Explainable artificial intelligence (XAI) enabled anomaly detection and fault classification of an industrial asset

Jeetesh Sharma (2019rme9528@mnit.ac.in)
Malaviya National Institute of Technology https://orcid.org/0000-0001-7013-8492
Murari Lal Mittal
MNIT Jaipur: Malaviya National Institute of Technology
Gunjan Soni
MNIT Jaipur: Malaviya National Institute of Technology

Research Article

Keywords: Anomaly detection, Fault classification, Explainable artificial intelligence, Machine learning

Posted Date: April 19th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2780708/v1

Explainable artificial intelligence (XAI) enabled anomaly detection and fault classification of an industrial asset

Jeetesh Sharma[1*], Murari Lal Mittal[1] and Gunjan Soni[1]

[1]Department of Mechanical Engineering, MNIT, J.L.N. marg, Jaipur, 302017, Rajasthan, India.

*Corresponding author(s). E-mail(s): 2019rme9528@mnit.ac.in; Contributing authors: mlmittal.mech@mnit.ac.in; gsoni.mech@mnit.ac.in;

Abstract

Predictive maintenance helps organizations to reduce equipment downtime, optimize maintenance schedules, and enhance operational efficiency. By leveraging machine learning algorithms to predict when equipment failure will likely occur, maintenance teams can proactively schedule maintenance activities and prevent unexpected breakdowns. Anomaly detection and fault classification are essential components of predictive maintenance. Anomaly detection involves analyzing sensor data collected from equipment to identify deviations from normal behavior. Fault classification, on the other hand, involves identifying the root cause of a fault or failure. A dataset of an industrial asset is used to evaluate the proposed study. Four distinct data-driven anomaly detection methodologies were employed after the pre-processing of the data, with the deep learning-based autoencoder producing the best results of all the techniques. Implementing machine learning-based fault categorization approaches revealed that Random Forest had the best results. Bayesian optimization and sequential model-based hyperparameter optimization technique is used for greater accuracy and optimized hyperparameters. Significant progress has been made in anomaly detection and fault classification using machine learning, but the degree of their explainability is significantly limited by the “black-box” character of some machine learning techniques. Less emphasis has been placed on explainable artificial intelligence (XAI) approaches in the domain of maintenance. Therefore, the XAI tools have been used to acknowledge the extent of the variables to analyze the influence of respective features. A stability metric has been included to improve the explanation's overall quality. The findings of this article suggest that the utilization of eXplainable Artificial Intelligence (XAI) can offer significant contributions in terms of insights and solutions for addressing critical maintenance issues. As a result, decision-making processes can become more informed and effective.

Keywords: Anomaly detection, Fault classification, Explainable artificial intelligence, Machine learning

1 Introduction

The condition of the assets and consistency of production are the key elements of the manufacturing systems that affect product quality and ensure production continuity. Potential failures of equipment/machines could result not only in the production of inferior products; it could lead in compromised safety and dependability. Hence, all the equipment and machines in a factory should be appropriately maintained. Assessment and monitoring of industrial equipment are the two primary activities for developing maintenance strategies/plans. Maintenance refers to the routine activities and procedures to conserve and restore the optimal operational state of equipment. Maintenance aims to prevent equipment failure, extend its useful life, and guarantee safe and efficient functioning. Effective maintenance reduces downtime, enhances reliability, and boosts production. There are four approaches for managing the maintenance process listed below (Bevilacqua & Braglia, 2000):

- Corrective maintenance- The primary characteristic of corrective maintenance is that actions are only performed when a machine malfunctions. Interventions are not made until a failure has taken place.
- Preventive maintenance- Preventive maintenance is performed at predetermined schedules to proactively address issues before they cause breakdowns. It reduces the need for corrective maintenance and increases asset reliability.
- Condition-based maintenance (CBM)- CBM permits maintenance to occur only when required, determined by the actual state of the equipment, instead of following a fixed schedule.
- Predictive maintenance (PdM)- Employs prescient, data-driven maintenance procedures to evaluate equipment status and identify if maintenance has to be carried out.

Industry 4.0 requires industrial enterprises to keep an eye on their equipment and assets using cyber-physical systems that gather and upload data to the cloud (Iqbal et al., 2020). An essential step towards Industry 4.0 is promoting analytical tools that provide extra information on asset health. AI and Machine learning (ML) are essential technologies driving Industry 4.0, which focuses on integrating advanced technologies to optimize industrial processes and increase productivity. AI and ML algorithms can analyze enormous quantities of equipment data to forecast when maintenance is required for predictive maintenance. This is accomplished by examining data patterns, such as temperature, vibration, and pressure variations, to predict when equipment will likely fail. The responsibilities of AI and ML in predictive maintenance are as follows:

- To gather and process a vast amount of data from various sensors to provide a thorough overview of the operation.
- To analyze the collected data and find patterns, trends, and anomalies that could point to equipment failure.
- To examine equipment performance data to identify faults, predict failures, and help maintenance teams to identify and tackle issues effectively.
- Optimizing maintenance planning reduces the need for unscheduled maintenance and downtime.

Predictive maintenance typically consists of four steps: anomaly detection, fault diagnosis, prognosis, and mitigation. The ability of anomaly detection algorithms to identify the very first indication of asset breakdown is vital for predictive maintenance solutions since it allows for the planning of performing maintenance actions before the asset undergoes significant damage. Anomaly detection is a highly significant skill that reduces unplanned downtime and unnecessary maintenance enabling more effective management of critical components. A detailed examination of the existing literature reveals that publications using AI/ML are on the rise and are predicted to do so in the future (Dalzochio et al., 2020). This suggests that the technology has the interest of the maintenance community.

Understanding and tracing the steps taken by algorithms is a common challenge with artificial intelligence systems, particularly as they become more complex and autonomous. The entire calculation process is transformed into what is known as a "black box", which is difficult to understand. A black box model might be considered a limitation in the context of AI and ML algorithms because no one, not even the engineers who created the algorithm, can explain or describe what is occurring inside them or how the AI algorithm arrived at a certain result. Therefore, it is worthwhile to address the following questions to clarify a significant aspect of this stage:

- Why does a model forecast in the manner it does?
- How should we interpret typical model predictions?
- How can I believe a model's predictions?

These inquiries are fundamental to integrating AI/ML into engineering disciplines that typically value transparent techniques (Tjoa & Guan, 2020). Most AI techniques present a challenge of opaqueness to decision-makers in sectors where safety and dependability are paramount. Trust in decisions based on ML outputs requires decision-makers to interpret how ML techniques work and how they generate their outputs to accomplish AI/ML deployments within their industries. With the aid of explainable AI (XAI), these methods can be better explained and more transparent to decision-makers. XAI is most frequently applied in industries requiring high decision-making accuracies and accountability levels, such as healthcare (Faust et al., 2023; Rivera et al., 2023; Z. Wang et al., 2023; Zhao, Ren, Zhang, Wu, & Lyu, 2023) and management (Angelotti & Díaz-Rodríguez, 2023; Langer & König, 2023; Lee, Jung, Lee, Kim, & Park, 2021). There is relatively less research in the area of explainable artificial intelligence in predictive maintenance. Therefore, it is decisive to research and apply XAI in the area of predictive maintenance.

The following are the contributions that we make throughout this article:

- We applied four different anomaly detection algorithms, Interquartile Range (statistical-based approach), K means clustering (cluster-based approach), Isolation Forest (tree-based approach), and Autoencoder (deep learning-based approach).
- Support vector machine (SVM), Artificial neural network (ANN), and Random forest (RF) are implemented for fault classification.
- Explainable artificial intelligence (XAI) tools are employed to understand the black box model better
- A stability metric has been added for the quality of the explanation.

This article is structured as follows; Literature review is discussed in 2. Section 3 discusses the overview of data-driven techniques for anomaly detection. Dataset description, anomaly detection approaches, and fault classification approaches have been discussed in detail in section 4. Explainable artificial intelligence results are described in section 5. Section 6 summarizes, and section 7 concludes the article.

