool for AHUs. The technique combined a steady-state sensor and a regression model to make the diagnostic method easier to understand. An average F-measure of 0.97 was used to demonstrate the superior diagnostic performance of this technique.

XGBoost is a gradient-boosting DTree ensemble ML algorithm. When big datasets are not accessible and the input qualities are not visual, extreme gradient boosting could be a viable solution for IFD [36]. For instance, Zhang et al. [5] offered a unique signal-processing method based on the XGB algorithm that solely relies on phase voltage and current information. Preprocessed data and wavelet analysis were combined to create the XGBoost approach, which extracts attributes with more than 90% accuracy. In a different research

[32], an XGB method was introduced to increase fault ID precision, combining an improved genetic algorithm with the XGBoost to create a hybrid diagnostic network. Alfarizi et al. [37] also proposed an integrated fault diagnosis system using the extreme gradient boosting algorithm for a fuse test bench line. The article concludes that the proposed method achieves high classification accuracy, fast diagnosis time, and interpretable root cause analysis. Experimental results show that the proposed algorithm outperforms several common fault diagnostic approaches.

Another ensemble learning technique, RF, combines several ineffective models to get a more reliable model. Each tree produces a "vote," a classification for that class. A forest selects the categorization that obtains the highest number of votes in regression. RF classifiers excel in industrial settings where large datasets are frequently unavailable for diagnostic training models. A dependable method for detecting multiclass faults in spur gears was developed by Cerrada et al. [4] using an evolutionary algorithm. Zhang et al. [28] also used an RF classifier to identify mechanical issues with induction motor bearings. Additionally, RFs were employed while locating nonmechanical flaws. For instance, Puggini et al. [38] created an unsupervised RF method to recognize injured wafers using chemical fingerprints.

### D. DEEP NEURAL NETWORKS (DNN)

A DNN is a computer program that creates predictions and corrects data errors using complex algorithms. Such a network needs accurate constraints to perform appropriately, including the number of hidden layers and neurons in each [39]. Although data-collecting techniques for obtaining the data required for DL algorithms are yet essential, the feature extraction phase must be included in implementing deep learning models [40]. Therefore, DNNs revolutionized ML models and gained popularity recently [41]. DL-based analysis employs attributes learned from the inputs to identify appliance health issues. DL algorithms can solve shortcomings in current intelligent fault detection systems by learning feature hierarchies utilizing features from higher levels of the order established by the composition of lower-level characteristics [19].

Deep learning-based diagnostic methods incorporate attributes they have learned from input data to identify machine health issues. These models use hierarchical networks like fully-connected (FC) layers [42], [43], [44], [45], CNNs [46], [47], [48], [49], [50], deep belief networks [17], [51], [52], RNNs [53], [54], and multilayered auto-encoders [52], [55], [56] to find essential traits. After learning to relate these qualities to other classes in future layers, the model generates its output. After each training cycle, the diagnostic models' training parameters are updated via backpropagation. FC models cannot be the best at everything. However, they perform better when representing more intricate functions, and they have proven to outperform similar methods when multiple deep algorithms are coupled, and sufficient amounts of data are available.

## III. METHODOLOGY

The current study proposes an approach split into four major stages: preprocessing, determining the importance of features, ML model construction, and evaluation based on assessment criteria.

### A. DATASET

The data from the PHM challenge [11] is utilized in this experiment to evaluate several machine-learning architectures. This dataset collaborated with the Swiss Center for Electronics and Microtechnology to gain access to a dataset from a real-world industrial testbed. This system, which includes robotic arms, conveyor belt motors, and an infrared camera, enables continuous testing of electrical components. With the aid of subject experts, the dataset was collected in errorless work settings and supervised circumstances utilizing a range of seeded vulnerabilities. The collected data has 50 signals representing how a critical variable has changed. Every signal has fields attached to it that specify various signal qualities obtained from it utilizing an autonomous data collection method [11].

The PHM 2021 dataset contains four fault classes, each containing eight types of mechanical faults. There are 8000 data points in the dataset, with 2000 data points for each fault class. The dataset comprises 24 features, including statistical, spectral, and time-domain features, acceleration, temperature, and pressure measurements. Additionally, the dataset includes information about the motor's operating conditions, such as voltage and current values [11].

### B. DATA PREPROCESSING

This study's preprocessing includes removing ineffective columns and replacing NaN values with zero. The dataset is then split into input and output sections, transformed using quantile transformer to achieve a uniform and normal distribution, and divided into 80/20 train and test sets.

### C. FEATURE IMPORTANCE

A dataset's significance of features can be utilized to understand it better. The relative ratings may determine the target's priorities, showing which attributes are not essential. Understanding the association between the characteristics and the goal variable is made more accessible by the relevance of the features [57]. It also helps to identify the qualities that are irrelevant to the model. By calculating scores for each component, feature importance can assist in understanding which characteristics contribute most to a model's ability to forecast. Checking the significance score while making a forecast provides information about that particular model, including which components are most and least important to it [58], [59]. The present study employed scikit-learn's built-in functions for determining feature importance. It is important to note that deep learning models are similar to black boxes and do not provide information on the essential features that contribute to their results. As such, feature importance calculations were not conducted for the DNN, which served as one of the baseline methods. The entire dataset was trained once

for the other models, and then feature importance functions were applied to identify the top 15 features. These features were saved for each model, and subsequent training was performed using datasets that only included the top 15 features. It is worth noting that the remaining attributes varied across the different algorithms [57].

