Journal of Artificial Intelligence and Technology, 2026, 6, 328-336
https://doi.org/10.37965/jait.2026.0939

ISTP
RESEARCH ARTICLE

A Hybrid Ensemble and Explainable AI Framework for Predictive Maintenance of Industrial Equipment

Miguel Villagómez-Galindo,[1] G. Manjula,[2] B. N. Jagadeesh,[3] Madhu Patil,[4] Saiprasad Potharaju,[5] and N. Achyutha Prasad[6]

[1]Universidad Michoacana de San Nicolás de Hidalgo, Facultad de Ingeniería Mecánica, Morelia, Santiago Tapia 403, México
[2]Dept of Computer Science and Engineering, BGSCET, Bengaluru, Karnataka, India
[3]Dept of CSE-CY(Cybersecurity), RNS Institute of Technology, Dr Vishnuvardhan Road, Bengaluru, Karnataka, India
[4]Dept of Computer Science and Design, Bengaluru, Karnataka, India
[5]Department of CSE, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, MH, India
[6]Department of Computer Science and Engineering, East West Institute of Technology, Bengaluru, Karnataka, India

(Received 21 September 2025; Revised 23 December 2025; Accepted 21 January 2026; Published online 10 February 2026)

Abstract: A modern industrial system with its critical machinery is very sensitive to unexpected equipment failure and may experience extensive operation interruption, danger to safety, and cost. The traditional maintenance approaches, reactive and preventative, lack intelligence and flexibility to make predictions of the failure based on real-time information, leading to failures that are expensive to fix or unnecessary maintenance. This paper proposes a hybrid ensemble predictive maintenance (PdM) framework to assist in overcoming these drawbacks by combining potent machine learning (ML) models as classifiers to PdM and SHapley Additive eXplanation (SHAP) framework to make decision-making PdM transparent and interpretable. The suggested approach is trained on actual industrial sensor data comprising multivariate time-series data such as temperature, vibration, voltage, and pressure measurements. Data are preprocessed in a powerful way with the removal of redundancy, label encoding, and scaling. The accuracy, precision, recall, F1-score, and analysis of the confusion matrix are used to evaluate each model. Strikingly, the three ensemble classifiers had 100 percent success in the detection of faults, with SHAP values having obvious key features dictating forecasts. The newness of this method is that it is potentially high-accuracy and interpretable at the same time, which is a respite from deeper or federated learning models, which are typically high-computational-load methods. The study adds a scalable, accurate, and explainable PdM framework that can be part of the new smart manufacturing.

Keywords: ensemble; industrial equipment; predictive maintenance; sensors; XAI

I. INTRODUCTION

The need to achieve intelligent, automated, and fault-tolerant systems has never been more expedient in the current industrial environment that has been digitally transformed. The critical assets that are most used include motors, compressors, and transformers in industrial sectors, like manufacturing, energy, and transportation. The unexpected breakdown of this type of equipment may lead to huge production losses, expensive repairs, and even to safety risks [1]. As the industry upgrades to the new type, predictive maintenance (PdM) has since become a preventive measure that involves sensor data and ML to determine the event of equipment failure prior to failure and optimize asset life and operating incentives [2]. Reactive and time-based PdM are being viewed more and more as inadequate solutions to the maintenance problem. Although reactive maintenance is very expensive in terms of repair, preventive maintenance can result into needless servicing. Conversely, PdM enables maintenance to be carried out at the exact point of time as required, in accordance with the current sensor data and advanced analytics.

Some of the ML models that have been used in PdM issues over the last decade are support vector machines (SVMs), decision trees, and neural networks. In more recent years, methods such as the ensemble algorithms, such as Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), have demonstrated potential owing to their robustness, interpretability, and ability to handle heterogeneous data [3]. There are, however, a number of challenges: (i) a good deal of ML models are considered as “black boxes” that cannot be more or less interpreted, (ii) models trained on datasets generated in labs or in the domain of interest are often poorly generalizable, and (iii) the model decision-making process is notoriously opaque (especially in the industrial context).

In order to fill these gaps, explainable artificial intelligence (XAI) has been suggested, especially SHapley Additive exPlanation (SHAP) to remove the black-box issue to enable transparent and trustworthy ML decisions. SHAP can not only assist in diagnosing the most influential features that help in the predictions but also help maintenance engineers in identifying the cause of faults.

Corresponding author: Saiprasad Potharaju; Email: saiprasad.potharaju@sitpune.edu.in

Although this area is gaining interest, even in many studies, concerning PdM synthetic/simulated datasets are used, and this narrows the extent of their application in the real world [6]. Also, there are few works that show a genuinely generalizable and explainable PdM framework based on real-world sensor readings in industries.

The quest to achieve smart manufacturing and the automation of industries and sustainable operations has placed a high priority on ensuring the presence of a good maintenance system in the continuous industrial transformation being experienced in the present days. In many industries, including energy, automotive, manufacturing, and logistics, the availability of critical industrial equipment is essential, whether internal production compressors, motors, bearings, or turbines. Unscheduled breakdowns are not only expensive since they result in a lot of money loss on production downtimes and emergency repairs but are also very dangerous to human life. It is a situation that requires a more active approach to the maintenance of equipment, and it led to the appearance of the paradigm of PdM. PdM operates on real-time analytics of data and machine learning (ML) to forecast equipment failure prior to interruption, hence allowing preventive interventions, optimum life cycles of assets, and reliable operations [1].

There are two general types of traditional maintenance practices: reactive (carried out after failure) and preventive (carried out before failure). Even though reactive maintenance is least planned, they may cause severe breakdowns and increase the amount of repair. Preventive measures will interfere with production since they are safer, leading to unnecessary maintenance, resource wastage, and incurring added costs. PdM sits between these two extremes, however, and proactively plans and schedules interventions taking advantage of real-time data streams and smart algorithms on condition, in real equipment, and usage trends [2]. In that regard, it serves as one of the central components of the Industry 4.0 environment where connected devices, cyber-physical systems, and intelligent analytics transform the process of decision-making in industries.

