Journal of Environmental Informatics Letters 14(2) 124-134 (2025)

Journal of Environmental Informatics Letters

www.iseis.org/jeil

Resilience Engineering Framework for Crude Oil Refineries

S. R. Dhanushkodi[1] *, J. C. Zaveri[1], A. Srinivasan[2], and N. Mahinpey[3]

[1] Dhanushkodi Research group, Department of Chemical Engineering, Vellore Institute of Technology, Tamil Nadu 632014, India
[2] Boost Environmental Systems Inc., Vancouver, British Columbia, Canada
[3] Department of Chemical Engineering, University of Calgary, Calgary, Alberta T2N 1N4, Canada

Received 05 July 2025; revised 17 November 2025; accepted 23 December 2025; published online 31 December 2025

ABSTRACT. Crude oil refineries generate vast, multidimensional data streams that are critical for maintaining real-time operational control and ensuring compliance with environmental regulations. Distillation columns, in particular, produce intricate and evolving datasets that create significant challenges for early failure detection, escalating the risk of environmentally harmful incidents. To effectively tackle this pressing issue, we advocate for a robust resilience-engineering framework that seamlessly integrates the Lagrangian Support Vector Machine (LSVM) with the Functional Resonance Analysis Method (FRAM) within cutting-edge environmental informatics architecture. The LSVM component rigorously analyzes refinery operational data — such as pressure, temperature, flow rates, and reflux ratios using an 80:20 train-test split and consistently achieves a remarkable 96% accuracy rate in classifying failure events. Beyond merely identifying anomalies, this environmental informatics approach leverages additional data sources, including alarm logs and weather information, to enable proactive predictive analytics for forecasting environmental risks. The FRAM operator then decisively evaluates the anomalies identified by the LSVM through functional resonance modeling, which effectively traces system-level variability and uncovers potential cascading failure pathways, such as tray flooding, over-pressurization, or off-spec discharges. This integrative approach not only enhances interpretability and facilitates early intervention, but it also fortifies environmental protection by linking data-driven fault detection with actionable decision-making insights for proactive risk mitigation. Ultimately, this proposed framework firmly advances operational resilience and reinforces sustainability objectives while ensuring unwavering regulatory compliance within refinery processes.

Keywords: failure analysis, functional resonance analysis method, distillation column, support vector machine

1. Introduction

Refineries around the world are transforming through the power of industrial automation. Distillation columns are indispensable in these operations, performing critical roles in component separation. However, real-time monitoring and control of distillates present challenges due to the nonlinear, multivariable, and dynamic nature of the processes involved. The purity of both top and bottom products is significantly influenced by operational parameters such as reflux ratio, boil-up rate, and feed composition. As a result, precise and timely calculations of these internal states are paramount for optimizing the performance and efficiency of distillation columns. In the pursuit of digitalization and automation, refineries face significant hurdles, including aging infrastructure and a shortage of skilled operators. These challenges have led to frequent failures and operational issues in distillation columns, directly impacting product quality. Additionally, factors such as thermal coupling, heat integration, and gravitational effects further challenge system efficiency. Operational instabilities like flooding and weeping also complicate both operations and maintenance. To maximize productivity in distilleries, it is imperative to extend the time between plant turnarounds. Maintaining distillation columns during these critical periods is not just challenging. Issues such as thermal coupling between the pre-fractionator and the main column, the necessity for improved heat integration, high gravity field requirements, and the recycling of unused reactants all play a vital role in enhancing distillation efficiency. Given the complexities of these operations, it's clear that managing challenges such as flooding, weeping, or over-pressurization is crucial to maintaining optimal column performance.

Historically, fault detection methods like Fault Tree Analysis (FTA) and Event Tree Analysis (ETA) have been employed to evaluate safety and conduct post-incident investigations in distillation columns (Besnard and Hollnagel, 2014; Sarbaye et al., 2019; Saghir et al., 2024; Wang et al., 2024; Pani and Soofastaei, 2025; Wan et al., 2025). While these methods can be effective in failure analysis, they fall short in dynamic, data-rich environments. Their static and predefined structures are not equipped to capture the nonlinear and time-varying behaviors that characterize modern distillation processes. Modern

* Corresponding author. Tel.: +919626845903.
E-mail address: shankarraman.d@vit.ac.in (S. R. Dhanushkodi).

ISSN: 2663-6859 print/2663-6867 online
 doi:10.3808/jeil.202400167

Distributed Control Systems (DCS) in refineries generate massive amount of datasets, which include real-time measurements of temperature, pressure, reflux ratio, and feed composition. These data are useful in monitoring and controlling the distillation process, but identifying and extracting insights from such large-scale datasets remains a challenge for engineers. Often noisy and inconsistent data led to sensor faults, pump failure, or accelerated flooding. In the past, operators applied heuristic strategies to interpret the data. Aging infrastructure and complex operational conditions affect perception-driven decisions on daily base events in the distillation column. Traditional physicochemical models, which rely on first-principles equations, also fail to capture the dynamic and complex nature of modern distillation columns, particularly in real-time monitoring scenarios. Both perception driven and model-based execution leads to biases and errors in distillation columns. Understanding the failure events on the flow characteristics in a column is significant in the refinery operation. No physicochemical distillation column models are accurate in simultaneous capturing of the hydrodynamic charac-teristics and the column failure events (Saghir et al., 2024; Jusoh et al., 2025; Kumar et al., 2025; Pani and Soofastaei, 2025). Therefore, implementing a data-driven decision-making approach would be a good option to minimize operator biases. But it requires learning from the daily base events, recording the sequence of the events, and recognizing the pattern associated with the process. Meticulous assessment of the entire dataset collected in the database using a data classifier and the supervised learning models can be the first step towards the development (Zheng et al., 2024)

