

Article

Predicting Furnace Tube Rupture Using Multiclass Decision Forest with Explainable Risk-Based Lead-Time Classification

Saharudin Haron, Muhammad Taqiuddin Baharum *  and Shamimimraphay Shahul Hameed

Faculty of Chemical and Energy Engineering, University Teknologi Malaysia, Johor Bahru 81310, Malaysia; saharudinh@utm.my (S.H.)

* Correspondence: taqiuddinbaharum@yahoo.com

Abstract

Furnace tube rupture is a critical safety and reliability issue in high-temperature industrial systems, often leading to unplanned shutdowns, severe economic losses, and safety incidents. Conventional monitoring approaches based on threshold alarms and single-variable diagnostics are frequently inadequate for detecting early degradation under complex multivariable operating conditions. This study proposes a multiclass predictive maintenance framework using a decision forest algorithm to predict furnace tube rupture severity from industrial operational data. The dataset comprised 10,957 observations and 56 furnace operating parameters collected from industrial historian systems. Following preprocessing and Pearson correlation-based feature selection, 19 significant parameters were retained for model development. The proposed framework integrates data preprocessing, feature selection, hyperparameter optimization, and validation using unseen operational data from 2021. The results demonstrate that the model effectively captures rupture-risk trends and provides early warning signals more than 14 days before rupture events, with high-risk classifications exceeding 70% predictive probability. Unlike conventional binary classification methods, the proposed multiclass framework enables risk-based maintenance decision-making, including targeted inspections, load reduction, and scheduled shutdown planning. The findings, based on validation against a single industrial furnace system, highlight the effectiveness of ensemble machine learning techniques in improving predictive maintenance, operational safety, and reliability engineering for this class of industrial furnace; broader generalization across furnace configurations and sites remains to be confirmed through multi-installation validation.

Keywords: predictive maintenance; furnace tube rupture; decision forest; machine learning; industrial safety; feature selection



Academic Editors: Maja Trstenjak,
Hrvoje Cajner, Tihomir Opetuk and
Paolo Renna

Received: 20 May 2026

Revised: 16 July 2026

Accepted: 27 July 2026

Published: 7 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Industrial furnaces are widely used as thermal processing systems in the petrochemical refining, metallurgical processing, and manufacturing industries. They directly control productivity, product quality, and energy efficiency in their operation. Furnace tubes are one of the most vulnerable parts of these systems that are subjected to high-temperature creep, oxidation, carburization, sulfidation, and cyclic thermal fatigue [1]. The damage caused by these degradation mechanisms causes the loss of structural integrity of the tube and can lead to catastrophic rupture events [2,3], which cause unplanned shutdowns, substantial economic losses, equipment damage, and safety incidents.

The operational consequences of furnace tube rupture are severe and significant, with impacts that include economic losses, environmental risks, equipment damage, and safety risks. In November 2023, a furnace tube burst during startup operations at the Marathon Martinez Renewables Facility in California, causing a major fire and an extended plant shutdown [4]. Investigation reports found that tube degradation was a major contributing factor and that inadequate monitoring practices were a significant cause of the failure [5]. Likewise, the U.S. Chemical Safety and Hazard Investigation Board (CSB) reported instances of tubes deforming too much before the outer diameter exceeded 3% of the original outer diameter before being identified, meaning the degradation was not detected in time to prevent failure [6], as shown in Figure 1. These incidents expose a fundamental limitation of conventional monitoring approaches, which remain largely reactive and rely on periodic visual inspections, single-variable temperature monitoring, and threshold-based alarm systems that are structurally incapable of detecting early-stage degradation under complex, multivariable operating conditions [7,8].



Figure 1. Surveillance footage of the heater fire source [6].

Operational data from distributed control systems (DCS), supervisory control and data acquisition (SCADA) systems, and plant historians are generated continuously and on a large scale in modern industrial plants. However, a significant portion of this data remains underutilized for predictive reliability applications [9]. Machine learning (ML) has become a promising approach for uncovering predictive intelligence from high-dimensional industrial data, capturing nonlinear relationships among process variables that traditional statistical methods do not model well [10,11]. In the industry, ensemble learning techniques such as decision forest models have been found effective for fault prediction tasks because they are robust to noisy data, can work with process variables different scales, and do not suffer from overfitting [12,13].

Most conventional furnace monitoring approaches remain reactive rather than predictive. Traditional maintenance systems are mainly based on scheduled inspections or monitoring of a single variable, like the tube metal temperature (TMT) measurement. While widely adopted in industrial practice, the TMT-based monitoring generally detects abnormalities only after significant degradation has already occurred [14]. Similarly, rule-based alarm systems are not able to perform well in identifying nonlinear and complex relationships amongst operational variables that can indicate early failure mechanisms.

Machine learning (ML) has improved the field of industrial process monitoring and predictive maintenance over the last few years. In particular, machine learning techniques are well-suited for the analysis of high-dimensional datasets from the industry, for the detection of indicated correlations, and for the modeling of the nonlinear behavior of processes, which cannot be described adequately by conventional statistical methods [10]. In the field of chemical engineering applications, the ML models are good predictors for process optimization, fault detection, soft sensing, and assessment of the reliability of equipment [11]. In particular, the use of artificial neural network (ANN) models to

predict olefin yields in hydrocarbon cracking processes has been demonstrated to be successful, where the complex relationship between the operating conditions and the reaction performance is learned [15]. These studies highlight the growing importance of data-driven intelligence in industrial thermal systems.

In recent years, efforts have been made to expand machine learning applications for furnaces. Ref. [14] showed that the accuracy of the tube metal temperature estimation was higher than that of the traditional monitoring methods by using a soft-sensing method based on neural networks. Likewise, ensemble learning algorithms, including the random forest model, the gradient boosting model, and decision forest models, have been found to perform well on industrial fault prediction problems due to their robustness, interpretability, and their ability to deal with process variables of mixed scales [12]. Ensemble tree-based methods are especially well-suited for industrial process data applications, where these nonlinear interactions, noise, and multicollinearity are typical issues.

In addition to thermal systems, predictive machine learning models have proven useful in various engineering and industrial applications. In industrial settings where Industry 4.0 is being adopted, ML-based predictive maintenance systems have enhanced operational efficiency by providing a way to diagnose faults in systems early, optimize industrial processes, and manage industrial assets intelligently [16,17]. Random forest and artificial neural network (ANN) models provided better performance than formula-based methods in predicting the optimal tracheal tube depth for a pediatric population, highlighting the flexibility of machine learning for safety-critical prediction tasks in healthcare engineering applications [18]. The interdisciplinary achievements indicate the potential of machine learning frameworks for making a substantial contribution to the reliability management of industrial furnaces.

Despite these advances, existing machine learning approaches for furnace monitoring share a critical limitation: Most research uses a binary furnace fault classification, which only categorizes faults as failures and normal working conditions. The operational resolution issue in binary models is that industrial degradation is a continuous and progressive process rather than a discrete event [19]. Furthermore, binary models often trigger too many false alarms or insufficient warning periods for desired maintenance scheduling. The practical applicability of existing models is further constrained by the absence of prospective validation on independent unseen industrial operational data [3,9]. This study proposes a multiclass decision forest model to address these limitations for furnace tube rupture prediction. Rather than conventional binary classification, the proposed framework classifies operational risk into three actionable severity levels: low-, medium-, and high-risk rupture levels defined by quantitative lead time to failure thresholds [20]. These categories correspond to specific maintenance intervention strategies such as routine monitoring, targeted inspections, and planned shutdown [21].