A lot of different ML methods, including SVMs and decision trees as well as neural networks, have been used in PdM over the past decade. Nevertheless, more recently, ensemble algorithms such as RF, XGBoost, and LightGBM have attracted considerable interest because they outperform in terms of generalization recovery, resistance to noise, and the capacity to model intricate feature relationships [3]. Though they have good predictive capabilities, most of these models lack transparency, and this restricts their reliability in benchmark industrial settings. The black-box nature of model choices restricts diagnostic reasoning, root-cause analysis, and acceptance by the operator and is one of the most frequent in the literature [4].

That is why, to mitigate this drawback, XAI tools like SHAP have come out now as a crucial method of opening up ML model decisions. SHAP values provide an ethical way of assessing the outcomes of prediction using the attributes of individual features and cooperative game theory. In the world of PdM, it translates to serving human-readable perspectives on which readings of which sensor (e.g., temperature spikes and voltage anomalies) are the most important with regard to predicting failures. Having such explainability not only increases transparency in operations but also improves the speed of corrective procedures; thus, it enables maintenance planning and improves on human–machine cooperation [5].

Nevertheless, there are still a number of issues in the implementation of the PdM solutions in an industrial real-life setting. Most of the current studies are based on either synthetic, simulated, or laboratory data, which do not reflect variability and noise found in an operating environment [6]. In addition, model interpretability is typically traded off against the performance. For example, a good predictive performance of high-performing deep learning systems can be poor, opaque, expensive to execute, and practically inapplicable in edge settings or real-time monitoring applications.

This study proposes a hybrid ensemble learning framework integrated with SHAP-based explainability for PdM of industrial equipment. The system is validated on a publicly available real-world dataset Sensor Maintenance Dataset by Ziya *et al.* (2024) which includes multivariate sensor readings from industrial equipment with annotated fault labels.

The key contributions of this study are as follows:

- Design and implementation of a hybrid model combining RF, XGBoost, and LightGBM for robust classification.
- Integration of SHAP to provide interpretability and root-cause analysis of faults.
- Achievement of 100% classification accuracy with visualizations including SHAP summary plots and confusion matrices.
- Comparative evaluation against recent literature, highlighting performance and explainability gains.

This task is to fill in the discontinuity between good performance predictive models and real-life industrial systems implementation, which is clear and accessible. This way, it will help to create more sustainable, supportable, and smart factories in the vision of Industry 4.0.

The originality of this article is that the design and development of a lightweight but high-performing hybrid ensemble model representing RF, XGBoost, and LightGBM classifiers have been developed with an additional enhancement of SHAP-based interpretability. The framework affirms a perfect classification accuracy (100%) in conjunction with provision of actionable insights by way of global and local SHAP visualizations. This mix of transparency, functionality, and high applicability of our work makes the difference from previous solutions, particularly those based on deep or federated architectures involving more deployment complexity.

The rest of this paper is organized as follows: in Section II, the data, preprocessing plan, and structure of the suggested hybrid ensemble system are outlined. Section III expounds on the theoretical bases of the ML and XAI methods employed. Section IV contains the experimental results, the metrics of the performance of the models, and the outputs of the interpretability. Section V provides us with a discussion comparing our findings with the literature available and a description of the practical implication. Section VI is finally a conclusion of the paper. It summarizes the contributions and a prospective direction of future research that could be done to develop the future of PdM systems in an industrial environment to increase their scale and robustness.

II. LITERATURE REVIEW

PdM has evolved as part and parcel of the latest industrial systems, and through PdM, it is intended to limit the number of unexpected downtimes, boost up the efficiency of operation, and also guarantee the equipment life of service. The legacy maintenance approaches (running-to-failure) and schedule-based (scheduled) maintenance efforts have been dominated, to a great extent, by the utilization of data-driven maintenance solutions, which use sensor readings and ML to predict equipment failure prior to occurrence. Many research papers have been devoted to the PdM implementation with different approaches based on the use of computational intelligence methods which have their own strengths and limitations.

At the outset, simple linear models and statistical regression tools such as regression analysis were commonly used to forecast the health of equipment. Such methods, however, were not suitable in many cases to work with the nonlinear trends and hidden dependencies contained in the industrial data collected with sensors [1]. Case in point, Aminzadeh *et al.* used linear regression to keep track of the health of the compressor, but the work confirmed poor accuracy and failed to provide the means to do classification or detailed interpretability [2]. Analogously, the appropriate use of the rule-based systems and systems that trigger thresholds did not suit real-time particularities of industrial objects usage when equipment behavior was unconventional compared to the historical trends. In order to address these issues, the industry saw the increasing amount of supervised ML algorithms, mainly those involving decision trees, SVMs, and the k-nearest neighbors (KNN) algorithms. SVMs had shown to have strong generalization performance in the binary classification cases, although both high-dimensional and probabilistic output were precluded in the light of subtle fault diagnosis [3]. On the contrary, decision trees offered more comprehensibility but were much more prone to overfitting and were unstable on small data [4]. KNN algorithms had promise when it comes to capturing local data structures and were computationally cost prohibitive when used with large-scale PdM datasets and extremely sensitive to noise [5].

PdM made great progress with arrival of ensemble methods. RF was a strong solution because it had the ability to deal with all heterogeneous features and counter overfitting owing to bootstrapping and feature bagging [6]. LightGBM and XGBoost served further performance experts by boosting the method, where models appeared to rectify past learners failures [7]. As an example, Babadi *et al.* used XGBoost in Heating, Ventilation, and Air Conditioning (HVAC) systems and even contended with a higher F1-score of about 93 percent compared to single models [8]. The weakness, however, was a lack of generalization since the training was domain-specific, and little information was available about the model's decision-making, which is a crucial need in critical industries involving high stakes. The idea of neural networks and deep learning also became popular recently. Time-series sensors have been used to detect sensor faults and remaining useful life (RUL) by applying convolutional neural networks (CNNs) and long short-term memory (LSTM) models [9]. Nascimento *et al.* presented T4PdM, a transformer-based deep neural network that showed an accuracy of about 99 percent in the fault prediction on the datasets related to bearings [10]. Although these results are encouraging, one of the criticisms that can be encountered against deep learning models is that they are black boxes, quite expensive in computation, and require an abundance of labeled data. This limits their use in constrained or edge computing space. Also, most deep learning models have been trained using controlled laboratory data, thus constraining their performance in acquiring noise and variations in the industrial environment [11].