### 1) SHAP FEATURE IMPORTANCE

SHAP (SHapley Additive exPlanations) is a revolutionary technique for calculating feature importance in ML interpretation [60]. This value, operating on many ML models like SVMs and DTree algorithms, is helpful for regression and classification tasks. Using Shapley values, we may determine how to distribute the "payout fairly" (also known as the forecast) across the attributes [59], [60]. The Shapley additive explanations values in this study are estimated using the SHAP Python package. Except for the deep learning baseline model, it is applied to all algorithms and identifies the top 15 significant characteristics for retraining. This indicates that just a few essential traits are chosen for additional training.

This analysis is crucial in determining the root cause of mechanical faults, as it helps identify the underlying factors that lead to the defects. By examining the impact of each feature on the model's predictive performance, we can determine which features are most critical for classification and analyze the root causes.

### D. PROPOSED ALGORITHM

According to preliminary analyses conducted for this study, tree-based ML structures like RF and eXtreme gradient boosting can perform better than identical models [37]. DNNs and ANNs in general also showed encouraging outcomes with increased computing load. As a result, adopting tree-based algorithms is advantageous regarding assessment metrics, accuracy, and computing efficiency [61]. Many tree-based ML models were trained and evaluated to find the top solution on the preprocessed dataset from the PHM [62]. These studies were conducted to reduce the training time and hardware resources needed while increasing the suggested method's accuracy. An ET classification algorithm was chosen as the suggested method in this research since the preliminary tests showed more promising results than similar tree-based approaches.

The ETs method, a member of the ensemble learning algorithms, amalgamates DTs, drawing resemblance to both RFs and bootstrap aggregation (bagging) techniques [61]. Unlike conventional DTs or RFs, the ETs algorithm generates unpruned DTs using the training dataset. Notably, at each split point within a DT, the method employs random sampling of features, similar to RFs, enhancing diversity and reducing overfitting potential [63], [64]. Unlike the RF's greedy method of selecting the optimal division point, the ET classifier opts for a random selection strategy for division points. This characteristic differentiates the ET algorithm from other
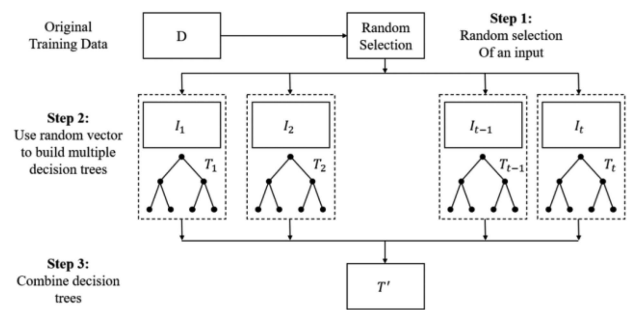


**FIGURE 1.** Visual process of an ET classifier [53].

tree-based models, contributing to its robustness and adaptability. This ET classifier's procedure is shown in Fig. 1.

The implementation of the ET algorithm, facilitated through the scikit-learn package, involves a specific set of hyperparameters crucial in shaping its functionality. The choice of n_estimators=100 delineates the number of trees forming the forest; a higher count can potentially enhance model robustness but might also lead to increased computational demands. The criterion=gini parameter signifies the metric employed to assess the quality of splits in the DTs. Gini, measuring impurity, drives the algorithm to create child nodes by discerning significant differences in label probability distributions within nodes, thus influencing the tree's branching for optimal classification accuracy. Additionally, the min_samples_split=2 parameter denotes the minimum number of examples necessary to divide an internal node. Adjusting this parameter might impact the depth and breadth of DTs, affecting the algorithm's sensitivity to individual data points and, consequently, its overall performance. In this study, the utilization of Gini to evaluate node impurity aligns with the goal of minimizing misclassifications, a pivotal aspect in optimizing the resulting DTs for accurate fault diagnosis and maintenance prediction.

In this implementation, all other arguments have default values. These parameters were improved using a Bayesian optimization approach to achieve the finest outcomes on the PHM's dataset. A systematic optimizing method using the Bayes theorem, known as Bayesian optimization. This method is beneficial in cases where the objective function is expensive to evaluate, and the parameter space is large and complex. In the case of the PHM dataset, the classification problem involves many features and complex relationships, making it challenging to optimize the model's hyperparameters [65]. Therefore, Bayesian optimization is a suitable choice as it effectively balances exploration and exploitation of the parameter space while minimizing the number of evaluations required to reach the optimal set of hyperparameters. It functions by building a substitute or stochastic framework for the target function, which thereafter is efficiently examined through an acquisition function prior to prospective samples being picked to assess the objective function. This method directs a reasonable investigation into a global optimization issue. The model's hyperparameters are optimized via Bayesian

optimization employing the validation data set, which makes up 20% of the initial data. This method efficiently refines the model's hyperparameters, presenting a nuanced and optimized approach that mitigates the complexity inherent in determining these values.