This study makes three practical contributions. First, the proposed framework categorizes furnace operating conditions into three desirable levels of actionable risk, based on quantitative lead-time-to-failure thresholds that were determined using historical furnace rupture data, and directly supports risk-based maintenance decision-making. Second, the framework incorporates feature selection using Pearson correlation and explainability analysis using XGBoost (version 2.1.4), which ensures that the model predictions are physically interpretable and traceable to the known furnace degradation mechanisms, such as overheating, creep damage, oxidation, and thermal fatigue. Third, the model was validated on a second, independent set of industry operational data for 2021, demonstrating that its predictive performance holds beyond the original training periods for this furnace system. The model offered an early warning lead time of more than 14 days before a rupture event, with high-risk classifications having predictive probabilities of more than 70%, which is

operationally useful for maintenance scheduling and/or safety planning. As discussed in the limitation section, these contributions reflect the applied combination of established feature selection and ensemble learning techniques with a lead time-based multiclass risk scheme, rather than a new algorithmic method.

2. Methodology

2.1. Research Framework and Data Description

This study proposed a machine learning-based predictive maintenance model to predict the severity of industrial furnace tube ruptures. The proposed methodology combined historical operational data, feature engineering, multiclass classification, and predictive validation to form an intelligent early warning system for furnace operation. The data were obtained from furnace operating conditions recorded by the PI Process Book system, which operates in an industrial environment. Operational parameters were measured every four hours, yielding 10,957 observations and 56 initial process parameters. These variables were the key furnace operating parameters related to tube integrity, thermal behavior, combustion stability, and process efficiency. The overall process was divided into six key steps: data acquisition, preprocessing, feature selection, model development, hyperparameter tuning, and predictive validation.

The framework was developed specifically to categorize the severity of any rupture into one of the following actionable operational risk categories to aid predictive maintenance planning as shown Figure 2.

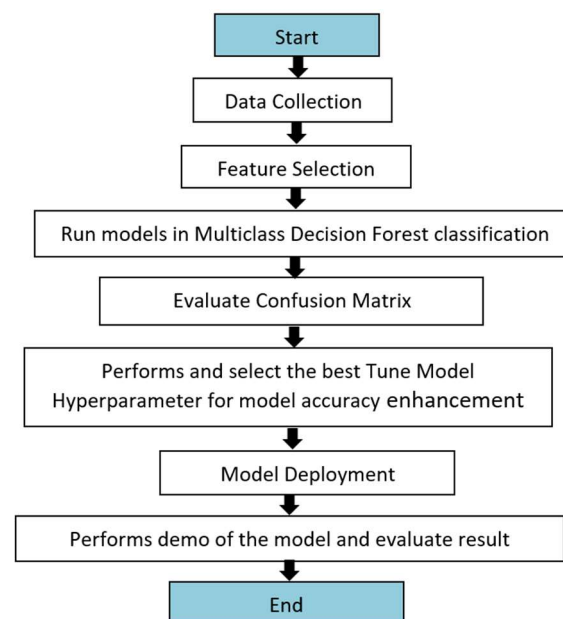


Figure 2. Flowchart of model deployment and prediction process.

2.2. Data Preparation and Preprocessing

The dataset was pre-processed thoroughly before developing the model to deal with the data analytically and to make the model reliable. Industrial process data are often incomplete, have inconsistent sensing data, and contain noise in the operational signals because of sensor limitations and interruptions in the process. Hence, preprocessing of data for cleaning and normalization was essential before the application of ML algorithms.

In the event that missing observations were initially flagged using placeholder identifiers during data extraction, they were subsequently treated during preprocessing to avoid bias in model training. The dataset was then filtered for inconsistencies and abnormal

process data. To minimize the number of dimensions for calculation and performance efficiency, feature selection was performed using the “compute linear correlation” function in Azure Machine Learning Studio. The relationship between furnace parameters and the rupture conditions was analyzed by correlation analysis to identify and understand the correlation between the parameters and the most influential parameters in the furnace that are associated with the furnace rupture condition. A Pearson correlation coefficient was computed as in Equation (1):

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum [\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}]} \quad (1)$$

The mean values of x and y are denoted by $(\bar{x})$ and $(\bar{y})$, respectively, where (x_i) and (y_i) refer to individual observations of these variables, and (n) indicates the number of observations. The correlation threshold of $|r| \geq 0.30$ was applied to identify parameters with meaningful linear association with the rupture-risk label. This threshold reduced the feature set from 56 to 19 operational parameters, retaining those most physically and statistically meaningful for rupture-risk classification.

2.3. Feature Selection and Explainability

Feature importance analysis was conducted using the XGBoost classifier to improve model interpretability and support practical deployment. The results indicated that furnace air register control is the most significant predictor of the rupture-risk classification, followed by TLE outlet temperature, TLE A inlet pressure, north cross over pressure, and floor burner north flow. These variables are directly linked to combustion stability, heat transfer efficiency, thermal loading, and pressure fluctuations in the furnace system. The dominance of temperature and pressure parameters confirms that abnormal conditions in these variables are strongly associated with degradation mechanisms in furnaces. While Pearson correlation captures linear associations between operational parameters and the rupture-risk label, it is acknowledged that process variables may also exhibit nonlinear interactions. To address this, XGBoost-based feature importance analysis was also employed to validate the selected features from a nonlinear ensemble perspective. The XGBoost importance scores confirmed that the 19 parameters identified through Pearson correlation are also the top predictors in the nonlinear model, providing cross-validation of the feature selection outcome. Future work will explore mutual information and SHAP-based methods to further capture nonlinear variable interactions as shown in Table 1.

Table 1. Selected operational parameters, measurement units, and relevance to tube degradation.

No.	Parameter Name	Unit	Location	Relevance to Tube Degradation
1	Furnace Draft Control	mmH ₂ O	Furnace draft	Abnormal draft pressure disrupts combustion airflow, causing localized overheating and uneven heat distribution across tube coils.
2	North COT Control	MW	Coil outlet	Coil outlet temperature (COT) directly reflects thermal load on process coils; sustained elevation accelerates creep damage in tube walls.
3	TLE Outlet Temperature	°C	TLE outlet	Elevated transfer line exchanger (TLE) outlet temperature indicates insufficient heat recovery, contributing to downstream thermal stress.
4	HHPS Product Flow	t/h	Steam header	High-pressure steam product flow reflects furnace heat balance; deviations indicate energy imbalance and potential overloading.

Table 1. Cont.