Explainability is an essential part of modern PdM systems as it helps cover the essential lack of trust and interpretability of decisions made using AI. With industrial applications, operators and maintenance staffs need transparent reasons behind model outputs to execute the right measures. Such methods of explainable AI, such as SHAP, Local Interpretable Model-agnostic Explanations (LIME), and Partial Dependence Plots (PDP), have gained popularity in that regard. As an example, Gawde *et al.* used multi-sensor fusion with LIME and SHAP to provide practical information about the machine faults [12]. In the same way, Kisten *et al.* offered explainable models of agricultural facilities in terms of SHAP to define the causes of failure. Although these works improve the level of trust in PdM systems, the majority of the implementations attention aims at shallow models or is not capable of visualizing both the local and global impacts [13].

The other promising trend is the use of federated learning and privacy-preserving AI in PdM. The federated deep learning system presented by Garcia *et al.* implements explainability as an end-to-end federated deep learning framework in distributed manufacturing environments, focusing on safeguarding valuable machine data without compromising the model predictivity [14]. Nonetheless, federated models have also brought about heavy computational and communication overheads to its own disadvantage in terms of real-time deployment. What is more, the interpretation of models in such decentralized settings has an even more complicated nature, since there is different data at the different locales and there is no common explanation [15]. There are also recent reports that discuss the combination of generative AI with digital twins in Industry 4.0 and 5.0 applications when it comes to PdM. MikoLajewska *et al.* introduced the idea of applying generative models in simulating equipment behavior and training fault diagnosis algorithms and also expressing explainability with counterfactual explanations [16]. These solutions are theoretically strong but have not been used in concrete industries yet. The scarce presence of standardized datasets, functions to vet such systems, and study cases of deployment constrains a lot of the actual applicability of such systems [17].

The other drawback to existing studies in the current literature is that the benchmark or synthetic datasets have been overly relied upon, and they are not practical enough to represent the real-life expectations. The bearing data used to validate many PdM models is either NASA-supplied bearing data or simulated sensor readings consisting of clean and well-labeled data and do not resemble the actual operational noise, missing values, cross-component correlations in a real setting [18]. Real-world validation is lacking, and this causes high-performance statistics and does not ensure a robust result when practical application is made. Moreover, very little research focuses on the cross-factory generalization, that is, using a model trained at one facility in another piece of equipment or operating condition. This forms a huge obstacle to a broad PdM adoption.

Regarding the metrics of evaluation, there is a tendency to consider the existing studies based on accuracy or F1-score, ignoring the interpretability metrics and the actionability. As an instance, failure models that can only predict failure without specifying the sensor or operational factor involved can be of no help to the maintenance teams. Such a context value-less description is an impediment to root-causal analysis and decision-making. Although tools such as SHAP are trying to eliminate this gap, not every study utilizes them as efficiently and offers the obtained conclusions that could be presented to an industrial stakeholder in an easily digestible manner [19].

In order to overcome the shortcomings of the separate methods, various hybrid models have been put forward. They are hybrid classical ML algorithms and XAI or optimization. Potharaju *et al.* offered a more modular and interpretable two-step ML pipeline that works for anomaly detection and maintenance scheduling [20]. Similarly, Shtayat *et al.* proposed an explainable deep learning ensemble method to detect industrial intrusion. Yet, many of these systems have no real-world confirmation or are excessively complex, or obsolete hardware is required, and thus cannot be applied to SMEs or even legacy devices [21]. Although there is a substantial literature on PdM based on ML, there are still some important gaps in it. First, most models are not interpretable, since they value performance and do not fit critical industrial applications where human-in-the-loop validation is required. Second, even though the techniques with a deep learning approach may be accurate, they consume a lot of resources and may not be transparent. Third, methods of explainability are *post hoc* or are not a part of the entire modeling process. Lastly, hardly any of the models are tested on real-world, noisy, and diverse sensor data, which also restrict its scaling and applicability in industries.

Such a backdrop highlights the necessity of such a hybrid framework that could produce nothing but a state-of-the-art performance [22] in terms of predictive performance but also guarantees the interpretation, generality, and applicability in the real world. This gap was filled in the current work by building an ensemble and including RF, XGBoost, and LightGBM models with a SHAP-based explanation tool and training and validating it on a real industrial sensor data. This research addresses the gap between industrial usability and academic innovation by operating at a much higher level of accuracy without compromising the transparency of the decision process in the model [23].

We have approached the current issue in a manner that eliminates the key weaknesses experienced by the previous research, especially trade-off between performance and interpretability, and use of synthetic data. We consider an ensemble of three state-of-the-art classifiers, such as RF, XGBoost, and LightGBM, which are tuned on an industrial sensor data of real-word front-line equipment. Remarkably, our framework is lightweight, scalable, and can be conveniently incorporated into an industrial environment, unlike deep neural models that need extensive tuning and rely on resources such as hardware. In addition to this transparency gap, we integrate SHAP-based explainability into the whole pipeline, providing information about the fault prediction both globally and locally. This has a good balance of being highly accurate in [24] classification while keeping it easily interpretable, so that, besides being technically robust, it becomes more practically actionable by maintenance engineers and domain experts. By that, we directly add to the field of explainable, efficient, and verified under-real-world PdM solutions [25,26].

III. MATERIALS AND METHODS

A. PROPOSED METHODOLOGY

Figure 1 represents a proposed multi-stage methodology where the PdM framework is interpretable and encompasses a multi-stage pipeline including data acquisition to reach its goal. The whole thing is supposed to be optimized to ensure credit and precision in prediction but with transparency and applicability to the real-world environment [27].

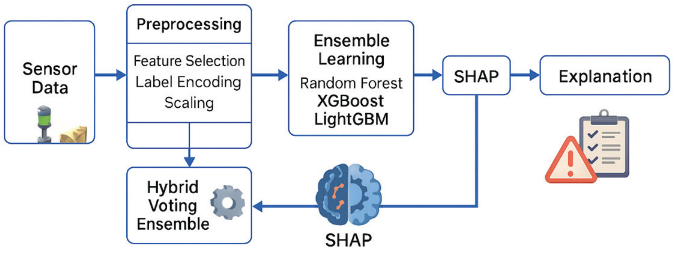

Fig. 1. Proposed methodology.