### E. BASELINE ALGORITHMS

The proposed method's results are compared against four baseline algorithms to evaluate its effectiveness from various aspects using different metrics. Since tree-based models performed better than other ML approaches, three baseline models are categorized as tree-based ML algorithms: XGBoost, CATBoost, and Hist gradient boosting classifier. A DNN is also fine-tuned as the fourth baseline since it demonstrated promising results. Similar to the proposed algorithm, these methods are implemented using the Python programming language and its libraries, including scikit-learn, TensorFlow, and Keras.

#### 1) XGBOOST

It has been demonstrated that the XGBoost method is highly adaptable in various learning scenarios, quicker than gradient boosting, and supports regularization techniques. Additionally, parallel processing delivers quicker results in time-sensitive circumstances [32]. This classifier was developed using Python programming and the "xgboost" package. The model predicts that the input data will fall into the category with the maximum likelihood quantity. Additionally, the gb-tree booster was selected as the booster core for the classifier, and after doing numerous tests with various values, the learning rate was set to 0.3.

First, the dataset is scaled and encoded using multiple methods to find the best way to represent features using automated ML (AutoML). For a predictive modeling task, AutoML approaches are strategies to find a high-performing ML model pipeline automatically. Hyperopt-Sklearn and TPOT were the two main AutoML libraries in Sklearn utilized in the current research. Hyperopt-Sklearn uses Bayesian optimization to search through model configurations, while TPOT uses genetic programming to explore a large space of possible pipelines. They are used in this study to find the best way to represent features and improve the performance of the predictive modeling task. As a result of the findings, which indicated that encoding approaches did not significantly enhance classification quality, they were not applied in the first implementations to reduce model time complexity. In other words, the models trained on the dataset did not exhibit a significant difference in performance between the various encoding methods used. Additionally, the classification quality depended more on the choice of the ML algorithm and hyperparameters used in the model rather than the encoding method applied to the data.

#### 2) CATBOOST

The CatBoost ML algorithm from Yandex was just released as open source. It generates cutting-edge results excluding the extensive training dataset required by traditional ML techniques. CatBoost uses several statistics to convert categorical input to numerical values [66]. The "catboost" package of the Python programming language, which includes a CatBoost-Classifier function, is used to build this model. Since earlier sets of hyper-parameters produced poor results, the parameters in this implementation are all set to their default values.

#### 3) HIST GRADIENT BOOSTING CLASSIFIER

Gradient boosting is a statistical framework that expands the capabilities of boosting algorithms such as AdaBoost. It can be used with any loss function, and its ensembles are suitable for solving structured predictive modeling problems, reducing the number of distinct values for each attribute [67]. Gradient boosting, including histogram-based gradient boosting [68], can be used with DT ensembles to speed up the creation of single DTs on large datasets. This study uses the histogram-based gradient boosting approach implemented by scikit-learn, which can be tuned using several parameters such as learning rate, max depth, max iter, and l2 regularization. The loss function used is categorical cross entropy, which is suitable for multilabel classification.

#### 4) DEEP NEURAL NETWORK

As a representation of deep learning techniques, a DNN is used to carry out the categorization. Preprocessing and feature importance steps are omitted, allowing the model to extract features independently. The training procedure is repeated ten times to counteract the impact of random weight initializations. Then the models were compared, demonstrating that random initializations do not affect the final results significantly and that the model can perform the classification task regardless of the initialized weights. Therefore, we specified the random state and generated the final model. An FC neural network with seven hidden layers and nine neurons in the output layer with SoftMax activation is used, having over $180\,000$ trainable parameters. An early stopping mechanism tracks the loss function on the validation set, comprising 20% of the training data, for up to 20 epochs, with maximum training epochs of 150.

## IV. RESULTS AND DISCUSSION

The evaluation technique may assess the models' efficiency once implementations have been completed using the proposed and baseline methods. The introduction of the evaluation criteria for such evaluations, together with an estimation of each model's performance using these metrics, are covered in this section. In the end, it is described how the proposed technique performs better than the baselines regarding accuracy metrics and computing expenses.