No.	Parameter Name	Unit	Location	Relevance to Tube Degradation
5	Furnace Coil A Outlet Temperature to TLE	°C	Coil A exit	Coil A-specific outlet temperature provides localized thermal monitoring; hotspots here correlate with early-stage tube degradation.
6	Wall Burner North Flow	kg/h	North wall burner	Imbalanced burner fuel flow can create uneven heat flux distribution across the radiant section, causing localized tube overheating.
7	Furnace SQE CG Outlet Temperature	°C	SQE outlet	Secondary quench exchanger (SQE) crack gas outlet temperature indicates cracking severity; high values correlate with excessive tube temperatures.
8	Furnace Steam Drum Level	%	Steam drum	Steam drum level fluctuations can indicate unstable steam generation, which affects tube cooling and thermal equilibrium.
9	Furnace Air Register Control	%	Air register	Air register position controls combustion air supply; incorrect settings produce incomplete combustion and localized high-temperature zones.
10	North Crossover Pressure	barg	North crossover	Crossover pressure reflects process fluid flow resistance; elevated values may indicate coking or mechanical restriction inside tubes.
11	TLE A Inlet Pressure	barg	TLE A inlet	TLE inlet pressure changes indicate flow restrictions or fouling upstream in Coil A, signaling accelerated tube degradation.
12	Furnace Steam Drum Pressure	barg	Steam drum	Steam drum pressure is directly linked to heat generation rate; deviations from set-point indicate process imbalance and thermal stress.
13	Furnace DMDS to Dilution Steam Flow Control	%	Dilution steam	DMDS (dimethyl disulfide) and dilution steam flow control the hydrocarbon partial pressure; deviations affect coke deposition rate and tube wall temperature.
14	Coil A Delta Pressure	ratio	Coil A	Coil pressure differential is a direct indicator of coke buildup inside tubes; increasing ΔP signals progressive internal fouling and flow restriction.
15	TLE A Outlet Temperature	°C	TLE A outlet	TLE A outlet temperature monitors heat recovery efficiency; an abnormal rise indicates tube fouling or degradation in the TLE section.
16	HPS Superheater Outlet Temperature Control	°C	Superheater outlet	Superheater outlet temperature reflects the overall furnace thermal output; sustained high values indicate the furnace operating beyond design limits.
17	Floor Burner North Flow	t/h	North floor burner	Floor burner fuel flow distribution affects radiant section heat uniformity; imbalanced floor burner flow creates high heat flux zones on lower tube sections.
18	Furnace Coil C Outlet Temperature to TLE	°C	Coil C exit	Coil C outlet temperature enables cross-coil thermal comparison; divergence from Coil A values identifies asymmetric degradation patterns.
19	Furnace Draft Control (Output Signal)	%	Furnace draft	The control output signal complements the measured draft value, providing insight into controller response and draft regulation behavior.

2.4. Model Development and Algorithm Selection

After data preprocessing and feature selection, the data were fed into various machine learning classifiers, and their performance was compared in terms of classification accu-

racy. The comparative analysis was conducted using multiclass decision forest, random forest, SVM, and XGBoost classifiers to evaluate their suitability for rupture-risk prediction. Studies in the industrial field have also shown that ensemble decision forest classifiers are superior to some traditional classifiers for predictive maintenance systems because of their robustness and ability to handle nonlinear relationships in the processes. The developed model was trained to classify the operating condition of the furnace into low-, medium-, and high-risk categories for furnace ruptures. These severity classes were correlated with thresholds for lead-time-to-failure to support maintenance scheduling and operational decisions. Historical process parameters were used as inputs to the model, and the levels of rupture severity were used as outputs for classification as show in Table 2.

Table 2. Definition of rupture-risk classes and maintenance responses.

Risk Class	Meaning	Maintenance Action
Low Risk	Normal operating condition; lead time > 30 days to rupture; no immediate concerns	Routine monitoring (4 h intervals)
Medium Risk	Early degradation indicators detected; lead time 7–30 days to rupture	Targeted inspection; reduce load 10–15%; notify maintenance
High Risk	High rupture probability; lead time within 7 days to rupture; immediate action is required.	Initiates controlled shutdown emergency inspection; ISO 31000 protocol [22]

This multiclass classification structure provided significant operational advantages over the conventional binary fault detection methods. The three risk classes were determined using quantitative thresholds for lead-time-to-failure, derived from a historical rupture database with the assistance of the plant engineers. Observations taken over 30 days before the confirmed rupture event were deemed low risk and considered normal operating conditions with no immediate degradation concern. Observations made within 7 to 30 days of the rupture event were classified as medium risk, indicating early signs of degradation and a need for more specific inspection. Observations made within 7 days of a rupture event were classified as high risk, indicating a high likelihood of rupture and a need for immediate maintenance action. Requirements for plant operational maintenance scheduling determined these thresholds, which were then verified against the PI Process Book historian system’s historical rupture information. All labels were assigned by domain experts using known rupture timestamps, and a reproducible, physically meaningful classification scheme was used, consistent with ISO 31000 risk management principles [23].

2.5. Multiclass Decision Forest Architecture

In this study, the core predictive model was provided by the algorithm known as Multiclass Decision Forest, a standard ensemble learning component available in Azure Machine Learning Studio SDK V2. Ensemble learning methods, such as decision forests, are methods of combining multiple decision trees to increase classification accuracy and generalization performance [20]. During training, many decision trees are created using various subsets of the operational data. Every tree reveals a unique pattern and relationships among the process variables.

The final classification is done by combining the results of all the decision trees in the ensemble using a majority voting approach rather than using only one tree for prediction. Such an approach is very robust and less prone to overfitting, especially in the case of high-dimensional industrial process data with high nonlinear interactions and noisy operating conditions. The multiclass decision forest model has one main advantage in that it can handle complex industrial data with mixed-size operational variables, as shown in Figure 3.

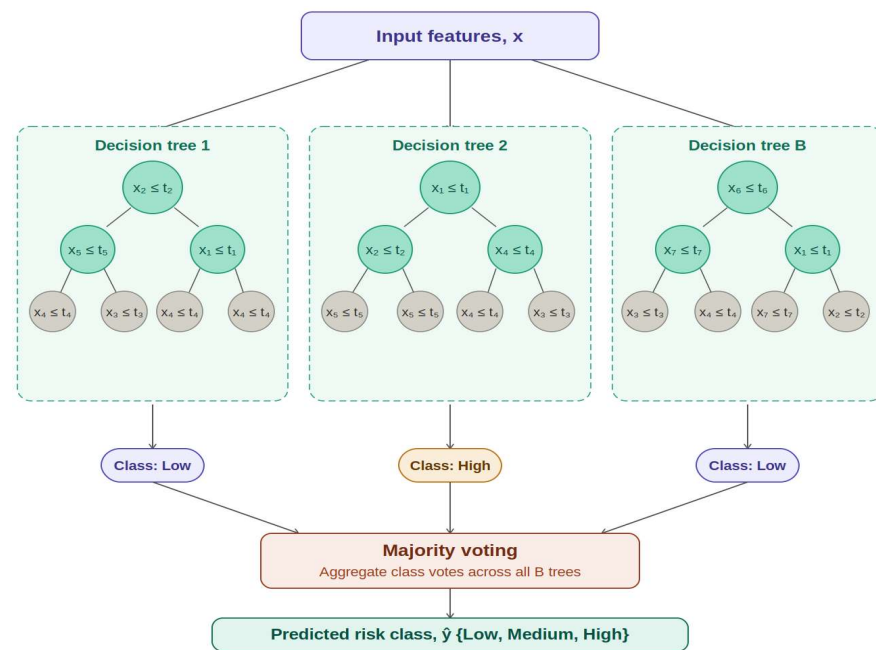


Figure 3. Architecture of the proposed multiclass decision forest framework for rupture-risk classification.