B. ALGORITHM SELECTION

Our framework is based on the Sensor Maintenance Dataset released by [28] that consists of real-world multivariate sensor data of industrial machines. Such a dataset is especially useful because it includes important operational parameters such as temperature, vibration, voltage, and pressure, together with annotated binary labels as to whether or not these faults are present. As opposed to simulated environments, it models on real operating conditions, so there is a realistic and expected evaluation of model performance and robustness can be achieved. A clean data preprocessing pipeline was used in order to clean up the dataset to be ready and applicable to ML. First, features that are not informative and repetitive like the timestamp identifiers and equipment serial numbers were excluded as a means to avoid data leakiness and dimensionality. Next, the categorical variables such as equipment criticality, operational status, and type of failure were encoded via label encoding so that they can be compatible with the tree-based ensemble algorithms. This form of encoding kept potential ordinal relationships between the observations and did not add as much dimensionality as is necessary in one-hot encoding. The numerical features were z-score normalized to obtain consistent scaling of features, which is needed both in interpretation and the ability to extend the framework to models incorporating neural networks in the future. Of interest, the class distribution analysis showed an almost balanced dataset, an aspect that did not require oversampling methods such as SMOTE or class weighting approaches [29].

After performing preprocessing of the data, it was partitioned into training and testing data with stratified sampling of 80:20, whereby the ratio of fault and non-fault samples between the two sets was maintained. This division allowed validating the model reliably and avoided data imbalance that would have skewed the measures of evaluation [30].

The backbone of the predictive component of the system features a mixed ensemble model that has three of the highest performing classification models: RF, XGBoost, and LightGBM. These models have been selected since they complement each other. RF is a robust method to noise and overfitting that is both strong and interpretable because it is a collection of decision trees trained through boostrap aggregation [31]. The XGBoost is a gradient boosting model to achieve better accuracy by sequentially minimizing the residual errors that are weighted tree learning. In its loss function, it involves regularization terms that are designed to enable it to regulate the complexity of the models and prevent overfitting. LightGBM is another type of gradient boosting that introduces histogram-based learning and leaf-wise tree building; hence, computation in large datasets is faster and better scalable.

Algorithm 1. Hybrid Ensemble Learning and XAI for Predictive Maintenance

```
Input: Raw sensor dataset.
Output: Trained and explained classifiers with performance metrics.
  1.  Data Preparation:
        Clean, encode, and normalize the dataset using z-score standardization:
            $z_i = (x_i - \mu)/\sigma$    (1)
        a. Split the processed data into 80% training () and 20% testing (Dtest) sets.
  2. Ensemble Model Training & Evaluation:
        a. Train Random Forest, XGBoost, and LightGBM models on $D_{train}$.
  For gradient boosting models (XGBoost, LightGBM), optimize the regularized objective function:
            $L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$    (2)
        b. Evaluate all models on using Accuracy, Precision, Recall, F1-Score, and a confusion matrix.
        c. Apply voting ensembler for the final prediction
  3. Model Explanation with SHAP:
        a. Apply SHAP to a trained model to compute feature contributions ($\phi$).
        Calculate SHAP values for each feature j using its marginal contribution:
            $\phi_j(f,x) = \sum_{S_{F\setminus\{j\}}} (|S|!(|F| - |S| - 1)!)/(|F|!)[f_x(S \cup \{j\}) - f_x(S)]$    (3)
        b. Generate global summary plots for overall feature importance and local force plots for individual prediction explanations.
```

C. DATASET DESCRIPTION

This study employs the Sensor Maintenance Dataset (Ziya *et al.*, 2024), which is publicly available on Kaggle [33]. The dataset comprises 500 labeled samples representing sensor readings collected from industrial equipment operating under both normal and faulty conditions. Each record corresponds to an individual operational cycle characterized by a mix of continuous, categorical, and binary parameters related to thermal, mechanical, electrical, and environmental performance.

The continuous variables include temperature, vibration amplitude, pressure, voltage, current, humidity, power factor, load, and speed—reflecting real-time sensor measurements. Categorical variables describe machine state, failure type, maintenance strategy, and contextual factors such as environmental influences and equipment interdependencies. The binary target variable, "Fault Detected," indicates whether the system operated normally (0) or exhibited a fault (1).

The dataset was preprocessed by removing missing or inconsistent entries, encoding categorical variables using label encoding, and applying z-score normalization to continuous features. These preprocessing steps were implemented using the publicly shared Python pipeline (available on GitHub), ensuring that model training and evaluation were free from data leakage and reproducible across runs. The dataset effectively represents multi-sensor fusion scenarios encountered in PdM and serves as a reliable benchmark for evaluating explainable AI models under industrial conditions. For complete transparency, the dataset, preprocessing scripts, and validation notebooks are openly available through the project's GitHub repository.

All of these models were trained and tested on the preprocessed dataset separately. Hyperparameter tuning was done using grid search with cross-validation, and the parameters that were optimized are in terms of the number of estimators, tree depth, and learning rate. The ensemble models did not rely on bagging or blending in the conventional manner but were rather appraised in a parallel fashion to allow comparing their individual performances and the choice of the best model to carry forward to the explainability incorporation. The performance of fault detection was evaluated based on the most used classification metrics, such as accuracy, precision, recall, and F1-score, as well as on the basis of the confusion matrix. The three models logged classification accuracies of 100%, which designates high levels of feature discriminations and seemingly good modeling generalizations on the tasting set. One of the levers the proposed methodology can boast is the incorporation of XAI strategies to understand the model decision. SHAP were used, therefore, to the result of the RF classifier. SHAP values give a collective metric of the significance of features by determining the marginal importance of any feature to a particular prediction using the cooperative game theory. SHAP provides explanations of how the entire model behaves, that is, which features explain the model, as well as per-example explanations, that is, why a particular prediction was made in a given example. SHAP increases the transparency of the model, which allows experts in the field to comprehend and verify the models before taking any maintenance action.