2 Literature Review

The significance of condition monitoring and predictive maintenance for the industry's efficient and effective operations has been extensively acknowledged. However, an efficient procedure is required to convert complex multidimensional data into results humans can understand, especially in this day and age of Industry 4.0, when vast amounts of data can be quickly gathered through digital devices. Numerous research studies have investigated various data-based intelligent systems to develop the advanced diagnostic capability for condition monitoring in rapidly evolving environments (Althubaiti, Elasha, & Teixeira, 2022; Tiddens, 2018). This section examines the research on anomaly detection, fault classification, and PdM-XAI works.
Condition monitoring and fault diagnostics generally leverage machine, process, or system data for failure identification or prediction Tiddens (2018). Chandola, Banerjee, and Kumar (2009) reviewed various anomaly detection strategies and evaluated their efficacy. Goldstein and Uchida (2016) assessed the effectiveness of multivariate unsupervised anomaly detection methods.

Distance-based, statistical-based, classification-based, artificial neural networks, and deep learning methods are standard anomaly detection techniques applied in the literature. Bai, Liu, Chai, Zhao, and Yu (2020) proposed a unique approach for gas turbine anomaly identification utilizing just normal data from the standpoint of data-driven strategies and gas turbine pre-existing knowledge combination. Fu, Hu, and Tan (2005) suggested a hierarchical clustering-based anomaly detection approach to categorize vehicle motion patterns. Caliva et al. (2018) integrated the convolutional neural networks (CNN), k means clustering and denoising autoencoder (DAE) algorithms in a deep learning-based solution for anomaly detection in nuclear reactors. In semiconductor manufacturing, Puggini and McLoone (2018) introduced dimensionality-reducing attribute selection and isolation forest-based anomaly detection. Xuyun, Hui, Zhong, and Lin (2019) proposed a CNN-DAE based deep learning method for aircraft engine fault detection. C. Zhang, Hu, and Yang (2022) presented an anomaly detection approach based on deep learning and a diagnostic system for wind turbines based on LSTM, stacked DAE, and XGBoost. Several research studies have used machine learning algorithms for fault diagnosis in various domains; for instance, induction motors (Widodo, Yang, & Han, 2007; B.-S. Yang, Di, & Han, 2008), rolling-element bearings (Samanta & Al-Balushi, 2003; J. Yang, Zhang, & Zhu, 2007), spur gears (Cerrada et al., 2016), and ball bearing (Kankar, Sharma, & Harsha, 2011). These methods make it possible to classify normal and abnormal classes.

AI-based techniques will account for a substantial portion of the improvements in this field. As a result, promoting AI/ML integration is extremely important (Zhu, 2015). From this vantage point, a literature review reveals that AI/ML has been used for various maintenance issues (Carvalho et al., 2019; Çınar et al., 2020; Prytz, Nowaczyk, Rögnvaldsson, & Byttner, 2015; Ruiz-Sarmiento et al., 2020; Sharma, Mittal, & Soni, 2022; Susto, Schirru, Pampuri, McLoone, & Beghi, 2014; W. Zhang, Yang, & Wang, 2019).

It is crucial to have technologies that can explain the decision-making processes, especially in cases where inaccurate predictions can have severe repercussions, given the rising deployment of artificial intelligence across various fields. The terms “explainability” and “interpretability” are frequently used interchangeably in machine learning and AI (Došilović, Brčić, & Hlupić, 2018). Explainability measures how well people can understand a complicated machine learning system’s internal functioning. On the other hand, “interpretability” refers to the extent to which we can predict what would occur if the input parameters of a model were altered. In other words, interpretability is focused on comprehending the inner workings of an AI system, whereas explainability is concerned with providing stakeholders with clear and comprehensible explanations of the model’s judgments.

A few techniques are available to investigate the logic behind an AI/ML model’s predictions, despite XAI/Interpretable machine learning (IML) being a relatively new study field (Molnar, 2020). The methods to model interpretability, or the features to consider, are organized around three main criteria- "by the model" ("intrinsic or post-hoc"), "by scope" ("local or global"), and "by method" ("model-specific or model-agnostic"). Intrinsically interpretable machine learning models have a simple structure, such as linear models and decision trees. The term "post hoc interpretable model" refers to the use of explanation techniques following model training and is unrelated to the internal design of the model, such as the significance of permutation features. Model-specific explanation tools are limited to some specific model classes, whereas model-agnostic methods can be used with any machine learning model. Global explanations encompass the entire model, whereas local interpretable techniques focus on a single prediction (Du, Liu, & Hu, 2019).

There are few studies in the field of explainable AI for predictive maintenance, as this is a relatively new application of AI in the industry. Hajgató et al. (2022) proposed a comprehensive framework for data exploration in predictive maintenance. Human professionals also analyzed condition monitoring systems to forecast oil degradation in industrial gearboxes. Garouani et al. (2022) proposed a technique based on automated machine learning by allowing for a better selection of the combination of artificial intelligence algorithm and hyperparameter configuration. The acceptability of the models is enhanced by the explainability of the final algorithm provided by this method. To monitor the wear and tear of rollers in a hot strip mill at a steel production facility, Jakubowski, Stanisz, Bobek, and Nalepa (2021) utilized a variational autoencoder (VAE). The Shapley additive explanations (SHAP) technique was utilized to comprehend the model's predictions. Kuzlu, Cali, Sharma, and Güler (2020) presented several examples of solar photovoltaic predictions utilizing machine learning and XAI tools, including SHAP, Local interpretable model-agnostic explanations (LIME), and Explain like I'm 5 (ELI5) for smart grid application. The use of XAI techniques enhances solar photovoltaic prediction models. Roelofs, Lutz, Faulstich, and Vogt (2021) suggested an innovative method for autoencoder anomaly root cause analysis for wind turbines. An intentional inaccuracy was introduced into the wind speed sensor measurements to provide a controlled testing environment and apply this method to an open data set of wind turbine sensor data. Alfeo, Cimino, and Vaglini (2022) proposed a combination of a feature learning approach with a health stage classifier. Features were automatically created from minimally processed data using feature learning. High-quality features were successfully retrieved by analyzing several input signals, and insightful knowledge about the input signals' most informative domain transformations was gained.

3 Anomaly Detection for Predictive Maintenance

A class of methods for recognizing observations that differ from the majority of data is known as anomaly detection or outlier identification. Anomaly detection is founded on the principle that a model can learn to classify new observations by comparing them to a previously known scheme by studying examples from "normal" operating conditions. Since detecting anomalies is a widespread challenge in industries, a wide range of methodologies are available to detect anomalies, from conventional statistical to machine learning techniques.

3.1 Data-driven techniques for anomaly detection

Data-driven approaches are usually learning-based methods in which the absence of a solid mathematical model is compensated by the accessibility of vast quantities of data from which valuable information can be gained. Machine learning encompasses a broad scope and has a variety of potential applications. It is divided into three categories: Supervised, unsupervised, and semi-supervised learning. Deep learning is also on the rise as a result of technological improvements. Many machine learning approaches are being deepened or integrated with deep learning. Similarly, Chandola et al. (2009) recognizes three basic methods for the problem of outlier detection, namely:

- Supervised: Techniques learned in supervised mode presuppose the availability of a training data set with examples labeled for the normal and anomalous classes. Supervised techniques are highly trustworthy since pre-labeled data represents the ground truth. However, this kind of data is unavailable or incomplete in many practical systems (Vilalta & Ma, 2002). Semi-supervised and unsupervised techniques fill this void. Some popular supervised algorithms are Neural Networks, Support Vector Machine (SVM), k-Nearest Neighbors (kNN), and Decision Trees (Omar, Ngadi, & Jebur, 2013).
- Unsupervised: Unsupervised techniques do not require labeled data and are the most extensively applicable. The strategies in this division are based on the implicit premise that normal instances outnumber anomalies in the test data. Clustering is a popular approach since it groups data points into groups based on similarity. Anomalies are data instances that do not cluster or have substantially smaller clusters than the rest of the data. Probabilistic modelling methods are the basis of other unsupervised learning algorithms, where anomalies are uncovered by the estimation of the likelihood of each data instance.
- Semi-supervised: Semi-supervised learning techniques assume that the training data only comprises instances of the normal class that have been labeled. They have a more comprehensive range of applicability than supervised approaches since they do not require labeling for the abnormal class. One-class support vector machines (Y. Wang, Wong, & Miner, 2004) and autoencoders (Sakurada & Yairi, 2014) are some well-known semi-supervised anomaly detection algorithms.

