Enhanced Fault Detection and Diagnosis in Industrial Distillation Column Using Explainable Artificial Intelligence and Machine Learning

Sumana Roy[1], Somasish Saha[2], Bitopama Modak[3], Fahim Ahmed[4], Aditi Mahto[5], Sandip Kumar Lahiri[6*]

[1,2,5,6]*National Institute of Technology Durgapur, Mahatma Gandhi Avenue, Durgapur – 713209, West Bengal, India*
sr.19ch1501@phd.nitdgp.ac.in
ss.19ch1502@phd.nitdgp.ac.in
am.22u10882@btech.nitdgp.ac.in
**Corresponding author:sklahiri.che@nitdgp.ac.in*

[3,4]*Jadavpur University, 188, Raja S.C. Mallick Road, Kolkata – 700032, West Bengal, India*
mbitopama@gmail.com
fahimahamed0576@gmail.com

Abstract

This study presents a comprehensive methodology for developing a Fault Detection and Diagnosis (FDD) system for an industrial distillation column using advanced machine learning algorithms. Steady state and dynamic simulations in Aspen Plus® generate extensive datasets under normal and faulty conditions. Feature engineering, using the Minimum Redundancy Maximum Relevance (MRMR) algorithm, selects the most relevant features for fault detection. Various machine learning models, including Decision Trees, Support Vector Machines, k-nearest Neighbours, and Neural Networks, were trained and evaluated based on performance metrics such as accuracy, recall, precision, and F1 score.

The top models were integrated into a stacked classifier system with a voting mechanism to enhance fault detection reliability. Explainable Artificial Intelligence (XAI) techniques, such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP), were incorporated to improve model interpretability, allowing engineers to understand and validate the FDD system's decision-making process.

Simulation results confirm that the proposed methodology accurately identifies and classifies faults. By integrating dynamic simulations, advanced machine learning, and XAI techniques, a robust and scalable solution is achieved for fault detection in distillation columns, improving operational reliability, safety, and reducing downtime. Future work could extend this approach to other industrial processes and explore additional machine learning algorithms to further enhance performance.

1. Introduction

The growing global competition has driven chemical process industries to minimize downtime, improve equipment availability, and enhance operational profitability. Fault Detection and Diagnosis (FDD) systems are essential in managing the complexity and dynamic nature of these processes, where manual fault detection is increasingly difficult due to numerous interacting variables.

Effective FDD systems offer real-time monitoring, early fault detection, and accurate diagnosis, leading to enhanced process safety, operational reliability, reduced environmental impact, and improved product quality (Isermann, 2006). Faults are inevitable in complex process systems. Various faults in process plants can arise from changes in process parameters, actuator issues, sensor malfunctions, or external disturbances (Chiang et al., 2000; Kesavan & Lee, 1997). Fault detection involves identifying the process variables most relevant to a fault, whereas fault diagnosis determines the root cause and location of the fault. Ignored faults can lead to system failures, and minor faults can escalate into severe issues, disrupting overall processes (Chiang et al., 2000; Chiang & Pell, 2004; Tidriri et al., 2016; Zhou et al., 2014; Isermann & Balle, 1997).

The vast amount of operational data collected every second in process industries offers an opportunity to develop advanced fault detection systems. Although large volumes of data are collected, their underutilization often leads to a gap between data availability and actionable insights. Leveraging

https://doi.org/10.36001/IJPHM.2025.v16i2.4416

historical process data, it is possible to build data-driven fault detection systems that transform raw data into actionable insights. Given the nonlinearity and increasing complexity in modern process industries, there is a growing demand for data-driven approaches (Dash et al., 2003; Meng et al., 2019; Copertaro et al., 2019; Venkatasubramanian, Rengaswamy, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, & Yin 2003; Ge et al., 2013; Chetouani, 2011). Although data-driven methods require large amounts of historical data, this is less of an issue today due to the widespread use of distributed control systems (DCS) and soft computing technologies, which facilitate data collection (Ge et al., 2017). Moreover, database and data mining technologies support the development of data-driven modelling methods in industrial processes.

In literature, FDD have been applied to various chemical processes, including the Tennessee Eastman process (TEP) (Yu & Zhang, 2020; D. Liu et al., 2020), reactor systems (Zio et al., 2009 ; Nawaz et al., 2020), distillation columns (Chetouani, 2011; Chetouani, 2007; Manssouri et al., 2008; S. A. Taqvi et al., 2018b; S. A. Taqvi et al., 2018a; Mujtaba et al., 2020; Amiruddin et al., 2020; S. A. A. Taqvi, Zabiri, Tufa, Uddin, et al., 2020; S. A. A. Taqvi, Zabiri, Tufa, Fatima, et al., 2020), bearing faults (Rajakarunakaran et al., 2008), crude and gas mixture pipelines (Amiruddin et al., 2020; Mujtaba et al., 2020), industrial gas turbines (Nozari et al., 2012), heating furnaces (Schubert et al., 2011), water-cooled centrifugal chillers (Zhao et al., 2013), biochemical wastewater treatment plants (X. Zhang & Hoo, 2011), controlled two-tank systems (Weber et al., 2006), and fluid catalytic cracking units (Vedam et al., 1999).

Distillation is a key and widely used separation technique in process industries, known for its high-energy consumption. Industrial distillation columns, essential in chemical processes, are sensitive to faults like sensor failures, actuator malfunctions, and process disturbances, which can affect product purity and energy efficiency. Monitoring these columns is difficult due to their nonlinearity, disturbances, nonstationary behaviour, and multivariable coupling. The high dimensionality, multiple operational modes, and slow dynamics further complicate fault diagnosis. Therefore, developing a robust FDD system is critical to ensure optimal operation, prevent downtime, and avoid safety incidents (Rengaswamy & Venkatasubramanian, 2001).

Traditional approaches for FDD systems in distillation columns include Principal Component Analysis (PCA), Partial Least Squares (PLS), and Independent Component Analysis (ICA), widely used for fault detection in multivariate data. PCA detects sensor faults by analysing principal components in process data (W. Li et al., 2000). PLS monitors processes by modelling input-output relationships (MacGregor & Kourti, 1995) , while ICA separates mixed signals into independent components for fault detection (J. M. Lee et al., 2004). However, these methods assume linearity, limiting their effectiveness with complex, non-linear data. Consequently, data-driven machine learning methods, treating fault diagnosis as a classification task, are now preferred.

Machine learning (ML) algorithms for fault diagnosis have been extensively explored, particularly supervised learning techniques trained on historical data to detect faults. In literature, techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Bayesian Networks (BN), Random Forests (RF), Decision Trees (DT), and k-nearest Neighbours (KNN) are commonly used. Decision Trees (DT) are popular for process monitoring and fault detection in industries (Quinlan, 1979; Quinlan, 2014), with hybrid models like Artificial Neural Networks Decision Trees (ANN-DT) and Support Vector Machines Decision Trees (SVM-DT) improving fault classification and reliability. (Ma & Wang, 2009; He et al., 2013; Kuo & Lin, 2010; Demetgul, 2013; Aydin et al., 2014; Karabadji et al., 2014).

KNN is a non-parametric method applied to various fault detection scenarios (Altman, 1992; Andre et al., 2013; Nguyen & Lee, 2010; Tudón-Martínez & Morales-Menendez, 2015; Hasan & Kim, 2019; Y. Li & Zhang, 2014; S. Zhang et al., 2017). Fisher Discriminant Analysis (FDA) is used for data classification and dimensionality reduction (Nor et al., 2020). FDA has been extensively applied in process monitoring and fault classification (Jiang et al., 2006; Z. Xu et al., 2006). Advanced methods combining FDA with statistical analysis, genetic algorithms, and kernel techniques have shown improved fault detection and classification performance (Nor et al., 2020; Chiang & Pell, 2004; Zhu & Song, 2010; X. Zhang et al., 2007; Ge et al., 2016).

Artificial Neural Networks (ANN) map complex input-output relationships and have been widely used in chemical engineering for fault detection and diagnosis (Amiruddin et al., 2020). Various ANNs, including Multilayer Perceptron (MLP), Radial Basis Function (RBF), and dynamic fuzzy neural networks, have been applied in process monitoring, emission monitoring, fault location estimation, and tool condition monitoring (Sharma et al., 2004; M. W. Lee et al., 2005; Iliyas et al., 2013; Gonzaga et al., 2009; Jamil et al., 2014; Lu & Xue, 2014; S. Xu & Liu, 2014; Pani & Mohanta, 2015). Comparative studies have highlighted ANN's effectiveness in fault classification tasks (Hwang et al., 1993; Nagpal & Brar, 2014; Venkatasubramanian & Chan, 1989; Watanabe et al., 1989; Ungar et al., 1990; Hoskins et al., 1991).

SVM is another popular tool for process monitoring and has shown superior performance in fault diagnosis, often enhanced by hybrid models (S. A. Taqvi et al., 2017).

The literature review highlights the versatility of machine learning algorithms such as Decision Trees, KNN, FDA, ANNs, and SVMs in addressing complex fault diagnosis challenges in chemical processes.

There are two key gaps in the literature regarding the application of Machine Learning techniques in FDD for chemical process industries. While ML techniques offer a promising approach for FDD systems by effectively modelling non-linear relationships, traditional ML models are often perceived as black boxes, lacking transparency, which is crucial in safety-sensitive environments like chemical process industries (J. Zhang et al., 2019). Another challenge is the limited availability of fault data in well-maintained plants, which can be addressed through dynamic simulations using tools like Aspen Dynamics to create diverse fault scenarios (Maurya et al., 2007).

Dynamic simulation enables the creation of extensive fault datasets, aiding in the training and validation of ML models for FDD systems. In this paper, we simulate normal and fault data for an industrial distillation column and apply various ML algorithms to develop an FDD system. The best algorithms are selected based on accuracy and robustness. To improve model interpretability, we use Explainable Artificial Intelligence (XAI) techniques like Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP), which allow plant engineers to understand and validate the decisions made by the FDD system (Molnar, 2019).

LIME provides insights into how specific features impact predictions by approximating black-box models locally (Molnar, 2019). SHAP, based on cooperative game theory, consistently measures feature importance by evaluating all possible feature combinations (Lundberg et al., 2017). Both LIME and SHAP help engineers interpret ML model decisions for fault detection. By combining first-principle methods with data-driven approaches, we aim to create robust fault diagnosis systems that improve safety, efficiency, and cost savings by reducing downtime and preventing failures.

The following sections cover the methodology, results, and application of advanced techniques for distillation column fault detection. Section 2 details the process of generating simulation data, using the Minimum-Redundancy-Maximum-Relevance (MRMR) algorithm for feature selection, and applying Machine Learning models, with Explainable AI techniques enhancing interpretability. Section 3 presents the results from dynamic simulations and fault classification, showing improved accuracy and efficiency through feature selection and LIME/SHAP analysis. Section 4 concludes by highlighting the integration of machine learning, simulations, and Explainable AI for effective fault detection and offers recommendations for future research in real-time FDD systems.