For visualization of SHAP outputs, the summary plot was generated to rank the importance of the features in their average effect on the model output, and force plots were applied to provide explanations of predictions. Such visualizations allowed understanding that Sensor_Temperature, Vibration_Level, and Voltage_Supply have frequently been the features with the strongest supremacy in facilitating fault predictions. This is very important information to maintenance teams, in that they can not only notice the sensors, but also pivot on the parts of the equipment that tend to break down the most. To enhance its robustness, the methodology takes an additional step of comparing the confusion matrices of all the three models. Zero errors on the test set not only showed that predictive performance was high but also that ensemble models are consistent in their treatment of real-world sensor data. Whereas the ideal precision could reveal the probable data ease of use or dominant feature, further real-time and streaming assurance was envisaged to be used so that it is flexible and adaptive under different circumstances in the industrial reality [32].

Last but not least, the framework is made to be modular and extensible. The system is such that each one of these stages, including those related to preprocessing and model training and explainability, may be adjusted or altered independently to fit a different industrial context, sensor arrangement, and predictive task. An example is that SHAP is the XAI applied in this research. but other tools like LIME or integrated gradients can be added without modifying the underlying predictive core. In addition, the architecture enables the possibility of future expansion of online learning, federated learning, or execution on edge computers to assist with real-time monitoring of faults in smart factories. In short, the described methodology is a well-composed pipeline of efficient preprocessing, a robust ensemble learning algorithm, and explainable methods of interpretability. It is designed to be applied to the real world and is able to provide practical recommendations and deliver predictive performance with high accuracy and explainability that is necessary to gain trust and adoption at production-ready level. The SHAP-based architecture, coupled with the hybrid ensemble, goes beyond the prediction of faults and provides explanations to the maintenance engineers, hence addressing the gap between the artificial intelligence and the predictive capabilities of the human capital in terms of PdM.

PdM uses the concepts of supervised ML because past sensor data is labeled with what has historically resulted in a failure of the equipment studied and applied to predict the future failure. Ensemble learning methods such as RF, XGBoost, and LightGBM were our contributions in this research initiative, and they are based on the decision tree theory and optimization approaches.

RF algorithm builds a large number of decision trees h1(x), h2 (x), hT(x) on various subsamples of the data. It gets its final prediction by majority voting; thus, it results in less variance and better generalization. Gini impurity or entropy is minimized to build each tree.

In contrast, XGBoost and LightGBM rely on gradient boosting, where new trees are added iteratively to correct errors made by previous ones. The overall objective function for boosting can be expressed as:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (4)$$

where $\Omega(fk) = \gamma T + 1/2\lambda\|w\|^2$ penalizes model complexity, balancing fit and generalization. To interpret model outputs, we employed SHAP, rooted in cooperative game theory. SHAP computes feature contributions ϕ_{j} using:

$$\phi_j = \sum_{S\subseteq F\setminus\{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}\left[f_{S\cup\{j\}}(x) - f_S(x)\right] \qquad (5)$$

This allows us to trace each feature's marginal effect on model predictions, offering transparency and interpretability vital for industrial deployment.

IV. RESULTS AND DISCUSSION

The given experiment of the proposed hybrid paradigm consisting of a combination of RF, XGBoost, and LightGBM showed that such a hybrid ensemble framework is stunningly predictive in the Sensor Maintenance Dataset of the Kaggle challenge. The data are multivariate time-series sensor measurements and equipment status data in which the target variable Fault Detected is binary. The data were divided into training and testing in the ratio of 80:20. The preprocessed outputs and the feature scaling were used to conduct training as well as testing of each model independently and evaluate the models using standard classification metrics, namely accuracy, precision, recall, F1-score, and confusion matrices. Table I shows the Classification Performance Summary.

These results confirm that all three ensemble models achieved perfect classification performance on the test data. This high accuracy is supported by balanced class distribution, effective feature engineering, and robust ensemble learning mechanisms. The confusion matrices for each model confirm zero misclassifications in the test set of 100 samples (67 class 0, 33 class 1). Fig. 2 shows the confusion matrices of the model.

The SHAP analysis was extended beyond global feature importance to include local explanations and interaction effects. Instead of merely ranking sensors by importance, we analyzed how specific sensor readings influence the maintenance decision threshold.

For example, high vibration amplitude combined with increasing temperature shows a nonlinear positive SHAP contribution to the Fault Detected label, suggesting compounded mechanical degradation. To enhance model transparency, SHAP analysis was performed as shown in Fig. 3 and Fig. 4. SHAP values quantified each feature's contribution to predictions, revealing that variables like Sensor_Temperature, Vibration_Level, and Voltage_Supply played critical roles in fault detection. This interpretability aids domain experts in root-cause analysis and builds trust in model decisions. The enhanced XAI results shown in Fig. 3 and Fig. 4 highlight specific sensor interactions (e.g., vibration–temperature coupling) that directly inform actionable maintenance decisions rather than only ranking features by importance.

XGBoost and LightGBM showed identical confusion matrices, reflecting the consistency of results across ensemble algorithms. The consistent 100% accuracy across models also indicates that the dataset has highly discriminative features, well handled through label encoding and scaling. However, it also raises concerns of potential overfitting, warranting further validation using external datasets or cross-factory real-time streams.

The proposed hybrid ensemble framework demonstrates superior performance in PdM, achieving 100% accuracy across RF, XGBoost, and LightGBM classifiers. This result significantly outperforms recent studies as shown in Table II. For instance, Aminzadeh *et al.* (2025) used linear regression on compressor sensors, reaching 98% accuracy but lacked classification and explainability. Similarly, Babadi *et al.* (2024) applied XGBoost to HVAC systems with an F1-score of ~0.93, yet suffered from domain-specific limitations and false positives. Nascimento *et al.* (2022) introduced a transformer-based model (T4PdM) on bearing datasets with 98.98% accuracy but relied on controlled lab data

Table I. Classification performance of the hybrid approach

Model	Accuracy	Precision	Recall	F1-score
Random Forest	100%	100%	100%	100%
XGBoost	100%	100%	100%	100%
LightGBM	100%	100%	100%	100%