**TABLE 1.** Numerical Comparison Between the Implementations on the PHM Challenge's Dataset Among All Models and Regarding all Metrics

| | Accuracy | Precision | Recall | F1 score | Kappa | ROC | Matthew's |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0.997 | 0.996 | 0.995 | 0.995 | 0.994 | 0.999 | 0.994 |
| **CATBoost** | 0.996 | 0.994 | 0.992 | 0.993 | 0.992 | 0.999 | 0.992 |
| **HistGradientBoosting** | 0.997 | 0.996 | 0.995 | 0.995 | 0.994 | 0.999 | 0.994 |
| **Deep neural net** | 0.988 | 0.985 | 0.979 | 0.982 | 0.978 | 0.982 | 0.978 |
| **ExtraTrees Classifier** | 0.995 | 0.997 | 0.989 | 0.993 | 0.991 | 0.999 | 0.991 |
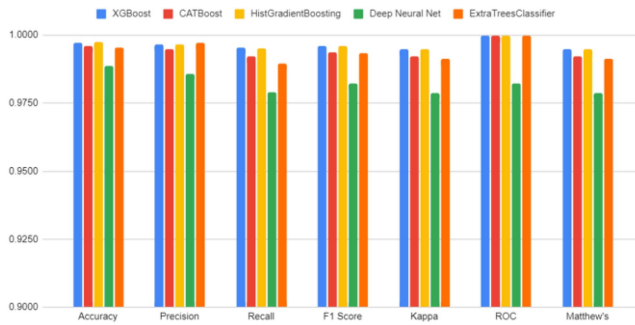


**FIGURE 2.** Visual comparisons of all models' implementations and metrics for all models.

## A. EVALUATION CRITERIA

The preprocessed data from the preceding part was divided into two sets for train and testing in this experiment, with an 80/20 split between the two sets. Each model is then given the training set, allowing them to understand the characteristics of the attributes and map these features into various categories. Once the training phase is through, the resulted algorithms are next fed with the test data. The test set is then sent to each algorithm, and it is assumed that it would use the training data to predict the outcome of each input sample. For each set of inputs, these predictions are compared with the actual results, and the effectiveness of each model is evaluated. Accuracy, precision/recall, the F1-score, ROC, the kappa values, and MATTEW are standard metrics used to assess how well ML models do categorization. The accuracy metric measures the model's overall performance, while precision and recall help identify the models' false-positive and false-negative rates. The F1 score is a harmonic mean of precision and recall, while the ROC curve, kappa, and MATTHEW measure agreement between the predicted and actual class labels. These metrics are appropriate for this research as they objectively evaluate models' performance in categorizing mechanical faults.

## B. EVALUATION RESULTS

The contrast of all models' implementations on the dataset from the PHM challenge is summarized in Table 1 and visualized in Fig. 2. The results show that, with an accuracy of over 99%, the presented ET classification model did remarkably well on the classification problem of this research. Their results for precision, recall, and F1 measures also show
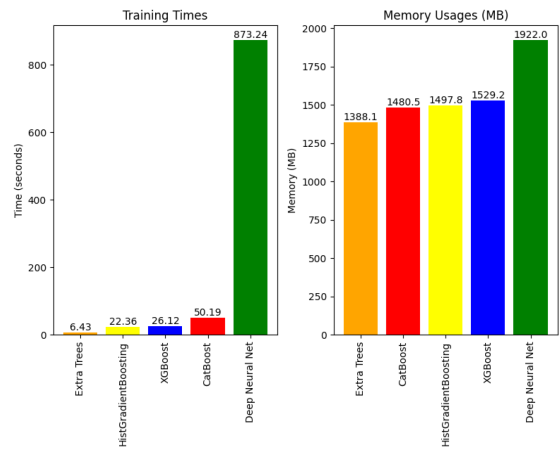


**FIGURE 3.** Evaluation of the training times (seconds) and memory usage (megabytes) for each implementation.

that the model can accurately classify a significant portion of the total relevant outcomes. The Kappa value's proximity to one also suggests a satisfactory association between the classes this model categorizes and those it predicts. The presented technique performs significantly well in precision and achieves equivalent results in the rest of the criteria, despite the proposed models and baseline algorithms producing almost identical results.

The proposed algorithm's key benefits are its quick learning curve and low training resource requirements. Fig. 3 illustrates how the upgraded dataset with the top 15 important characteristics can train the suggested algorithm faster than seven seconds. The following best outcome is devoted to the Hist gradient boosting, which takes 3.5 times as long as the suggested algorithm, demonstrating that the ET methodology delivers equivalent faster. The deep learning model took almost 870 s to perform the same operation as the ET classifier. When considering resource efficiency, examining memory usage provides valuable insights. The memory usage values in megabytes for each model are as follows: ETs (1388.09 MB), CatBoost (1480.48 MB), HistGradientBoosting (1497.77 MB), XGBoost (1529.22 MB), and deep neural net (1922.04 MB). Notably, the ETs model stands out as the most memory-efficient, using significantly less memory compared with other models. This demonstrates that the ETs classifier excels not only in training speed but also in terms of resource usage, making it a robust choice for efficient fault diagnosis. The causation of each failure is investigated using a root cause analysis, which reveals that many of these flaws are caused by minor problems like humidity and temperature. This stage demonstrates that the proposed model does not operate as a black box and that users may identify the causes of each failure, which is another benefit over DNNs.