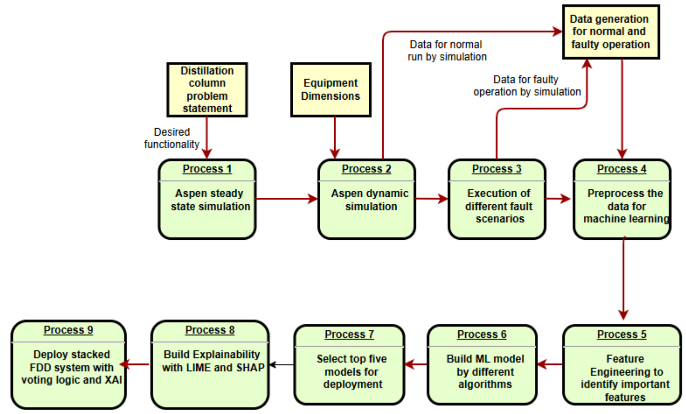

Figure 1. An overview of the methodology used to develop the FDD system for a distillation column using machine learning algorithms

2. METHODOLOGY ADOPTED TO BUILD THE FAULT DETECTION AND DIAGNOSIS SYSTEM

2.1. Overview

Figure 1 provides an overview of the methodology used to develop the FDD system for a distillation column using machine learning algorithms. The process starts with steady-state and dynamic simulations in Aspen Plus® to generate datasets from various fault scenarios. Afterward, data pre-processing and feature selection are performed using the MRMR algorithm to identify key features. Several machine learning models, including Decision Trees, Discriminant Analysis, Logistic Regression, SVMs, k-NN, Ensemble Methods, Naive Bayes, and Neural Networks, are trained and evaluated. The top five models are combined into a stacked classifier with a voting mechanism to enhance reliability and robustness. Explainable AI techniques like LIME and SHAP are also applied to improve model interpretability, helping plant engineers quickly diagnose and address faults. Table 1 summarizes the step-by-step methodology adopted for FDD system development.

2.2. Step 1: Aspen steady state simulation

The first step involves creating a steady-state simulation of the distillation column using Aspen Plus®. This process uses a mixture of propane and isobutene with a (60-40) mole % composition as the feed. The Rigorous Distillation Simulation Module (RADFRAC model) in Aspen Plus which includes a column, condenser, and Reboiler, is employed. The column has an overall height of 5.5 meters, an internal diameter of 0.15 meters, and a tray spacing of 0.35 meters as calculated from Aspen hydraulic calculations.

The reflux ratio and reboiler heat duty are varied to ensure the top and bottom compositions contain 99 mole% propane and 99 mole% isobutanes, respectively. The steady-state simulation results are validated against plant operating conditions, reporting top and bottom temperatures of 37.9 °C and 78.9 °C, respectively, and a column pressure of 13 bar (absolute). The specifications and sizing data for the industrial-scale distillation column (located in Saudi Arabia) are detailed in Table 2.

Step	Description
1. Generate Dataset	Use Aspen Plus dynamic simulation to create datasets for normal and faulty operations.
2. Prepare Dataset	Organize data into predictors (X) and class labels (Y). "X" contains observed and derived variables, while "Y" holds class labels (e.g., normal, fault type 1, fault type 2).
3. Feature Selection	Apply the MRMR algorithm to identify and select the top-ranked features.
4. Define Cross-Validation Method	Use k-fold cross-validation: partition data into k sets, train on out-of-fold data, evaluate on in-fold data, and average the error across all folds.
5. Choose Classification Algorithm	Select algorithms like Decision Trees, Ensemble Methods, Neural Networks, etc., as listed in Table 4.
6. Train Classifier	Train the classifier using the simulation data with selected features.
7. Test Classifier	Test the classifier on the prepared dataset to assess its performance.
8. Evaluate Performance	Evaluate performance using metrics like accuracy, precision, recall, and F1 score (see Table 7).
9. Implement a Stack Classifier System	Choose the top 5 classifiers and implement a voting logic system to classify faults based on the collective output, improving reliability.
10. Enhance Model Interpretability with Explainable AI (XAI)	Use LIME and SHAP to explain the top classifier's decisions, helping engineers diagnose faults and improve system trust.

Table 1. Fault Diagnosis Algorithm Using Machine Learning Techniques

2.3. Step 2: Aspen dynamic simulation

After the steady-state simulation is run without errors, it is exported to Aspen Plus® Dynamics for further control studies and data generation. Various controllers for feed flow, top and bottom compositions, pressure, and level are installed in the distillation column. Six controllers are installed in a dynamic simulation environment, summarized in Table 3 and shown in Figure 2. The feed flow controller ensures a consistent feed rate, the pressure and level controllers maintain the appropriate pressure and liquid levels within the column and the composition controllers adjust the reflux and reboiler heat duty to achieve the target purities for propane and isobutene.

The RADFRAC model in Aspen Plus® Dynamics accounts for the pressure drop across each stage due to liquid and vapour flow resistances and allows for stage hydraulics modelling (refer Table 2). The column pressure is slightly increased to 13.24 bar for the dynamic simulation to accommodate additional inlet/outlet valves and associated pressure drops across each tray. Different control functions in the distillation column are presented in Table 3.

2.4. Step 3: Execution of different fault scenarios

In this step, the dynamic simulation of the distillation column is used to generate both normal and faulty datasets. Regular operation is simulated for 20 hours, followed by introducing faults after every 5 hours of stable operation. Eight types of faults are introduced to represent various operational issues. In industrial distillation columns, the occurrence of these faults is not uncommon, as they stem from both mechanical issues and operational inefficiencies. Faults like tray efficiency loss (F1, F5, F6, F7) can result from wear, corrosion, or fouling, leading to poor separation and product quality.

Specifications	Value
Feed specifications	
Feed flow rate	100 kilo mole/hour
Feed composition	Propane (40 mole%), Isobutane (60 mole%)
Feed temperature	40 °C
Feed pressure	18 bar
Column specifications	
Column height	5.5 meter
Colum diameter	0.15 meter
Type of tray	Sieve
No. of trays	31
Distillate flow rate	40 kilo mole/hour
Bottom flow rate	60 kilo mole/hour
Steady-state operating condition	
Top temperature	37.96
Bottom temperature	78.92
Reboiler duty	0.647 Million kilocalorie/hour
Condenser duty	-0.556 Million kilocalorie/hour
Column pressure	13 bar
Distillate composition	99% propane, 1% isobutene (molecule)
Bottom Composition	99 % isobutene, 1% propane(molecule)

Table 2. Specifications and Sizing Data for the Distillation Column Dynamics Simulation

Feed flow and composition variations (F2, F4) may arise due to equipment malfunctions or inconsistent raw material supply, while temperature fluctuations (F3) could be caused by heat exchanger issues. Valve stiction (F8, F9), a prevalent problem, often occurs due to aging control valves, causing oscillatory behaviour that disrupts the column's stability. Addressing these faults is critical for maintaining process efficiency, product quality, and safety in industrial distillation operations. While this study primarily considers step-type failures due to their well-defined impact and ease of simulation, it is acknowledged that real-world faults may also manifest as gradual or incipient failures. Step faults allow for clear benchmarking and classification in early development of FDD systems. However, future work will incorporate gradual degradation patterns to better represent real-life industrial conditions. This is especially important in soft-fault scenarios like sensor drift, valve wear, or slow fouling, where early detection remains a challenge. Inclusion of such patterns will further improve robustness and practical deployment of FDD systems. These fault scenarios provide a comprehensive dataset to evaluate the fault detection and diagnosis methods. Details of the introduced faults are summarized in Table 4.

2.5. Step 4: Data pre-processing for machine learning

The time series data generated from normal and faulty operations in dynamic simulations is pre-processed to make it suitable for machine learning applications. This process involves data cleaning, normalization of input and output variables to the range [0, 1], and splitting the data into training, validation, and testing sets in a (60-20-20) percentage ratio.

Additionally, pseudo-random binary sequences (PRBS) excite the plant under normal conditions in dynamic simulations, and zero mean normal distributed noise is added to all measured variables to simulate sensor noise. The variables shortlisted for fault diagnosis are listed in Table 5.

2.6. Step 4: Data pre-processing for machine learning

Feature engineering is applied to extract the most relevant variables using the Minimum Redundancy Maximum Relevance (MRMR) algorithm. MRMR optimizes feature selection by balancing relevance to the target variable and minimizing redundancy between features, based on mutual information. This ensures that each feature contributes unique and valuable information, enhancing model performance.

The algorithm employs a forward selection method to incrementally add features, significantly reducing computational complexity. Features are ranked based on their Mutual Information Quotient (MIQ), with high scores indicating strong predictors. This process improves classification accuracy by focusing only on the most critical features, reducing model overfitting, and simplifying data analysis.

MRMR is particularly effective for handling high-dimensional datasets, making it an essential tool for improving the accuracy and robustness of machine learning models in fault diagnosis. Its scalability ensures efficient processing of large datasets, enabling precise and reliable fault detection in complex systems.

2.7. Step 6: Build ML model by different algorithms

Several machine learning algorithms were explored to develop an optimal fault diagnosis system, each with unique strengths suited for different aspects of fault classification.

Extensive literature survey is done to shortlist strong algorithms which are successfully used in fault diagnosis. Decision Trees (DT) were used for their interpretability and ability to handle non-linear relationships with minimal pre-processing. Discriminant Analysis (DA) was selected for its robustness in classifying linearly or quadratically separable classes. Logistic Regression (LR) offered simplicity and efficiency in binary classification, providing probabilistic interpretations. Support Vector Machines (SVM) handled high-dimensional data and were effective in complex fault scenarios. K-Nearest Neighbours (k-NN) excelled in cases with undefined decision boundaries and large datasets.

Ensemble Methods improved accuracy by combining classifiers, while Naive Bayes (NB) provided fast probabilistic fault detection. Neural Networks (NN) captured complex data patterns, ideal for diagnosing dynamic, non-linear faults. Applying different machine learning algorithms in FDD allows leveraging the unique strengths of each model, improving overall system performance. Algorithms like Decision Trees and SVMs offer interpretability and handling of complex relationships, while Ensemble Methods and Neural Networks enhance accuracy by combining models or capturing intricate patterns in dynamic systems. The top five models were shortlisted by evaluating the performance of these diverse algorithms and their variants based on their accuracy, precision, recall, and F1 score. This comprehensive evaluation ensures the selection of the most effective and reliable models for fault diagnosis in the distillation column, leveraging the unique strengths of each algorithm to enhance system performance.

The different machine learning algorithms and their advantages and limitations in fault diagnosis are summarized in Table 6.