The architecture illustrates the flow of operational data from preprocessing and feature selection through multiclass decision forest classification to rupture-risk prediction.

2.6. Hyperparameter Optimization and Accuracy Enhancement

Hyperparameter optimization was applied to improve the model's predictive performance. Hyperparameters directly influence the model's behavior and have a considerable effect on classification accuracy, robustness, and computational efficiency. Various optimization techniques were explored, including grid search, random grid search, and random sweep methods. Grid search systematically evaluates combinations of hyperparameters across predefined ranges. This method is computationally intensive but can yield the optimal parameter configuration. Random grid search, on the other hand, explores the parameter space randomly, thus providing less computation with still very competitive results. Random sweep optimization is a hybrid method that intelligently combines systematic and random exploration to ensure faster computation of the results, as shown in Figure 4.

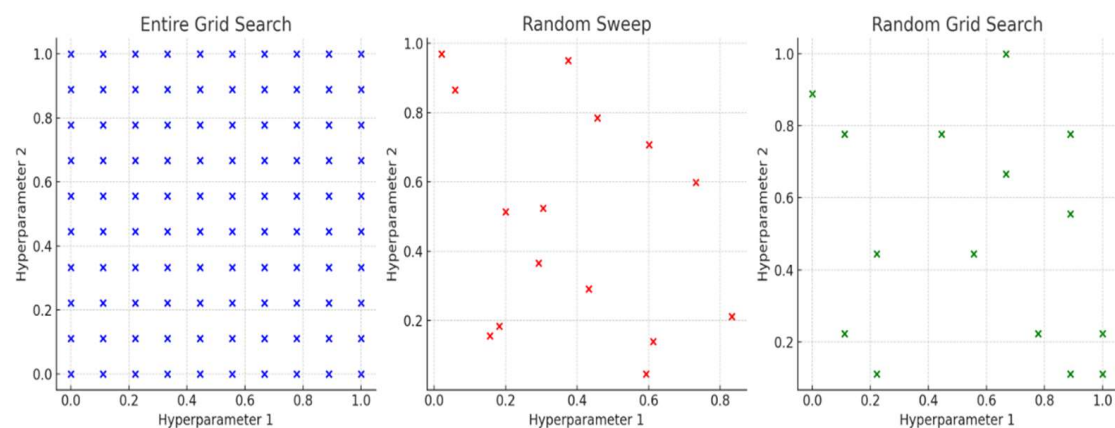


Figure 4. Hyperparameter tuning strategies: (Left) entire grid search, (Center) random sweep, and (right) random grid search.

The parameters optimized were related to the ensemble size, the depth of trees, the split criterion for the tree, and the selection of features in the behavior. Hyperparameter tuning improved the accuracy and reliability of predictions of the models by maximizing their capacity to generalize and minimizing classification errors.

The multiclass decision forest model employed 200 decision trees with an 80:20 training–testing split and a random state of 42. Hyperparameter optimization was conducted using grid search, random grid search, and random sweep strategies, with the final configuration selected based on validation accuracy and generalization performance as reported in Table 3.

Table 3. Experimental settings and comparative model parameters.

Parameter	Value
Training-Test Split	80:20
Random State	42
SVM Kernel	RBF
Number of Decision Trees	200
Maximum Tree Depth	Auto (controlled by min leaf size)
Minimum Samples per Leaf	1
Feature Fraction per Split	All features
Optimization Method	Random Sweep
XGBoost Objective	Multiclass Classification
Evaluation Metric	mlogloss
Feature Scaling	StandardScaler

3. Result and Discussion

3.1. Comparative Performance of Machine Learning Models

The comparative performance of all evaluated models is presented in Table 4. Among the baseline classifiers, XGBoost (version 2.1.4) achieved the highest accuracy of 99.28%, outperforming random forest version 1.5.2 (98.19%), LSTM Tensorflow version 2.17.0 (98.05%), and SVM scikit-learn version 1.5.2 (95.31%). The boosting mechanism of XGBoost sequentially reduces prediction error and considers complex nonlinear relationships among furnace operating parameters, explaining its strong performance among the comparison models. However, the proposed multiclass decision forest achieved a perfect classification on the available test dataset (100% accuracy) with zero false alarms, outperforming all comparison models, including the deep learning LSTM baseline. The ensemble majority-voting structure of the decision forest provided superior generalization across all three risk classes, including the minority high-risk class, where detection failure has the most severe operational consequences.

Table 4. Performance comparison of the evaluated machine learning models.

Model	Accuracy	Precision	Recall	F1-Score	FAR
SVM	95.31	95.56	95.31	94.77	--
Random Forest	98.19	98.23	98.19	98.08	--
XGBoost	99.28	99.28	99.28	99.26	0.72
LSTM (Deep Learning Baseline)	98.05	97.89	98.05	97.96	1.95
Multiclass Decision Forest (Proposed)	100.00	100.00	100.00	100.00	0.00

The outcomes confirm the suitability of ensemble-based methods for predictive maintenance in furnace systems for industrial applications.

3.2. Confusion Matrix and False Alarm Analysis

The confusion matrix presented in Figure 5 indicates that the proposed multiclass decision forest model achieved excellent classification performance across the three rupture-risk categories (low, medium, and high) on the available test dataset. The model correctly classified all observations in the test set, resulting in class-wise precision, recall, and F1-scores of 1.00, as summarized in Table 5. However, because the high-risk category contained only four observations, these performance estimates should be interpreted cautiously and require validation using larger datasets with a greater number of rupture events.

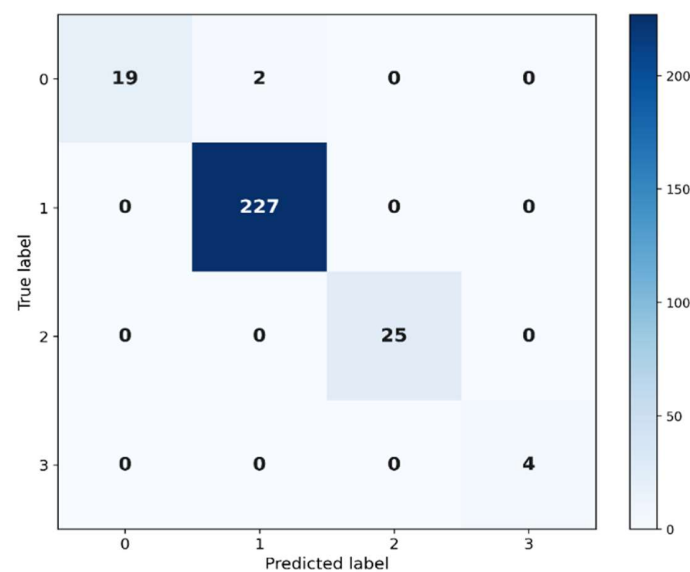


Figure 5. Confusion Matrix for Multiclass Risk Classification (XGBoost).