## C. ROOT CAUSE ANALYSIS

Root cause analysis is a method to locate and examine the causes of problems. To determine which factors relate to what
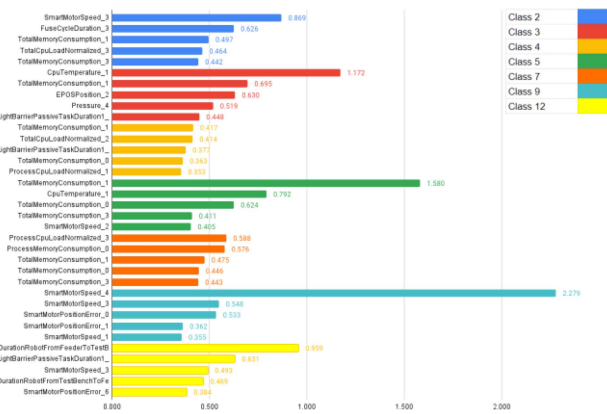
**FIGURE 4.** Top five characteristics in defective categories based on the relevance of the SHAP features.

**TABLE 2.** Most Influential Signals for Each Class

| Class | Most important signals | Fault interpretation |
|-------|------------------------|----------------------|
| 2 | FuseCycleDuration_3<br>FeederAction2_0 | The fuses of this category have shorter cycles |
| 3 | CpuTemperature_1<br>Pressure_4 | The controller and data collection machine's average CPU temperature is higher than typical |
| 4 | TotalCpuLoadNormalized_2 | Greater CPU trends based on derivatives |
| 5 | TotalMemoryConsumption_0<br>ProcessMemoryConsumption_0 | Variations in the amount of data that are acquired for memory utilization |
| 7 | TotalMemoryConsumption_1<br>Humidity_0<br>Temperature_0 | The distinction between environmental variables and memory usage earned value |
| 9 | SmartMotorSpeed_0<br>SmartMotorPositionError_0 | Decrease in the amount of information collected for motor speed |
| 11 | SmartMotorSpeed_3<br>SmartMotorPositionError_3<br>DurationRobotFromFeederToTestBench_6 | Increase in motor speed value; as a result, movement of the feeder to the test bench takes less time |
| 12 | DurationRobotFromFeederToTestBench_6 | A longer measuring period with a higher sampling rate |

issues and which might cause the problem's source, ML models group the problems and examine the reasons that impact them. Based on the characteristics causing mistakes in each category of faults, the underlying causes are considered in this study. Visualizations show that several common problems arise in various faults, and several errors in each category cause defects. The fundamental causes were examined in a class-based technique, and the most significant factors were identified. The top five qualities that significantly impact the defective classes are presented in Fig. 4 to conclude this part. Common elements like humidity and temperature are excluded except for the classes on which they have the most significant influence, and unique characteristics are considered in each defect category. Therefore, for each fault class, only distinctive characteristics are considered.

### D. FAULT DIAGNOSIS

This article's main objective is identifying errors with intelligent methods. Now that the suggested algorithm's strength has been shown, this approach may be used to diagnose faults. These conclusions were reached based on the feature significance values computed for each defect class and the root cause analysis. In other words, the underlying causes of a problem are indicated by the crucial signals in the feature importance investigation. More data analysis is being done to comprehend better the issues' grounds and how they are physically interpreted, especially for the most important signals. These essential signals for each fault category are listed in Table 2. It lists these signals responsible for the greatest number of cases in each class's examples. Even though specific signals may have high significance levels, they often do not cause errors. For instance, standard signals like humidity cannot be the primary factor in any class since they are present in all categories. Only class 7 exhibits this phenomenon; even then, temperature and humidity are the main fault-causing factors. In addition, the feature importance analysis demonstrated that humidity is not among the most critical signals for any other class, implying that other signals have a higher impact on

the occurrence of defects. Overall, these results highlight the complexity of fault diagnosis and emphasize the importance of considering the interrelationships among different signals to identify the underlying causes of defects accurately.

It is noteworthy that root cause analysis and fault diagnosis are distinct but complementary processes in the field of predictive maintenance. Fault diagnosis involves identifying the type and location of a fault in a system, while root cause analysis seeks to determine the underlying cause or causes. In this article, the feature significance analysis and root cause analysis results are used to support the fault diagnosis process, providing insights into the most critical signals for each fault class.

### V. CONCLUSION

This study suggested using an ET classifier to identify mechanical failures. In the experimental section of this article, various models were tested to evaluate their performance in identifying defects in the manufacturing process. These experiments showed that tree-based and deep-learning methods are able to obtain the most promising outcomes between all potential techniques. Therefore, based on the experimental findings, it was concluded that these two types of models could obtain the best results among all possible methods. The outcomes support these experiments, with each model's accuracy levels above 99%. Accordingly, four tree-based algorithms trees were chosen: the eXtreme gradient boosting, CATBoost, histogram gradient boosting, and the ET classification model; and the SHAP (SHapley Additive exPlanations) technique was used to preprocess the data fed to them. An FC DNN was correspondingly developed to complete the identical job deprived of the preprocessing steps. An assessment procedure was created to evaluate the models' performance concerning multiple indicators. Investigations proved that the five selected strategies could attain more than 99% accuracy and that their outcomes are generally equivalent.