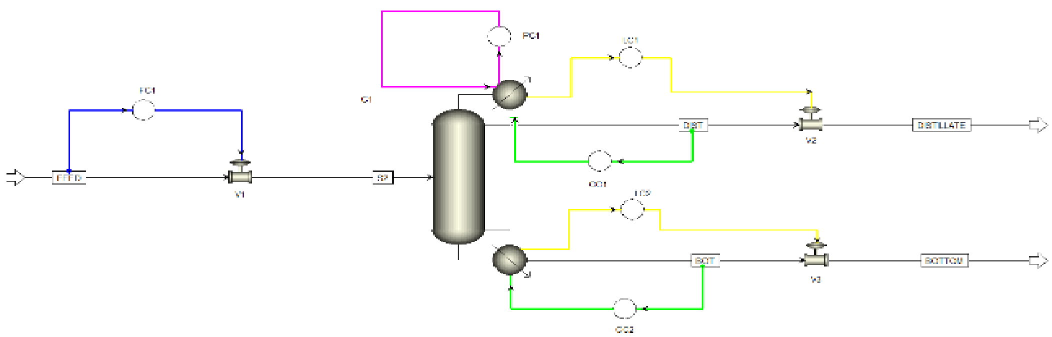

Figure 2. Controllers Implemented in Aspen

SL No	Controller name	Controller function	Operation
1	Feed Flow Controller (FC1)	Regulates the feed flow rate into the distillation column.	It ensures a consistent and controlled feed rate, crucial for stable column operation and accurate separation of components.
2	Pressure Controller (PC1)	Maintains the pressure within the column.	Adjusts the condenser heat duty for maintaining column pressure to ensure the correct boiling points are maintained for separating propane and isobutene, which is critical for achieving the desired product purity.
3	Level Controller (LC1)	Controls the liquid level in the reflux drum.	Regulates the flow of the distillate to maintain the liquid level within the column, preventing overflows or dry trays, which can affect separation efficiency.
4	Composition Controller (CC1)	Maintains the purity of the propane in the top product.	Manipulates the reflux flow rate to ensure the top product achieves 99-mole% propane purity. This controller adjusts the amount of liquid returned to the column versus taken off as the product.
5	Level Controller (LC2)	Manages the liquid level at the bottom of the distillation column.	Regulates the bottom product flow rate to maintain the liquid level, ensuring stable operation and effective separation of components.
6	Composition Controller (CC2)	Maintains the purity of the isobutene in the bottom product.	Manipulates the reboiler heat duty by adjusting the steam flow to ensure the bottom product achieves 99 mole% isobutene purity. This controller ensures that the correct amount of heat is supplied to the reboiler for optimal separation.

Table 3. Functions of Controllers in the Distillation Column

Fault ID	Fault Description	Fault Type	Magnitude	Impacts
F1	Feed tray efficiency loss	Step Fault	Efficiency reduced from normal to 1 Percent	Significant reduction in tray efficiency, causing weeping and decreased liquid levels on the feed tray.
F2	Significant feed loss	Step Fault	Feed control valve opening reduced from 50 Percent to 8 Percent	Severe reduction in feed flow, leading to operational instability, potentially from valve malfunction.
F3	Feed temperature drop	Step Fault	Feed temperature drop from(68-30)°C	Drastic temperature reduction impacting top and bottom product compositions and flow rates.
F4	Feed composition fluctuations	Random Variation	Isobutane mole% fluctuates between 3.33% to +16.67 % from its normal value	Variability in feed composition affects reflux rate, reboiler duty, and overall product consistency.
F5	Top tray efficiency loss	Step Fault	Efficiency reduced to 1 Percent from normal	Similar to F1 but impacts the top (2nd) tray, resulting in poor separation and product quality.
F6	Bottom tray efficiency loss	Step Fault	Efficiency reduced to 1 Percent from normal	Affects the bottom (30th) tray, reducing vapor-liquid separation efficiency and impacting product purity.
F7	Refining section efficiency loss	Step Fault	Efficiency for all trays above the feed tray reduced to 1 Percent	Impairs the refining section’s efficiency, compromising overall column performance.
F8	Reflux valve stiction	Random Variation	Oscillation amplitude ±35 Percent of the nominal valve position	Valve stiction causes large fluctuations in reflux flow, affecting product quality and energy efficiency.
F9	Reboiler steam valve stiction	Random Variation	Oscillation amplitude ±45 Percent of the nominal valve position	Steam flow irregularities caused by stiction affect the bottom product composition and energy usage.

Table 4. Fault Scenarios and Their Descriptions

The top five models were shortlisted by evaluating the performance of these diverse algorithms and their variants based on their accuracy, precision, recall, and F1 score. This comprehensive evaluation ensures the selection of the most effective and reliable models for fault diagnosis in the distillation column, leveraging the unique strengths of each algorithm to enhance system performance. The different machine learning algorithms and their advantages and limitations in fault diagnosis are summarized in Table 6.

2.8. Step 6: Build ML model by different algorithms

The performance of various machine learning models for fault diagnosis is evaluated using statistical measures like true positive (TP), true negative (TN), false negative (FN), and false positive (FP). Key metrics include accuracy, recall, precision, and F1 score, which assess each model's effectiveness. Accuracy indicates the overall correct predictions, recall measures the ability to identify faults, precision evaluates the correctness of detected faults, and F1 score balances recall and precision. A detailed breakdown of these metrics is provided in Table 7.

2.9. Step 8: Build and deploy fault detection and diagnosis system by best performing models

The top five best-performing machine learning models are selected to build and deploy the Fault Detection and Diagnosis (FDD) system. This system is integrated into the distillation column's control framework to monitor operations and continuously diagnose faults in real-time. Fresh data from the distillation column is input into each of the five classifiers, which classify the fault independently. A voting logic is then applied to aggregate the outputs of all classifiers, making the final fault classification based on the majority vote. This stacked classifier approach enhances the reliability and robustness of the FDD system, ensuring prompt and accurate fault detection and diagnosis. The deployment of this advanced FDD system significantly improves operational reliability and safety.

Sl no	Selected variables for FDD (global level) without feature engineering	Shortlisted top five variables after feature engineering
1	Feed Composition (Isobutane mole fraction)	Feed Flow, kilo mole/hour
2	Feed Flow, kilo mole/hour	Feed temperature, °C
3	Feed temperature, °C	Feed Composition (Isobutane mole fraction)
4	Top Composition, Propane mole fraction	Reflux Flow, kilo mole/hour
5	Bottom Composition, isobutene mole fraction	
6	Reflux drum level, meter	
7	Bottom sump level, meter	
8	Top Temperature, °C	
9	Bottom Temperature, °C	
10	Top distillate Flow, kilo mole/hour	
11	Bottom flow, kilo mole/hour	
12	Column Pressure, bar	
13	Reflux Flow, kilo mole/hour	
14	Reboiler Duty, million kilocalories /hour	
15	Condenser Duty, million kilocalories /hour	

Table 5. Selected Variables for Fault Diagnosis

2.10. Step 9: enhance model interpretability with explainable artificial intelligence

To enhance the interpretability of the fault diagnosis models, we incorporate Explainable AI (XAI) techniques, such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP). XAI is crucial for building trust and understanding in machine learning models, allowing plant engineers to interpret and validate the decisions made by the FDD system. By applying LIME to the most accurate model among the top-performing models, we can provide detailed insights into which process parameters are responsible when a fault occurs. SHAP values further explain the contribution of each feature to the model's predictions. These techniques guide plant engineers in quickly diagnosing the root cause of faults, facilitating timely preventive and corrective actions to maintain operational safety and efficiency.

3. RESULTS AND DISCUSSION

3.1. Results of dynamic simulations

This section presents the results of various fault scenarios simulated using Aspen Plus Dynamics for a distillation column, revealing the system's response and control strategy effectiveness under different conditions. All the controllers are put in auto mode during dynamic simulation to allow them to take corrective actions when a fault occurs. Blue colour dotted line in figure 3-11 represents the parameters in operation and solid orange colour line represents in faulty operation in closed loop.

Figure 3 illustrates the impact of feed temperature change on various distillation column variables. A step decrease in feed temperature leads to reduced top product purity and increased reboiler heat duty, indicating higher energy demand. The control system adjusts the reflux flow to stabilize the process, emphasizing the need for robust temperature control to maintain product quality and stability.

Figure 4 shows the effect of feed loss on process variables. A significant reduction in feed flow causes sharp declines in product purity and reduced reboiler heat duty. The control system exhibits limited ability to compensate, emphasizing the importance of reliable feed flow to prevent process disruptions.

Figure 5 presents the impact of feed tray efficiency reduction on distillation column performance. Reduced tray efficiency results in lower product purities and increased variability in reboiler heat duty and reflux flow, underscoring the need to maintain optimal tray efficiency for consistent operation.

Figure 6 demonstrates the effect of feed composition changes on process variables. Variations in feed composition cause significant fluctuations in product purity, reboiler duty, and reflux flow. Monitoring and adjusting feed composition are crucial for maintaining process efficiency.

Figure 7 depicts the impact of top tray efficiency reduction on the distillation column. A decrease in top tray efficiency leads to lower product purity and increased variability in process variables, emphasizing the importance of regular maintenance and monitoring to prevent major disruptions.

Figure 8 shows the effect of bottom tray efficiency reduction on process performance.

Serial No	Key performance indices	Calculation	Significance
1	**Accuracy**	$Accuracy = \frac{true\ positive\ (TP) + true\ negative\ (TN)}{[true\ positive\ (TP) + false\ positive\ (FP) + true\ negative\ (TN) + false\ negative\ (FN)]}$	This metric represents the ratio of correctly predicted observations to the total number of observations.
2	**Specificity**	$Specificity = \frac{\text{true negative (TN)}}{\text{true negative (TN), +false positive (FP)}}$	It measures the proportion of actual negatives that are correctly identified as such by the classifier.
3	**Recall**	$Recall = \frac{\text{true positive (TP)}}{\text{true positive (TP) + false negative (FN)}}$	Also known as, sensitivity, recall is the conditional probability of correctly identifying a fault, given that the sample is faulty.
4	**Precision**	$Precision = \frac{\text{true positive (TP)}}{\text{true positive (TP) + false positive (FP)}}$	Precision indicates the conditional probability of a detected fault being correct.
5	**F1 Score**	$F1 = 2 \times \frac{Precision\ \times Recall}{Precision + recall}$	The F1 score is a measure that balances recall and precision, providing a single metric for evaluating the overall performance of the fault diagnosis system. It is the harmonic mean of recall and precision
6	**G-mean**	$G\ mean = \sqrt{Sensitivity\ \times Specificity}$ Where Specificity is already defined above and $Sensitivity = \frac{\text{true positive (TP)}}{\text{true positive (TP) + false negative (FN)}}$	It provides a balance between the sensitivity (recall for the positive class) and the specificity (recall for the negative class), aiming to maximize both while ensuring that one does not significantly overshadow the other.

Table 7**.** Key Performance Indices for Machine Learning Algorithm

Bottom tray inefficiency causes a decline in bottom product purity and variability in reboiler duty, highlighting the need for efficient operation across all trays for maintaining overall performance.

Figure 9 illustrates the impact of reduced refining section tray efficiency. A decrease in efficiency leads to lower product purities and variability in reboiler duty and reflux flow, stressing the importance of effective monitoring and maintenance of the refining section.

Figure 10 presents the effect of reflux valve stiction on process variables. Valve stiction causes oscillations in product purity, reboiler duty, reflux flow, and product flow rate, demonstrating the challenges posed by valve issues and the need for regular valve maintenance.

Figure 11 shows the effect of reboiler steam valve stiction on the distillation column. Like reflux valve stiction, steam valve stiction causes fluctuations in product purity and energy balance, reinforcing the importance of control strategies and valve maintenance.

These dynamic simulation results provide valuable insights into how various fault conditions impact distillation column performance. The findings emphasize the need for robust control strategies and regular maintenance to ensure process stability and product quality, informing the development of advanced fault detection and diagnosis methods for greater reliability in chemical processes.

3.2. Results of feature engineering

After collecting time series data for 15 parameters (see Table 5), fault numbers were assigned as the output in the last column. The data was then processed using the MRMR algorithm for feature selection, reducing redundancy and enhancing feature relevance. This improved computational efficiency and often boosted classification accuracy. Figure 12 and table 5 highlights the top five dominant features, which were selected for model training. The reduced feature set enabled faster training and inference, making the fault detection system more suitable for real-time use. MRMR proved valuable in maintaining high accuracy while reducing computational load.