Resilience engineering (RE) is a new concept used in modern industries to assess the failure in the process equipment (Bartlet and Hurdle, 2009) It uses supervised learning algorithms to recognize the failure patterns and machine learning models to predict the performance of the system. Although the information related to the failure events could be either linear or nonlinear, RE has the intrinsic ability to absorb, avoid, evade, and detect the failure events. It is a rule-based model and must be created based on one of the following algorithms: SVM, decision tree, and deep learning. Since demystifying the parameters related to failure events in the large data domain is not a direct approach, the SVM that uses a Lagrangian approach is efficient to capture, store and visualize the volumes of data collected in the industry. First, it assesses and discards the imbalances and inaccuracies in the dataset using its algorithm. A distributed file system is created to test its algorithm to report if any deformation is observed in the dataset continuously. Therefore, it can be used as a diagnostic approach to identify the root cause of the failure, evaluate the structural deterioration in the dataset, figure out communication error-induced system failures, and sense the transducer faults from the operating room. It is utilized as a mathematical tool to find the operational hazards and identify the faults in semi-automated industries (Bartlett and Kelly, 2009; Kumar et al., 2025; Qi et al., 2025; Samanta and Bhattcherjee, 2025). To implement the action plans for industry 4.0, a method that can treat successful events as the flip side of failure events is needed to depict accurate details on the failure and day-to-day events. The functional resonance analysis method (FRAM) model is one such technique that considers both past and possible future events in its approach. It is a bi-directional method that differentiates the accident investigation from the risk assessment. The model has a hexagon that assumes that the failures and successes are from the same origin; daily operation details get adjusted to match the analysis conditions. The model identifies functions, variability in the present performance, and strategies for how variability could be combined and improved. The output from the hexagon-based approach is the emergent basis, not the resultant (Motamedzade et al., 2016)

Fault detection and diagnosis (FDD) in complex industrial processes, such as distillation, has been studied via data science-driven, knowledge-driven, and hybrid models. Traditional machine-learning techniques, including Support Vector Machines (SVM), multi-kernel SVM (MK-SVM), Principal Component Analysis (PCA), Partial Least Squares (PLS), and neural-network-based models, have shown robust performance in classifying abnormal process dynamics and predicting respective equipment failures. Specifically, in distillation columns, SVM have been applied for process monitoring, figuring out faulty tray temperatures, process pressure deviations, and anomalies in reflux ratio, underscoring the model fitment of SVM for high-dimensional, nonlinear industrial data. Although, the conventional SVM methods encounter scalability constraints with large datasets, lack transparency in root-cause interpretation remains challenging. The Lagrangian Support Vector Machine (LSVM) addresses these limitations by reformulating training through a linear system rather than quadratic optimization, thereby enabling significantly faster computation and potential for real-time integration. Concurrently, resilience-engineering methodologies have been accepted for elucidating systemic failures, particularly those arising from interactions amongst human, organizational, and process functions. The Functional Resonance Analysis Method (FRAM) has been applied to domains such as healthcare, aviation, transportation, and nuclear safety, demonstrating strong capability in modeling functional variability and emergent behavior in socio-technical systems. However, existing FRAM applications are predominantly retrospective, with limited deployment in continuous process industries and almost no reported frameworks integrating FRAM with real-time machine-learning-based detection systems. To the best of our knowledge, no published work has combined LSVM with FRAM for early failure detection and operational resilience analysis in crude-oil distillation columns. The novelty of this research lies in bridging quantitative LSVM-based anomaly detection with qualitative FRAM-based systemic explanation, enabling both improved detection performance and enhanced interpretability for operators and safety teams. This integrated framework supports resilience-based decision-making and offers practical applicability to refinery operations where process abnormalities propagate through tightly coupled technical and human workflows.

Concurrently, resilience-engineering methodologies have been accepted for elucidating systemic failures, particularly those arising from interactions amongst human, organizational, and process functions. The FRAM has been applied to domains such as healthcare, aviation, transportation, and nuclear safety, demonstrating strong capability in modeling functional variability and emergent behavior in socio-technical systems. However, existing FRAM applications are predominantly retrospective, with limited deployment in continuous process industries and almost no reported frameworks integrating FRAM with real-time machine-learning-based detection systems. To the best of our knowledge, no published work has combined LSVM with FRAM for early failure detection and operational resilience analysis in crude-oil distillation columns. Integrating environmental informatics into the SVM+FRAM framework significantly strengthens its capability to monitor environmental hazards. By incorporating failure event data, the framework not only enhances operational environmental safety but also ensures alignment with broader environmental sustainability objectives. This integration allows for real-time tracking of potential risks, such as chemical spills, hazardous emissions, resource overconsumption, and waste generation, providing operators with actionable insights to prevent these risks before they escalate. Furthermore, the use of environmental informatics allows for a more comprehensive risk assessment by analyzing the interaction between operational processes and environmental factors.