The proposed method has advantages in low processing resource usage and quick training times. Compared with the DNN, which took almost 870 s, the ET classifier only took less than 7 s. Additionally, the proposed algorithm is reliable in classification quality and effective in computing power. A root cause analysis revealed minor problems like humidity and temperature as the cause of many failures, demonstrating that the suggested algorithm is not a black-box and allows for identifying failure causes, another benefit over DNNs. However, it is essential to recognize the proposed method's limitations in practical applications. The effectiveness of this approach heavily depends on the quality and quantity of available data. In situations with limited data, the performance might not be as robust. Additionally, as with many ML techniques, the complexity of the model can increase with a high number of features, potentially affecting its efficiency. Also, while our approach offers causality analysis, it may not be as interpretable as simpler models like simple DTs with lower accuracies.

The forthcoming IFD studies are set on a route that uses more data. Generative models might be utilized to enhance the quantity of data the models acquire and extend the datasets. To benefit from robust models' expertise with the IFD process, transfer learning methodologies can also be used with pretrained models. Unsupervised learning techniques are also able to be used with supervised procedures to enhance the examination of every fault's cause and the consequent setups. Data samples from each class may be analyzed using clustering algorithms, which group them according to the structures that can fail.

Furthermore, researchers can use Bayesian optimization techniques to enhance the model's hyperparameters. However, accomplishing these techniques might be challenging and impractical for mechanical engineers with no technical foundation in ML engineering. In this area, AutoML techniques might be helpful to speed up the ML procedure and reduce the difficulties for mechanical professionals.

## REFERENCES

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

[2] T. Van Tung and B. Yang, "Machine fault diagnosis and prognosis: The state of the art," *Int. J. Fluid Machinery Syst.*, vol. 2, no. 1, pp. 61–71, 2009.

[3] 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.

[4] M. Cerrada, G. Zurita, D. Cabrera, R. V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mech. Syst. Signal Process.*, vol. 70–71, pp. 87–103, 2016, doi: 10.1016/j.ymssp.2015.08.030.

[5] J. Zhang, Y. Li, Z. Zhang, Z. Zeng, Y. Duan, and Y. Cao, "XG-Boost classifier for fault identification in low voltage neutral point ungrounded system," in *Proc. IEEE Sustain. Power Energy Conf.*, 2019, pp. 1767–1771.

[6] Z. M. Çınar, A. A. Nuhu, Q. Zeeshan, and O. Korhan, "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," *Sustainability*, vol. 12, 2020, Art. no. 8211.

[7] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 72–73, pp. 92–104, May 2016, doi: 10.1016/j.ymssp.2015.11.014.

[8] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017, doi: 10.1109/TIE.2016.2627020.

[9] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, Dec. 2017, doi: 10.1016/j.measurement.2017.07.017.

[10] F. Calabrese, A. Regattieri, M. Bortolini, M. Gamberi, and F. Pilati, "Predictive maintenance: A novel framework for a data-driven, semi-supervised, and partially online prognostic health management application in industries," *Appl. Sci.*, vol. 11, no. 8, pp. 1–28, 2021, doi: 10.3390/app11083380.

[11] L. Biggio, M. Russi, S. Bigdeli, I. Kastanis, D. Giordano, and D. Gagar, "PHME data challenge," in *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, 2021. [Online]. Available: https://github.com/PHMEDatachallenge/Data-Challenge-2021#citation

[12] W. D. Yingjie Song, G. Zhao, B. Zhang, H. Chen, and W. Deng, "An enhanced distributed differential evolution algorithm for portfolio optimization problems," *Eng. Appl. Artif. Intell.*, vol. 121, 2023, Art. no. 106004.

[13] W. D. Huayue Chen, T. Wang, and T. Chen, "Hyperspectral image classification based on fusing S3-PCA, 2D-SSA and random patch network," *Remote Sens.*, vol. 15, no. 13, 2023, Art. no. 3402.

[14] W. D. Xiangbing Zhou, X. Cai, H. Zhang, Z. Zhang, T. Jin, and H. Chen, "Multi-strategy competitive-cooperative co-evolutionary algorithm and its application," *Inf. Sci.*, vol. 635, pp. 328–344, 2023.

[15] X. Li, H. Zhao, and W. Deng, "BFOD: Blockchain-based privacy protection and security sharing scheme of flight operation data," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2023.3296460.

[16] G. Dorgo, A. Palazoglu, and J. Abonyi, "Decision trees for informative process alarm definition and alarm-based fault classification," *Process Saf. Environ. Protection*, vol. 149, pp. 312–324, 2021, doi: 10.1016/j.psep.2020.10.024.

[17] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Rel. Eng. Syst. Saf.*, vol. 115, pp. 124–135, 2013, doi: 10.1016/j.ress.2013.02.022.

[18] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, 2017, Art. no. 425, doi: 10.3390/s17020425.

[19] 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, 2018, doi: 10.1016/j.ymssp.2018.02.016.