3.3. Fault classification accuracy

This section evaluates the fault classification accuracy of various machine learning algorithms using both the full set of 15 parameters and a reduced set of five key parameters selected via MRMR feature engineering. The focus is on the performance of models with the reduced parameters and identifying the top-performing models.

3.3.1. Accuracy with reduced parameters:

Table 8 summarizes Fault Classification Accuracy of different machine learning algorithm. After feature reduction, the Fine Tree model in the Decision Trees category achieved an outstanding accuracy of 99.80 Percent, surpassing the Medium Tree (98.20 Percent) and Coarse Tree (40.10 Percent) models. Discriminant Analysis models, LDA and QDA, performed well, both achieving around 87 Percent accuracy. However, Logistic Regression saw a notable accuracy drop from 95.34 Percent to 79.30 Percent, indicating sensitivity to feature reduction.

Naive Bayes models, especially Gaussian and Kernel, improved to 97.40 Percent. SVM models remained strong, with Quadratic SVM reaching 99.80% and other variants maintaining accuracies above 96 Percent. KNN classifiers performed exceptionally well, with Fine and Weighted KNN models both achieving 99.80 Percent accuracy, while Medium and Cosine KNN models were slightly lower at 99.60 Percent and 99.50 Percent.

Among Ensemble classifiers, the Bagged Trees model achieved the highest accuracy of 99.90 Percent, with Boosted Trees and Subspace k-NN following closely at 99.70 Percent and 98.70 Percent. Neural Networks also showed excellent results, with Medium NN and Wide NN achieving 99.70 Percent and 99.80 Percent, while Narrow and Bilayered NN models exceeded 99 Percent.

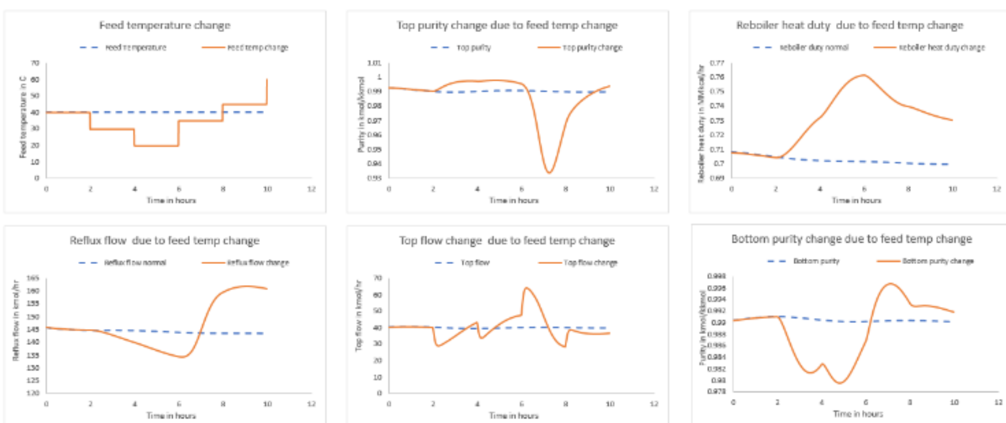

Figure 3. Impact of Feed Temperature Change on Distillation Column Variables (A decrease in feed temperature results in reduced product purity and increased reboiler duty. The control system compensates by increasing reflux flow)

3.3.2. Identification of top models and building a robust stack FDD system

When comparing the average accuracy between the full and reduced parameter sets, performance remained stable or improved for most models. The reduction from 15 to five parameters did not significantly affect accuracy, demonstrating the effectiveness of the MRMR feature selection method.Using the reduced parameter set, the top five models identified are Bagged Trees (99.90 Percent accuracy), Quadratic SVM (99.80 Percent), Fine KNN (99.80 Percent), Wide Neural Network (99.80 Percent), and Fine Tree (99.80 Percent). Figure 13 shows the confusion matrix of Ensemble bagged tree model with 99.9 Percent accuracy. Table 9 shows the Comparison of Classification Metrics for top 5 models.

These models proved to be the most reliable for fault detection in the distillation column. Fresh data from the distillation column is fed into each of the five classifiers, which independently classify the fault. The final fault classification is then determined using a voting mechanism that aggregates the outputs, with the majority vote deciding the result.

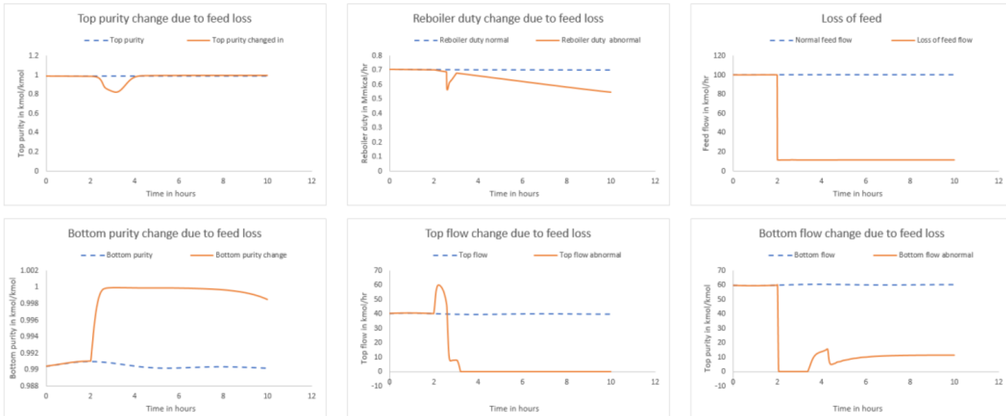

Figure 4. Impact of Feed Loss on Distillation Column Variables (Sudden feed reduction leads to instability in product flow and purity. The system struggles to maintain setpoints, highlighting vulnerability to feed disruptions)

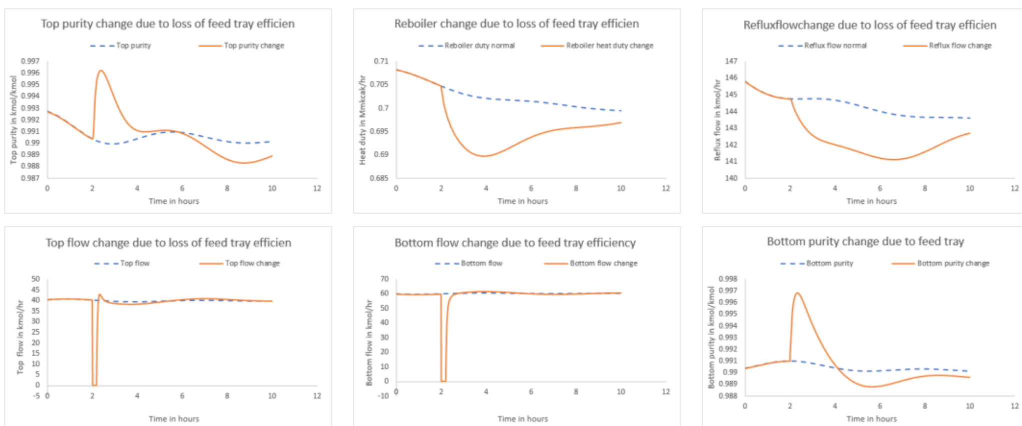

Figure 5. Impact of Feed Tray Efficiency Reduction on Distillation Column Variables (Decreased tray efficiency causes deteriorated separation, visible in reduced product purities and unstable energy demand)

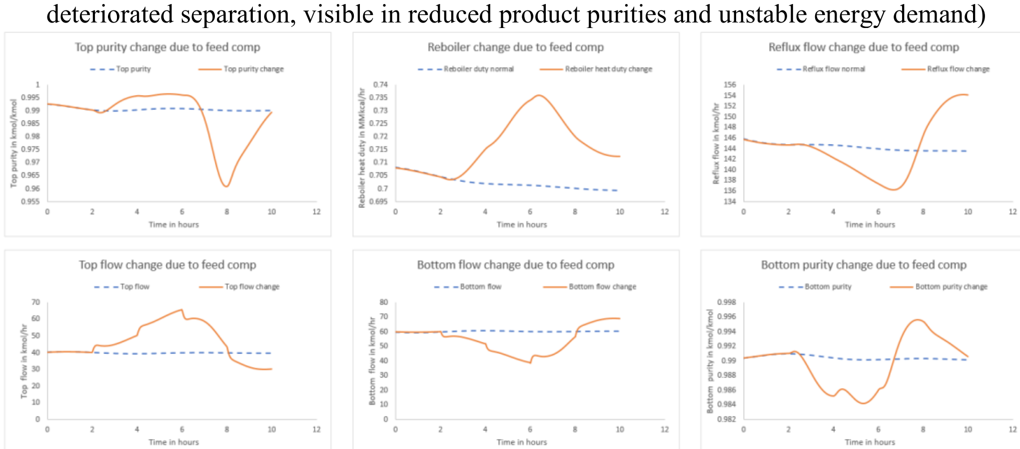

Figure 6. Impact of Feed Composition Change on Distillation Column Variables (Variations in feed composition lead to oscillations in reflux and reboiler duty, showing the sensitivity of the process to raw material changes)

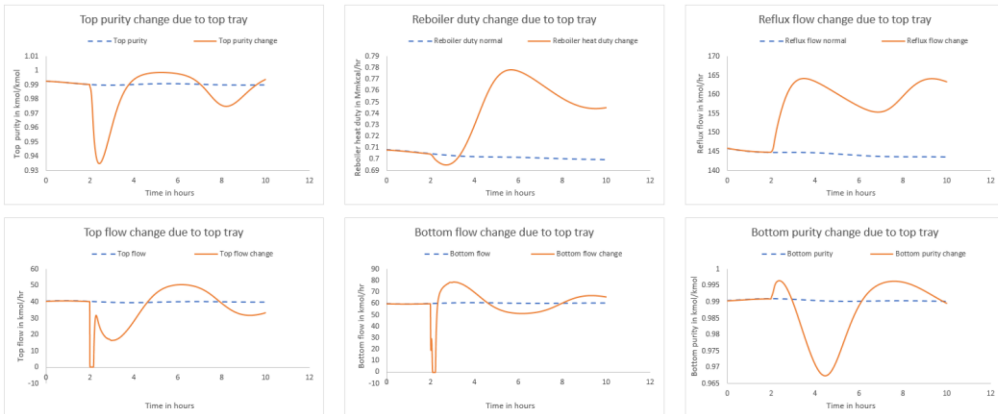

Figure 7. Impact of Top Tray Efficiency Reduction on Distillation Column Variables (Loss of top tray efficiency results in impaired light-end separation, impacting top product purity and increasing energy consumption)