Safety in industry 4.0 demands focusing on the events that happen every minute. To meet this high demand, the probability of ignorance when the plant is in operation must be reduced. By combining the SVM with FRAM methods in distillation column could provide a novel approach to (a) assess the characteristic functions of the operation, (b) monitor the system response when operating guidelines are violated, (c) implement the big data to study the failure event, and (d) anticipate the failure events at all-time. Therefore, SVM+FRAM can act as a loop to control the data flow and use the output data from the framework to regulate the chemical process. If the parameters such as temperature, pressure, or re-flux, or variable such as crude composition are too low, the model can switch on the regulating algorithm to prevent catastrophic failure of the system. It alerts the operator and suggests a new approach to handle the variability of the entire process during the everyday performance by prompting the change in operating conditions. However, the application of the resilience-engineering (SVM+FRAM) concept to address a failure in unit operations is still in the infantry state. The availability of real-time data and of databases that can provide the details on the operator knowledge on the past event failure hinders the implementation of RE theories in the industry environment. To address these two issues, the following research objectives were formulated. They are:

- Developing a novel SVM model using a supervised learning algorithm.
- Test the RE (SVM+FRAM) algorithm using a model.
- Provide strategies to identify the failsafe mode.
- Establish a correlation between the SVM+FRAM models and the failure modes.

The novelty of this research is on\bridging quantitative LSVM-based anomaly detection with qualitative FRAM-based systemic explanation, enabling both improved detection performance and enhanced interpretability for operators and safety teams. This integrated framework would support resilience-based decision-making and offers practical applicability to refinery operations where process abnormalities propagate through tightly coupled technical and human workflows.

2. Theory

The model scheme applied in the study is shown in Figure 1. The support vectors here are the imaginary line passing through the points at the boundary, and the solid line represents the hyperplane formulated by maximizing the distance between the support vector line and hyperplane.

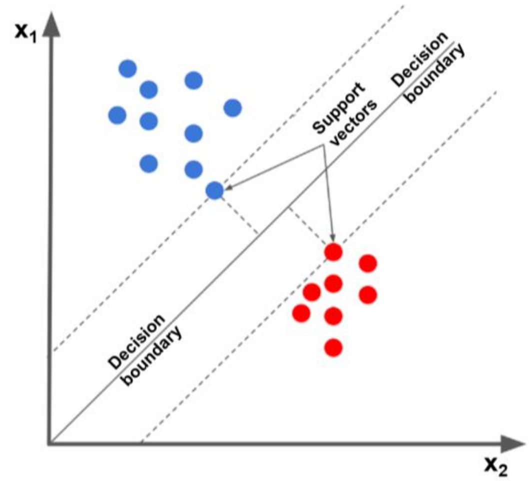

Figure 1. Optimal hyperplane and support vector boundaries.

2.1. Lagrangian Support Vector Machine (LSVM)

The model used in the study has been conceptualized as matrix expansion based on a method described by Mangasarian and Musicant. It is in binary non-linear classification and regression form. LSVM method occupies intensive mathematical resources. It is similar to the approach used in the SVM. A baseline hyperplane acts as a classifier to detect the failure points in the distillation column. The classifier is generated using the Lagrangian Equation (1). The plane generated by the LSVM acts as a decision boundary to differentiate the normal operating points from the failure data in the distillation column.

$$L = \frac{1}{2}|w|^2 - \sum_i \alpha_i \left(Y_{svi} \left(\vec{w}.\overrightarrow{X_{svi}} + b \right) - 1 \right) \tag{1}$$

where w vector is the weights, b vector is the bias, X_{svi} is the features, α_i is the Lagrange multiplier, and Y_{svi} is the labels. The above equation is deducted to formulate the best fit hyperplane. The linear separability can be achieved for the realtime refinery data using SVM. Therefore, SVM can be directly applied to the higher dimensional space. The LSVM has an inbuilt kernel function that can transform the vector space to feature space:

$$K(X_1, X_2) = \exp(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}) \tag{2}$$

where σ is the variance and the hyperparameter $\|x_1 - x_2\|$ is the Euclidean distance between two data points x_1 and x_2. Coding is written in Python based on Eequations (1) and (2). A trick code in the program transforms the data to a higher dimension and generates plots. The performances of the LSVM are better than the methods such as logistic regression and deep learning method.

