

Machine Learning–Driven Predictive Maintenance for Smart Beverage Manufacturing Systems

Peerawit Pongtanankorn¹ and Kunlapat Thongkaew^{1,2,*}

¹ Department of Industrial and Manufacturing Engineering, Faculty of Engineering, Prince of Songkla University, Kho Hong, Hat Yai, Songkhla, 90110, Thailand

² Smart Industrial Research Centre, Faculty of Engineering, Prince of Songkla University, Kho Hong, Hat Yai, Songkhla, 90110, Thailand

*Corresponding Author E-mail: kunlapat.t@psu.ac.th

Received: Mar 12, 2026; Revised: May 18, 2026; Accepted: Jun 05, 2026

Abstract

This study proposed a machine learning-based predictive maintenance system in a beverage manufacturing process to enhance failure prediction accuracy and support improved maintenance planning. The approach focused on the beverage filling line, captured short-term temporal process dynamics using rolling-window feature extraction, and addressed the strong class imbalance between normal operation and breakdown events. Significant sensor variables were initially identified through correlation analysis, and the selected sensors and breakdown criteria were subsequently analyzed using various classification models. A 60-minute decision horizon was defined as a classification-based early-warning window derived from the time-to-failure (TTF) variable. Observations were labeled breakdown-proximate state when the next breakdown occurred within 60 minutes ($TTF \leq 60$ min) and as normal operation when the next breakdown occurred after 60 minutes ($TTF > 60$ min). Therefore, the model classifies a breakdown-proximate state rather than predicting the exact remaining time to failure. Model performance was evaluated using precision, recall, F1-score, ROC-AUC, and chronological walk-forward validation. Based on correlation analysis and process relevance, 15 sensors were selected as key features associated with process breakdown. The Random Forest model was the best performer, achieving an accuracy of 0.990, precision of 0.918, recall of 0.722, F1-score of 0.808, and ROC-AUC of 0.995 at the optimized threshold of 0.45. The recall result indicates acceptable detection of breakdown-proximate states, although 27.8% was missed as a false negative. Considering the balance between recall and precision, the precision exceeded 90%, showing that false alarms were maintained at a low level. These results demonstrate a practical trade-off between reducing missed breakdown warnings and avoiding excessive unnecessary inspections. Thus, the Random Forest is suitable for predicting breakdowns with accuracy, stability and robustness. Future work will extend the predictive maintenance results linked with failure-risk predictions to reliability metrics of time between failures (TBF) and mean time between failures (MTBF), improved alignment between maintenance actions and actual machine conditions.

Keywords: Maintenance, Predictive Maintenance (PdM), Beverage Manufacturing Process, Machine Learning

1. Introduction

Beverage production operations are frequently automated and managed by machinery, which works under a range of circumstances monitored by several sensors. Despite comprehensive sensing and monitoring, system alarms and process abnormalities are often hard to detect with certainty. Even when there are no actual mechanical problems, this might cause unanticipated process interruptions or system breakdowns, resulting in increased downtime and production expenses. As a result, beverage manufacturing systems in industrial production environments must meet stringent quality requirements while operating at high throughput rates. Even minor machinery faults in such severe circumstances may quickly cause production losses and quality deviations, frequently resulting in equipment breakdowns that affect the entire operation. Comprehensive maintenance planning is therefore critical to avoiding unplanned downtime and maintaining optimal operating efficiency. Corrective Maintenance (CM), Preventive Maintenance (PM), and Predictive Maintenance (PdM) are three common types of industrial maintenance techniques. CM causes unpredictable downtime and high maintenance costs, especially for continuous operations.

PM uses established routines or usage-based intervals, which might result in wasteful maintenance or disregarding condition-dependent breakdown under variable operating circumstances. Due to these restrictions, PdM has grown in popularity, as it forecasts rates of failure based on equipment condition rather than time intervals. [1–4].

Predictive maintenance (PdM) has developed as a critical strategy for increasing operating dependability by using operational and sensor data to forecast the probability of failures leading to equipment breakdowns. Within the Industry 4.0 paradigm, the integration of cyber-physical systems, artificial intelligence (AI), and machine learning (ML) has facilitated the change from reactive and time-based maintenance policies to data-driven decision-making frameworks, enabling continuous condition monitoring and timely intervention [5–8].

Despite substantial progress, deploying ML-based PdM systems in real-world manufacturing environments remains difficult. The process data in beverage manufacturing is generally complex, nonlinear, and time-dependent, with significant interactions between important operational variables and machine dynamics. Such characteristics impede quick breakdown detection,

necessitating multivariate, data-driven monitoring systems designed specifically for food and beverage manufacturing facilities [9–11]. Furthermore, many high-performance models, particularly complex ML and deep learning approaches, can operate as black boxes, providing limited insight into the underlying process mechanisms associated with equipment failures and limiting industrial adoption where transparency and trust are required [12–14].

Recent research has focused on explicable artificial intelligence (XAI) as a method of integrating predicted performance with interpretability in PdM systems. Explainability approaches are increasingly being used to ascribe failure-risk estimates to influential process factors, which aid diagnostic reasoning and human-machine decision-making [12–14]. For example, Sinnah et al. [15] have shown how applying predictive models with explainability could decrease process safety hazards in safety-critical industrial equipment. Nonetheless, much present research concentrates on generic industrial assets, with limited domain-specific adaptation to beverage production processes [9–11]. Furthermore, temporally inconsistent validation techniques are still frequent, which can lead to unrealistic performance predictions that do not represent actual deployment conditions [15–21]. Consequently, the lack of PdM frameworks that satisfy these criteria for beverage production systems is an obvious research gap.

2. Literature Review

2.1 Applications of Predictive Maintenance in Manufacturing Systems

Predictive maintenance is growing increasingly popular in production systems as a data-driven strategy to predict equipment breakdowns and improve operational performance. PdM is especially important in beverage manufacturing considering the high production throughput and tightly linked process conditions, where multivariate interactions among operational variables impact both equipment condition and product quality [9–11]. Continuous operation under such conditions exposes filling and packing equipment to dynamic impacts, frequent operational changes, and progressive wear, preventing the identification of issues early when using established maintenance methods [1–2],[4]. According to studies in the food and beverage industry, breakdown events seldom occur because of rapid changes in individual variables; instead, they occur as a result of related adjustments across many process signals. This behavior has motivated the use of multivariate, sensor-based monitoring systems able to identify interactions between process variables [9–11]. As a result, new predictive maintenance studies focus more importance on data-centric methods for detecting early-stage decline trends before failure occurrences [7],[10],[22]. Nonetheless, the challenges of continuous filling operations and deployment-oriented validation remain unsatisfied, necessitating the implementation of application-focused predictive maintenance studies specific to beverage production systems [9],[11].

2.2 Challenges of Machine Learning-Based Predictive Maintenance

There are various ML techniques applied with maintenance strategies, such as, Decision Tree, Random Forest, Nearest Neighbors, XGBoost (eXtreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), Logistic Regression, Extra Trees, and a MLP (Multilayer Perceptron) [11]. Recent research shows that ensemble learning approaches, such as Random Forest and boosting-based models, are well suited for industrial predictive maintenance due to their noise opposition, capacity for expressing nonlinear connections, and ability to capture complex feature interactions [7],[17–18]. Advanced machine learning algorithms have outperformed conventional techniques for defect identification and condition monitoring in a variety of industrial systems [15],[17–21],[23]. However, effective PdM performance is dependent on algorithm selection and the design of an integrated learning framework that involves feature engineering, class imbalance handling, decision threshold calibration, and temporally consistent validation [15–21].