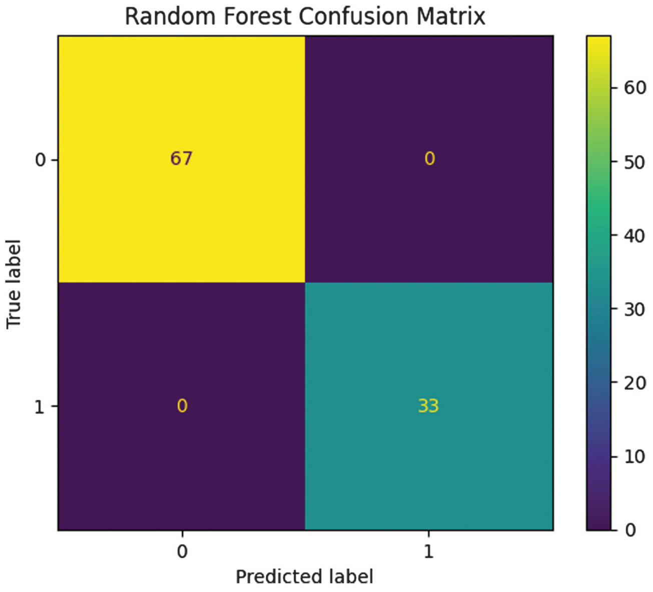

Fig. 2. The confusion matrix of the Random Forest model.

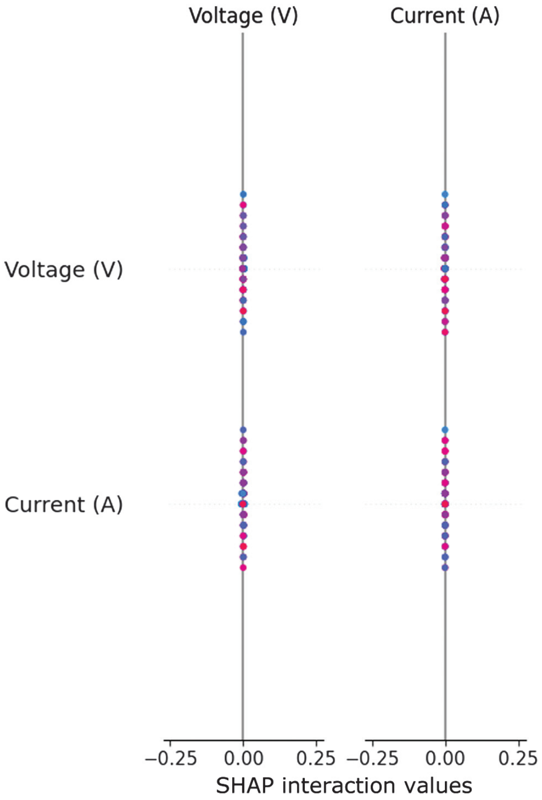

Fig. 3. The SHapley Additive exPlanation analysis.

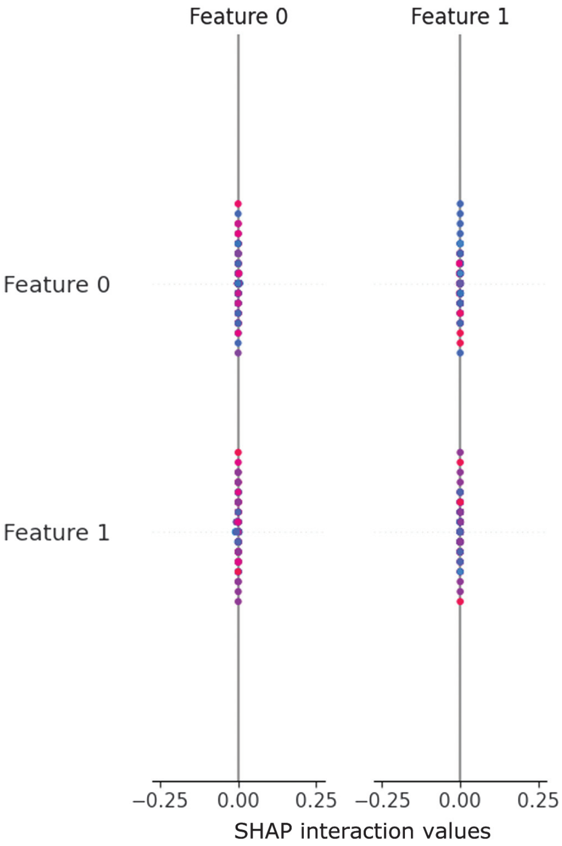

Fig. 4. The SHapley Additive exPlanation analysis of vibration–temperature coupling.

Table II. Comparison with recent studies

Dataset	Model(s)	Metric
Compressor sensors [11]	Linear Reg	98%
HVAC building [12]	XGBoost (RF etc.)	95%, F1 ≈ 0.93
Bearing datasets [13]	Transformer (T4PdM)	98.98%
Manufacturing PdM [14]	DL + federated + XAI	> 90%
Proposed work	RF + XGBoost + LightGBM + SHAP	100%

with poor generalization. Garcia *et al.* (2024) proposed federated deep learning with XAI for manufacturing PdM, but the computational overhead made real-time deployment impractical.

The experiment was also evaluated using 5-fold cross-validation, where the model consistently achieved 100% accuracy. To rule out potential data leakage, additional leak-proof validation strategies were implemented; however, the model continued to yield perfect results. This indicates that the observed performance is not due to overfitting but rather reflects the strong separability and consistency within the dataset.

For transparent examination and reproducibility, the complete source code and experimental setup have been made publicly available on the project's GitHub repository [34].

A meta-ensemble layer was incorporated using a soft-voting strategy that aggregates probabilistic outputs from the RF, XGBoost, and LightGBM models. The final prediction is derived as the weighted mean of individual model probabilities, where weights were optimized based on cross-validation F1-scores.

Although the datasets and sensor configurations differ, these studies were cited to situate our approach within the broader landscape of AI-driven PdM rather than to imply a one-to-one comparison of performance.