In many data-driven scenarios, supervised, semi-supervised, and unsupervised methods perform admirably. The main problem with such approaches is that they do not respond proactively to changes unless explicitly instructed. To achieve this goal, cutting-edge techniques based on the capacity for autonomous learning have been created, such as reinforcement learning and deep learning.

4 Dataset description

This section provides details about the case study database used in the rest of this work. A data set of an industrial asset is used to evaluate the proposed study. The data has been divided into 52 sensor columns, one timestamp column, and one target machine status column. The data is recorded in incremental steps of one minute, and it is unclear the parameters each sensor is tracking. The label data contains three machine status values. Broken indicates that the machine has failed. Recovering denotes a machine attempting to recover from a failing state. Normal suggests that the machine is in normal operation. Normal is by far the most common, which makes sense since the machine should normally work most of the time. The minority classes are Recovering and Broken. There are missing values, an empty column, and a timestamp with an invalid data type in the data. Table 1 shows the percentage of missing values in the first ten sensors.

Table 1: Percentage of missing values in the first ten sensors

Sensors	s_0	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
Percentage of missing values	4.633	0.167	0.008	0.008	0.008	0.008	2.177	2.474	2.317	2.085

Some procedures we applied to transform raw data into a valuable and efficient format are eliminating unnecessary columns and duplicates, managing missing values, and converting data types to the correct data type. Exploratory Data Analysis (EDA) is used to identify odd patterns and trends. It's intriguing to see the sensor data displayed over time with the broken and recovering machine state highlighted in red and yellow, respectively, on the same graph (Figure 1). In this manner, we can observe when the asset fails and how it affects the sensor values.

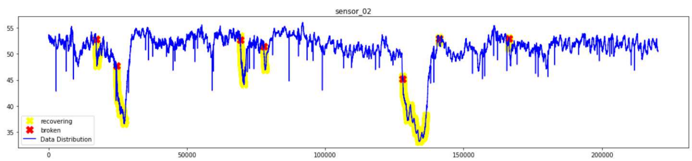

Fig. 1: Sensor 02 measurements with broken and recovering machine condition

It is computationally complex and inefficient to train models with all 52 sensors. Principal Component Analysis (PCA) extracts the most relevant information from data and compresses the size of the data set by retaining only the most essential information to simplify the data set's description. The first principal component has the most significant variance, and the second has to be orthogonal to the first component (Abdi & Williams, 2010).

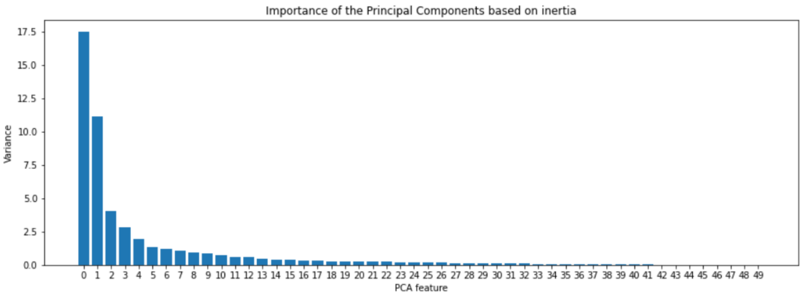

Fig. 2: Importance of principal components based on inertia

According to the features retrieved by PCA (Figure 2), the first two principal components are the most relevant, and as a result, they will be employed as features in the training of the models. For time series analysis, it is crucial to check the data is stationary. The augmented dickey fuller test determines the stationarity of time series data. The Augmented Dickey-Fuller Test is an expanded version of the basic Dickey-Fuller Test. An augmented Dickey-Fuller test evaluates the null hypothesis that a time series sample comprises a unit root (Mushtaq, 2011).

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + ... + \delta_{p-1} \Delta_{t-p+1} + \varepsilon_t \tag{1}$$

y is the variable of interest, t is the time index, α is a constant, β is the time trend coefficient, ε_t denotes error, and p is the autoregressive process's lag order. The unit root test is then performed with the null hypothesis $\gamma = 0$ versus the alternative hypothesis gamma $\gamma < 0$. The p-value for the Augmented Dickey-Fuller test on the first principal component is 0.000165, which is less than 0.05. Hence, the null hypothesis ought to be rejected, and it should be stated that the data is stationary and there is no unit root; the second principal component yielded the same result. Thus, both principal components are stationary.

4.1 Anomaly detection approaches

In this subsection, we will examine various data-driven and unsupervised anomaly detection approaches and briefly summarize each. Four anomaly detection approaches have been applied: 1. IQR (statistical anomaly detection), 2. K Means Clustering (Cluster-based anomaly detection), 3. Isolation

Forest (Tree-based anomaly detection), 4. Autoencoder (Deep learning-based anomaly detection)

4.1.1 Interquartile Range (IQR)

The interquartile range (IQR) is a mathematical spread measure that spans 50% of the data series and indicates the difference between the third (Q_3) and first (Q_1) intervals in a data set with values sorted from small to large.

$$IQR = Q_3 - Q_1 \tag{2}$$

where Q_3 is the third quartile, and Q_1 is the first quartile, or the 75th and 25th percentiles, respectively (Dekking, Kraaikamp, Lopuhaä, & Meester, 2005). Suppose there is a value in a data series that is less than the lower limit value of Q_1-1.5 x IQR and more prominent than the upper bound value of Q_3+1.5 x IQR. In that case, these values are anomalous data points that do not match the measurement distribution. Figure 3 shows the anomalies predicted by the IQR of sensor 11. The red part shows the anomalies identified, and the blue part shows normal data points.

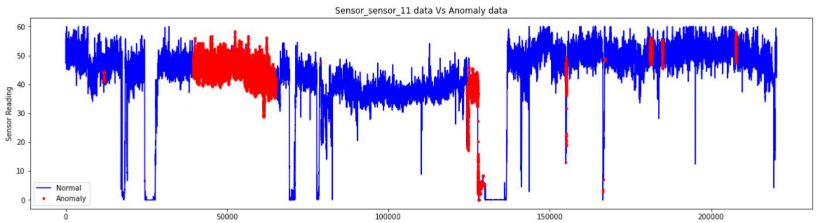

Fig. 3: Interquartile range anomaly detection results of Sensor 11

4.1.2 K-Means Clustering

The k-means algorithm utilizes clustering to group objects based on their feature values into K-distinct clusters. Comparable feature values exist among items belonging to the same cluster. A predetermined positive integer value, K, is used to specify the number of clusters that must be identified (Xiao, Shao, & Liu, 2006). A threshold value can be used to identify anomalies. If a data point's distance from its nearest centroid exceeds the threshold value, the data point is considered to be an anomaly. Using Euclidean distance, a common K-Means Clustering technique computes the cluster centroid for each K cluster. Then, until it exceeds the threshold value, each observation is assigned to the cluster whose center is closest to it using Euclidean distance. Figure 4 depicts the anomalous and normal data points of the sensor 11 by K-means clustering.