The quality and structure of industrial data provide an important obstacle for ML-based PdM. Sensor measurements are often noisy, incomplete, and affected by drift, while failure labels are confusing because of gradual breakdown processes [4],[7]. Feature engineering is therefore crucial, with time-aggregated and rolling-window features extensively utilized to capture loss dynamics that are not obvious in immediate sensor measurements [4],[16–17]. A further frequent problem is a major imbalance between normal operation and failure events, which, if not addressed, can lead to false results. Imbalance-aware learning algorithms and evaluation metrics, including accuracy, recall, and F1-score, are essential for effective rare event identification and false alarm management [7],[15–21].

Finally, interpretability and practicality are critical for industry acceptance. To make intelligent choices, maintenance engineers must understand the factors that contribute to greater failure-risk forecasts. Black-box models have the potential to undermine reliability, especially in safety-critical scenarios. As a result, explainable artificial intelligence (XAI) solutions are increasing in popularity for enhancing transparency, human-centered maintenance planning and practical factors involving computing effectiveness and integration with current industrial systems [12–14],[22–24].

2.3 Research Gap and Study Contributions

Despite significant progress, current PdM research focuses primarily on generic industrial assets and fails to address the combined challenges of multivariate sensor integration, temporal dependency, interpretability, and deployment-oriented validation in beverage manufacturing environments. To fill this gap, this study presents a machine learning-based predictive maintenance system developed particularly for beverage production operations. The suggested technique combines multi-sensor data collection, structured

database-driven data mining, ensemble machine learning models, and IoT-enabled alert systems. The purpose of this research is to construct and assess a predictive maintenance system capable of improving failure-risk prediction accuracy under realistic operating conditions for beverage production systems. The subsequent sections provide a full overview of the research methodology and framework.

3. Research Methodology

This paper provides a predictive maintenance strategy to prevent breakdowns in an industrial beverage filling system. The methodology is intended to mimic real-world operational situations and employs a systematic, time-based, deployment-focused approach. It combines process knowledge, breakdown description, multivariate sensor data analysis, machine learning-based breakdown classification, and validation under realistic time limitations. **Figure 1** depicts an overview of the research methodology, including the beverage filling process, data collecting, breakdown definition, machine learning modeling, and deployment-oriented validation.

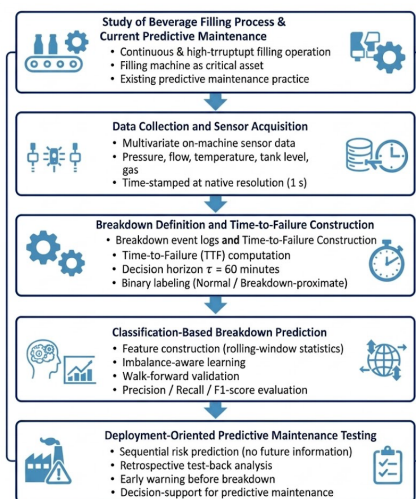


Figure 1 Framework of the proposed predictive maintenance method for the beverage filling system.

This research used Python for data preprocessing, feature construction, model training, threshold optimization, validation and visualization. The main

libraries selected for data analysis and vitalization were *Pandas* and *NumPy*, while *Scikit-learn* was implemented for preprocessing [25]. Python was also used for model development and evaluation.

3.1 Beverage Filling Process

This study focuses on an industrial beverage filling line that operates under continuous, high-throughput manufacturing circumstances. The line is made up of several sequential and interdependent equipment that perform depalletizing, empty container inspection, cleaning, filling, sealing, pasteurization, drying, inspection, packaging, and weight checking prior to storage and distribution. Because of its central location and continuous operation, the filling machine has a considerable impact on line stability, production throughput, and product quality. **Figure 2** displays a diagram of the beverage filling process, focusing on the filling machine.

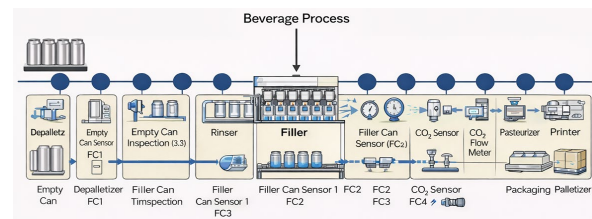


Figure 2 Beverage filling line process with emphasis on the filling machine

3.2 Data Collection and Breakdown Definition

3.2.1 Sensor Data Collection

Real-time sensor measurements from the filling machine are the focus of the operational data gathered from an industrial beverage filling line. All sensor signals are continuously collected through typical manufacturing processes and time-stamped to ensure exact temporal timing. A total of 18 input sensor variables were collected from the filling machine and related subsystems, as illustrated in **Figure 3**. These sensors relate to monitor five key operational characteristic groups of the filling process, including speed, pressure, flow rate, temperature, and gas parameters. Thus, these variables were used to analyze breakdown indicators, which will be discussed in the next section on breakdown definition and labeling.

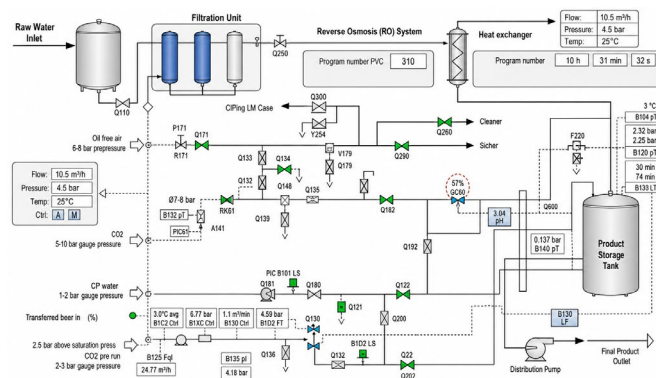


Figure 3 Sensor layout of the filling machine

3.2.2 Breakdown Definition and Labeling

In this study, a machine breakdown was defined as an unexpected stoppage of a filling machine that affected normal output and required corrective action. Manufacturing systems recorded such events, with each breakdown assigned a unique start timestamp. This timestamp served as a reference point for failure analysis, and timestamp synchronization was applied to align breakdown records with sensor data. Each time-stamped sensor observation was associated with a time-to-failure (TTF) variable to aid predictive modeling. TTF was defined as the time between the observation time and the start of the next record breakdown event [4],[22]. This transformation converted raw breakdown logs into a supervised learning target that measures the time between breakdowns. Predictive maintenance was considered as a binary classification based on the estimated TTF variable [18],[22].