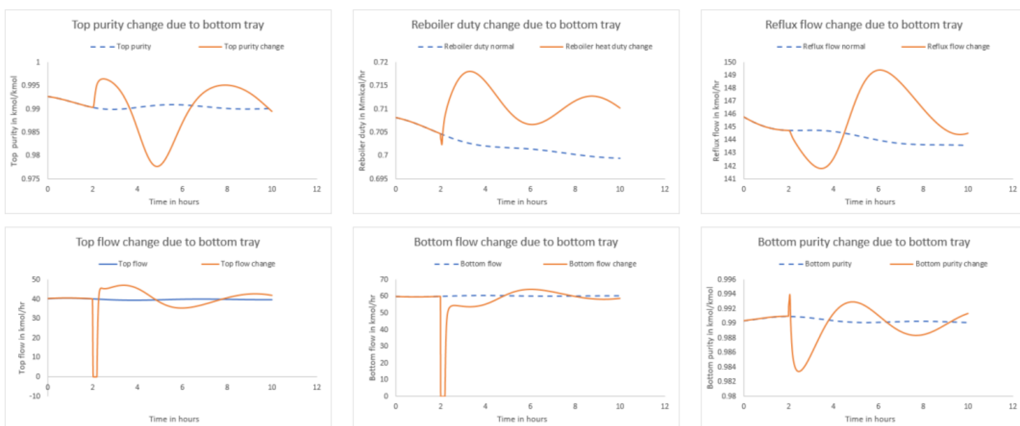

Figure 8. Impact of Bottom Tray Efficiency Reduction on Distillation Column Variables (Bottom tray inefficiency causes variation in bottom product quality and reboiler duty, demonstrating poor stripping performance)

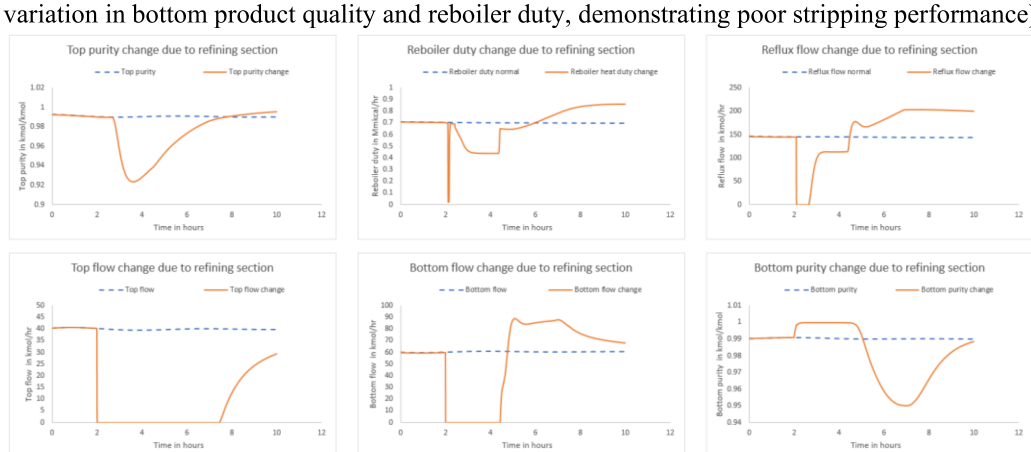

Figure 9**.** Impact of Refining Section Tray Efficiency Reduction on Distillation Column Variables (Global reduction in refining section efficiency impacts overall separation, leading to broad product quality degradation)

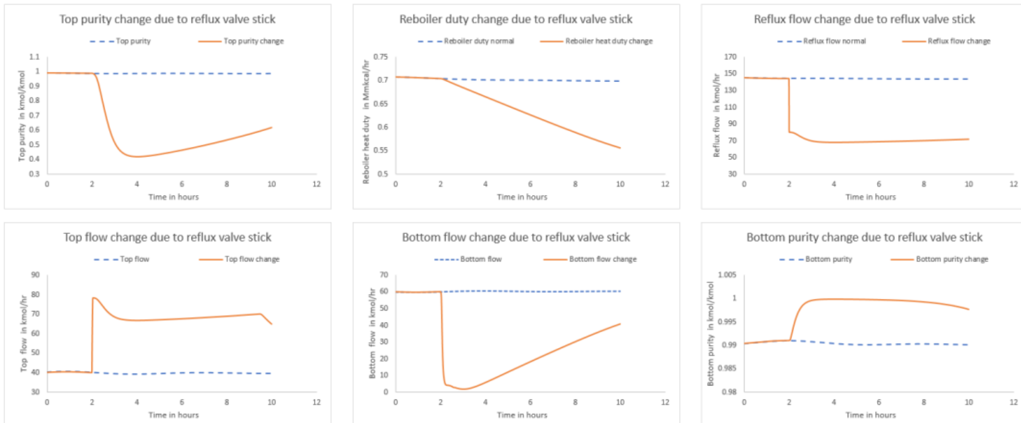

Figure 10**.** Impact of Reflux Valve Stiction on Distillation Column Variables (Reflux flow oscillations caused by stiction lead to fluctuating purities and energy usage, indicating control loop performance degradation)

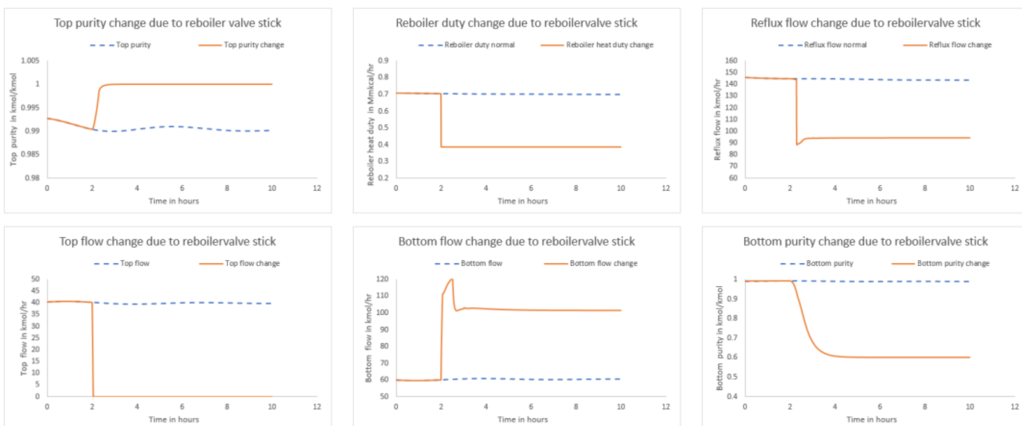

Figure 11. Impact of Reboiler Steam Valve Stiction on Distillation Column Variables (Steam valve stiction induces irregular heat input, seen as product purity swings and reboiler duty instability)

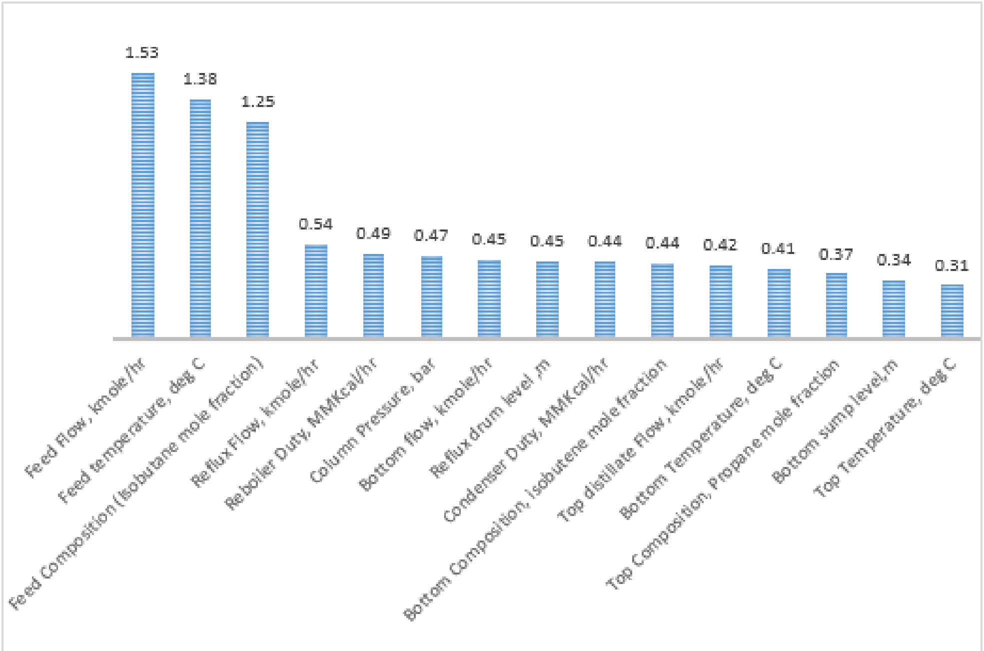

Figure 12**.** Relative Importance of Process Variables in Distillation Column Operation as per MRMR scores (Reboiler duty, feed flow, reflux flow, and feed composition rank highest in fault detection relevance)

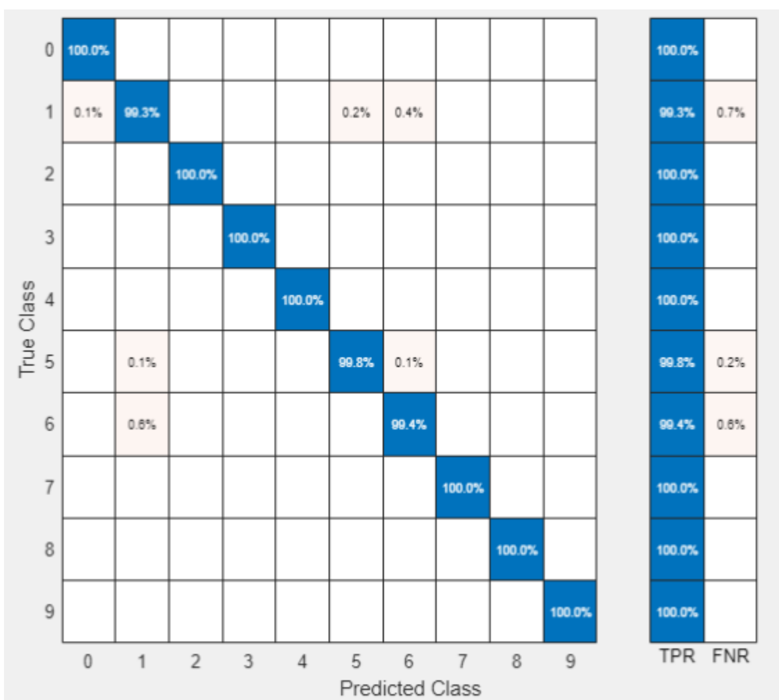

Figure 13. Confusion matrix of ensemble bagged tree model (99.9% accuracy) (Perfect or near-perfect classification accuracy achieved for all fault categories using ensemble-based classifier)

Sr No	Classification Algorithms	Accuracy (%) with all 15 parameters	Accuracy (%) with five parameters
1	**Decision Trees**		
1.1	Fine Tree	99.73	99.80
1.2	Medium Tree	98.69	98.20
1.3	Coarse Tree	47.52	40.10
2	**Discriminant Analysis**		
2.1	Linear Discriminant Analysis (LDA)	84.09	87.00
2.2	Quadratic Discriminant Analysis (QDA)	84.10	87.10
3	**Logistic Regression**		
3.1	Efficient logistic regression	95.34	79.30
4	**Naive Bayes**		
4.1	Gaussian Naïve Bayes	95.16	97.40
4.2	Kernel Naïve Bayes	95.16	97.40
5	**Support Vector Machine (SVM)**		
5.1	linear	98.86	94.50
5.2	Quadratic	99.95	99.80
5.3	Cubic	99.93	99.80
5.4	Fine Gaussian	99.67	99.80
5.5	Medium Gaussian	99.84	99.60
5.6	Coarse Gaussian	99.62	96.60
6	**k-Nearest Neighbour (KNN)**		
6.1	Fine KNN	99.76	99.80
6.2	Medium KNN	99.54	99.60
6.3	Coarse KNN	98.89	98.20
6.4	Cosine KNN	99.55	99.50
6.5	Cubic KNN	99.53	99.60
6.6	Weighted KNN	99.67	99.80
7	**Ensemble Classifier**		
7.1	Boosted trees	99.79	99.70
7.2	Bagged trees	99.89	99.90
7.3	Subspace discriminant	85.88	87.90
7.4	Subspace can	99.65	98.70
7.5	RUS Boosted Tree	98.10	98.20
8	**Neural Network Classification Models**		
8.1	Narrow NN	99.85	99.30
8.2	Medium NN	99.90	99.70
8.3	Wide NN	99.90	99.80
8.4	Bilayered NN	99.83	99.80
8.5	Trilayered NN	99.81	99.70

Table 8. Fault Classification Accuracy of different machine learning algorithm

3.4. Fault classification accuracy

The use of Explainable AI (XAI) methods, specifically LIME and SHAP, provides valuable insights into how machine learning models make predictions. LIME offers faster, localized explanations by simplifying complex models into interpretable components. It is ideal for gaining quick insights into how specific features affect individual predictions. On the other hand, SHAP provides a more theoretically grounded, consistent measure of feature importance across both local and global models. Although SHAP is more computationally intensive, its consistency and reliability across model instances make it crucial for broader insights.