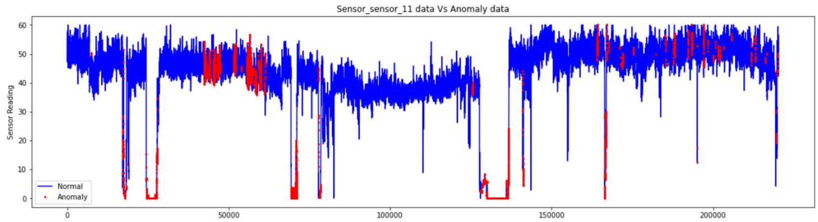

Fig. 4: K-Means clustering anomaly detection results for sensor 11

4.1.3 Isolation Forest

Isolation forests use the concept of isolation to discover anomalies rather than traditional methods such as measuring distance or density. For a given dataset, this method creates an ensemble of isolation trees; anomalies are those occurrences with short average path lengths on the isolation trees (Liu, Ting, & Zhou, 2012). The travel from the root of an isolation tree to a leaf node will be shorter for outliers than for inliers. The anomaly score will be proportionate to this path length; the shorter the distance, the more abnormal the data point is. Figure 5 represents the anomalies identified by the isolation forest of sensor 11.

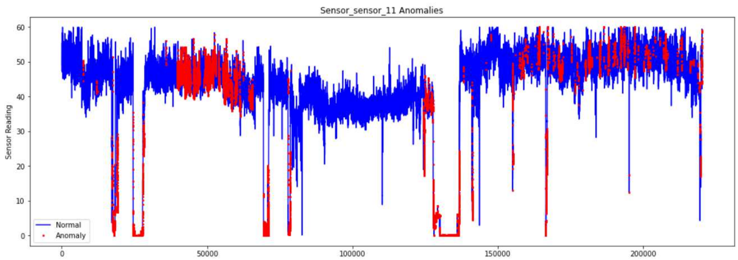

Fig. 5: Isolation Forest anomaly detection result for sensor 11

4.1.4 Autoencoder

An autoencoder is a specific type of neural network design that reconstructs input values using semi-supervised learning (Borghesi, Bartolini, Lombardi, Milano, & Benini, 2019). Simply expressed, it aims to accurately replicate the input value X supplied to a neural network (Figure 6). Autoencoders can discover low-dimensional representations of high-dimensional data, which allows them to reconstruct the input from the output.

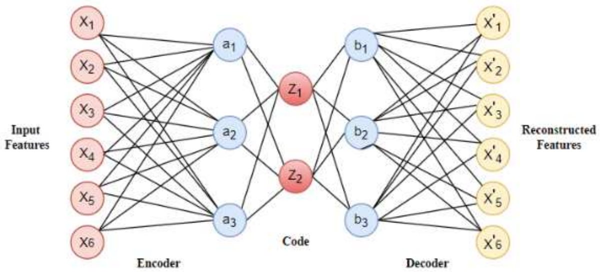

Fig. 6: Schematic diagram of autoencoder

An encoder creates a low-dimensional dataset from a high-dimensional one. The input sent into the decoder is reduced and contained in the code. The low-dimensional data is expanded into high dimensions via a decoder. Autoencoders use a unique aspect of a neural network to achieve efficient methods of training neural networks to acquire normal behavior. When an outlier data point is encountered, the auto-encoder cannot effectively encode it. It learned to express patterns absent from the data set. Attempting to rebuild the original data from its compact form will result in a reconstruction that does not resemble the original data. Thus, aiding in the detection of anomalies as they arise.

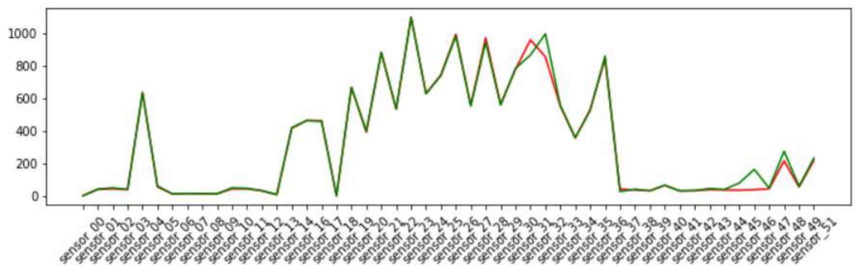

Fig. 7: Model performance on normal data

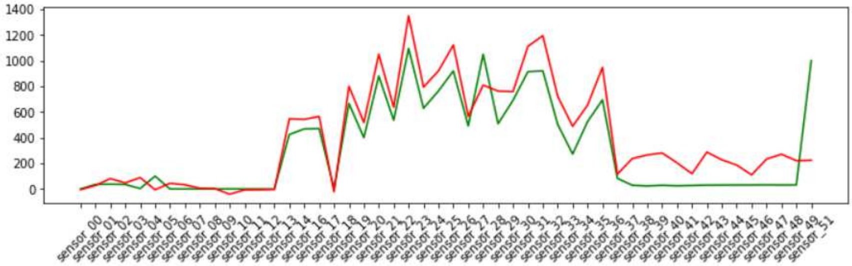

Fig. 8: Model performance on anomaly data

Figure 7 shows a small reconstruction error on normal data. The distinctiveness between the green and red lines is negligible. The green represents anomaly test data, and the red represents decoder output. We can observe that the reconstruction error is relatively large in Figure 8. We take the mean of the training loss and multiply it by the second standard deviation to establish a threshold from which we can determine that values above it are anomalies and values below it is normal data. Figure 9 shows the normal training loss and the anomaly loss in separate histograms on a single graph and a vertical line to represent the threshold for better visualization. The autoencoder then achieves a finding anomaly class accuracy of 99%.

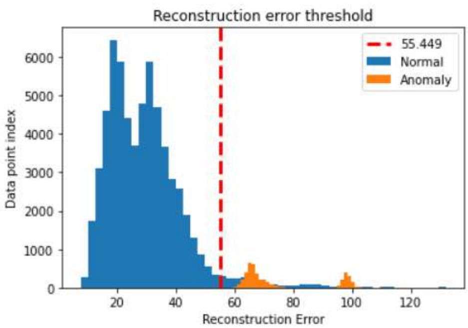

Fig. 9: Reconstruction error threshold for normal and anomaly class

The accuracy, precision, recall, and F1 score of the approaches for detecting anomalies using the Interquartile Range, K Means Clustering, and Isolation Forests are shown in Table 2. Neural network-based Autoencoder obtained the maximum 99% accuracy while statistical method-based Interquartile Range has an accuracy rate of 81%.

Table 2: Accuracy, Precision, Recall, and F1 Score of IQR, K Means Clustering, and Isolation Forest anomaly detection techniques

Algorithm	Interquartile Range	K Means Clustering	Isolation Forest
Accuracy (%)	80.93	92.58	91.43
Precision	0.48	0.73	0.7
Recall	0.46	0.93	0.85
F1 Score	0.46	0.79	0.75

4.2 Fault classification

The complex models' Support vector machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) are used to develop the model to attain a high prediction performance. Data partitioning is accomplished via SVM, a convex optimization problem, by exploring the hyperplane with maximum intervals (Widodo & Yang, 2007). The ensemble of decision trees produces RF. Due to its speed, excellent accuracy, and resilience, it is frequently employed in analysis (Breiman, 2001). The inclusion of SVM, ANN, and RF algorithms in this study is due to the fact that, according to past research, these are some of the most often utilized approaches with superior performance.