Horizon-based labeling strategy is commonly adopted to convert time-to-failure prediction into a practical classification framework [19]. The decision horizon (τ) directly influences class imbalance and predictive performance [20]. A shorter horizon provided a more immediate warning window but generated fewer positive samples, whereas a longer horizon increased the number of positive samples but included observations farther from the actual breakdown event [20]. In this study, three levels of τ including 30, 60 and 90 were investigated under identical preprocessing procedures, rolling-window feature construction, and chronological validation settings to ensure comparability across alternative early-warning windows. The dataset was conducted from one month of operational data collected in March 2025, comprising a total of 2,676,013 time-indexed observations. From the analysis, the number of positive samples increases as τ expands, yielding 111,919; 135,507; and 257,697 positive samples, respectively. The τ of 60 minutes was selected for this study as implying a balanced compromise between operational utility and modelling robustness. It also provides a sufficiently practical early-warning period in the filling process, while preserving time relevance for maintenance planning, consistent with [21]. From **Table 1**, observations with TTF values exceeding a predefined decision horizon ($TTF > 60$ min) are classified as Class 0 (normal operation), whereas Class 1 (breakdown-proximate state) refers to those having TTF values less than or equal the decision horizon ($TTF \leq 60$ min). For clarity, the decision is intended to identify a breakdown-proximate state within 60 minutes rather than to directly predict the exact failure time. Based on the selection of $\tau = 60$ minutes, the resulting class distribution is imbalanced with breakdown-proximate observations accounting for 5.06% of the dataset. Although this imbalance could rise issues of modelling, it is acceptable for actual production situations, where breakdown events occur rarely.

Table 1 Class distribution based on TTF analysis at $\tau = 60$ minutes.

Class	Description	Operational State	Samples	Percentage
0	Negative Class	Normal operation ($TTF > 60$ min)	2,540,506	94.94%
1	Positive Class	Breakdown-proximate state ($TTF \leq 60$ min)	135,507	5.06%
Total			2,676,013	100%

3.3 Classification-Based Breakdown Prediction Using Machine Learning

Machine learning was employed as a unified decision-support system for predicting breakdown-proximate states of the filling machine. The proposed approach establishes a complete classification pipeline that integrates feature construction, preprocessing, model training, probability estimation, and decision calibration, as shown in **Figure 4**.

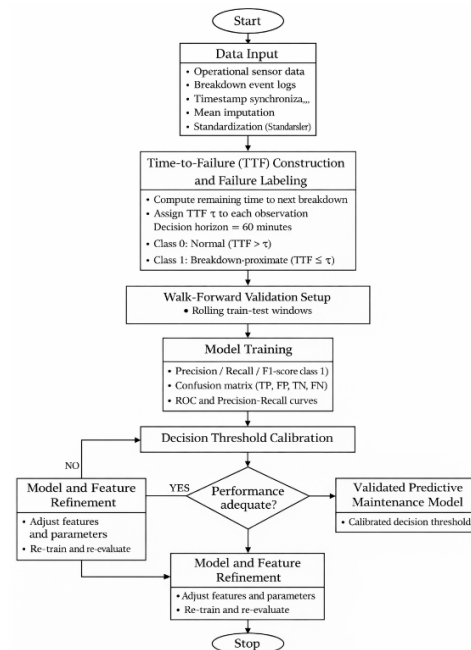


Figure 4 Predictive maintenance modeling and validation workflow

3.3.1 Feature Construction and Selection

Feature construction and selection were guided by a combination of process knowledge and exploration analysis of sensor behavior prior to failure [26]. Correlation analysis was performed on the initial set of 18 sensor variables to examine the relationship among the recorded sensor data before examining their association with the breakdown indicators. The correlation matrix in **Figure 5** uses color intensity to represent the direction and strength of associations among these variables, where warmer colors indicating positive correlations and cooler colors indicating negative correlations [4],[22]. Based on

Figure 5, three variables, namely setpoint speed, product counter, and operating hours, were not selected for breakdown analysis, as they have low correlations and mainly reflect control settings, cumulative production output, and cumulative operating time rather than direct real-time process behaviors.

In order to identify breakdown indications, the remaining 15 sensor variables (**Table 2**) that cover

five important behaviors (speed-related, pressure-related, flow-related, temperature-related, and gas-related variables) were examined for their association with TTF, as presented in **Figure 6**. This analysis was conducted to determine variables exhibiting systematic correlations with breakdown-proximate events using Pearson correlation analysis because all variables are continuous [27].

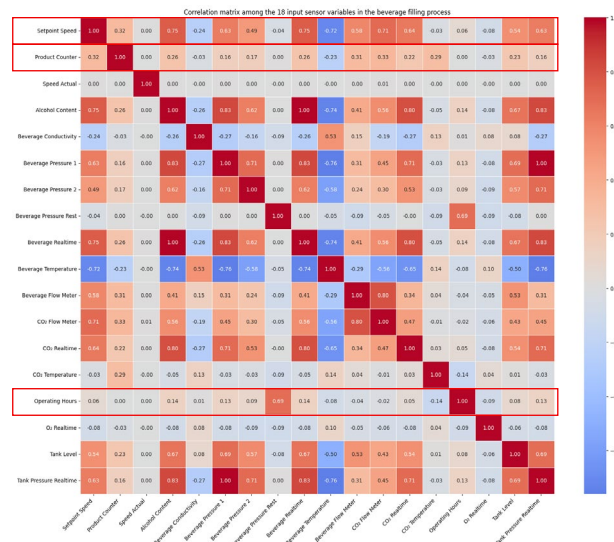


Figure 5 Correlation matrix of the input sensors in the beverage filling process.

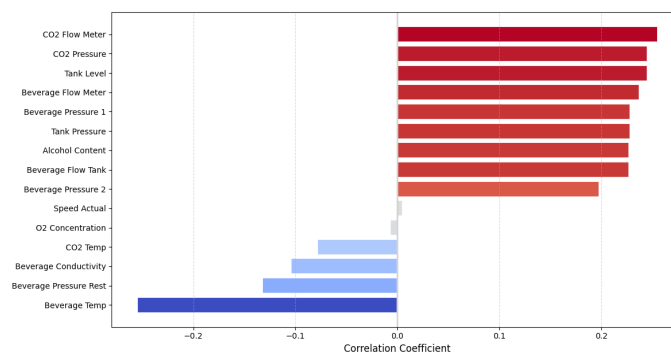


Figure 6 Correlation analysis between selected sensors and TTF breakdown indicators.

Table 2. Lists of selected sensors used as model inputs

Sensor	Variable	Type	Unit
1	Speed Actual	Numerical	can/min
2	Beverage Flow Meter	Numerical	L/min
3	Beverage Flow Tank	Numerical	L/min
4	Beverage Pressure 1	Numerical	bar
5	Beverage Pressure 2	Numerical	bar
6	Beverage Pressure Rest	Numerical	bar
7	Tank Pressure	Numerical	bar
8	CO ₂ Pressure	Numerical	bar
9	CO ₂ Flow Meter	Numerical	L/min
10	CO ₂ Temp	Numerical	°C
11	Beverage Temp	Numerical	°C
12	Beverage Conductivity	Numerical	μs/cm
13	O ₂ Concentration	Numerical	ppm
14	Alcohol Content	Numerical	%ABV
15	Tank Level	Numerical	%

Figure 6 shows that the variables' speed and O₂ concentration had low correlation coefficients; however, they remained for the training of the predictive models. The speed actual directly reflects the continuity and stopping behavior of the machine, while the O₂ concentration implies fluctuations in gas-balance inside the liquid. As a result, all 15 sensor variables were retained as inputs for the final predictive model development. In summary, speed variables denote machine continuity and stoppage behavior; pressure and flow variables reflect supply and filling stability; temperature variables refer to thermal operating conditions; gas variables (CO₂ and O₂) display gas-balance disturbances. These characteristics reveal behavior indication of potential breakdown. In addition, rolling-window statistical features were employed to capture short-term

temporal dynamics prior to breakdown events in later sections to confirm whether low-correlation variables should be avoided [4],[18].