Both LIME and SHAP were implemented in Matrix Laboratory MATLAB to enhance model transparency. As seen from the bar plots (Figures 14 and 15), these methods enabled a detailed understanding of the fault prediction process by highlighting which variables had the greatest influence on the model's decisions. Table 9 summarizes all the insights getting from LIME and SHAP values in figure 14 and 15. This clarity allowed for improved model validation, debugging, and optimization. By providing this level of interpretability, engineers can better trust the decisions made by the machine learning models.

Shapley plots (Figure 14) illustrate the contribution of features to model predictions. Key variables such as *reboiler duty*, *feed flow*, and *reflux flow* consistently appear as the most influential factors in fault scenarios. For instance, in Fault 1 and Fault 2, *reboiler duty* and *feed flow* dominate, highlighting the need for close monitoring of these parameters to prevent faults. Shapley values provide detailed root cause insights, helping engineers take targeted actions to improve reliability and efficiency. The frequent appearance of *reboiler duty* and *feed flow* as dominant predictors across faults shows their critical role in maintaining distillation column performance. Shapley values (Table 9) help identify these key predictors, allowing for focused control and improved fault management.

LIME plots (Figure 15) provide local explanations for fault predictions, with *reboiler duty* consistently emerging as the top predictor across all faults. This reinforces its significance for column stability. While LIME offers quick insights into individual faults, it complements the broader view provided by SHAP. By combining LIME and SHAP, both local and global insights are obtained. LIME offers fast, specific insights, while SHAP ensures consistency across the model. Together, they identify *reboiler duty* as the most critical factor for fault detection, improving process monitoring and optimizing performance.

Metric	**Ensemble Bagged Trees**	**Quadratic SVM**	**Fine KNN**	**Wide Neural Network**	**Fine Tree**
Accuracy	1.000	0.999	0.999	0.999	0.999
Specificity	1.000	1.000	1.000	1.000	1.000
Recall	0.998	0.997	0.997	0.997	0.997
Precision	0.998	0.997	0.997	0.997	0.997
F1 Score	0.998	0.997	0.997	0.997	0.997
G-mean	0.999	0.998	0.998	0.998	0.998

Table 9. Comparison of Classification Metrics for top 5 models

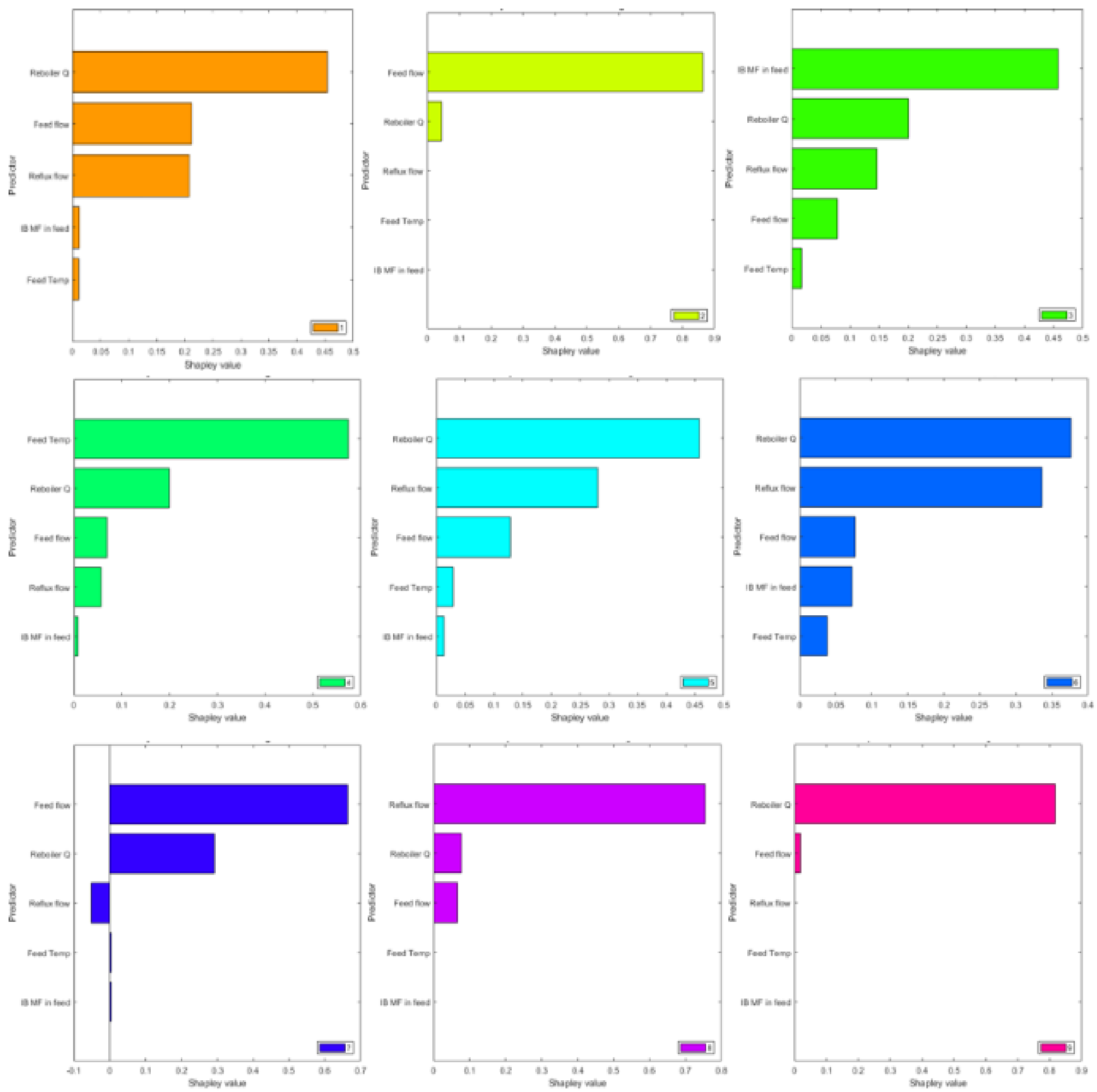

Figure 14. Shapley Value Explanations for Key Predictors across nine (Fault Scenarios: Reboiler duty emerges as the dominant feature across multiple faults, followed by feed flow and reflux flow.) Fault 1 (Orange), Fault 2 (Yellow), Fault 3 (Green), Fault 4 (Light Green), Fault 5 (Cyan), Fault 6 (Blue), Fault 7 (Dark Blue), Fault 8 (Magenta), Fault 9 (Pink)

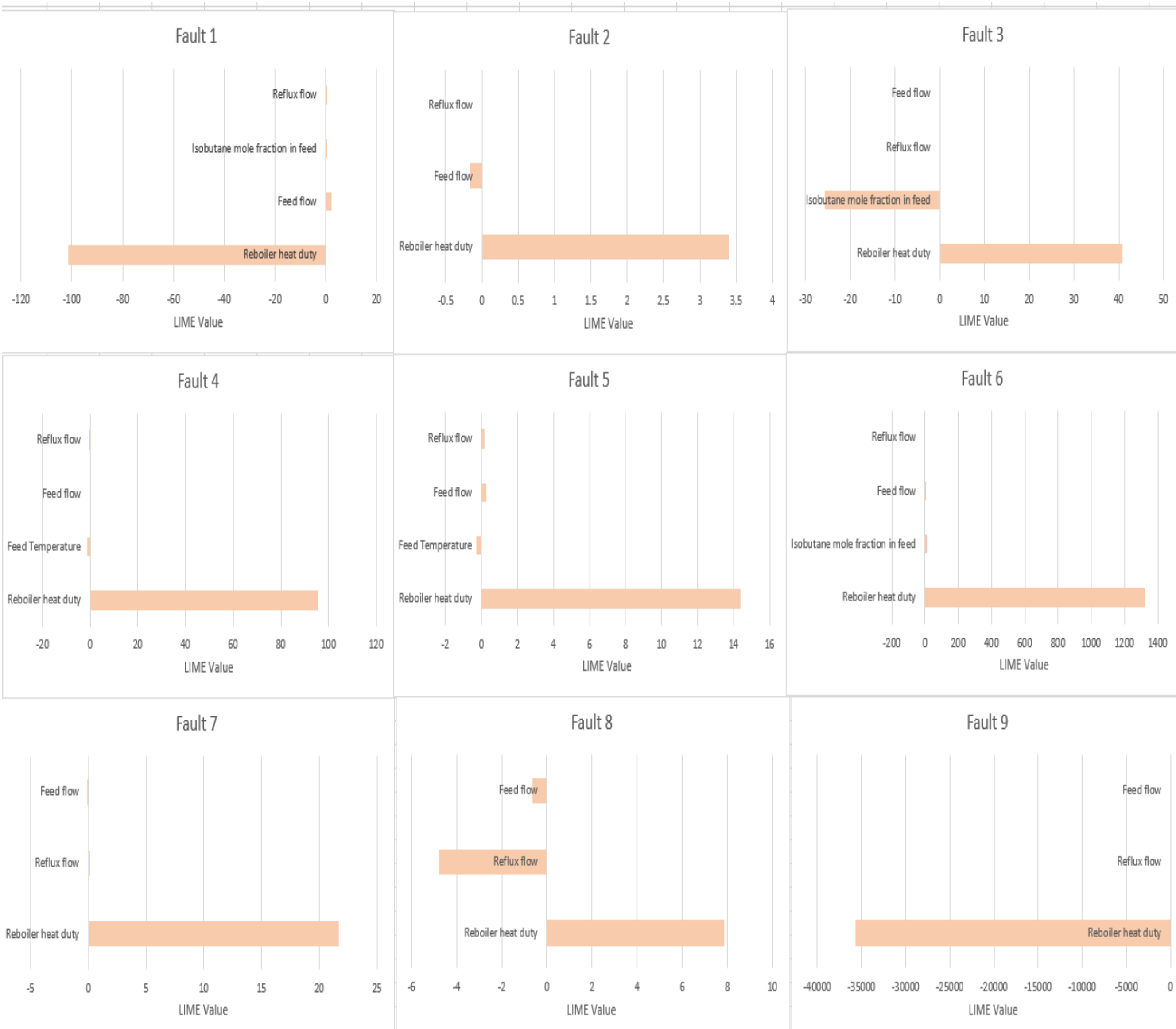

Figure 15. LIME Explanations for Key Predictors across nine Fault Scenarios in Distillation Column Process (LIME confirms reboiler duty as the most influential predictor, providing localized insights aligned with SHAP findings)