4.2.1 Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm that is based on statistical learning. It's an effective method for fault classification. Figure 10 depicts triangles and circles as the two sample point types. Mainly SVM locate the maximum-margin (middle) line. The hyperplane is one of the planes that separate two classes. A flat subspace with dimension n-1 is a hyperplane in n-dimensional space. Margin refers to the shortest distance between the hyperplane and the observation. Two parallel planes denoted by dashed lines are parallel to the hyperplane and pass through the sample points in these two classes closest to the hyperplane. The maximal margin hyperplane has the most significant distance between itself and the training observations, resulting in minimal generalization error (Kankar et al., 2011). This hyperplane allows for the classification of testing data.
A training set of n points is provided in the form

$$(x_1, y_1), \dots, (x_n, y_n) \tag{3}$$

where y_i is either 1 or - 1, corresponding to the class to which point x_i belongs. Any hyperplane can be expressed as

$$w^T x - b = 0 \tag{4}$$

The normal vector, w, of the hyperplane determines its orientation, while the offset of the hyperplane from the origin along the normal vector, w, is dictated by $\frac{b}{||w||}$.

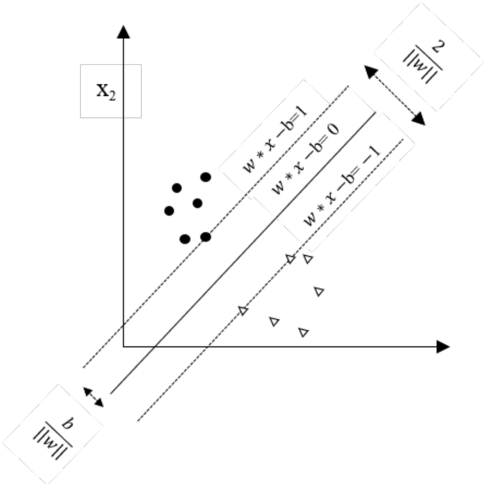

Fig. 10: Hyperplane classifying two classess

4.2.2 Artificial Neural Network (ANN)

A network of connected artificial neurons is known as an artificial neural network (ANN). These neurons process information using a computational model. An adaptive system called an ANN modifies its structure in response to information passing through the network (Zurada, 1992). Each output acts as an input for the following function, and neurons within the network connect with neurons in the adjacent layer. Every function, including the initial neuron, employs an internalized function that produces a numeric output by adding a bias term specific to each neuron. The output is then multiplied by the corresponding weight to generate the numeric input for the subsequent layer's function. This process is repeated until the network provides its final output. Figure 11 shows a single neuron neural network for basic understanding.

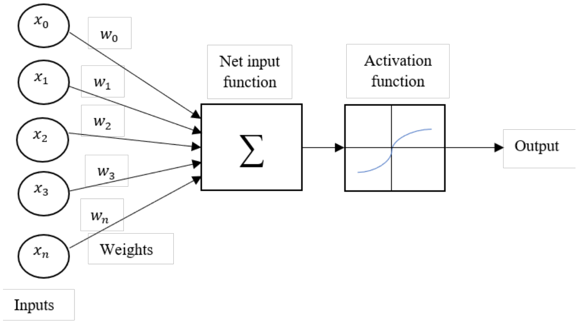

Fig. 11: Single neuron neural network

4.2.3 Random Forest (RF)

Breiman (2001) introduced Random Forest algorithm (RF) is a collective term for ensemble approaches utilizing classifiers of the tree type. By employing bagging, a meta-algorithm to enhance classification model accuracy, RF creates a significant number of decision trees from a subset of a single original training set. The RF algorithm produces a prediction for each tree, and the final prediction of the model is determined by selecting the class with the highest number of votes among all the tree predictions (Figure 12).

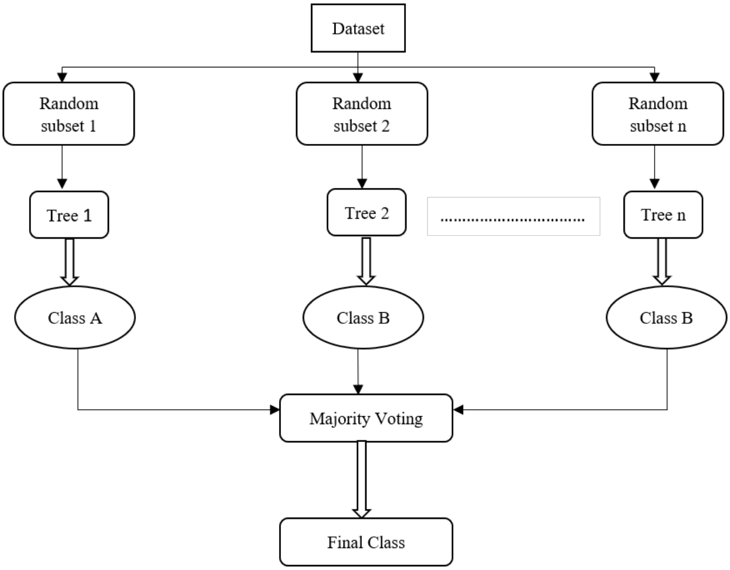

Fig. 12: Random Forests classification

4.3 Hyperparameter Optimization

Hyperparameter optimization is a fundamental aspect of machine learning that involves finding the set of hyperparameter values for a specific algorithm that results in the best performance on a validation set. This is typically accomplished through an iterative search process, where various combinations of hyperparameter values are tested and evaluated to identify the optimal configuration. Hyperparameter optimization aims to improve the machine learning model's performance and achieve the best possible results. For example, the maximum depth of a tree, i.e., the longest path from the root node to a leaf node, and n estimators, i.e., how many distinct decision trees will be built, are some hyperparameters of the random forest algorithm.

HyperOpt is a software tool designed to facilitate the process of hyperparameter optimization in machine learning algorithms. It leverages the advanced Bayesian Optimization and Sequential Model-Based Global Optimization techniques to automate the search for the best possible configuration of hyperparameters (Bergstra, Yamins, & Cox, 2013). For optimizing hyperparameters, HyperOpt requires four fundamental components: the search space, the loss function, the optimization algorithm, and a database for recording the history. A convex function determines the search space. The function to be optimized is the loss function derived by evaluating the model with the configuration.

After fitting with the best hyperparameters obtained from the HyperOpt following are the classification algorithms (SVM, RF, ANN) results.

Table 3: Accuracy, Precision, Recall, and F1 score of SVM, ANN, and RF classification algorithms

Method	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Random Forest (RF)
Accuracy (%)	0.92	0.97	0.99
Precision	0.56	0.98	0.98
Recall	0.74	0.85	0.98
F1 Score	0.59	0.90	0.99

Table 3 shows that Random Forest gives promising results among the approaches that have been applied. Typically, Random Forests is viewed as a black box. The decision-making process of the random forest model is difficult to comprehend due to the vast number of deep trees, each of which is trained on a subset of randomly selected features and bagged data. It is not feasible to gain a complete understanding of the model by evaluating each tree. Additionally, even if we focus on a single tree, we can only interpret its decision-making process if it has a limited number of features and shallow depth. Further Post-hoc Explainable Artificial Intelligence (XAI) methods have been applied to the Random Forest algorithm for a better and more elaborated understanding of the model.

5 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) is a branch of study that tries to make the outcomes of AI systems more understandable to humans (Adadi & Berrada, 2018). There has been a renewed interest in XAI among academia and practitioners due to the significant increase in research related to explainable artificial intelligence.

Explainable AI (XAI) is important in predictive maintenance because it provides transparency and understanding of how AI models make decisions, which is critical for ensuring the reliability and trustworthiness of the predictions made. In predictive maintenance, AI models identify potential equipment failures and predict when maintenance is required. It's crucial to understand the reasons behind these predictions to make informed decisions and avoid false alarms or missed failures. XAI can provide this insight by allowing users to understand the factors that influence predictions and the importance of each factor. This leads to improved decision-making and increased trust in the AI system.

Two popular post-hoc (evaluate the model afterward training) model-agnostic (can be applied to any machine learning model) frameworks, SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), were used to explain the results of the Random Forest approach. SHAP is a game-theoretic explanation for any machine learning model's output. It relates optimal credit allocation to local explanations utilizing the classical Shapley values from game theory and their associated extensions (Lundberg & Lee, 2017). LIME is a method that aims to explain individual prediction by approximating any black box machine learning model with a local, interpretable model (Ribeiro, Singh, & Guestrin, 2016).