2.2. Functional Resonance Analysis Method (FRAM)

The model is structured as a hexagon, as depicted in Figure 2, where each vertex represents a distinct function: input, precondition, resource, output, control, and time. These six functions are interconnected by strings that define their relationships. The roles of these functions are as follows: (a) Input refers to the elements that process or transform the function, (b) Output signifies the result or change in the function's state, (c) Preconditions are the conditions that must exist before the function operates, (d) Resources are the necessary execution conditions that are consumed to produce the output, (e) Time represents the temporal constraints that affect the function, and (f) Control governs how the function must be monitored. Unlike conventional methods, the FRAM model is constructed in advance of any failure, with each vertex following the 'one-to-many' and 'many-to-one' rule based on the input provided. There are no restrictions on how hexagons can be connected. The model is applied to survey the risk-related events and complexities in distillation column operations, relying on input from operators. Given that the FRAM model uses a static hexagonal graphical structure, the functional version of this geometry is used to analyze failure events. Several assumptions are considered during model construction to account for variability. The variability in the output of the hexagon model arises from factors such as (1) the nature of the function (e.g., crude, heat, or distillate), (2) the operational environment of the distillation unit, (3) influences from upstream functions (e.g., salt levels in crude), (4) the type of coupling between pressure, temperature, and reflux ratio, (5) timing, speed, and duration of operations, (6) fluid force, pipe distance, and direction, and (7) the comments used to manage the plant balance.

Model Assumptions:

1. Describing the Event: Pressure valve faults were assumed as the failure-causing event.

2. Detecting the event's failures: What should happen in the event, both ideally and non-ideally, was described. The data was collected from various research topics on pressure valve faults to perform this step.

3. Input to FRAM: The outcome of this second step (b) was converted into FRAM. A model and the potential column failure are linked to many events.

4. Predicting the future based on the past event: The specific information on distillation failure, such as weeping, is considered as a metric to evaluate the failure events.

The assumptions outlined above form the foundation for the survey questions used in the FRAM analysis. These questions are designed to explore various aspects of an event within the system, focusing on identifying the event, understanding its causes, describing its occurrence, evaluating its consequences, assessing the likelihood of recurrence, and proposing preventive measures. By addressing these critical areas, the survey offers a thorough framework for analyzing system behavior and mitigating potential risks.

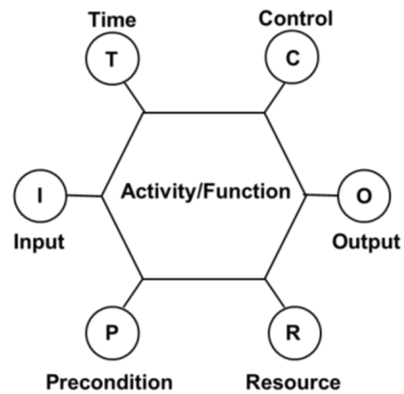

Figure 2. Details of the FRAM model, A hexagon with a function on each vertex.

3. DATA PROCESSING

3.1. SVM

Data for training, testing, and validating the SVM model was sourced from an Indian public sector oil and gas company in Mumbai, Maharashtra, comprising reflux, temperature, and pressure measurements from the top tray of a naphtha stabilizer. The dataset, with 2,511 rows and four columns, was split into training (80%) and testing (20%) sets, yielding 2,008 training points and 503 test points. Binary labels (1 for failure, 0 for normal operation) were assigned based on thresholds for pressure, temperature, and reflux levels. The SVM-based model was tuned for accuracy and sensitivity, with false positives removed to improve classification performance. A sensitivity analysis was conducted to optimize parameters, and no data normalization was performed, as the data points were contiguous. The model's output was further categorized for Fault Tree Analysis Modelling (FRAM), providing insight into system behavior and deviations in operating conditions.

3.2. FRAM

Prior to feeding data points into the model, the accuracy of critical parameters (such as feed compositions, preconditioning procedures, available resources, column output, control measures, and time) was validated against the raw dataset obtained from multiple runs. This process ensured the repeatability of the input. The performance variability functions and outcomes were presented within a four-dimensional time-space continuum. The variability in the output was linked to pressure valve conditions within the column. Output data from the Least Squares SVM (LSVM) was analyzed to identify causes of frequent failures, leading to the development of a new model based on various valve failure events.

Table 1. Literature Findings

Method	Highlights	Reference
Management of industrial safety	Provides details on the causation behind human errors, procedure compliance, protection and safety, and root cause analysis	(Sarbayev et al., 2019)
Quantitative resilience index	Presents resilience index in process safety	(Wan Jusoh et al., 2025)
SVM	Predictability of SVM with a small data set	(Pani and Soofastaei, 2025)
Safety science	Discusses the notions of the safety science	(Kumar et al., 2025)
SVM	Depicts the classification success rate of the SVM	(Hollnagel, 2017)
RE	Emphasizes the analysis related to the disturbances and interferences in the data set presents delayed learning methods	(Patriarca et al., 2020)
FRAM	Provides the database related to FRAM with case studies	(Xu et al., 2019)
ANN	Presents how ANN is mapped with a fault tree to assess dynamic failure	(Hong and Pai, 2006)
Incremental SVM	Demonstrate the resampling and misclassification rates in SVM	(Gryllias and Antoniadis, 2012)
SVM	Discusses nonlinear classification of SVM	(Suthaharan, 2016)

This analysis, incorporating pressure readings and expert input from daily valve performance observations, led to the creation of a hexagon-based model for managing and analyzing failure events. The model supports systematic analysis of pressure valve faults and over-pressurization failures (Hong and Pai, 2006). A questionnaire, designed from operator feedback and past incidents, was used to assess performance and realtime operational constraints. The probability of failure was computed from the dataset and cross-validated with the results from the SVM model.