Fault	Key Predictors (LIME - in order of influence)	Key Predictors (SHAP - in order of influence)	Summary
Fault 1	1. Reboiler Duty (Q) 2. IB MF in Feed 3. Feed Flow	1. Reboiler Duty(Q)	Reboiler Heat Duty (Q) is the dominant factor, with Isobutane Mole Fraction (IB MF) and Feed Flow contributing less. Both LIME and SHAP confirm the importance of Reboiler Heat Duty (Q).
Fault 2	1. Feed Flow 2. Reboiler Duty	1. Feed Flow 2. Reboiler Duty	Feed Flow plays the primary role in this fault, with Reboiler Heat Duty (Q) as a secondary factor. SHAP aligns with LIME's findings.
Fault 3	1. Reboiler Duty 2. IB MF in Feed	1. Reboiler Duty 2. IB MF in Feed	Both methods highlight Reboiler Heat Duty (Q) and IB MF in feed as critical to the fault.
Fault 4	1. Reboiler Duty 2. Feed Temperature 3. Feed Flow	1. Reboiler Duty	Reboiler heat duty consistently emerges as the most influential predictor with minimal contribution from Feed Temp and Flow. SHAP confirms this.
Fault 5	1. Reboiler Duty 2. Feed Flow 3. Feed Temperature	1. Reboiler Duty	Reboiler Heat Duty (Q) plays the most significant role, while Feed Flow and Temp have lesser impacts.
Fault 6	1. Reboiler Duty 2. IB MF in Feed 3. Feed Flow	1. Reboiler Duty	Reboiler Heat Duty (Q) remains the top predictor, with IB MF and Feed Flow playing lesser roles.
Fault 7	1. Reboiler Duty 2. Feed Flow	1. Reboiler Duty 2. Feed Flow	Reboiler Heat Duty (Q) is the primary predictor, with a smaller contribution from Feed Flow.
Fault 8	1. Reboiler Duty 2. Reflux Flow	1. Reboiler Duty 2. Reflux Flow	Reboiler Heat Duty (Q) and Reflux Flow are the most important factors for this fault, as shown by both methods.
Fault 9	1. Reboiler Duty 2. Feed Flow 3. Reflux Flow	1. Reboiler Duty	Reboiler Heat Duty (Q) is the leading factor, with smaller contributions from Feed Flow and Reflux Flow. SHAP supports this.

Table 10. Key Predictors for Fault Scenarios identified by LIME and SHAP in Order of Influence

4. CONCLUSION

This study outlines a comprehensive approach to developing a Fault Detection and Diagnosis (FDD) system for a distillation column using machine learning. By creating dynamic simulations and applying data pre-processing, feature engineering with the MRMR algorithm, and multiple machine learning models, accurate fault detection and diagnosis were achieved.

The results show that combining algorithms like Decision Trees, Discriminant Analysis, Logistic Regression, SVM, k-nearest Neighbours, Ensemble Methods, Naive Bayes, and Neural Networks significantly enhances system reliability. Implementing a stacked classifier system further improved fault detection accuracy and stability.

Incorporating Explainable AI (XAI) techniques, such as LIME and SHAP, enhanced the interpretability of the models, enabling plant engineers to understand and trust the system's decisions, facilitating timely corrective actions. This study underscores the value of combining simulation-based data with advanced machine learning and XAI techniques to build robust FDD systems, improving safety, efficiency, and reducing downtime. The novelty of this paper lies in integrating dynamic simulations for fault data generation, applying advanced ML algorithms for

FDD, and using XAI techniques to enhance model interpretability. This approach offers an effective solution for fault diagnosis in distillation columns, with future work focused on integrating ML models, historical data, and XAI for even better fault detection.

Although the simulated faults are primarily step-type, they reflect several common real-world operational disruptions such as sudden valve malfunction, tray damage, or feed composition jumps. Expanding the dataset to include gradual degradation profiles observed in actual plant history would further enhance system generalization and industrial relevance.

Future work could explore additional algorithms and refine the stacked classifier system. Expanding this approach to other real industrial processes would increase its applicability and advance fault detection technologies across sectors.

REFERENCES

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175–185. doi.org/10.1080/00031305.1992.10475879

Amiruddin, A. A. A. M., Zabiri, H., Taqvi, S. A. A., & Tufa, L. D. (2020). Fault detection and diagnosis in chemical processes. *Neural Computing and Applications*, *32*, 447–472. doi.org/10.1007/s00521-019-04167-8

Andre, A. B., Beltrame, E., & Wainer, J. (2013). Fault detection and diagnosis in chemical processes. *Applied ArtificialIntelligence*,*27*(1),3649.doi.org/10.1080/08839514.2013.748914

Aydin, I., Karakose, M., & Akin, E. (2014). Fault detection and diagnosis in chemical processes. *ISA Transactions*, *53*(2),220–229.doi.org/10.1016/j.isatra.2013.10.007

Chetouani, Y. (2007). Fault detection and diagnosis in chemical processes. *International Journal of Computer Science & Applications*, *4*(3), 119–133.

Chetouani, Y. (2011). Fault detection and diagnosis in chemical processes. *Systems Engineering Procedia*, *1*, 473–480. doi.org/10.1016/j.sepro.2011.08.072

Chiang, L. H., & Pell, R. J. (2004). Fault detection and diagnosis in chemical processes. *Journal of Process Control*,*14*(2),143-155. doi.org/10.1016/j.jprocont.2003.09.003

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2000). *Fault detection and diagnosis in industrial systems*. Springer. doi.org/10.1007/978-1-4471-0409-4

Copertaro, E., Chiariotti, P., Revel, G. M., & Paone, N. (2019). Fault detection and diagnosis in chemical processes. *Journal of Process Control*, *80*, 180–192. doi.org/10.1016/j.jprocont.2019.06.001

Dash, S., Rengaswamy, R., & Venkatasubramanian, V. (2003). Fuzzy-logic based trend classification for fault diagnosis of chemical processes. *Computers & Chemical Engineering*, *27*(3), 347–362. doi.org/10.1016/S0098-1354(02)00214-4

Demetgul, M. (2013). Fault detection and diagnosis in chemical processes. *International Journal of Advanced Manufacturing Technology*, *67*(9–12), 2183–2194. /doi.org/10.1007/s00170-012-4635-4

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Fault detection and diagnosis in chemical processes. *IEEE Access*,*5*,2059020616. doi.org/10.1109/ACCESS.2017.2756879

Ge, Z., Song, Z., & Gao, F. (2013). Fault detection and diagnosis in chemical processes. *Industrial & Engineering Chemistry Research*, *52*(10), 3543–3562. doi.org/10.1021/ie302071k

Ge, Z., Zhong, S., & Zhang, Y. (2016). Fault detection and diagnosis in chemical processes. *IEEE Transactions on Industrial Informatics*, *12*(4), 1403–1411. doi.org/10.1109/TII.2016.2552539

Gonzaga, J. C. B., Meleiro, L. a. C., Kiang, C., & Maciel Filho, R. (2009). ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. *Computers & Chemical Engineering*,*33*(1),43–49. doi.org/10.1016/j.compchemeng.2008.05.019

Hasan, M. J., & Kim, J.-M. (2019). Fault detection and diagnosis in chemical processes. *Energies*, *12*(6), 991. doi.org/10.3390/en12060991

He, S.-G., He, Z., & Wang, G. A. (2013). Fault detection and diagnosis in chemical processes. *Journal of Intelligent Manufacturing*, *24*(1), 25–34 . doi.org/10.1007/s10845-011-0541-4

Hoskins, J., Kaliyur, K., & Himmelblau, D. M. (1991). Fault detection and diagnosis in chemical processes. *AIChE Journal*,*37*(1), 137–141. doi.org/10.1002/aic.690370114

Hwang, B., Saif, M., & Jamshidi, M. (1993). Fault detection and diagnosis in chemical processes. *IFAC Proceedings*, *26*(2), 547550. doi.org/10.1016/S14746670(17)47358-1

Iliyas, S. A., Elshafei, M., Habib, M. A., & Adeniran, A. A. (2013). Fault detection and diagnosis in chemical processes. *Control Engineering Practice*, *21*(7), 962–970. doi.org/10.1016/j.conengprac.2013.03.004

Isermann, R. (2006). *Fault-diagnosis systems: An introduction from fault detection to fault tolerance*. Springer Science & Business Media. doi.org/10.1007/3-540-30368-5

Isermann, R., & Balle, P. (1997). Fault detection and diagnosis in chemical processes. *Control Engineering Practice*, *5*(5), 709–719. doi.org/10.1016/S0967-0661(97)00053-1

Jamil, M., Kalam, A., Ansari, A., & Rizwan, M. (2014). Fault detection and diagnosis in chemical processes. *Applied SoftComputing*,*19*,322–332. doi.org/10.1016/j.asoc.2014.02.003

Jiang, L., Lei, X., & Shuqing, W. (2006). Fault detection and diagnosis in chemical processes. *Chinese Journal of Chemical Engineering*, *14*(3), 343–348. doi.org/10.1016/S1004-9541(06)60072-8

Karabadji, N. E. I., Seridi, H., Khelf, I., Azizi, N., & Boulkroune, R. (2014). Fault detection and diagnosis in chemical processes. *Engineering Applications of ArtificialIntelligence*,*35*,71–83. doi.org/10.1016/j.engappai.2014.06.001

Kesavan, P., & Lee, J. H. (1997). Fault detection and diagnosis in chemical processes. *Industrial & Engineering Chemistry Research*, *36*(7), 2725–2738. doi.org/10.1021/ie960768m

Kuo, Y., & Lin, K.-P. (2010). Fault detection and diagnosis in chemical processes. *International Journal of Advanced Manufacturing Technology*, *50*(9–12), 1243–1251. doi.org/10.1007/s00170-010-2585-6

Lee, J. M., Yoo, C. K., Choi, S. W., Vanrolleghem, P. A., & Lee, I. B. (2004). Nonlinear process monitoring using kernel principal component analysis .*Chemical Engineering Science*, *59*(1), 223–234. doi.org/10.1016/j.ces.2003.09.022

Lee, M. W., Joung, J. Y., Lee, D. S., Park, J. M., & Woo, S. H. (2005). Fault detection and diagnosis in chemical processes. *Industrial & Engineering Chemistry Research*, *44*(11), 3973–3982. doi.org/10.1021/ie0489276

Li, W., Yue, H. H., Valle, S., Qin, S. J., & McAvoy, T. J. (2000). Recursive PCA for adaptive process monitoring. *Journal of Process Control*, *10*(5) ,471–486. doi.org/10.1016/S09591524(00)00022-6

Li, Y., & Zhang, X. (2014). Fault detection and diagnosis in chemical processes. *Chemometrics and Intelligent Laboratory Systems*, *136*, 47–57. doi.org/10.1016/j.chemolab.2014.04.005