5.1 Local Interpretable Model-agnostic Explanations (LIME)

LIME was introduced by Ribeiro et al. (2016). This algorithm provides interpretable explanations for individual predictions made by machine learning models. It works by creating a local explanation around the forecast instead of trying to understand the entire model. LIME generates an explanation by training a simple interpretable model on a small subset of the data points near the prediction of interest rather than trying to understand the more complex and globally applicable model. This allows for human-understandable and model-agnostic explanations, meaning they can be applied to any machine-learning model regardless of its internal structure. This makes LIME a valuable tool for understanding and explaining the predictions of complex machine learning models, particularly in domains like predictive maintenance, where the ability to understand and trust the predictions is critical. Local surrogate models with an interpretability condition can be expressed as follows:

$$e(x) = arg \min_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{5}$$

In equation 5, g is the explanation model, for instance, x, that seeks to minimize the loss L, which gauges the proximity of the explanation to the prediction made by the original model f. The model complexity, represented by $\Omega(g)$, is kept at a low level, and G represents the group of likely explanations. The proximity measure π_x determines the extent of the area surrounding instance x that is considered in generating the explanation.

5.2 SHapely Additive exPlanations (SHAP)

Lundberg and Lee (2017) first proposed the SHapley Additive exExplanations (SHAP). Individual predictions can be explained using the SHAP technique. SHAP values can be used to provide human-interpretable explanations for machine learning model predictions and to evaluate the importance of features across many instances and models. This makes SHAP an essential tool for comprehending and interpreting the predictions provided by machine learning models, especially in domains such as predictive maintenance, where the capacity to understand and rely on the predictions is crucial. SHAP uses the optimal Shapley values from coalitional game theory. SHAP defines the explanation as follows:

$$e\left(z'\right) = \phi_0 + \sum_{j=1}^{M} \phi_j \; z'_j \tag{6}$$

In equation 6 e denotes the explanatory model, $z' \in \{0,1\}^M$ is the coalition vector, M is the maximal coalition size, and $\phi_j \in R$ is the Shapley value for a feature j. A coalition vector value of 1 indicates the presence of the relevant feature value, while an entry of 0 indicates its absence. Significant characteristics have substantial absolute Shapley values. Since we desire global significance, we take the mean of the absolute Shapley values for each feature over all the data (Molnar, 2020).

$$P_j = \frac{1}{n} \sum_{j=1}^{n} \mid \phi_j^i \mid \tag{7}$$

Python's Shapash library has been used based on existing utilities to implement the XAI framework.

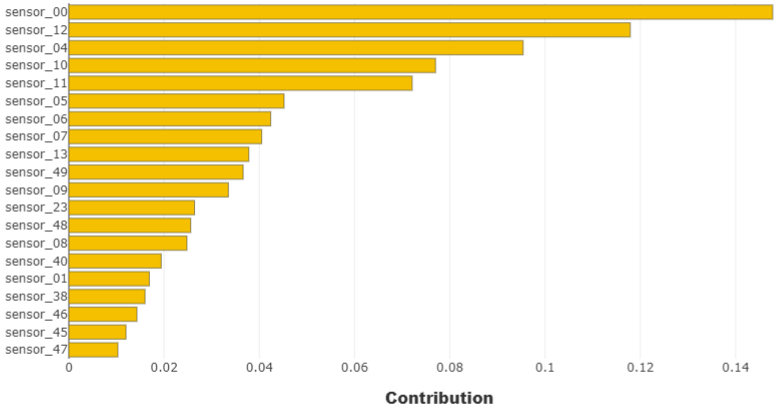

Fig. 13: Contribution of top 20 features

Figure 13 depicts how each of the top 20 features globally contributes to predicting the class label. Figure 14 and Figure 15 aids in answering questions such as how a particular feature influences the forecast, whether it contributes positively or negatively, and so on. Classes 0 and 1 indicate the normal and broken labels, respectively. Figure 14 depicts the respective index, predicted class 1 'broken', reading of sensor 00, and contribution of sensor 00 in predicting the class label.

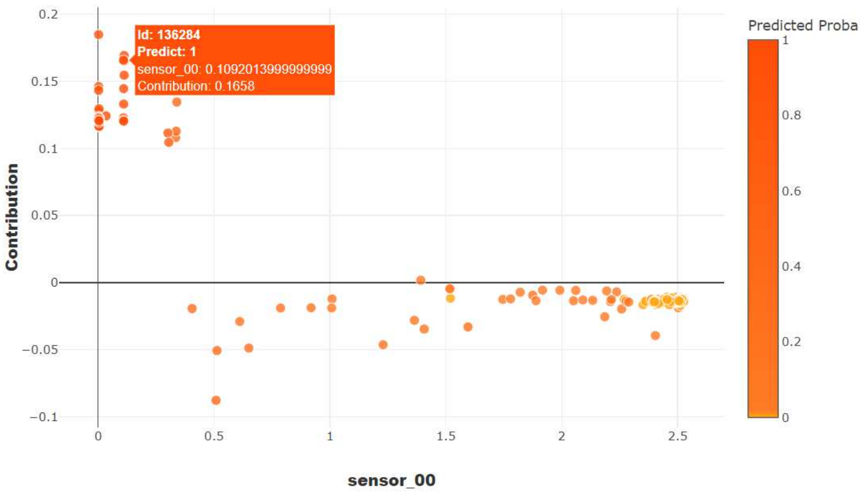

Fig. 14: Sensor 00 feature contribution plot

Figure 15 indicates the local explanation of index 131000. Features are arranged in terms of their importance from top to bottom on the y-axis, and x-axis shows the contribution of sensors. Figure 16 depicts the compare plot of index 131004, which indicates the broken class, and index 139930 shows the normal class. Sensor 0, sensor 12, and sensor 4 are the most important variables, shown on the y-axis, and their respective contributions are shown on the x-axis.

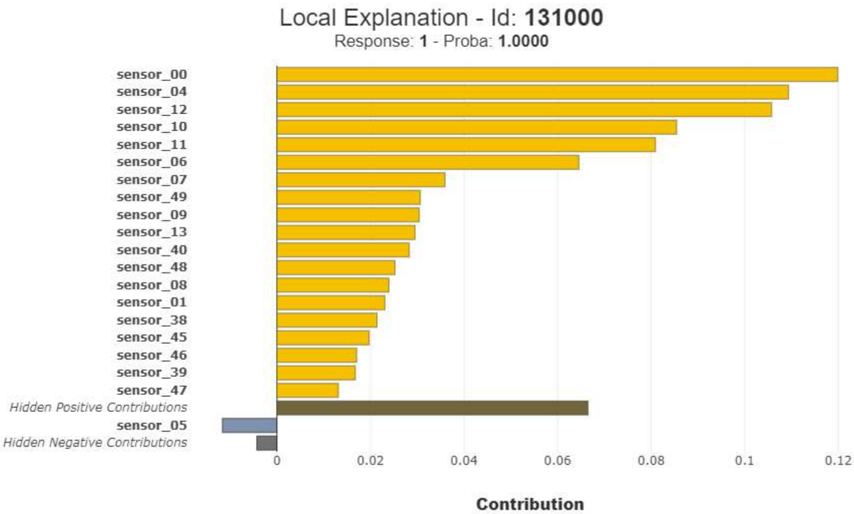

Fig. 15: Local explanation of index 131000

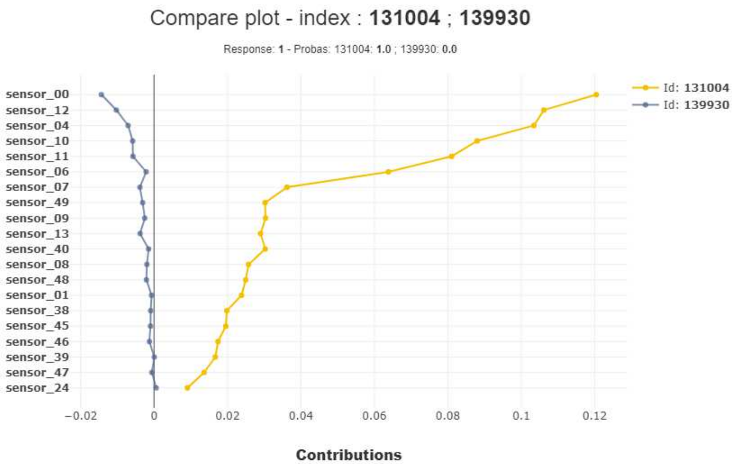

Fig. 16: Feature contribution compare plot of index 131004 and 139930

5.3 Quality of explanation - Stability

Stability is a way of enhancing trust in a method's explainability. For similar instances, similar explanations are intuitively anticipated. In other words, stability answers, " are the explanations identical for similar instances?".