Table 5. Class-wise performance metrics of the proposed multiclass decision forest model (proposed model, test set $n = 256$).

Class	Precision	Recall	F1-Score	Support
Low Risk	1.00	1.00	1.00	227
Medium Risk	1.00	1.00	1.00	25
High Risk	1.00	1.00	1.00	4
Overall/Weighted	1.00	1.00	1.00	256

The proposed multiclass decision forest achieved perfect classification performance on the available test dataset ($n = 256$). However, the high-risk class contained only four observations. The precision, recall, and F1-score that were obtained were all 1.00 and these performance measures must therefore be treated with some caution as they were calculated on a test set that only included four observations with a high risk classification. Thus, the metrics reported should not be viewed as general performance metrics, but rather as indicative of success on the set of available industrial data. The proposed framework, therefore, may also provide a significantly lower false alarm rate (FAR) without compromising detection of high-risk conditions for rupture. However, more validation with larger and more diverse industrial datasets is needed to validate this performance in more operating conditions.

The proposed model was further evaluated using an independent deployment dataset comprising 2021 operational data, as discussed in the model validation section below.

3.3. Explainable Feature Importance Analysis

The feature importance analysis of the XGBoost classifier was performed to enhance the transparency of the model and meet the explainability needs of industrial predictive maintenance systems. Figure 6 shows the 10 most important variables for rupture prediction in operation. Furnace air register control had the highest importance score of all variables, reflecting the importance of regulating combustion air in the furnace to ensure thermal stability. Due to improper air distribution, overheated areas may be formed, which will lead to the degradation of tube metal [24]. Temperature at the outlet of the TLE (TLE outlet temperature) and TLE inlet pressure were also found to be highly influential parameters; the former causes thermal stress on the furnace tubes, and the latter provides conditions that favor creep damage mechanisms. North cross over pressure and floor burner north flow were found to influence rupture-risk classification significantly [25]. The changes in crossing pressure may reflect the imbalance of the process and the uneven distribution of heat. At the same time, the abnormal burner flow rate may generate excessive heat flux, which will lead to the formation of hot spots. The following operating conditions are well known as factors that lead to tube thinning, metallurgical degradation, and potential tube rupture.

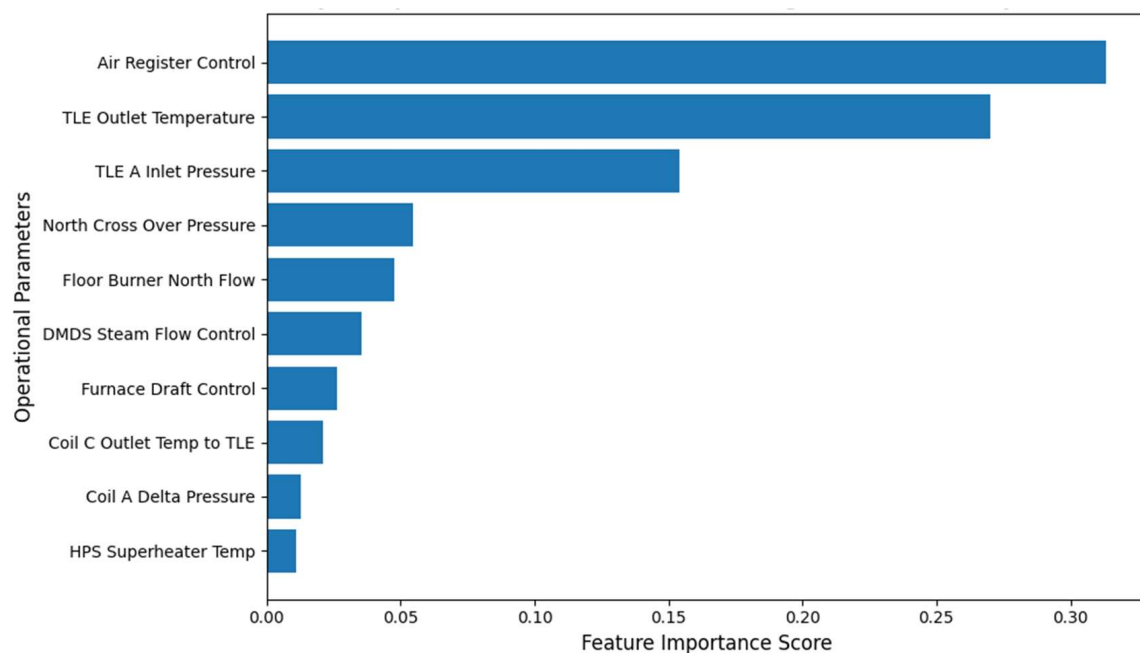


Figure 6. Top 10 operational parameters influencing furnace tube rupture-risk prediction. The x-axis represents the normalized XGBoost feature importance score (F-score). Parameters shown are: Air register control, TLE outlet temperature, TLE A inlet pressure, north crossover pressure, floor burner north flow, DMDS steam flow control, Furnace draft control, Coil C outlet temperature, Coil A delta pressure, and HPS Superheater outlet temperature.

The results demonstrate that the proposed framework captures meaningful operational patterns associated with known furnace degradation mechanisms rather than functioning as a black-box prediction system. Therefore, the model provides both predictive capability and engineering interpretability, supporting risk-informed maintenance planning and operational decision-making.

3.4. Multiclass Decision Forest Performance Evaluation

The algorithm found to be the best predictive model for classifying furnace tube ruptures is the multiclass decision forest algorithm. A comparative study with other machine learning classifiers showed that ensemble-based learning techniques are more accurate, robust, and generalize better. The model divided furnace operating conditions into three risk categories: low, medium, and high. It is a multiclass structure that is a significant improvement over traditional binary classification, which only distinguishes between failure and non-failure, as shown in Figure 7. Although the proposed multiclass decision forest achieved perfect classification on the available test dataset ($n = 256$), the high-risk class contained only four observations. Thus, the reported performance needs to be understood with caution, as the performance is based on a small minority class sample, which is not statistically overwhelming enough to demonstrate the actual predictive reliability in the real world. Given this, the findings are not definitive evidence of the predictive power, but rather preliminary evidence. These studies should be confirmed in future with larger data sets including more rupture events from various industrial furnace systems.