Liu, D., Shang, J., & Chen, M. (2020). Fault detection and diagnosis in chemical processes. *IEEE Transactions on Industrial Informatics*. doi.org/10.1109/TII.2020.3031496

Liu, T.-I., & Jolley, B. (2015). Fault detection and diagnosis in chemical processes. *International Journal of Advanced Manufacturing Technology*, *78*(9–12), 1999–2007. doi.org/10.1007/s00170-014-6765-7

Lu, W.-Z., & Xue, Y. (2014). Fault detection and diagnosis in chemical processes. *Building and Environment*, *78*, 111–117. doi.org/10.1016/j.buildenv.2014.04.014

Ma, C. Y., & Wang, X. Z. (2009). Fault detection and diagnosis in chemical processes. *Computers & Chemical Engineering*, *33*(10), 1602–1616. doi.org/10.1016/j.compchemeng.2009.05.004

MacGregor, J. F., & Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, *3*(3), 403–414. doi.org/10.1016/0967-0661(95)00014-L

Manssouri, I., Chetouani, Y., & Kihel, B. E. (2008). Fault detection and diagnosis in chemical processes. *International Journal of Computer Applications in Technology*, *32*(3), 181–186. doi.org/10.1504/IJCAT.2008.020953

Maurya, M. R., Rengaswamy, R., & Venkatasubramanian, V. (2007). Fault diagnosis using dynamic simulation and neural networks. *Journal of Process Control*, *17*(3), 272–289. doi.org/10.1016/j.jprocont.2006.10.004

Meng, Y., Lan, Q., Qin, J., Yu, S., Pang, H., & Zheng, K. (2019). Fault detection and diagnosis in chemical processes .*Journalof Food Engineering*, *241*, 159165. doi.org/10.1016/j.jfoodeng.2018.08.022

Molnar,C.(2019).*InterpretableMachine Learning*. Lulu.com.

Mujtaba, S. M., Lemma, T. A., Taqvi, S. A. A., Ofei, T. N., & Vandrangi, S. K. (2020). Fault detection and diagnosis in chemical processes. *Processes*, *8*(4), 474. doi.org/10.3390/pr8040474

Nagpal, T., & Brar, Y. S. (2014). Fault detection and diagnosis in chemical processes. *Neural Computing and Applications*, *25*(7–8), 1863–1870. /doi.org/10.1007/s00521-014-1658-3

Nawaz, M., Maulud, A. S., Zabiri, H., Taqvi, S. A. A., & Idris, A. (2020). Fault detection and diagnosis in chemical processes. *Chinese Journal of Chemical Engineering*. doi.org/10.1016/j.cjche.2020.08.035

Nguyen, N.-T., & Lee, H.-H. (2010). Fault detection and diagnosis in chemical processes. *International Journal of Electronics*, *97*(9), 1071–1081. doi.org/10.1080/00207217.2010.496673

Nor, N. M., Hassan, C. R. C., & Hussain, M. A. (2020). Fault detection and diagnosis in chemical processes. *Reviews in Chemical Engineering*, *36*(4), 513–553. doi.org/10.1515/revce-2019-0052

Nozari, H. A., Shoorehdeli, M. A., Simani, S., & Banadaki, H. D. (2012). Fault detection and diagnosis in chemical processes .*Neuro computing*, *91*, 29–47. doi.org/10.1016/j.neucom.2012.02.016

Pani, A. K., & Mohanta, H. K. (2015). Fault detection and diagnosis in chemical processes. *ISA Transactions*, *56*, 206–221. doi.org/10.1016/j.isatra.2014.10.004

Quinlan, J. R. (1979). *In Expert Systems in the Micro Electronics Age*. Edinburgh University Press.

Quinlan, J. R. (2014). *C4.5: Programs for Machine Learning*. Elsevier.

Rajakarunakaran, S., Venkumar, P., Devaraj, D., & Rao, K. S. P. (2008). Fault detection and diagnosis in chemical processes. *Applied Soft Computing*, *8*(1), 740–748. doi.org/10.1016/j.asoc.2007.07.002

Rengaswamy, R., & Venkatasubramanian, V. (2001). A syntactic pattern-recognition approach for process monitoring and fault diagnosis. *Engineering*

Applications of Artificial Intelligence, *14*(6), 785–796. doi.org/10.1016/S0952-1976(01)00030-8

Schubert, U., Kruger, U., Arellano-Garcia, H., de S. á Feital, T., & Wozny, G. (2011). Fault detection and diagnosis in chemicalprocesses.*Control Engineering Practice*, *19*(5),479–490. doi.org/10.1016/j.conengprac.2011.02.002

Shaik, N. B., Pedapati, S. R., Taqvi, S. A. A., Othman, A., & Dzubir, F. A. A. (2020). Fault detection and diagnosis in chemicalprocesses.*Processes*,*8*(6),661. doi.org/10.3390/pr8060661

Sharma, R., Singh, K., Singhal, D., & Ghosh, R. (2004). Fault detection and diagnosis in chemical processes. *Chemical Engineering and Processing: Process Intensification*, *43*(7), 841847. doi.org/10.1016/S02552701(03)00103-X

Taqvi, S. A. A., Zabiri, H., Tufa, L. D., Fatima, S. A., & Maulud, A. H. S. (2020). Distillation column: Review of major disturbances. *Proceedings of the 9th IEEE International Conference on Control System, Computing andEngineering*,1–6. doi.org/10.1109/ICCSCE50387.2020.9274987

Taqvi, S. A. A., Zabiri, H., Tufa, L. D., Uddin, F., Fatima, S. A., & Maulud, A. S.(2020). Fault detection and diagnosis in chemical processes. *Jurnal Teknologi*, *82*(5), 14350. doi.org/10.11113/jt.v82.14350

Taqvi, S. A., Tufa, L. D., Zabiri, H., Mahadzir, S., Maulud, A. S., & Uddin, F. (2017). Artificial neural network for anomalies detection in distillation column. In D. Modeling & S. of Systems Springer (Eds.), *Singapore* (pp. 447–472). Springer. doi.org/10.1007/978-981-10-6463-0_26

Taqvi, S. A., Tufa, L. D., Zabiri, H., Maulud, A. S., & Uddin, F. (2018a). Fault detection and diagnosis in chemical processes. *Neural Computing and Applications*, *32*, 3503–3519. doi.org/10.1007/s00521-018-3593-7

Taqvi, S. A., Tufa, L. D., Zabiri, H., Maulud, A. S., & Uddin, F.(2018b).Multiple Fault Diagnosis in Distillation Column Using Multikernel Support Vector Machine. *Industrial and Engineering Chemistry Research*, *57*(43), 14689–14706. doi.org/10.1021/acs.iecr.8b03360

Tidriri, K., Chatti, N., Verron, S., & Tiplica, T. (2016). Fault detection and diagnosis in chemical processes. *Annual Reviews in Control*, *42*, 63–81. doi.org/10.1016/j.arcontrol.2016.09.001

Tudón-Martínez, J. C., & Morales-Menendez, R. (2015). Fault detection and diagnosis in chemical processes. *IFAC-PapersOnLine*,*48*(21),1210–1215. doi.org/10.1016/j.ifacol.2015.09.372

Ungar, L., Powell, B., & Kamens, S. (1990). Fault detection and diagnosis in chemical processes. *Computers & ChemicalEngineering*,*14*(4–5),561–572. /doi.org/10.1016/0098-1354(90)87045-8

Vedam, H., Dash, S., & Venkatasubramanian, V. (1999). Fault detection and diagnosis in chemical processes. *Computers & Chemical Engineering*, *23*, S577–S580. doi.org/10.1016/S0098-1354(99)80097-1

Venkatasubramanian, V., & Chan, K. (1989). Fault detection and diagnosis in chemical processes. *AIChE Journal*, *35*(12), 1993–2002. doi.org/10.1002/aic.690351214

Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003). A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Computers & Chemical Engineering*, *27*(3), 313–326. doi.org/10.1016/S0098-1354(02)00161-8

Venkatasubramanian, V., Rengaswamy, R., & Yin, K. (2003). *A review of process fault detection and diagnosis Part I : Quantitative model-based methods*. *27*, 293–311.

Watanabe, K., Matsuura, I., Abe, M., Kubota, M., & Himmelblau, D. (1989). Fault detection and diagnosis in chemical processes. *AIChE Journal*, *35*(11), 1803–1812. doi.org/10.1002/aic.690351111

Weber, P., Theilliol, D., Aubrun, C., & Evsukoff, A. (2006). Fault detection and diagnosis in chemical processes. *IFAC Proceedings*, *39*(13), 90–95. doi.org/10.3182/20060821-3-FR-2920.00015

Xu, S., & Liu, X. (2014). Fault detection and diagnosis in chemical processes. *Neurocomputing*, *142*, 291–298. doi.org/10.1016/j.neucom.2014.04.046

Xu, Z., Weiwu, Y., & Huihe, S. (2006). Fault detection and diagnosis in chemical processes. *Chinese Journal of Chemical Engineering*, *14*(6), 759–764. doi.org/10.1016/S1004-9541(07)60007-1

Yu, J., & Zhang, C. (2020). Fault detection and diagnosis in chemical processes. *Journal of Process Control*, *92*, 119–136. doi.org/10.1016/j.jprocont.2020.03.001

Zhang, J., Li, J., & Wang, Y. (2019). A review on the research and applications of machine learning in fault diagnosis and detection for chemical process. *Processes*, *7*(12), 839. doi.org/10.3390/pr7120839

Zhang, S., Zhao, C., Wang, S., & Wang, F. (2017). Fault detection and diagnosis in chemical processes. *Industrial & Engineering Chemistry Research*, *56*(3), 728–740. doi.org/10.1021/acs.iecr.6b03845

Zhang, X., & Hoo, K. A. (2011). Fault detection and diagnosis in chemical processes. *Computers & Chemical Engineering*,*35*(1),132–148. doi.org/10.1016/j.compchemeng.2010.09.004

Zhang, X., Yan, W. W., Zhao, X., & Shao, H. H. (2007). Fault detection and diagnosis in chemical processes. *Chemical Engineering & Technology*, *30*(9), 1203–1211. doi.org/10.1002/ceat.200700140

Zhao, Y., Xiao, F., & Wang, S. (2013). Fault detection and diagnosis in chemical processes. *Energy and Buildings*, *57*, 278–288. doi.org/10.1016/j.enbuild.2012.11.008

Zhou, J., Guo, A., Celler, B., & Su, S. (2014). Fault detection and diagnosis in chemical processes. *Applied Soft Computing*,*14*,4–11. doi.org/10.1016/j.asoc.2013.12.001

Zhu, Z.-B., & Song, Z.-H. (2010). Fault detection and diagnosis in chemical processes. *Chemical Engineering Research and Design*, *88*(8), 936–951. doi.org/10.1016/j.cherd.2010.01.019

Zhu, Z.-B., & Song, Z.-H. (2011). Fault detection and diagnosis in chemical processes. *Expert Systems with Applications*,*38*(6),6895–6905. doi.org/10.1016/j.eswa.2010.12.038

Zio, E., Baraldi, P., & Popescu, I. C. (2009). Fault detection and diagnosis in chemical processes. *Annals of Nuclear Energy*, *36*(8),1159–1169. doi.org/10.1016/j.anucene.2009.03.004