3.3. Data Labelling

The thresholds for normal operations were defined based on input from refinery distillation specialists and control room engineers, as well as statistical analysis of five years of historical operational data, including fault logs. Parameters outside the normal operating range were labeled as “Failure” (1), while those within the range were classified as “Normal” (0). For instance, the normal range for column top pressure (P_top) was 1.5 ~ 2.2 bar, with failures occurring at pressures greater than 2.3 bar or less than 1.4 bar, based on historical data and safety limits. Column bottom temperature (T_bottom) was set between 108 and 120 °C, with failure thresholds above 122 °C or below 105 °C to maintain product purity. The reflux ratio (RR) was normally between 2.5 and 3.5, with values below 2.0 indicating potential flooding or dry-out. Finally, differential pressure across the column trays (ΔP_column) was maintained between 0.1 and 0.25 bar, with values exceeding 0.3 bar signaling tray flooding according to OEM specifications.

4. Result and Discussions

4.1. Training, Testing, and Validation of LSVM Models

The hyperplane created by LSVM has been transformed into a two-support vector, and indicators for the column failure (1) and success (0) are created. The support vectors are imaginary (dotted/dashed) lines and pass through the boundary. The distance between the support vector line and the hyperplane is maximized to get a solid line. The distance (d) between the hyperplane and the support vectors is intentionally kept maximum at the edge to create the hypothetical boundary. We can notice the LSVM technique uses the best features to construct the hyperplane and gives a good separation between the success and failure events that occur in the column. As reported in the Figure 1 topological structure of the plane has positive and negative sides where the positive quotient helps to find the best features associated with the regular column operation. One such limitation in the hyperplane domains’ ability is to distinguish the positive or negative events in big data are function overlap. When a dataset is treated in batch, we can treat the data points accumulated below the hyperplane or on the negative side of the plane as column failure events. To simplify the graphical presentation, we named the support vectors of success and failure events like 0 and 1, respectively. A shape of the hyperplane used as a metric to define the column failure is in good agreement with the literature, and preliminary dataset trained and tested are provided in Figure 1. The dataset obtained from different top trays is trained, tested, and validated with the data domain.

A three-dimensional space view provides behaviours of the three distillation parameters in a linear feature space. The colour coding is provided to easily Figure out the failure (red) and success (blue) in the Figure. Like Figure 4, the data points that are distributed above the plane are considered as positive or successful events. Therefore, we can safely assume that the distillation operation is well regulated and the temperature, pressure, and reflux are at an optimized level. The dataset presented below the plane could belong to the failure events. Since the validation with the real-time dataset shows good accuracy (R^2 above 85%), we can conclude that the hyperplane acts as the effecttive decision boundary to distinguish the column failure events from the normal operation. The value of R^2 supports the applicability of the model to detect the failure events at various operating conditions.

To increase the resolution of the model predictability data, critical data points relevant to the failure events are plotted against the pressure, temperature, and reflux in Figure 4. The features in the Y-axis represent the failures in the distillation column. Herein, the dataset presented at Y = 0 and Y = 1 is considered as success and failure events. The scattered data sets corresponding to pressure, temperature, and reflux are labelled in two cate-