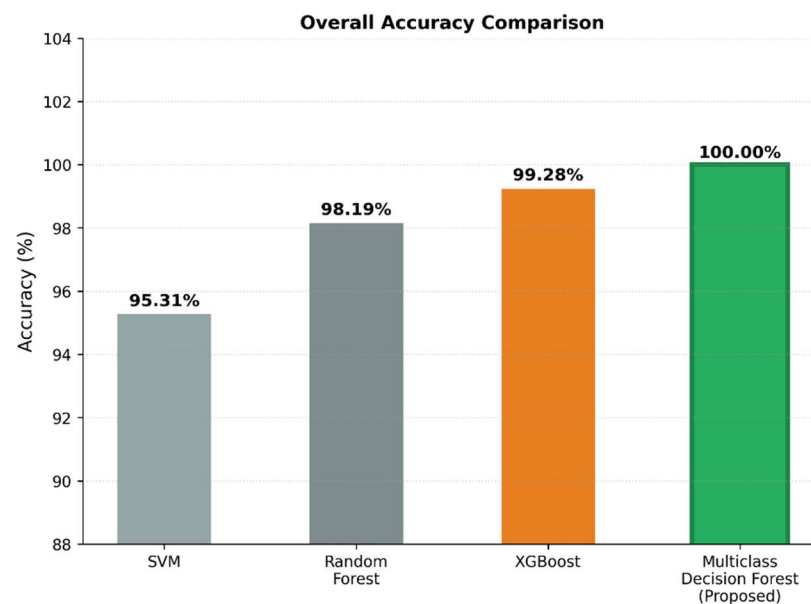


Figure 7. Comparative accuracy of SVM, random forest, XGBoost models, and multiclass decision forest.

Grading of the risk level of the degradation pattern allows detection earlier and provides actionable information for maintenance planning. Decision forest models have been extensively shown to be superior in industrial predictive maintenance tasks [21]. Likewise [13] argued that reducing variance and overfitting in the classification process can be achieved by combining multiple decision trees, thereby increasing the stability of the classification process, which is also known as ensemble learning.

The ensemble structure enabled the model to learn complex interactions between furnace variables that could not be detected by other statistical or linear approaches. The use of multiple voting in the decision trees greatly enhanced the reliability of prediction and robustness to noisy input data [19]. The results of this study provide strong empirical support that ensemble learning, especially based on the majority voting strategy, can solve uncertainty and changing operating conditions problems in an industrial setting.

3.5. Optimization of Hyperparameters and Enhancement of Models

Hyperparameter optimization was one of the key steps in improving the predictive capabilities of the proposed model. Three optimization methods were used in tuning: entire grid search, random grid search, and random sweep. Various techniques were employed to determine the best set of parameters for the models, such as their depth, number of estimators, and feature splitting criteria [2,26].

Grid search systematically searches every possible combination of hyperparameters within a specific search space. Although this guarantees an optimal configuration, it is very time-consuming, especially when dealing with large datasets [27]. Random grid search randomly samples parameter combinations, decreasing computational cost while still providing reasonable performance [28]. Random sweep is a hybrid compromise that also explores the parameter space in a structured manner.

Hyperparameter tuning has been emphasized as an effective approach to enhance machine learning performance in previous studies. Ref. [26] showed that better hyperparameter tuning can significantly improve the accuracy of the classifiers and the model stability using different machine learning algorithms. Likewise, for complex classification problems, tuning has been shown to lead to greater predictive accuracy, as demonstrated for high-dimensional data [29].

Random sweep proved to be the most effective approach in the current study, offering a good balance of computation and prediction. Results obtained from the optimized model showed more generalization performance and lower classification error rates, especially between the medium- and high-risk rupture conditions [20]. In industrial applications, early and accurate failure detection can be a key factor in operational safety and cost efficiency, making this improvement crucial.

3.6. Model Validation Using Independent Operational Data

The optimized model was validated using an independent set of 1365 observations collected from the furnace's operation in 2021, covering the entire year of operation, both at normal and verified rupture periods. The validation set comprised 1135 low-risk (83.2%), 123 medium-risk (9.0%), and 107 high-risk (7.8%) predicted observations. Seven event clusters of ruptures were identified through temporal analysis of high-risk predictions during the validation period. The highest number of observations in a single cluster (Cluster 4) amounted to 85, which was equal to 340 h of high-risk warnings (14.2 days) with a high-risk predictive probability of 86.3% on average. This shows that the proposed framework provides sustained early warning significantly higher than the one-week threshold; 76 out of 107 high-risk observations (71.0%) were greater than the 70% predictive confidence threshold.

Figure 8 presents the complete validation timeline for the model's predictions of risk probabilities and class assignments across all 1365 observations. These results indicate the practical applicability of the abovementioned framework for maintenance scheduling, load management, and planning of safety in this furnace operating environment. Ref. [30] emphasised that early failure prediction models support cost-effective maintenance solutions and minimize disruption in process industries. The proposed model extends conventional monitoring systems based on threshold alarms or single-variable monitoring, such as temperature, by enabling detection of degradation patterns that are generally not observed in conventional systems. The validation results indicate that, for this furnace system, the machine learning-based predictive maintenance approach performed more robustly than the conventional methods it was compared against [31]; confirming this advantage across other industrial assets will require further validation.

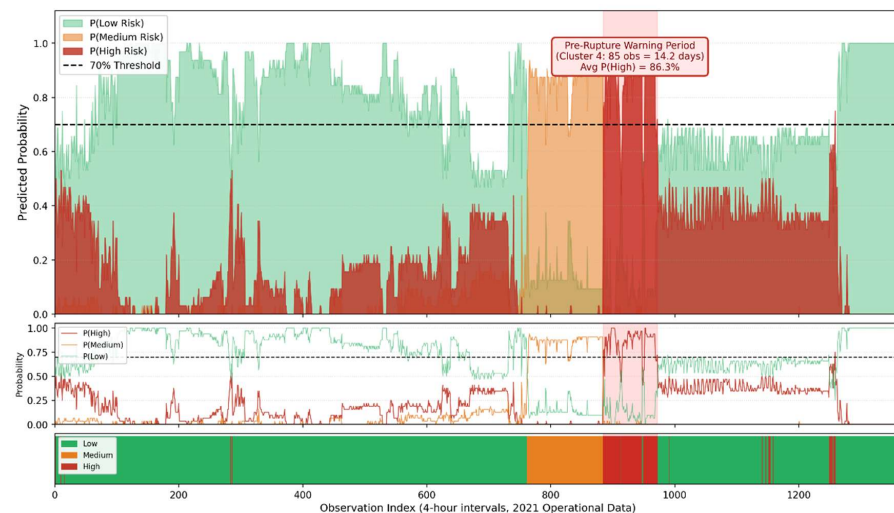


Figure 8. Validation timeline—decision forest predictions on 2021 independent operational data.

3.7. Industrial Implications of Predictive Maintenance Framework

This study's results suggest that machine learning can meaningfully enhance furnace monitoring in settings such as the furnace system evaluated here, driving enhanced efficiency, reduced downtime events, and improved safety [17]. Traditional maintenance methods, such as periodic inspections and rule-based alarm systems, are often insufficient for early-stage fault detection and typically identify failures only after significant damage has occurred [32]. The proposed machine learning framework addresses this gap by continuously processing operational data and dynamically predicting rupture risk, enabling proactive maintenance activities.

The multiclass classification system developed in this research offers a concrete operational benefit demonstrated on the study furnace: low-risk conditions trigger routine monitoring; medium-risk conditions prompt scheduled inspection and load reduction, and high-risk conditions trigger immediate controlled shutdown planning and safety escalation, in accordance with ISO 31000 risk management principles [30], as summarized in Figure 9.