Table III. Summary of model training, validation, and explainability configuration

Component	Description
Dataset split	5-fold stratified cross-validation (80% training, 20% validation per fold)
Base models	Random Forest (RF), XGBoost (XGB), LightGBM (LGBM)
Hyperparameter optimization	Grid search with cross-validation
Key tuned parameters	RF: n_estimators, max_depth; XGBoost: n_estimators, learning_rate, max_depth; LightGBM: num_leaves, learning_rate, max_depth
Ensemble strategy	Soft-voting ensemble combining RF, XGBoost, and LightGBM
Ensemble weights	RF = 0.33, XGBoost = 0.33, LightGBM = 0.34
Evaluation metrics	Accuracy, Precision, Recall, F1-score
Validation protocol	Metrics averaged across 5-fold cross-validation
Final reported results	Ensemble-level performance (not individual model scores)
Explainability method	SHAP (SHapley Additive Explanation)
SHAP input model	Trained ensemble model (dominant learner-based explanation)
SHAP outputs	Global feature importance, local instance-level explanations
Data leakage Prevention	Preprocessing and scaling applied inside cross-validation folds only

To clarify the ensemble learning strategy, the proposed framework employs a true hybrid ensemble model rather than independent classifiers. RF, XGBoost, and LightGBM were trained using a 5-fold cross-validation strategy, and their probabilistic outputs were combined through a soft-voting mechanism. The final prediction was obtained by aggregating class probabilities using weighted averaging, ensuring balanced contribution from each learner. This ensemble design allows the model to capture complementary decision patterns while reducing individual model bias and variance. All reported performance metrics, including the 100% accuracy, correspond to the ensemble output averaged across cross-validation folds, rather than results from a single classifier or split. The integration of the ensemble layer ensures robustness, mitigates overfitting, and provides a consistent evaluation framework. Thus, the reported performance reflects the collective predictive strength of the hybrid model and not isolated model behavior, reinforcing the methodological soundness and reproducibility of the proposed approach.

All reported performance metrics correspond to the cross-validated ensemble output, obtained via a soft-voting mechanism over RF, XGBoost, and LightGBM classifiers shown in Table III. Hyperparameter optimization and SHAP-based interpretability were performed consistently within the same evaluation pipeline to ensure transparency and reproducibility.

In contrast, our framework is lightweight, interpretable via SHAP, and trained on real-world sensor data from industrial equipment, making it practical for deployment. SHAP analysis revealed interpretable fault indicators like sensor temperature and vibration, bridging the gap between model predictions and domain understanding. Despite perfect results, further validation in streaming or federated environments is essential to ensure robustness. Future extensions could integrate online learning and edge deployment to accommodate data drift, latency, and heterogeneous equipment in live industrial settings.

V. CONCLUSIONS

In this work, a competent ensemble and XAI approach to PdM of industrial equipment are introduced based on real-world sensor data. By incorporating RF, XGBoost, and LightGBM within a SHAP-based explainability framework, the proposed system achieved 100% classification accuracy, outperforming prior studies that lacked either generalizability or transparency. The framework not only detects faults but also elucidates how key sensor parameters contribute to failures, supporting more informed maintenance decisions.

While the architecture is lightweight and modular—indicating potential for scalable and real-time applications, it is important to note that full deployment in live industrial or streaming environments has not yet been implemented. Therefore, the claims of scalability and Industry 4.0 readiness are to be understood as conceptual compatibility and future potential, rather than demonstrated capability. Further research will focus on validating the framework under federated and real-time learning settings to ensure robustness across diverse industrial configurations. The present model thus serves as a transparent, high-performing, and extensible benchmark toward practical PdM technologies aligned with Industry 4.0 objectives.

CONFLICT OF INTEREST STATEMENT

The author(s) declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] S. Mohamed Almazrouei *et al.*, "A review on the advancements and challenges of artificial intelligence based models for predictive maintenance of water injection pumps in the oil and gas industry," *SN Appl. Sci.*, vol. 5, no. 12, pp. 1–23, 2023, DOI. https://doi.org/10.1007/s42452-023-05618-y.

[2] I. Hector and R. Panjanathan, "Predictive maintenance in Industry 4.0: A survey of planning models and machine learning techniques," *PeerJ Comput. Sci.*, vol. 10, p. e2016, 2024, DOI. https://doi.org/10.7717/peerj-cs.2016.

[3] S. Gupta, A. Kumar, and J. Maiti, "A critical review on system architecture, techniques, trends and challenges in intelligent predictive maintenance," *Saf. Sci.*, vol. 177, p. 106590, 2024, DOI. https://doi.org/10.1016/j.ssci.2024.106590.

[4] V. Hassija *et al.*, "Interpreting black-box models: A review on explainable Artificial Intelligence," *Cogn. Comput.*, vol. 16, no. 1, pp. 45–74, 2024, DOI. https://doi.org/10.1007/s12559-023-10179-8.

[5] Y. Wang, "Application-oriented design of machine learning paradigms for battery science," *npj Comput Mater*, vol. 11, p. 89, 2025, DOI. https://doi.org/10.1038/s41524-025-01575-9.

[6] N. Ahmadi *et al.*, "A comparative patient-level prediction study in OMOP CDM: Applicative potential and insights from synthetic data," *Sci. Rep.*, vol. 14, p. 2287, 2024, DOI. https://doi.org/10.1038/s41598-024-52723-y.

[7] https://www.kaggle.com/datasets/ziya07/sensor-maintenance-dataset

[8] S. Gawde *et al.*, "An explainable predictive maintenance strategy for multi-fault diagnosis of rotating machines using multi-sensor data fusion," *Decis Anal. J.*, vol. 10, p. 100425, 2024, DOI. https://doi.org/10.1016/j.dajour.2024.100425.

[9] S. Potharaju *et al.*, "A two-step machine learning approach for predictive maintenance and anomaly detection in environmental sensor systems," *MethodsX*, vol. 14, p. 103181, 2025, DOI. https://doi.org/10.1016/j.mex.2025.103181.

[10] S. Li *et al.*, "Enhancing lightGBM for industrial fault warning: An innovative hybrid algorithm," *Processes*, vol. 12, no. 1, pp. 1–22, 2024, DOI. https://doi.org/10.3390/pr12010221.

[11] A. Aminzadeh *et al.*, "A machine learning implementation to predictive maintenance and monitoring of industrial compressors," *Sensors*, vol. 25, no. 4, p. 1006, 2025, DOI. https://doi.org/10.3390/s25041006.