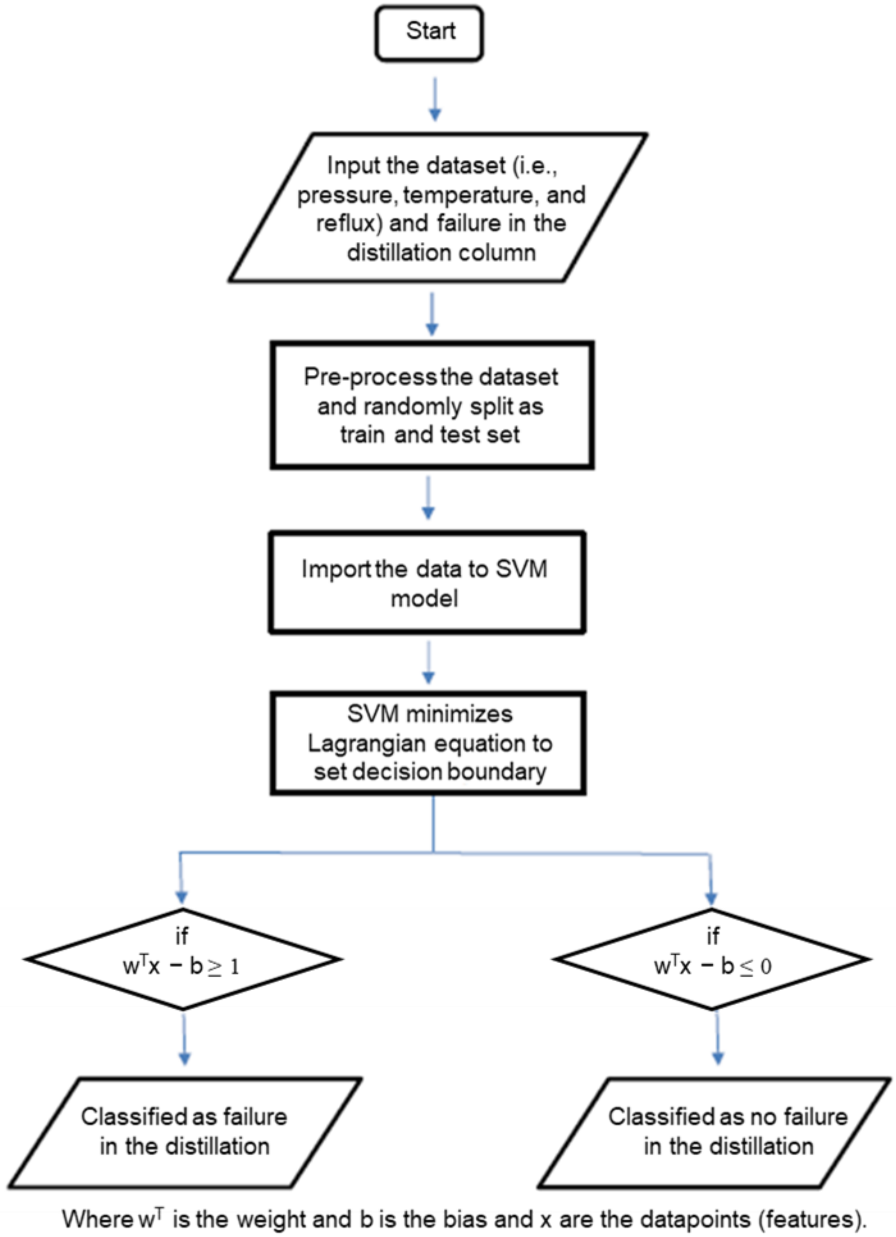

Figure 3. Support vector machine algorithm.

gories. They are the data set above the set value and below the set value. The data sets belong to three features and SVM classifiers obtained using the algorithm. Based on the hyperplane boundary, LSVM codes will automatically trigger the alarm to identify the event failures. During regular operation, the data indicator shows success (“Y = 0”). At this point, the values of the pressure, temperature, and reflux were varied from 7.9 to 9.1 kg/cm^2, the 57 to 63 °C, and 29 to 55 M.T./hr, respectively. Any value observed outside the specified range depicts the pos sibility of weeping, over-pressurization, and overheating of the column. This phenomenon is evident from the Figure. Furthermore, no failure is reported when the interaction effects of the P, T, and reflux are in the normal operating range. The variation of top reflux, temperature, and pressure are shown in Figure 6. The red dots represent the failure case, and the blue cross represents the distillation column’s success. The trend line for the graphs plotted in Figure 6 was also formulated, and the empirical expressions obtained from the equation are given in Table 2. Overall, the SVM model used to evaluate the parameters showed 96.02% efficiency, and it detects failure even if there was a small anoma-ly in the data set.

To quantify and assess the accuracy of the machine-learning model, several performance metrics were used. The counts of true positives (260) and true negatives (233) surpassed the false positives (5) and false negatives (15). Key metrics such as accuracy, precision, recall, F1-score, and receiver operating char-