3.3.2 Machine Learning Models

Six multiple classification models are selected for predictive maintenance analysis, including Logistic Regression, Random Forest, Extra Trees, XGBoost, LightGBM, and Multi-Layer Perceptron. Tree-based ensemble models are emphasized due to their ability to model nonlinear relationships, handle noisy sensor data, and perform robustly under severe class imbalance, which has been consistently reported in recent predictive maintenance and industrial machine learning studies [17],[27–28]. Logistic Regression is included as an interpretable baseline. The characteristics of each model are given in **Table 3**.

Table 3 Machine Learning Models

Model	Learning paradigm	Key characteristics	Relevance to PdM literature
Logistic Regression	Linear classification	Interpretable baseline, probabilistic output	Widely used as reference model in PdM studies [17],[27]
Random Forest	Bagging-based ensemble	Nonlinear modeling, variance reduction, robust to noise	Proven effectiveness in industrial PdM [17],[27]
Extra Trees	Bagging-based ensemble	Randomized splits, reduced correlation among trees	Effective for high-dimensional sensor data [27–28]
XGBoost	Gradient boosting	Strong bias reduction, scalable learning	Applied in complex PdM scenarios [28],[29]
LightGBM	Gradient boosting	Histogram-based efficiency, fast training	Suitable for large-scale industrial datasets [17],[28]
MLP	Neural network	Nonlinear representation learning	Explored in PdM with mixed stability results [18],[30]

3.3.3 Model Training and Parameter Settings

Model training was performed using a time-separated training set of 2,140,810 observations (80% of the total dataset) to ensure temporal integrity. **Table 4** shows the system-level parameters that were consistently applied across all models. Data preparation techniques, such as mean imputation, z-score normalization, and cost-sensitive learning to reduce information loss. The preprocessing workflow is summarized in **Table 5**, and the model-specific configurations are listed in **Table 6**.

Table 4 Parameter setup for model training

Parameter	Value	Description
Time resolution	1 second	Native resolution of sensor data
Decision horizon (τ)	60 minutes	Lead time for failure labeling
Rolling window size	60 minutes	Window for feature construction

Decision-threshold optimization was performed after model training by transforming predicted probabilities into binary decisions based on validation results without retraining the models [18],[22],[29]. Candidate probability thresholds were compared using precision, recall, and F1-score for the breakdown-proximate class, and the selected operating threshold was used for final evaluation.

Table 5 Data preprocessing steps

Step	Description / Strategy	Value in code
Missing values	Mean imputation	SimpleImputer (strategy = mean)
Scaling	Z-score normalization	StandardScaler
Class handling	Cost-sensitive learning	Class-weighted loss
Pipeline	Unified preprocessing pipeline	Prevents data leakage

Table 6 Machine learning model configurations

Parameter Setting	Machine Learning Approaches					
	Random Forest	Extra Trees	XGBoost	Light GBM	MLP	Logistic Regression
Model Type	Bagging	Bagging	Boosting	Boosting	Neural Net	Linear
Estimators / Iterations	200 trees	200 trees	300 rounds	400 rounds	300 epochs	—
Complexity Control	Max depth: None	Max depth: None	Max depth: 6	Num leaves: 31	Layers (128, 64)	C = 1.0 (L2)
Learning Rate	—	—	0.05	0.05	Adaptive	—
Imbalance Handling	Balanced	Balanced	Scale-pos-weight	Scale-pos-weight	—	Balanced
Key Property	Variance reduction	Random splits	Bias reduction	Histogram-based	Nonlinear	Interpretable baseline

3.4 Model Evaluation and Validation Strategy

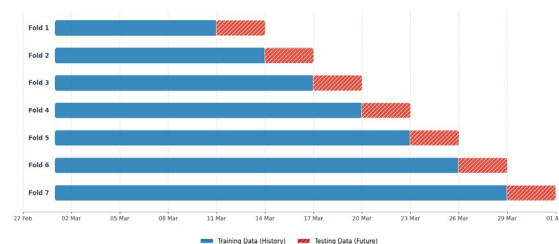
Model performance was evaluated on a chronologically separated test set comprising 535,203 observations (20% of the dataset) under a highly imbalanced breakdown-risk classification setting. Overall accuracy can be misleading because breakdown-proximate events are rare; therefore, accuracy, precision, recall, and F1-score for the breakdown class (Class 1) are emphasized as the primary evaluation metrics [16],[22]. These metrics are derived from the confusion matrix, which quantifies true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) and provides direct insight into the trade-off between missed breakdown warnings and false alarms. The evaluation metrics and their operational interpretation are given in **Table 7**.

Table 7 Evaluation of metrics used in this study

Metric (Class 1)	Description	Operational focus
Precision	Proportion of correctly predicted breakdown-proximate states among all predicted breakdown warnings	Control false alarms
Recall	Proportion of actual breakdown-proximate states correctly identified by the model	Reduce missed breakdowns
F1-score	Harmonic mean of precision and recall	Primary evaluation metric

Threshold-independent discrimination capability was further investigated using receiver operating characteristic (ROC) and precision-recall (PR) curves, which are recommended for severe class imbalance. A chronological walk-forward validation technique was used to simulate real-world deployment conditions [4],[22]. As shown in **Figure 7**, in each fold, the model was trained on earlier observations and evaluated on a later, unseen time window. No random shuffling was applied because sensor observations are time dependent, and adjacent observations may share similar operating conditions. This procedure preserves temporal causality,

reduces information leakage, and provides a more realistic estimate of model stability under changing production conditions. It can be seen from **Figure 7** that the seventh time fold was conducted as the final chronological validation step, where the training window covered the longest accumulated historical period and the model was evaluated on the subsequent unseen testing period.

**Figure 7** Walk-forward validation strategy.

3.5 Deployment-Oriented Evaluation of Predictive Maintenance

An evaluation of the proposed predictive maintenance system feasibility was performed through a deployment-oriented prospective testing study. The analysis employed the use of operational data collected in March 2025, which consisted of 2,676,013 observations recorded in time. This phase focused on analyzing the developed breakdown classification model's ability to forecast increasing breakdown risk prior to the occurrence of actual breakdowns, compared to only testing its forecasting accuracy.

4. Results and Discussion

This section describes the results obtained from applying the proposed machine learning-driven predictive maintenance framework to an industrial beverage filling system. The performance of Logistic Regression, Random Forest, Extra Trees, XGBoost, LightGBM, and MLP for breakdown classification was evaluated, with particular emphasis on reliability under class imbalance, temporal validation robustness, and deployment-oriented operational considerations.