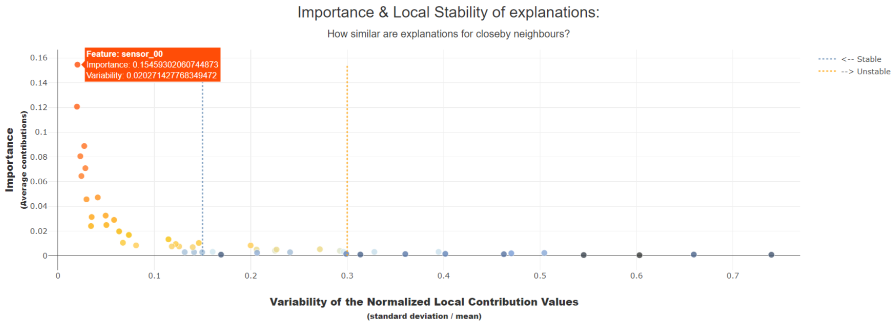

Fig. 17: Importance and local stability of explanations

The aforementioned plot examines the features and model results surrounding each instance provided. It displays the average importance of each feature throughout the dataset over its contributions on the y-axis and the average variability of each feature across the neighborhood of the instances on the x-axis. The features on the left are more stable than those on the right. Unlike lower features, the top ones are important. Here, features such as sensor 00, sensor 12, and sensor 04 on the left-hand side appear to have substantial and generally constant contributions; hence, they may be utilized with greater confidence as explanations. On the other hand, features such as sensor 37, sensor 30, and sensor 17 are far less stable, and we should use caution when evaluating explanations pertaining to these features.

6 Discussion

The article offers three different aspects:
1. Different anomaly detection techniques for an industrial asset dataset
2. Three popular machine learning-based fault classification approaches for the dataset.
3. An XAI framework for valuable insights and prospects for addressing significant maintenance challenges that will lead to better decision-making.
After data pre-processing steps, four different data-driven anomaly detection approaches were implemented, in which deep learning-based autoencoder had the best results among different methods. Machine learning-based fault classification techniques were implemented, in which Random Forest obtained the highest accuracy. For optimal hyperparameters of the algorithms used in

fault classification, Hyperopt (Bayesian optimization and sequential model-based optimization) hyperparameter optimization technique is used for better accuracy and optimized hyperparameters. Explainable frameworks, SHAP and LIME, were used. Shapash python library was used to implement the explainable artificial intelligence framework. Further, the stability metric was used to enhance the quality of the explanation and trust in the method's explainability.

7 Conclusion

This paper presents an explainable artificial intelligence-enabled anomaly detection and fault classification approach for an industrial asset. Despite machine learning techniques' notable success in industrial maintenance, their black-box character (poor interpretability) and generalization deficiency are substantial drawbacks. XAI has evolved into a study area focusing on machine learning interpretability to solve the issue of systems being unable to explain their judgments. By maintaining the level of predicted performance, this approach enables a more transparent AI. Mechanical issues result in hefty expenditures for production delays. By preventing problems from arising in maintenance, XAI reduces the effects of downtime and creates a more precise and transparent method to track its performance. Gaining the assurance and trust of clients requires XAI. By monitoring the system's performance, integrity, and correctness, XAI makes it more understandable and valuable. In predictive maintenance, understanding the importance of different features can help in identifying the key factors contributing to equipment failure and in developing effective maintenance strategies.
Explainable Artificial Intelligence (XAI) is a relatively new area of research, it is increasingly recognized as a critical requirement for the effective implementation of AI in real-world scenarios. Consequently, numerous scientific publications, workshops, and conferences are being held worldwide each year, proposing new XAI techniques and disseminating their findings. However, the resulting knowledge is often scattered. Most publications concentrate on XAI applications in natural language processing and computer vision, with relatively little research focusing on explaining methods used for time series data analysis. By leveraging ML And XAI tools, industries can gain valuable insights into potential issues before they occur, allowing them to take targeted action to address them. Despite the challenges that come with implementing these systems in industrial settings, the potential benefits are substantial. Like most research, this one provides valuable insights but has some limitations. This article discusses anomaly detection approaches and fault classification approaches. In later works, we plan to discuss interpretability in different remaining useful life (RUL) techniques.

Compliance with Ethical Standards

Funding: The authors declare that no financial support was received for this article's research, authorship, and/or publication.

Conflict of Interest: The authors declare no conflicts of interest.
Ethical approval: This article does not contain any studies with human participants performed by any of the authors.

References

Abdi, H., & Williams, L.J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433–459.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, *6*, 52138–52160.

Alfeo, A.L., Cimino, M.G., Vaglini, G. (2022). Degradation stage classification via interpretable feature learning. *Journal of Manufacturing Systems*, *62*, 972–983.

Althubaiti, A., Elasha, F., Teixeira, J.A. (2022). Fault diagnosis and health management of bearings in rotating equipment based on vibration analysis–a review. *Journal of Vibroengineering*, *24*(1), 46–74.

Angelotti, G., & Díaz-Rodríguez, N. (2023). Towards a more efficient computation of individual attribute and policy contribution for post-hoc explanation of cooperative multi-agent systems using myerson values. *Knowledge-Based Systems*, *260*, 110189.

Bai, M., Liu, J., Chai, J., Zhao, X., Yu, D. (2020). Anomaly detection of gas turbines based on normal pattern extraction. *Applied Thermal Engineering*, *166*, 114664.

Bergstra, J., Yamins, D., Cox, D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *International conference on machine learning* (pp. 115–123).

Bevilacqua, M., & Braglia, M. (2000). The analytic hierarchy process applied to maintenance strategy selection. *Reliability Engineering & System Safety*, *70*(1), 71–83.

Borghesi, A., Bartolini, A., Lombardi, M., Milano, M., Benini, L. (2019). A semisupervised autoencoder-based approach for anomaly detection in high performance computing systems. *Engineering Applications of Artificial Intelligence*, *85*, 634–644.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Caliva, F., De Ribeiro, F.S., Mylonakis, A., Demazi'ere, C., Vinai, P., Leontidis, G., Kollias, S. (2018). A deep learning approach to anomaly detection in nuclear reactors. *2018 international joint conference on neural networks (ijcnn)* (pp. 1–8).

Carvalho, T.P., Soares, F.A., Vita, R., Francisco, R.d.P., Basto, J.P., Alcalá, S.G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, 106024.

Cerrada, M., Zurita, G., Cabrera, D., Sánchez, R.-V., Artés, M., Li, C. (2016). Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mechanical Systems and Signal Processing*, *70*, 87–103.

Chandola, V., Banerjee, A., Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), 1–58.

Çınar, Z.M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability*, *12*(19), 8211.

Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., Barbosa, J. (2020). Machine learning and reasoning for predictive maintenance in industry 4.0: Current status and challenges. *Computers in Industry*, *123*, 103298.