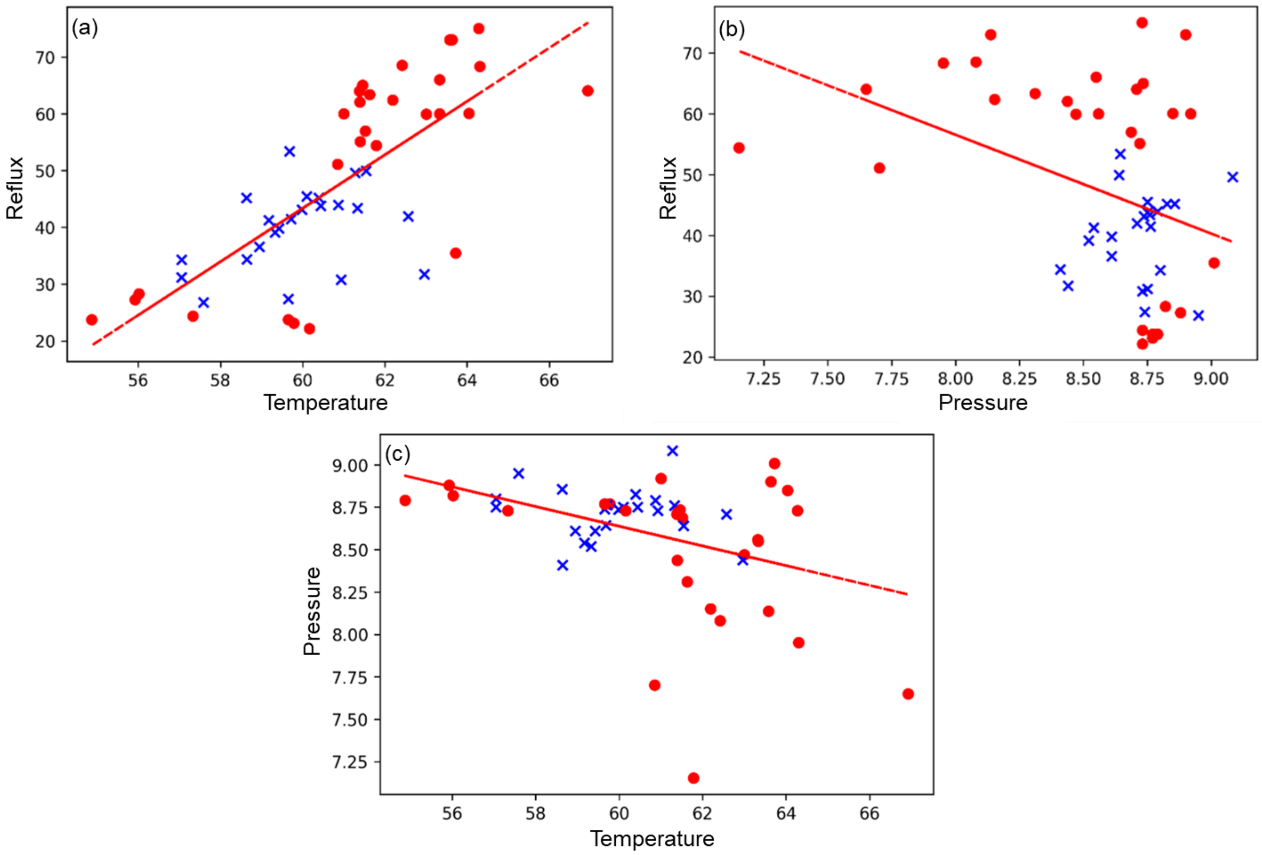

Figure 4. Graphs of (a) reflux vs. temperature, (b) reflux vs. pressure, and (c) pressure vs. temperature. The red dots show failure while the blue crosses show success.

acteristics (ROC) were calculated (Figure 5). Precision and recall values were derived from the counts of true positives, true negatives, false positives, and false negatives. The model's performance results are presented in Table 3.

ROC exhibits a higher measure of separability. In our case it can be used to quantify how accurately the models can quantify between the failure and success in the distillation column. The trade-off between the true positive rate and the false positive rate as the criterion for positivity is changed. The concave nature of the curve is due to the monotonically increasing likelihood ratio (distribution of the separator variable in failure and success). The area under the curve is the combined measure of specificity and sensitivity that relates to whether or not there is any failure in the distillation column. The greater the area under the ROC curve, the better is the model. In our case we can observe that the area under the curve is 0.99, which shows that our model is very efficient and can be used as an effective tool to identify complex failures of the column.

To assess the robustness of the LSVM model, we employed a k-fold cross-validation technique with k set to 10. This rigorous approach involved randomly partitioning the dataset into 10 equal subsets, ensuring comprehensive representation of both normal and fault cases. In each iteration, one subset was designated as the test set, while the remaining nine subsets were utilized for training. The results showcased remarkable stability across the folds, achieving a mean accuracy of 96.1 ± 0.8%, a mean recall of 95.4 ± 0.9%, and a mean precision of 96.7 ± 0.6%. These outcomes reveal minimal performance variability and highlight the strong generalization capability of our model. Additionally, we performed a sensitivity analysis with alternative training and testing splits (70/30, 75/25, and 80/20), consistently demonstrating results within a narrow ± 1.2% range of the mean values. The overall mean false-alarm rate was 96.7% ± 0.6%.

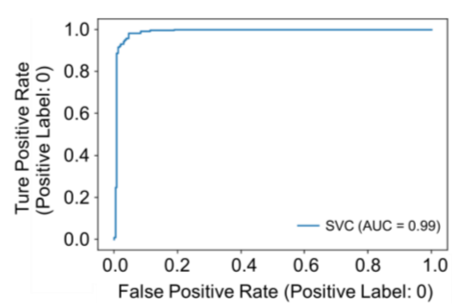

Figure 5. Receiver operating characteristic (ROC) curve.

For Set 1, the F1-score was 95.4% and the false-alarm rate was 3.1%. For Set 2, the F1-score was 96.3% and the false-alarm rate was 2.9%. For Set 3, the F1-score was 96.3% and the false-alarm rate was 2.6%.