4.1 Machine Learning Classification Results

This section describes the prediction behavior of machine learning classification models developed to detect malfunctions in the beverage filling process. All

investigated models can extract breakdown proximal patterns from multivariate sensor data, but with variable degrees of accuracy. Ensemble-based classifiers outperform linear and neural network models in terms of discrimination consistency and robustness under extreme class imbalance. This superior performance can be attributed to their ability to capture nonlinear feature interaction and complex decision boundaries [19–21]. **Table 8** presents the baseline confusion matrix results (TN, FP, FN, and TP) under the default threshold of 0.5, which show distinct variations in class-separation behavior among models.

Table 8. Baseline classification breakdown under default decision threshold (0.5)

Model	TN	FP	FN	TP
Random Forest Classifier	508,090	12	6	27,095
LGBM Classifier	508,078	24	10	27,091
XGBoost Classifier	508,088	14	12	27,089
Logistic Regression	508,021	81	18	27,083
Extra Trees Classifier	489,258	18,844	214	26,887
MLP Classifier	508,102	0	27,101	0

Tree-based ensemble techniques, such as Random Forest, Extra Trees, and LightGBM, show excellent discriminative abilities in the face of significant class imbalance and achieve remarkably low misclassification rates. Across the models, Random Forest demonstrates the most equitable separation structure (**Table 8**), providing roughly 12 false positives and 6 false negatives, indicative of consistent detection of breakdown-proximate states while maintaining strict control over false alarms. Logistic Regression has a recall-focused pattern, accurately recognizing the majority of breakdown-adjacent events while producing 18,844 false positives, thereby diminishing precision. In contrast, the MLP model fails to recognize any breakdown-proximate events leading to a total breakdown of the minority class. These findings further validate the efficiency of ensemble-based approaches for industrial breakdown classification tasks.

4.2 Model Evaluation Results

4.2.1 Baseline Breakdown Prediction Performance

This section evaluates the accuracy of baseline predictions made by machine learning algorithms when recognizing machine states that are close to breakdowns, using the TTF-based categorization framework described in Section 3. From **Table 7**, all models were first evaluated using their default decision thresholds to establish baseline. **Table 9** show that the predictive performance was measured using accuracy, precision, recall, and F1-score. The results indicate that ensemble-based models, particularly Random Forest, LightGBM, and XGBoost, achieved notably high F1-scores with default settings, implying an ability to capture breakdown-proximate patterns in multivariate sensor data.

Table 9. Baseline predictive performance before threshold optimization

Model	Thresh old	Precisio n	Recall	F1-Score
Random Forest Classifier	0.5	0.926	0.706	0.801
LGBM Classifier	0.5	0.899	0.694	0.783
XGBoost Classifier	0.5	0.810	0.714	0.759
Logistic Regression	0.5	0.637	0.894	0.744
Extra Trees Classifier	0.5	0.856	0.300	0.445
MLP Classifier	0.5	0.000	0.000	0.000

Logistic regression has a high recall but a lower precision, indicating a cautious decision-making approach that prioritizes sensitivity to breakdown-risk conditions, leading to a higher incidence of false alarms. In contrast, the Extra Trees model demonstrates limited recall, and the MLP fails to achieve significant accuracy in ordinary configuration. These baseline results highlight a critical issue in predictive maintenance: breakdown events are infrequent yet expensive. Models that emphasize sensitivity may generate several false alarms, while more cautious models may fail to detect early indicators of breakdown. This trade-off encourages precise of adjustment decision threshold to enhance the correlation of predictive behavior with operational objectives.

4.2.2 Decision Threshold Optimization of Predictive Models

The decision thresholds for each classifier were adjusted based on validation results to improve the alignment between model predictions and the imbalanced distribution of breakdown-proximate events. In predictive maintenance implementation, the decision threshold determines the transformation of predicted breakdown probabilities into practical maintenance warnings [8],[22]. Decreasing the threshold increases recall, thus improving early detection and reducing the probability of missed failures. Conversely, increasing the threshold reduces false alarms and restricts interventions to high-confidence cases, but increases the risk of delayed detection and unexpected failures [17],[22]. From **Table 9**, the recall values were mostly lower than those of other performance metrics; therefore, the threshold was decreased. With the post-training threshold optimization using validation results [29], the threshold of 0.45 was selected. **Table 10** shows that the predictive performance subsequent to the threshold adjustment can increase the recall values. Compared with the default threshold of 0.50 (**Table 9**) of Random Forest Classifier, the optimized threshold improved recall and F1-score to 0.722 and 0.808, while maintaining high precision at 0.918. This indicates that the selected operating point improved the detection of breakdown-proximate observations without introducing excessive false alarms.

Table 10. Model performance after threshold optimization.

Model	Threshold	Accuracy	Precision	Recall	F1-score	TN	FP	FN	TP
Random Forest Classifier	0.450	0.990	0.918	0.722	0.808	752,145	1,397	6,035	15,636
XGBoost Classifier	0.978	0.990	0.972	0.662	0.787	753,131	411	7,333	14,338
Logistic Regression	0.579	0.990	0.917	0.688	0.786	752,197	1,345	6,754	14,917
MLP Classifier	0.735	0.986	0.733	0.813	0.771	747,110	6,432	4,043	17,628
LightGBM Classifier	0.180	0.982	0.646	0.806	0.717	743,976	9,566	4,203	17,468
Extra Trees Classifier	0.047	0.139	0.031	1.000	0.061	86,400	667,142	0	21,671

The Random Forest classifier showed the most balanced overall performance among the evaluated approaches, with the highest F1-score of 0.808, precision of 0.918, and recall of 0.722. The recall demonstrates that the model properly recognized the actual breakdown-proximate states with acceptable performance, while 27.8% regarded as false negatives. In practice, the missed failures may be undesirable as undetected breakdowns may arise in unexpected downtime, production losses, and increased maintenance expenses. Improving recall may be achieved by increasing the amount of training data and including more representative breakdown events, thereby enhancing the model's ability to learn failure characteristics. The balance between recall and precision is considered according to operational requirements and risk tolerance. In this study, the predictive model achieved a high precision of 0.918, indicating strong discrimination between normal operation and breakdown states while limiting the cost of false alarms, particularly in cost-sensitive environments. These results demonstrate a practical trade-off between reducing missed breakdown warnings and avoiding excessive unnecessary inspections.

From a managerial perspective, the decision threshold can be interpreted as a control parameter for warning sensitivity. A lower threshold increases the opportunity to detect breakdown-proximate states earlier, but it may increase inspection workload and false alarms. In contrast, a higher threshold reduces unnecessary warnings but may increase the risk of missed breakdown-proximate events. Therefore, threshold selection should not be considered only as a statistical optimization step, but also as an operational decision that must balance missed breakdown warnings, false alarms, inspection capacity, maintenance cost, and organizational risk tolerance [8],[17].

4.2.3 Deployment Oriented Predictive Maintenance Testing

From a deployment-oriented perspective, predictive maintenance models should be assessed the stability of their warning behavior through different decision thresholds. In actual industrial settings, breakdown-risk classifiers require consistent and interpretable signals that effectively support in maintenance decisions. Therefore, receiver operating characteristic (ROC) and precision-recall (PR) curves were evaluated, as shown in **Figure 8**.