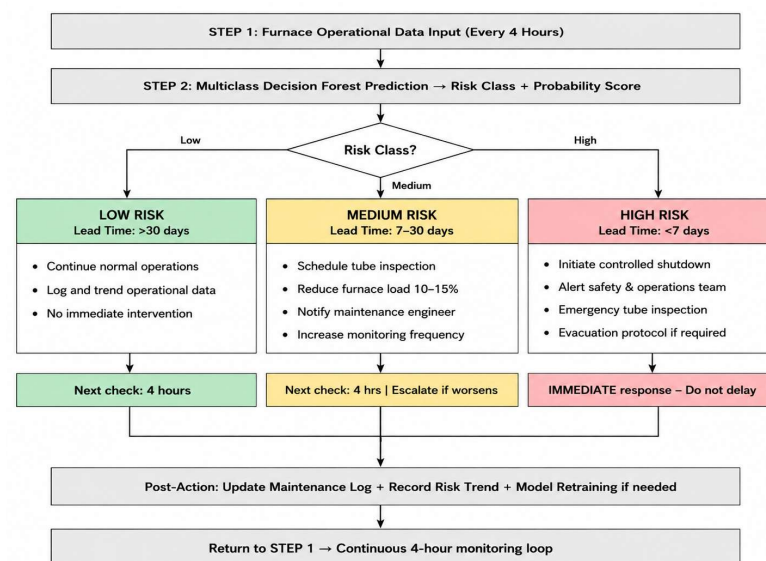


Figure 9. Risk-based maintenance decision flowchart linking each predicted risk class to recommended maintenance actions and lead-time ranges, in accordance with ISO 31000 Risk Management Principles [23].

This structured decision-making capability offers operational flexibility while reducing unnecessary maintenance and improving asset utilization, thereby decreasing production losses caused by unplanned downtime for the furnace system studied. Extending this benefit to other furnace types and industrial sites is identified as necessary future work rather than an established outcome of the present study. Future studies could combine failure mode and effects analysis (FMEA) with machine learning models to enhance risk assessment and maintenance strategies.

3.8. Strengths and Limitations of Proposed Framework

The proposed multiclass decision forest framework demonstrated strong capability for early prediction of furnace tube rupture under complex industrial operating conditions. A key strength of the framework is its ability to classify rupture severity into actionable operational risk categories rather than relying on conventional binary fault detection [33]. This multiclass structure enhances operational interpretability and supports proactive maintenance planning through risk-based intervention strategies. The integration of feature selection, ensemble learning, and hyperparameter optimization also improved computational efficiency, prediction stability, and model generalization performance when validated using unseen industrial operational data.

Another important strength of the framework is its practical industrial relevance. The ability to provide early rupture-risk warnings more than 14 days before failure events offers valuable lead time for maintenance scheduling, load management, and operational safety planning. Furthermore, the ensemble-based decision forest structure provides robustness against noisy and high-dimensional industrial process data, making the framework suitable for predictive maintenance applications within Industry 4.0 environments.

Despite these advantages, the study has several limitations. The predictive framework was developed using historical operational data from a single industrial furnace system, which may limit its generalizability across furnace configurations and industrial environments. In addition, the correlation-based feature selection approach may not fully capture complex nonlinear interactions among operational variables. It should also be emphasized that the fundamental techniques employed, such as Pearson correlation-based feature selection and a built-in decision forest ensemble classifier, are well-established and commonly used techniques, and that the novelty of this work lies in their applied combination with a lead time-based multiclass risk-labeling scheme and independent-year validation on real rupture data, rather than in algorithmic novelty. Future research could explore advanced feature extraction techniques, hybrid deep learning models, and real-time deployment architectures to improve predictive accuracy and industrial scalability further. Another limitation is the relatively small number of high-risk observations in the testing dataset. While the classification results are good, further testing on larger and more diverse industrial datasets is needed before the proposed framework can be generalized.

4. Conclusions

This study developed and validated a multiclass predictive maintenance system for furnace tube rupture prediction using a multiclass decision forest model. The proposed framework aimed to shift furnace monitoring on the studied system from reactive inspection practices towards a data-driven predictive maintenance approach that supports proactive operational decision-making in high-risk industrial environments. The model used industrial operational data comprising 56 furnace variables, which were reduced to 19 significant parameters through correlation-based feature selection to improve computational efficiency and interpretability. The multiclass classification structure categorized rupture conditions into low-, medium-, and high-risk levels, providing greater operational

relevance than conventional binary fault detection methods. Results demonstrated that the proposed model effectively identified rupture-risk patterns and provided early warning more than 14 days before potential failure events.

Integrating ensemble learning with hyperparameter optimization improved prediction robustness, model stability, and generalization performance under the complex operating conditions of the investigated furnace. The proposed framework shows that currently available ensemble learning tools enabled by a risk-based multiclass labeling scheme, which is triggered by the furnace's lead time, can support predictive maintenance planning for industrial furnace systems operating under conditions similar to those investigated in this study to improve operational safety and minimize unplanned downtime of such furnace systems. Despite these contributions, the study is limited by its reliance on historical data from a single industrial system and by the small size of the high-risk test sample. Future research should focus on real-time deployment, advanced feature engineering, validation on larger high-risk datasets, and validation across multiple furnace environments to improve scalability and generalizability.

Author Contributions: Conceptualization, M.T.B. and S.H.; methodology, M.T.B.; software, M.T.B.; validation, M.T.B. and S.H.; formal analysis, M.T.B.; investigation, M.T.B.; data curation, M.T.B.; writing—original draft preparation, M.T.B.; writing—review and editing, S.H. and S.S.H.; supervision, S.H. and S.S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are not publicly available due to industrial confidentiality and operational restrictions. However, they may be available from the corresponding author upon reasonable request and subject to organizational approval.

Acknowledgments: The authors would like to acknowledge the support provided by the Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia, for facilitating this research. The authors also express their appreciation to the industrial plant personnel for providing operational data and technical insights that contributed to this study.

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

References

1. Du, C.; Song, C.; Li, X.; Yuan, J.; Liu, C.; Xin, Y.; Lu, J. Oxidation behavior and performance deterioration of cracking furnace tubes in service. *Heliyon* **2024**, *10*, e24500. [CrossRef] [PubMed]
2. Bragg-Sitton, S.; Dickens, R.; Farmer, J.; Davis, J.; Adams, M.; Martin, J.; Webster, K. *Heater Development, Fabrication, and Testing: Analysis of Fabricated Heaters*; NASA/TM—2008-215466; National Aeronautics and Space Administration, Marshall Space Flight Center: Huntsville, AL, USA, 2008.
3. Li, W.; Li, T. Comparison of deep learning models for predictive maintenance in industrial manufacturing systems using sensor data. *Sci. Rep.* **2025**, *15*, 23545. [CrossRef] [PubMed]
4. Hamilton, V.A.; Mitra, B.; He, L. *Trends and Effects of Changes in Business Cases for Petroleum Refineries*; Pacific Northwest National Laboratory (PNNL): Richland, WA, USA, 2025.
5. Hehn, T.M.; Kooij, J.F.; Hamprecht, F.A. End-to-end learning of decision trees and forests. *Int. J. Comput. Vis.* **2020**, *128*, 997–1011.
6. CSB. U.S. Chemical Safety Board Issues Final Report. Available online: <https://www.csb.gov/us-chemical-safety-board-issues-final-report/> (accessed on 15 June 2025).
7. Chowdhury, H.; Chowdhury, T.; Hossain, N.; Chowdhury, P.; Salam, B.; Sait, S.M.; Mahlia, T.M.I. Exergetic sustainability analysis of industrial furnace: A case study. *Environ. Sci. Pollut. Res.* **2021**, *28*, 12881–12888. [CrossRef]
8. Pfeifer, H. Industrial furnaces-status and research challenges. *Energy Procedia* **2017**, *120*, 28–40. [CrossRef]
9. Liu, H.; Zhang, G. Enhancing Efficiency and Energy Optimization: Data-Driven Solutions in Process Industrial Manufacturing. *EAI Endorsed Trans. Energy Web* **2024**, *11*, 1–11. [CrossRef]