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

[21] E. T. Maddalena, M. W. F. Specq, V. L. Wisniewski, and C. N. Jones, "Embedded PWM predictive control of DC–DC power converters via piecewise-affine neural networks," *IEEE Open J. Ind. Electron. Soc.*, vol. 2, pp. 199–206, 2021, doi: 10.1109/OJIES.2021.3058411.

[22] A. Bernieri, M. D'Apuzzo, L. Sansone, and M. Savastano, "A neural network approach for identification and fault diagnosis on dynamic systems," *IEEE Trans. Instrum. Meas.*, vol. 43, no. 6, pp. 867–873, Dec. 1994.

[23] C. Kara Mostefa Khelil, B. Amrouche, A. Soufiane Benyoucef, K. Kara, and A. Chouder, "New intelligent fault diagnosis (IFD) approach for grid-connected photovoltaic systems," *Energy*, vol. 211, pp. 1–18, 2020, doi: 10.1016/j.energy.2020.118591.

[24] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015, doi: 10.1109/TIE.2015.2419013.

[25] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mech. Syst. Signal Process.*, vol. 18, pp. 625–644, 2004, doi: 10.1016/S0888-3270(03)00020-7.

[26] V. Ã. Sugumaran, V. Muralidharan, and K. I. Ramachandran, "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing," *Mech. Syst. Signal Process.*, vol. 21, pp. 930–942, 2007, doi: 10.1016/j.ymssp.2006.05.004.

[27] M. Piliougine, R. A. Guejia-Burbano, and G. Spagnuolo, "Detecting partial shadowing and mismatching phenomena in photovoltaic arrays by machine learning techniques," *IEEE Open J. Ind. Electron. Soc.*, vol. 3, pp. 507–521, 2022, doi: 10.1109/OJIES.2022.3208140.

[28] 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, 2015, doi: 10.1016/j.measurement.2015.03.017.

[29] S. Datta and S. Sarkar, "A review on different pipeline fault detection methods," *J. Loss Prevention Process Industries*, vol. 41, pp. 97–106, 2016, doi: 10.1016/j.jlp.2016.03.010.

[30] A. Stetco et al., "Machine learning methods for wind turbine condition monitoring: A review," *Renewable Energy*, vol. 133, pp. 620–635, Apr. 2019, doi: 10.1016/j.renene.2018.10.047.

[31] P. Akulwar, S. Pardeshi, and A. Kamble, "Survey on different data mining techniques for prediction," in *Proc. 2nd Int. Conf. I-SMAC (IoT Social, Mobile, Analytics, Cloud) (I-SMAC)*, 2018, pp. 513–519, doi: 10.1109/I-SMAC.2018.8653734.

[32] Z. Wu, M. Zhou, Z. Lin, X. Chen, and Y. Huang, "Improved genetic algorithm and XGBoost classifier for power transformer fault diagnosis," *Front. Energy Res.*, vol. 9, 2021, Art. no. 745744, doi: 10.3389/fenrg.2021.745744.

[33] H. Shafizadeh-Moghadam, A. Asghari, A. Tayyebi, and M. Taleai, "Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth," *Comput. Environ. Urban Syst.*, vol. 64, pp. 297–308, 2017, doi: 10.1016/j.compenvurbsys.2017.04.002.

[34] R. Yan, Z. Ma, Y. Zhao, and G. Kokogiannakis, "A decision tree based data-driven diagnostic strategy for air handling units," *Energy Buildings*, vol. 133, pp. 37–45, 2016, doi: 10.1016/j.enbuild.2016.09.039.

[35] Y. Zhao, L. Yang, B. Lehman, J.-F. de Palma, J. Mosesian, and R. Lyons, "Decision tree-based fault detection and classification in solar photovoltaic arrays," in *Proc. 27th Annu. IEEE Appl. Power Electron. Conf. Expo.*, 2012, pp. 93–99.

[36] J. Han, M. Kamber, and J. Pei, "Data mining: Concepts and techniques," Data Mining Concepts Tech., pp. 317–319, 2012, doi: 10.1016/C2009-0-61819-5.

[37] M. G. Alfarizi, J. Vatn, and S. Yin, "An extreme gradient boosting aided fault diagnosis approach: A case study of fuse test bench," *IEEE Trans. Artif. Intell.*, vol. 4, no. 4, pp. 661–668, Aug. 2023, doi: 10.1109/tai.2022.3165137.

[38] L. Puggini, J. Doyle, and S. McLoone, "Fault detection using random forest similarity distance," *IFAC-PapersOnline*, vol. 48, pp. 583–588, 2015, doi: 10.1016/j.ifacol.2015.09.589.

[39] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning," *Nature*, vol. 1, no. 2, pp. 185–186, 2016.

[40] C. Lu, Z. Wang, W. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, 2017, doi: 10.1016/j.sigpro.2016.07.028.

[41] S. Taheri and S. Aliakbary, "Research trend prediction in computer science publications: A deep neural network approach," *Scientometrics*, vol. 127, pp. 849–869, 2022, doi: 10.1007/s11192-021-04240-2.