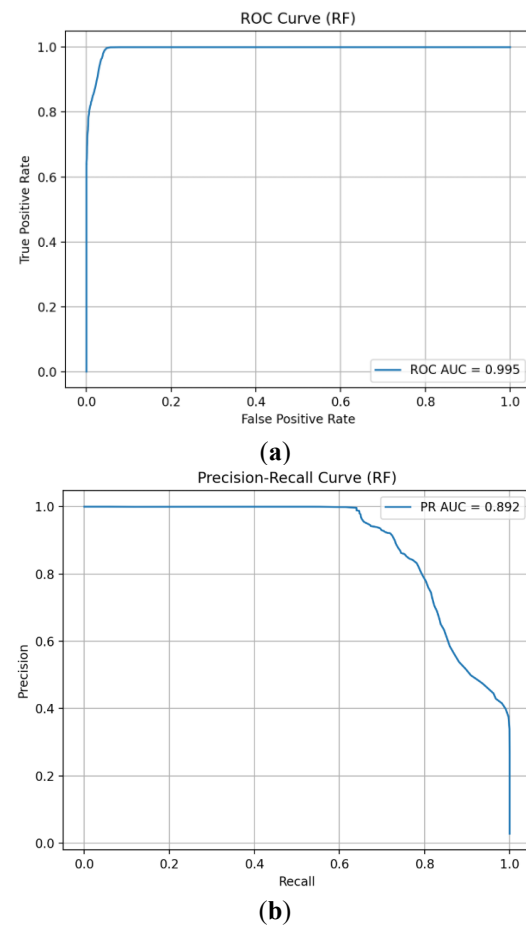


Figure 8 Stability evaluation (a) ROC curve and (b) PR curve of the Random Forest classifier for breakdown prediction.

The ROC curve indicates ROC-AUC (Receiver Operating Characteristic Area Under the Curve) = 0.995, resulting in robust class separability. Nonetheless, in the presence of PR-AUC (Precision-Recall Area Under the Curve) = 0.892 offers a more insightful assessment of predictive stability. The results show that ensemble-based models, in particular Random Forest, XGBoost, and LightGBM, exhibit resilient and dependable deployment behavior by maintaining comparatively steady precision-recall interactions across a wide range of thresholds. In contrast, the other models demonstrate inconsistent warning features, resulting in either an abundance of

false alarms or the failure to detect circumstances near breakdown due to threshold fluctuations.

4.3 Model Validation Using Walk-Forward Validation

Model performance was assessed across consecutive folds of the walk-forward validation to confirm robustness under shifting operational settings, as shown in **Figure 9**. The results show that ensemble-based models, in particular Random Forest, LightGBM, and XGBoost, continue to perform well and consistently across validation iterations. This stability reveals resilience to irregular situations,

including production rate changes, temporary disruptions, and short-term process variations. Overall performance trends show high prediction capabilities under changing data distributions, despite slight variances between folds. Models with rigid decision limits, however, show higher variability between folds, indicating a sensitivity to change operational conditions. These findings highlight the significance of time-consistent validation when selecting predictive maintenance models for industrial deployment [4],[11].

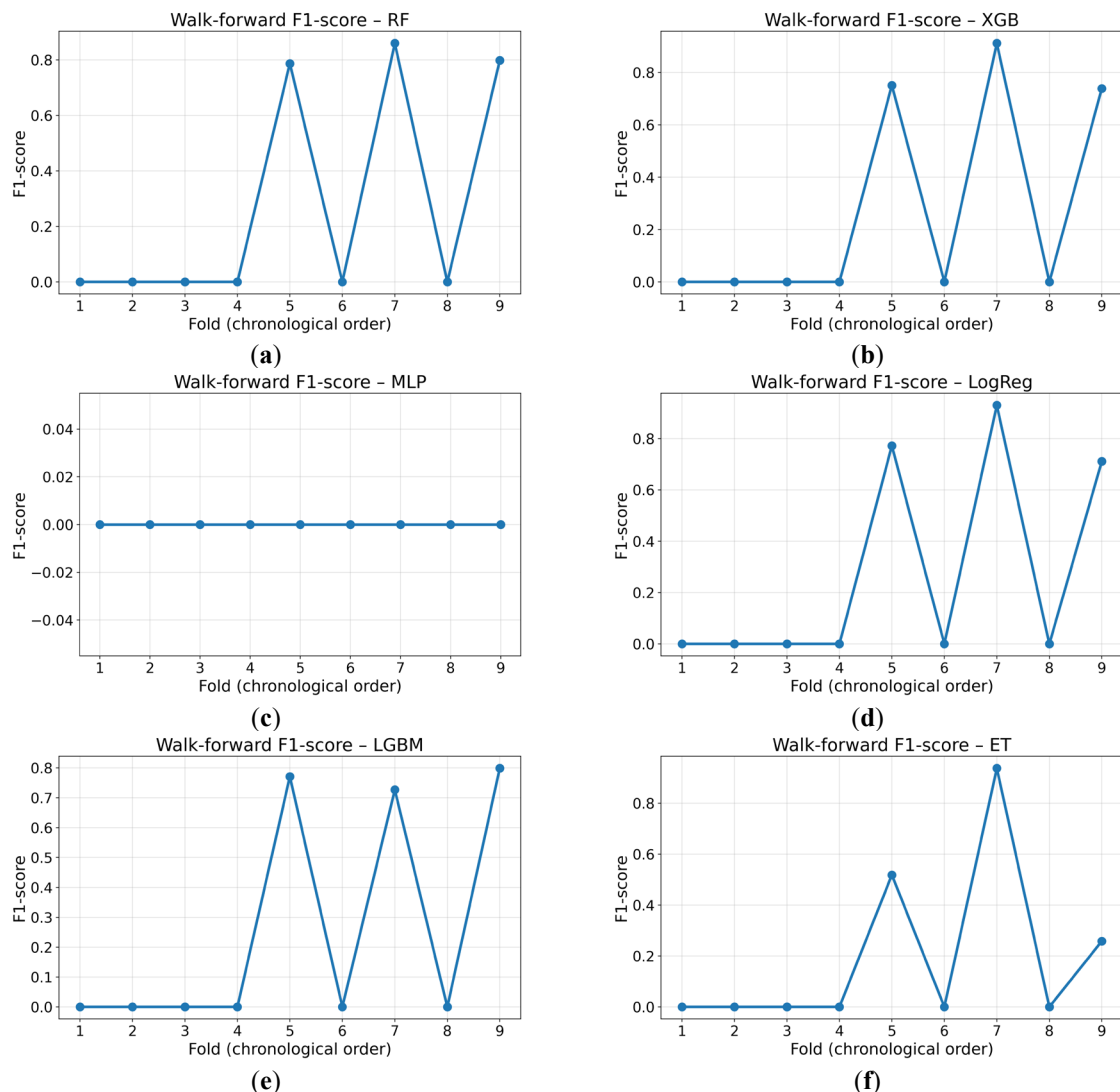


Figure 9. Walk-forward validation F1-score across folds for classification models: (a) Random Forest, (b) XGBoost, (c) MLP, (d) Logistic Regression, (e) LightGBM, and (f) Extra Trees.