Dekking, F.M., Kraaikamp, C., Lopuhaä, H.P., Meester, L.E. (2005). *A modern introduction to probability and statistics: Understanding why and how* (Vol. 488). Springer.

Došilović, F.K., Brčić, M., Hlupić, N. (2018). Explainable artificial intelligence: A survey. *2018 41st international convention on information and communication technology, electronics and microelectronics (mipro)* (pp. 0210–0215).

Du, M., Liu, N., Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, *63*(1), 68–77.

Faust, O., De Michele, S., Koh, J.E., Jahmunah, V., Lih, O.S., Kamath, A.P., ... others (2023). Automated analysis of small intestinal lamina propria to distinguish normal, celiac disease, and non-celiac duodenitis biopsy images. *Computer Methods and Programs in Biomedicine*, *230*, 107320.

Fu, Z., Hu, W., Tan, T. (2005). Similarity based vehicle trajectory clustering and anomaly detection. *Ieee international conference on image processing 2005* (Vol. 2, pp. II–602).

Garouani, M., Ahmad, A., Bouneffa, M., Hamlich, M., Bourguin, G., Lewandowski, A. (2022). Towards big industrial data mining through explainable automated machine learning. *The International Journal of Advanced Manufacturing Technology*, *120*(1-2), 1169–1188.

Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS one*, *11*(4), e0152173.

Hajgató, G., Wéber, R., Szilágyi, B., Tóthpál, B., Gyires-Tóth, B., Hős, C. (2022). Predmax: Predictive maintenance with explainable deep convolutional autoencoders. *Advanced Engineering Informatics*, *54*, 101778.

Iqbal, A., Zhao, G., Suhaimi, H., He, N., Hussain, G., Zhao, W. (2020). Readiness of subtractive and additive manufacturing and their sustainable amalgamation from the perspective of industry 4.0: A comprehensive review. *The International Journal of Advanced Manufacturing Technology*, *111*(9), 2475–2498.

Jakubowski, J., Stanisz, P., Bobek, S., Nalepa, G.J. (2021). Anomaly detection in asset degradation process using variational autoencoder and explanations. *Sensors*, *22*(1), 291.

Kankar, P.K., Sharma, S.C., Harsha, S.P. (2011). Fault diagnosis of ball bearings using machine learning methods. *Expert Systems with applications*, *38*(3), 1876–1886.

Kuzlu, M., Cali, U., Sharma, V., Güler, Ö. (2020). Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools. *IEEE Access*, *8*, 187814–187823.

Langer, M., & König, C.J. (2023). Introducing a multi-stakeholder perspective on opacity, transparency and strategies to reduce opacity in algorithm-based human resource management. *Human Resource Management Review*, *33*(1), 100881.

Lee, J., Jung, O., Lee, Y., Kim, O., Park, C. (2021). A comparison and interpretation of machine learning algorithm for the prediction of online purchase conversion. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(5), 1472–1491.

Liu, F.T., Ting, K.M., Zhou, Z.-H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *6*(1), 1–39.

Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30.*

Molnar, C. (2020). *Interpretable machine learning.* Lulu. com.

Mushtaq, R. (2011). Augmented dickey fuller test.

Omar, S., Ngadi, A., Jebur, H.H. (2013). Machine learning techniques for anomaly detection: an overview. *International Journal of Computer Applications*, *79*(2).

Prytz, R., Nowaczyk, S., Rögnvaldsson, T., Byttner, S. (2015). Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data. *Engineering applications of artificial intelligence*, *41*, 139–150.

Puggini, L., & McLoone, S. (2018). An enhanced variable selection and isolation forest based methodology for anomaly detection with oes data. *Engineering Applications of Artificial Intelligence*, *67*, 126–135.

Ribeiro, M.T., Singh, S., Guestrin, C. (2016). " why should i trust you?" explaining the predictions of any classifier. *Proceedings of the 22nd acm* *sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144).

Rivera, A., Muñoz, J.C., Pérez-Goody, M., de San Pedro, B.S., Charte, F., Elizondo, D., ... del Jesus, M. (2023). Xaire: An ensemble-based methodology for determining the relative importance of variables in regression tasks. application to a hospital emergency department. *Artificial Intelligence in Medicine*, 102494.

Roelofs, C.M., Lutz, M.-A., Faulstich, S., Vogt, S. (2021). Autoencoder-based anomaly root cause analysis for wind turbines. *Energy and AI*, *4*, 100065.

Ruiz-Sarmiento, J.-R., Monroy, J., Moreno, F.-A., Galindo, C., Bonelo, J.-M., Gonzalez-Jimenez, J. (2020). A predictive model for the maintenance of industrial machinery in the context of industry 4.0. *Engineering Applications of Artificial Intelligence*, *87*, 103289.

Sakurada, M., & Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. *Proceedings of the mlsda 2014 2nd workshop on machine learning for sensory data analysis* (pp. 4–11).

Samanta, B., & Al-Balushi, K. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical systems and signal processing*, *17*(2), 317–328.

Sharma, J., Mittal, M.L., Soni, G. (2022). Condition-based maintenance using machine learning and role of interpretability: a review. *International Journal of System Assurance Engineering and Management*, 1–16.

Susto, G.A., Schirru, A., Pampuri, S., McLoone, S., Beghi, A. (2014). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE transactions on industrial informatics*, *11*(3), 812–820.

Tiddens, W.W. (2018). Setting sail towards predictive maintenance: developing tools to conquer difficulties in the implementation of maintenance analytics.

Tjoa, E., & Guan, C. (2020). Quantifying explainability of saliency methods in deep neural networks. *arXiv preprint arXiv:2009.02899*.

Vilalta, R., & Ma, S. (2002). Predicting rare events in temporal domains. *2002 ieee international conference on data mining, 2002. proceedings.* (pp. 474–481).

Wang, Y., Wong, J., Miner, A. (2004). Anomaly intrusion detection using one class svm. *Proceedings from the fifth annual ieee smc information assurance workshop, 2004.* (pp. 358–364).

Wang, Z., Qian, K., Liu, H., Hu, B., Schuller, B.W., Yamamoto, Y. (2023). Exploring interpretable representations for heart sound abnormality detection. *Biomedical Signal Processing and Control*, *82*, 104569.

Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing*, *21*(6), 2560–2574.

Widodo, A., Yang, B.-S., Han, T. (2007). Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert systems with applications*, *32*(2), 299–312.

Xiao, L., Shao, Z., Liu, G. (2006). K-means algorithm based on particle swarm optimization algorithm for anomaly intrusion detection. *2006 6th world congress on intelligent control and automation* (Vol. 2, pp. 5854–5858).

Xuyun, F., Hui, L., Zhong, S., Lin, L. (2019). Aircraft engine fault detection based on grouped convolutional denoising autoencoders. *Chinese Journal of Aeronautics*, *32*(2), 296–307.

Yang, B.-S., Di, X., Han, T. (2008). Random forests classifier for machine fault diagnosis. *Journal of mechanical science and technology*, *22*(9), 1716–1725.

Yang, J., Zhang, Y., Zhu, Y. (2007). Intelligent fault diagnosis of rolling element bearing based on svms and fractal dimension. *Mechanical Systems and Signal Processing*, *21*(5), 2012–2024.

Zhang, C., Hu, D., Yang, T. (2022). Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and xgboost. *Reliability Engineering & System Safety*, *222*, 108445.

Zhang, W., Yang, D., Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal*, *13*(3), 2213–2227.

Zhao, Y., Ren, J., Zhang, B., Wu, J., Lyu, Y. (2023). An explainable attention-based tcn heartbeats classification model for arrhythmia detection. *Biomedical Signal Processing and Control*, *80*, 104337.

Zhu, X. (2015). Machine teaching: An inverse problem to machine learning and an approach toward optimal education. *Proceedings of the aaai conference on artificial intelligence* (Vol. 29).

Zurada, J. (1992). *Introduction to artificial neural systems.* West Publishing Co.