[42] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, 2019, doi: 10.1016/j.ymssp.2018.05.050.

[43] G. Hu, T. Zhou, and Q. Liu, "Data-driven machine learning for fault detection and diagnosis in nuclear power plants: A review," *Front. Energy Res.*, vol. 9, pp. 1–12, 2021, doi: 10.3389/fenrg.2021.663296.

[44] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vol. 72–73, pp. 303–315, May 2016, doi: 10.1016/j.ymssp.2015.10.025.

[45] A. Sharida, N. Kamal, H. Alnuweiri, S. Bayhan, and H. Abu-Rub, "Digital twin-based diagnosis and tolerant control of T-type three-level rectifiers," *IEEE Open J. Ind. Electron. Soc.*, vol. 4, pp. 230–241, 2023, doi: 10.1109/ojies.2023.3290169.

[46] Z. Chen, C. Li, and R. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock Vib.*, vol. 2015, 2015, Art. no. 390134.

[47] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018, doi: 10.1109/TIE.2017.2774777.

[48] C. Lu, Z. Wang, and B. Zhou, "Advanced engineering informatics intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Adv. Eng. Inform.*, vol. 32, pp. 139–151, 2017, doi: 10.1016/j.aei.2017.02.005.

[49] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, 2018, doi: 10.1016/j.ymssp.2017.06.022.

[50] X. Yang, L. Shu, K. Li, Z. Huo, and Y. Zhang, "SA1D-CNN: A separable and attention based lightweight sensor fault diagnosis method for solar insecticidal lamp Internet of Things," *IEEE Open J. Ind. Electron. Soc.*, vol. 3, pp. 291–303, 2022, doi: 10.1109/OJIES.2022.3172899.

[51] H. Zhao et al., "Intelligent diagnosis using continuous wavelet transform and Gauss convolutional deep belief network," *IEEE Trans. Rel.*, vol. 72, no. 2, pp. 692–702, Jun. 2023.

[52] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[53] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Trans.*, vol. 77, pp. 167–178, Jun. 2018, doi: 10.1016/j.isatra.2018.04.005.

[54] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, 2017, Art. no. 273, doi: 10.3390/s17020273.

[55] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, Jul. 2016, doi: 10.1016/j.measurement.2016.04.007.

[56] G. Liu, H. Bao, and B. Han, "A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis," *Math. Problems Eng.*, vol. 2018, 2018, Art. no. 5105709, doi: 10.1155/2018/5105709.

[57] A. Zien, N. Krämer, S. Sonnenburg, and G. Rätsch, "The feature importance ranking measure," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2009, vol. 5782, pp. 694–709, doi: 10.1007/978-3-642-04174-7_45.

[58] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[59] R. K. Patel and V. K. Giri, "Feature selection and classification of mechanical fault of an induction motor using random forest classifier," *Perspectives Sci.*, vol. 8, pp. 334–337, 2016, doi: 10.1016/j.pisc.2016.04.068.

[60] D. Bowen and L. Ungar, "Generalized SHAP: Generating multiple types of explanations in machine learning," 2020, *arXiv:2006.07155*.

[61] Y. A. Alsariera, V. E. Adeyemo, A. O. Balogun, and A. K. Alazzawi, "AI meta-learners and extra-trees algorithm for the detection of phishing websites," *IEEE Access*, vol. 8, pp. 142532–142542, 2020, doi: 10.1109/ACCESS.2020.3013699.

[62] B. Wu, B. Zhang, W. Li, and F. Jiang, "A novel method for remaining useful life prediction of bearing based on spectrum image similarity measures," *Mathematics*, vol. 10, no. 13, pp. 1–10, 2022, doi: 10.3390/math10132209.

[63] C. Chen, N. Wang, and M. Chen, "Prediction model of end-point phosphorus content in consteel electric furnace based on PCA-extra tree model," *ISIJ Int.*, vol. 61, no. 6, pp. 1908–1914, 2021, doi: 10.2355/isijinternational.ISIJINT-2020-615.

[64] N. Chakrabarty and S. Biswas, "Navo minority over-sampling technique (NMOTe): A consistent performance booster on imbalanced datasets," *J. Electron. Inform.*, vol. 2, pp. 96–136, 2020, doi: 10.36548/jei.2020.2.004.

[65] Y. Zhao, J. Wen, F. Xiao, X. Yang, and S. Wang, "Diagnostic Bayesian networks for diagnosing air handling units faults—Part I: Faults in dampers, fans, filters and sensors," *Appl. Thermal Eng.*, vol. 111, pp. 1272–1286, 2017, doi: 10.1016/j.applthermaleng.2015.09.121.

[66] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 6639–6649.

[67] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: Gradient boosting with categorical features support," 2018, *arXiv:1810.11363*.

[68] Y. Shi, J. Li, and Z. Li, "Gradient boosting with piece-wise linear regression trees," 2019, *arXiv:1802.05640*.