The walk-forward validation results show that Random Forest and XGBoost exhibit relatively stable F1-score trajectories across folds, indicating robustness under time-dependent operational shifts. Random Forest is recommended for deployment-oriented predictive maintenance due to high stable performance. LightGBM exhibits similar stability with moderate fluctuation. Logistic Regression shows higher variability across folds, reflecting sensitivity to temporal distribution changes. Extra Trees presents inconsistent performance,

while the MLP model fails to detect breakdown-proximate cases, resulting in consistently low F1 values. These findings highlight the importance of temporal validation when selecting predictive maintenance models for real industrial deployment.

4.4 Practical Validation of Predictive Maintenance

A deployment-oriented predictive maintenance was conducted to validate the practical applicability. The validated breakdown classification model of Random Forest with the optimum threshold optimization of 0.450

in **Table 10** was systematically applied to historical sensor data in March 2025 (2,676,013 time-indexed observations). This aims to investigate whether higher risk signals often precede actual breakdown occurrences and how they might assist preventive maintenance decisions in contrast to traditional fixed-interval scheduling, rather than reevaluating prediction accuracy.

The deployment-oriented retrospective test-back results was illustrated in **Table 11** to assess how the validated model provides effective early-warning

signals prior to actual breakdown events. In **Table 11**, five breakdown IDs associated with key sensor groups were discovered. Breakdown ID BD-01 is related to pressure measurement (i.e. Beverage Pressure 1 and 2, Tank Pressure Realtime) and flow rate measurement (i.e. Beverage Flow Meter). Breakdown ID BD-02 refers to CO₂-measurement (CO₂ Realtime), and speed measurement (Speed Actual), while BD-03 concerned with temperature measurement (Beverage Temp).

Table 11. Deployment-oriented predictive maintenance retrospective analysis

Breakdown ID	Key Sensor Group	Actual Breakdown Date	Predicted Breakdown Date	Relative Lead-Time Category	Alignment with Actual Breakdown
BD-01	Beverage Pressure 1/2, Tank Pressure Realtime, Beverage Flow Meter	3/03/25	3/03/25	Short	Same-day detection
BD-02	CO ₂ Realtime, CO ₂ Flow Meter, Speed Actual	6/03/25	5-6/03/25	Medium	Prior detection
BD-03	Pressure, Flow, Beverage Temp variables	17/03/25	17/03/25	Medium	Same-day detection
BD-04	Beverage Pressure 1/2, Tank Level	21/03/25	21/03/25	Short	Same-day detection
BD-05	-	29/03/25	Not detected	—	No precursor

The retrospective test-back analysis indicates that the predicted breakdowns were detected either on the same date (BD-01 and BD-03) or within short lead-time intervals prior to the actual breakdown events (BD-02). Only BD-05 exhibited no clear precursor signal, suggesting that certain sudden stoppages may not present measurable early-warning behavior under the current sensor configuration. **Table 12** presents the breakdown event descriptions and dominant process symptoms according to each Breakdown ID for maintenance actions. For example, BD-01 exhibits process symptoms of pressure instability and flow

deviation with respected to the results of beverage pressure, tank pressure and beverage flow meter. BD-02 shows CO₂ imbalance combined with machine speed fluctuation, while BD-03 demonstrates multivariate deviation across pressure, flow, and temperature variables prior to stoppage. Overall, the findings show that the proposed predictive maintenance approach is suitable for practice, although the developed model is limited to binary classification of normal operation and a breakdown-proximate, rather than directly predicting the exact failure time.

Table 12. Breakdown event descriptions and dominant process symptoms.

Breakdown ID	Key Sensor Groups	Breakdown Description	Dominant Process Symptoms
BD-01	Beverage Pressure 1/2, Tank Pressure Realtime, Beverage Flow Meter	Unplanned machine stoppage	Pressure instability, flow deviation
BD-02	CO ₂ Realtime, CO ₂ Flow Meter, Speed Actual	Filling process interruption	CO ₂ imbalance, speed fluctuation
BD-03	Pressure, Flow, Beverage Temp variables	Process-related breakdown	Multivariate deviation before stoppage
BD-04	Beverage Pressure 1/2, Tank Level	Short-duration breakdown	Transient pressure fluctuation
BD-05	-	Sudden stoppage	No clear precursor observed

The deployment-oriented framework was validated by operational experts, and it is scheduled for implementation in more machines and the actual production environment. The predictive maintenance model exhibits feasibility from managerial perspective; however, additional improvement is advised in advance of extensive deployment. Extended data collection, particularly about breakdown behavior, could enhance the model's performance to predictive maintenance.

Considering the balance between normal operations and breakdown states is crucial, as it assist the manufacturer to decrease both missed breakdown signals and unnecessary maintenance. In addition, the improving breakdown indicators, such as dynamic time stamps could enhance model performance and improve preventive maintenance, thereby reducing machine breakdown. In order to improve maintenance planning and operational decision-making, additional work will integrate

breakdown-risk estimates with reliability metrics, such as time between failures (TBF) and mean time between failures (MTBF). This framework can be further applied to other machines in the filling line and across different production lines (e.g., capping, labeling, packaging processes), as well as other beverage manufacturing facilities with similar sensor configurations. Future research may involve integration with computerized maintenance management systems to enable automated work orders derived from risk assessments, thus enhancing scalability and operational adaptability.

5. Conclusion

This study developed a machine learning-based predictive maintenance framework for beverage manufacturing, focusing on the filling process to overcome the shortcomings of experience-driven maintenance planning. A month of high-resolution operational data (2,676,013 time-indexed observations from 18 sensors) was employed to improve predictive performance. Correlation analysis was performed to identify critical sensors related to breakdown occurrences, leading to the selection of 15 significant sensors for classification modelling. Six classification models, including Logistic Regression, Random Forest, Extra Trees, XGBoost, LightGBM, and MLP, were evaluated within a 60-minute decision horizon. Among the evaluated models, the Random Forest demonstrated the best performance. With the optimized operating threshold of 0.45, the model achieved a well-balanced performance (accuracy = 0.990, precision = 0.918, recall = 0.722, F1-score = 0.808, ROC-AUC = 0.995), effectively balancing early breakdown detection with false alarm control.

The developed model, although restricted to binary classification, distinguishes between normal operation and breakdown-proximate state with a 60-minute decision horizon. The results indicate that the proposed data-driven predictive maintenance framework can support the identification of breakdown-proximate states and provide useful information for maintenance decision. The manufacturer can reduce both missed failure signals and excessive maintenance by managing a balance between normal operations and breakdown states. More collecting data of a longer period and other machines is suggested for achieving full process deployment. Subsequent research will associate breakdown-risk predictions with reliability measures, including TBF and MTBF, to improve maintenance planning and operational decision-making.

6. Acknowledgments

The work was supported by Industrial and Manufacturing Engineering Department, Faculty of Engineering, Prince of Songkla University. The authors express their gratitude to the management of the beverage manufacturing plant for providing access to production data utilized for model validation.