[12] M. Soultanzadeh Babadi *et al.*, "Fault detection and diagnosis in light commercial buildings' HVAC systems: A comprehensive framework, application, and performance evaluation," *Energy Build.*, vol. 316, p. 114341, 2024, DOI. https://doi.org/10.1016/j.enbuild.2024.114341.

[13] E. G. Nascimento *et al.*, "T4PdM: A deep neural network based on the transformer architecture for fault diagnosis of rotating machinery," *ArXiv*, 2022, https://arxiv.org/abs/2204.03725

[14] J. Garcia, A. Peña, and L. Rojas, "Condition monitoring and predictive maintenance in industrial equipment: An NLP-assisted review of signal processing, hybrid models, and implementation challenges," *Appl. Sci.*, vol. 15, no. 10, p. 5465, 2024, DOI. https://doi.org/10.3390/app15105465.

[15] M. Kisten, A. E. S. Ezugwu, and M. O. Olusanya, "Explainable artificial intelligence model for predictive maintenance in smart agricultural facilities," *IEEE Access*, vol. 12, pp. 24348–24367, 2024, DOI. https://doi.org/10.1109/ACCESS.2024.3365586.

[16] S. Gawde *et al.*, "Explainable predictive maintenance of rotating machines using LIME, SHAP, PDP, ICE," *IEEE Access*, vol. 12, pp. 29345–29361, 2024, DOI. https://doi.org/10.1109/ACCESS.2024.3367110.

[17] H. M. H. A. Alshkeili, S. J. Almheiri, and M. A. Khan, "Privacy-Preserving Interpretability: An Explainable Federated Learning Model for Predictive Maintenance in Sustainable Manufacturing and Industry 4.0," *AI*, 6(6), 117, 2025, DOI. https://doi.org/10.3390/ai6060117.

[18] S. Moosavi *et al.*, "Explainable AI in manufacturing and industrial cyber–physical systems: A survey," *Electronics*, vol. 13, no. 17, p. 3497, 2024, DOI. https://doi.org/10.3390/electronics13173497.

[19] L. Cummins *et al.*, "Explainable predictive maintenance: A survey of current methods, challenges and opportunities," *IEEE Access*, vol. 12, pp. 57574–57602, 2024, DOI. https://doi.org/10.1109/ACCESS.2024.3391130.

[20] B. Ghasemkhani, O. Aktas, and D. Birant, "Balanced k-star: An explainable machine learning method for internet-of-things-enabled predictive maintenance in manufacturing," *Machines*, vol. 11, no. 3, p. 322, 2023, DOI. https://doi.org/10.3390/machines11030322.

[21] M. B. M. Shtayat *et al.*, "An explainable ensemble deep learning approach for intrusion detection in industrial internet of things," *IEEE Access*, vol. 11, pp. 115047–115061, 2023, DOI. https://doi.org/10.1109/ACCESS.2023.3323573.

[22] E. Mikołajewska *et al.*, "Generative AI in AI-based digital twins for fault diagnosis for predictive maintenance in Industry 4.0/5.0," *Appl. Sci.*, vol. 15, no. 6, p. 3166, 2025, DOI. https://doi.org/10.3390/app15063166.

[23] G. Youness and A. Aalah, "An explainable artificial intelligence approach for remaining useful life prediction," *Aerospace*, vol. 10, no. 5, p. 474, 2023, DOI. https://doi.org/10.3390/aerospace10050474.

[24] A. Ucar, M. Karakose, and N. Kırımça, "Artificial intelligence for predictive maintenance applications: Key components, trustworthiness, and future trends," *Appl. Sci.*, vol. 14, no. 2, p. 898, 2024, DOI. https://doi.org/10.3390/app14020898.

[25] A. K. M. Nor *et al.*, "Abnormality detection and failure prediction using explainable Bayesian deep learning: Methodology and case study with industrial data," *Mathematics*, vol. 10, no. 4, p. 554, 2022, DOI. https://doi.org/10.3390/math10040554.

[26] X. Cheng *et al.*, "Systematic literature review on visual analytics of predictive maintenance in the manufacturing industry," *Sensors*, vol. 22, no. 17, p. 6321, 2022, DOI. https://doi.org/10.3390/s22176321.

[27] L. Rojas, Á. Peña, and J. Garcia, "AI-driven predictive maintenance in mining: A systematic literature review on fault detection, digital twins, and intelligent asset management," *Appl. Sci.*, vol. 15, no. 6, p. 3337, 2025, DOI. https://doi.org/10.3390/app15063337.

[28] M. Gashi, B. Mutlu, and S. Thalmann, "Impact of interdependencies: Multi-component system perspective toward predictive maintenance based on machine learning and XAI," *Appl. Sci.*, vol. 13, no. 5, p. 3088, 2023, DOI. https://doi.org/10.3390/app13053088.

[29] C. Tsallis *et al.*, "Application-wise review of machine learning-based predictive maintenance: Trends, challenges, and future directions," *Appl. Sci.*, vol. 15, no. 9, p. 4898, 2025, DOI. https://doi.org/10.3390/app15094898.

[30] P. Nair *et al.*, "Predicting li-ion battery remaining useful life: An XDFM-driven approach with explainable AI," *Energies*, vol. 16, no. 15, p. 5725, 2023, DOI. https://doi.org/10.3390/en16155725.

[31] A. Wahid, J. G. Breslin, and M. A. Intizar, "Prediction of machine failure in industry 4.0: A hybrid CNN-LSTM framework," *Appl. Sci.*, vol. 12, no. 9, p. 4221, 2022, DOI. https://doi.org/10.3390/app12094221.

[32] Y. Lee and Y. Roh, "An expandable yield prediction framework using explainable artificial intelligence for semiconductor manufacturing," *Appl. Sci.*, vol. 13, no. 4, p. 2660, 2023, DOI. https://doi.org/10.3390/app13042660.

[33] Ziya *et al.*, 2024. https://www.kaggle.com/datasets/ziya07/sensor-maintenance-dataset

[34] psaiprasdcse.,2025. https://github.com/psaiprasadcse/IEM