10. Arnal, Á.J.; Díaz-Ramírez, M.; Acevedo, L.; Ferreira, V.J.; García-Armingol, T.; López-Sabirón, A.M.; Ferreira, G. Multicriteria analysis for retrofitting of natural gas melting and heating furnaces for sustainable manufacturing and industry 4.0. *J. Energy Resour. Technol.* **2020**, *142*, 022203.
11. Dobbelaere, M.R.; Plehiers, P.P.; Van de Vijver, R.; Stevens, C.V.; Van Geem, K.M. Machine learning in chemical engineering: Strengths, weaknesses, opportunities, and threats. *Engineering* **2021**, *7*, 1201–1211. [\[CrossRef\]](#)
12. Mahadevan, S.; Prakash, C.; Dixit, S.; Murali, G.; Vatin, N.I.; Epifantsev, K.; Kumar, K. Machine learning for prediction of heat pipe effectiveness. *Energies* **2022**, *15*, 3276. [\[CrossRef\]](#)
13. Rokach, L. Decision forest: Twenty years of research. *Inf. Fusion* **2016**, *27*, 111–125. [\[CrossRef\]](#)
14. Zhao, J.; Peng, Z.; Cui, D.; Li, Q.; He, J.; Qiu, J. A method for measuring tube metal temperature of ethylene cracking furnace tubes based on machine learning and neural network. *IEEE Access* **2019**, *7*, 158643–158654. [\[CrossRef\]](#)
15. Zhu, W.; Liu, X.; Hou, X.; Hu, J.; Diao, Z. Application of machine learning to process simulation of n-pentane cracking to produce ethylene and propene. *Chin. J. Chem. Eng.* **2020**, *28*, 1832–1839. [\[CrossRef\]](#)
16. Rai, A.; Constantinides, P.; Sarker, S. Editor's comments: Next-generation digital platforms: Toward human–AI hybrids. *MIS Q.* **2019**, *43*, iii–ix.
17. Nota, G.; Nota, F.D.; Toro, A.; Nastasia, M. A framework for unsupervised learning and predictive maintenance in Industry 4.0. *Int. J. Ind. Eng. Manag.* **2024**, *15*, 304–319. [\[CrossRef\]](#)
18. Shim, J.-G.; Lee, E.K.; Oh, E.J.; Cho, E.-A.; Park, J.; Lee, J.-H.; Ahn, J.H. Predicting the risk of inappropriate depth of endotracheal intubation in pediatric patients using machine learning approaches. *Sci. Rep.* **2023**, *13*, 5156. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Moosavi, S. An Explainable Artificial Intelligence-Based Framework for Diagnosing Faults in an Induction Furnace. Master's Thesis, University of Windsor, Windsor, ON, Canada, 2022. Available online: <https://scholar.uwindsor.ca/etd/9014/> (accessed on 25 July 2026).
20. Kapavarapu, B.B.; Kumar, K.S.; Nethi, V.N.; Shaji, S.; Anand, T.Y.S.S.; Krishna, P.B.M. Decision Tree and Random Forest Applications for the Oil and Gas Field. In Proceedings of the 2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI), Chennai, India, 17–18 April 2024; pp. 1–5.
21. Shyrokikh, K.; Girnyk, M.; Dellmuth, L. Short text classification with machine learning in the social sciences: The case of climate change on Twitter. *PLoS ONE* **2023**, *18*, e0290762. [\[CrossRef\]](#) [\[PubMed\]](#)
22. ISO 31000:2018; Risk Management Guidelines. International Organization for Standardization: Geneva, Switzerland, 2018. Available online: <https://www.iso.org/standard/65694.html> (accessed on 25 July 2025).
23. Fernández, P.M.-G.; López, A.J.G.; Márquez, A.C.; Fernández, J.F.G.; Marcos, J.A. Dynamic Risk Assessment for CBM-based adaptation of maintenance planning. *Reliab. Eng. Syst. Saf.* **2022**, *223*, 108359. [\[CrossRef\]](#)
24. Sinha, A.; Das, D. Data-Driven Techniques for Fault Diagnosis and Predictive Maintenance. In *Applied Artificial Intelligence and Machine Learning Techniques for Engineering Applications*; CRC Press: Boca Raton, FL, USA, 2025; pp. 30–49.
25. Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature selection in machine learning: A new perspective. *Neurocomputing* **2018**, *300*, 70–79. [\[CrossRef\]](#)
26. Elgeldawi, E.; Sayed, A.; Galal, A.R.; Zaki, A.M. Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. *Informatics* **2021**, *8*, 79. [\[CrossRef\]](#)
27. Agrawal, P. A survey on hyperparameter optimization of machine learning models. In Proceedings of the 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 15–16 March 2024; pp. 11–15.
28. Malakouti, S.M.; Menhaj, M.B.; Suratgar, A.A. Applying grid search, random search, Bayesian optimization, genetic algorithm, and particle swarm optimization to fine-tune the hyperparameters of the ensemble of ML models enhances its predictive accuracy for mud loss. *Res. Sq.* **2024**, preprint. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Deiss, L.; Margenot, A.J.; Culman, S.W.; Demyan, M.S. Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma* **2020**, *365*, 114227. [\[CrossRef\]](#)
30. Steurtewagen, B.; Van den Poel, D. Adding interpretability to predictive maintenance by machine learning on sensor data. *Comput. Chem. Eng.* **2021**, *152*, 107381. [\[CrossRef\]](#)
31. Tiddens, W.; Braaksma, J.; Tinga, T. Exploring predictive maintenance applications in industry. *J. Qual. Maint. Eng.* **2022**, *28*, 68–85. [\[CrossRef\]](#)
32. Bortey, L. *Predictive Modelling of Incident Risk to Pre-Empty Risk in Highway Operations: A Machine Learning Approach*; Birmingham City University: Birmingham, UK, 2025.
33. Arena, S.; Manca, G.; Murru, S.; Orrù, P.F.; Perna, R.; Reforgiato Recupero, D. Data science application for failure data management and failure prediction in the oil and gas industry: A case study. *Appl. Sci.* **2022**, *12*, 10617. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.