7. References

- [1] J. Moubray, "Proactive Maintenance 1: Preventive Tasks," in *Reliability-Centered Maintenance*, 2nd ed. New York, NY, USA: Industrial Press Inc., 1997, ch. 6, secs. 6.1–6.6, pp. 129–143.
- [2] J. Moubray, "Proactive Maintenance 2: Predictive Tasks," in *Reliability-Centered Maintenance*, 2nd ed. New York, NY, USA: Industrial Press Inc., 1997, ch. 7, secs. 7.1–7.9, pp. 144–169.
- [3] R. K. Mobley, "Impact of Maintenance," in *An Introduction to Predictive Maintenance*, 2nd ed., Boston, MA, USA: Butterworth-Heinemann, 2002, ch. 1, sec. 1.1, pp. 2–9.
- [4] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006, doi: 10.1016/j.ymssp.2005.09.012.
- [5] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009, doi: 10.1016/j.ymssp.2008.06.009.
- [6] J. Lee, B. Bagheri, and H. -A. Kao, "A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol. 3, pp. 18–23, 2015, doi: 10.1016/j.mfglet.2014.12.001.
- [7] M. Compare, P. Baraldi, and E. Zio, "Challenges to IoT-Enabled Predictive Maintenance for Industry 4.0," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4585–4597, 2019, doi: 10.1109/JIOT.2019.2957029.
- [8] A. Bousdekis, D. Apostolou, and G. Mentzas, "Predictive Maintenance in the 4th Industrial Revolution: Benefits, Business Opportunities, and Managerial Implications," *IEEE Engineering Management Review*, vol. 48, no. 1, pp. 57–62, 2019, doi: 10.1109/EMR.2019.2958037.
- [9] B. Hrnjica, and S. Softic, "Explainable AI in Manufacturing: A Predictive Maintenance Case Study," in *IFIP WG 5.7 International Conference*, Novi Sad, Serbia, Aug. 30–Sep. 3, 2020, pp. 66–73, 2020.
- [10] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Computers & industrial engineering*, vol. 150, 2020, Art. no. 106889, doi: 10.1016/j.cie.2020.106889.
- [11] S. Ayvaz and K. Alpay, "Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real time," *Expert Systems with Applications*, vol. 173, 2021, Art. no. 114598, doi: 10.1016/j.eswa.2021.114598.
- [12] Y. S. Cheong, C. S. Seah, Y. X. Loh, and L. H. Loh, "Artificial Intelligence (Ai) In The Food

- And Beverage Industry: Improves The Customer Experience,” in *2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, Ipoh, Malaysia, Sep. 8–9, 2021, pp. 1–6, doi: 10.1109/AiDAS53897.2021.9574261.
- [13] I. Ahmed, G. Jeon, and F. Piccialli, “From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5031–5042, 2022, doi: 10.1109/TII.2022.3146552.
- [14] V. Terziyan, and O. Vitko, “Explainable AI for Industry 4.0: Semantic Representation of Deep Learning Models,” *Procedia Computer Science*, vol. 200, pp. 216–226, 2022, doi: 10.1016/j.procs.2022.01.220.
- [15] Z. A. B. Sinnah, “Explainable AI-driven predictive maintenance for mitigating process safety risks in safety-critical industrial equipment,” *Journal of Loss Prevention in the Process Industries*, vol. 100, 2025, Art. no. 105907, doi: 10.1016/j.jlp.2025.105907.
- [16] C. Walker, V. Agarwal, L. Lin, A. Hall, R. Hill, T. Mortenson, and N. Lybeck, “Explainable Artificial Intelligence Technology for Predictive Maintenance,” Idaho National Laboratory, Idaho Falls, ID, USA, Rep. INL/RPT-23-74159, 2023.
- [17] A. Ucar, M. Karakose, and N. Kırımca, “Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends,” *Applied Sciences*, vol. 14, no. 2, 2024, Art. no. 898, doi: 10.3390/app14020898.
- [18] T. Akyaz and D. Engin, “Machine Learning-Based Predictive Maintenance System for Artificial Yarn Machines,” *IEEE Access*, vol. 12, pp. 125446–125461, 2024, doi: 10.1109/ACCESS.2024.3454548.
- [19] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, “Machinery health prognostics: A systematic review from data acquisition to RUL prediction,” *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018, doi: 10.1016/j.ymssp.2017.11.016.
- [20] W. Zhang, D. Yang, and H. Wang, “Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey,” *IEEE Systems Journal*, vol. 13, no. 3, pp. 2213–2227, 2019, doi: 10.1109/JSYST.2019.2905565.
- [21] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, “A systematic literature review of machine learning methods applied to predictive maintenance,” *Computers & Industrial Engineering*, vol. 137, 2019, Art. no. 106024, doi: 10.1016/j.cie.2019.106024.
- [22] P. Nunes, J. Santos, and E. Rocha, “Challenges in Predictive Maintenance – A Review,” *CIRP Journal of Manufacturing Science and Technology*, vol. 40, pp. 53–67, 2023, doi: 10.1016/j.cirpj.2022.11.004.
- [23] H. M. H. A. Alshkeili, S. J. Almheiri, and M. A. Khan, “Privacy-Preserving Interpretability: An Explainable Federated Learning Model for Predictive Maintenance in Sustainable Manufacturing and Industry 4.0,” *AI*, vol. 6, No. 6, 2025, Art. no. 117, doi: 10.3390/ai6060117.
- [24] M. Achouch, M. Dimitrova, R. Dhoubi, H. Ibrahim, M. Adda, S. S. Karganroudi, K. Ziane, and A. Aminzadeh, “Predictive Maintenance and Fault Monitoring Enabled by Machine Learning: Experimental Analysis of a TA-48 Multistage Centrifugal Plant Compressor,” *Applied Sciences*, vol. 13, no. 3, 2023, Art. no. 1790, doi: 10.3390/app13031790.
- [25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. VanderPlas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [26] F. E. Bezerra, G. C. de Oliveira Neto, G. M. Cervi, R. F. Mazetto, A. M. de Faria, M. Vidua, G. A. Lima, S. A. de Araújo, M. Sampaio, and M. Sampaio, “Impacts of Feature Selection on Predicting Machine Failures by Machine Learning Algorithms,” *Applied Sciences*, vol. 14, no. 8, 2024, Art. no. 3337, doi: 10.3390/app14083337.
- [27] E. Lwele, A. Shenfield, and C. E. Silva, “AI-Based Surrogate Models for the Food and Drink Manufacturing Industry: A Comprehensive Review,” *Processes*, vol. 13, no. 9, 2025, Art. no. 2929, doi: 10.3390/pr13092929.
- [28] K. G. Liakos, V. Athanasiadis, E. Bozinou, and S. I. Lalas, “Machine Learning for Quality Control in the Food Industry: A Review,” *Foods*, vol. 14, no. 19, 2025, Art. no. 3424, doi: 10.3390/foods14193424.
- [29] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, “Optimal Thresholding of Classifiers to Maximize F1 Measure,” in *European Conference, ECML PKDD 2014*, Nancy, France, Sep. 15–19, 2014, pp. 225–239, doi: 10.1007/978-3-662-44851-9_15.
- [30] A. Aminzadeh, S. S. Karganroudi, S. Majidi, C. Dabompre, K. Azaiez, C. Mitride, and E. Sénéchal, “A Machine Learning Implementation to Predictive Maintenance and Monitoring of Industrial Compressors,” *Sensors*, vol. 25, no. 4, 2025, Art. no. 1006, doi: 10.3390/s25041006.