Table 2. Empirical Expression from the Model Process Data

Graph of Pressure vs. Temperature	y = −0.058056x + 12.121081
Graph of Reflux vs. Temperature	y = 4.704938x − 238.943022
Graph of Reflux vs. Pressure	y = −16.258381x + 186.603348

To substantiate the efficacy of the proposed LSVM framework, we conducted comparative simulations with plant data using three reputable baseline models are included. The models such as Logistic Regression, a Decision Tree Classifier, and LSVM are rested with different volumes of data. All models were trained and tested on the same dataset using identical train-test splits and cross-validation procedures. According to the performance metrics, LSVM achieved the highest F1-Score at 96.0%, surpassing Standard SVM (92.1%), Decision Tree (88.4%), and Logistic Regression (85.3%). LSVM also produced the lowest false-alarm rate at 2.9%, compared with 4.2% for SVM, 5.9% for the Decision Tree, and 6.8% for Logistic Regression. In terms of computational efficiency, LSVM required only 0.73 seconds for training, considerably faster than Standard SVM, which took 1.98 seconds. Logistic Regression and the Decision Tree were quicker at 0.42 and 0.51 seconds respectively, but they exhibited markedly lower detection performance. These results confirm that LSVM delivers a more robust and computationally efficient solution for early fault detection in crude oil distillation operations.

4.2. Evaluation of SVM Datapoints in the Functional Resonance Analysis

The description of the hexagon model and its functions is intricately tied to its geometric connections. A semi-explicit stop switch ensures model completeness and consistency. To streamline analysis and reduce confusion, the textual descripttions of events and functions are labeled consistently with Figure 6. In this framework, the output from Hexagon 1, labeled ‘pressure valve functioning’, feeds into the input of Hexagon 2, termed ‘date of the pressure valve’. The data points from Hexagon 1 (pres-sure valve functioning) serve as the input to construct the model. Since there is no significant variability in the foreground function, the provided descriptions adequately represent the function. Any fault in the data at Hexagon 2 immediately impacts the model's execution and performance. Beyond testing nodes, the big data flow across the hexagons must remain accurate and up-to-date. Consequently, data packets traveling through the hexagons act as a mathematical sensor, allowing for the assessment of potential column failures. Upon input, the model triggers the “Analysis of pressure data” command, and if a failure is detected, it activates the safety alarm. Machine learning (ML) can then initiate subsequent actions to mitigate the failure. A crucial link within the hexagon model is where the data from the pressure valve acts as a tool to precondition the input stream. This is referred to as the Pre-set data from literature, processes, and SVM. Both the SVM and FRAM models are connected at this stage. Any failure due to over-pressurization can be detected and managed through the “pressure valve functioning” option. By mapping failures within the hexagon, the model enhances the predictability of events. All functions and background variabilities within the model are continuously monitored to ensure analysis stays within defined parameters.

The performance of the distillation column hinges on the proper functioning of the pressure valve system and the maintenance of a temperature gradient across the trays. Successful distillate transfer between trays contributes to risk assessment strategy formulation. However, if process parameters fluctuate outside acceptable ranges, the risk assessment index becomes unstable, increasing the likelihood of accidents. While components like temperature or pressure regulators can mitigate these issues, their own variability can leave the distillate output as the only reliable indicator. A drop in distillate output or compositional changes may signal an impending out-of-control situation. It is essential to synchronize parameter settings and measurements through algorithms, though this can be challenging when inaccurate data is provided. Therefore, a second iteration of the model should expand on the pressure valve functioning and its variability. Insights from operators can offer valuable guidance. Given the sensitivity of distillation operations to pressure, temperature, and reflux, a more detailed understanding of these functions, as facilitated by the FRAM, could provide greater insight into fault management. When integrated with the SVM+FRAM framework, the model triggers alarms within the Distributed Control System (DCS), enabling the operator to initiate failure prevention protocols and avert catastrophic outcomes. Monitoring the communication between hexagons ultimately enhances resilience and supports failure prevention.

4.3. Comparison amongst ET, FT, and SVM+FRAM

Unlike FTA (Fault Tree Analysis) and ETA (Event Tree Analysis), the SVM (Support Vector Machine) lacks a structured logical, temporal, or physical framework to represent the relationships between various events in a distillation column. In ETA, events are connected through failure probabilities, with conditions tied to the operational state of the system. RE-ML-based techniques, on the other hand, evaluate the process and identify potential hazards by analyzing events that lead up to them using trained data, based on SVM analysis. A key feature of the SVM approach is its straightforward graphical tool for event analysis. This framework helps recommend solutions for operational issues in distillation columns by mapping the cause, effect, and mitigation strategies. During real-time operation, unforeseen events may arise that cannot be represented by traditional FTA or ETA, such as incidents like flooding or weeping, which require more detailed event information. In these cases, RE/ML techniques that utilize SVMs can report the failure probability, whether or not the basic events are explicitly defined. The SVM’s graphical output serves as a direct alert for operators, helping them anticipate potential issues. Thus, this method proves superior in predicting failure probabilities, particularly for pressure-related events, and in triggering alarms while activating preventive control measures. The algorithm is particularly effective at detecting issues like over-pressurization in the column. In contrast, both FTA and ETA lack the ability to use virtual machines to trigger accident or failure prevention control algorithms.

Table 3. Evaluation Matrix for the Machine-Learning Model

	Support Vector Machine			
	Precision	Recall	F1-Score	Support
0(Success)	0.95	0.98	0.96	265
1(Failure)	0.98	0.94	0.96	238
Accuracy			0.96	503
Macro Average	0.96	0.96	0.96	503
Weighted Average	0.96	0.96	0.96	503

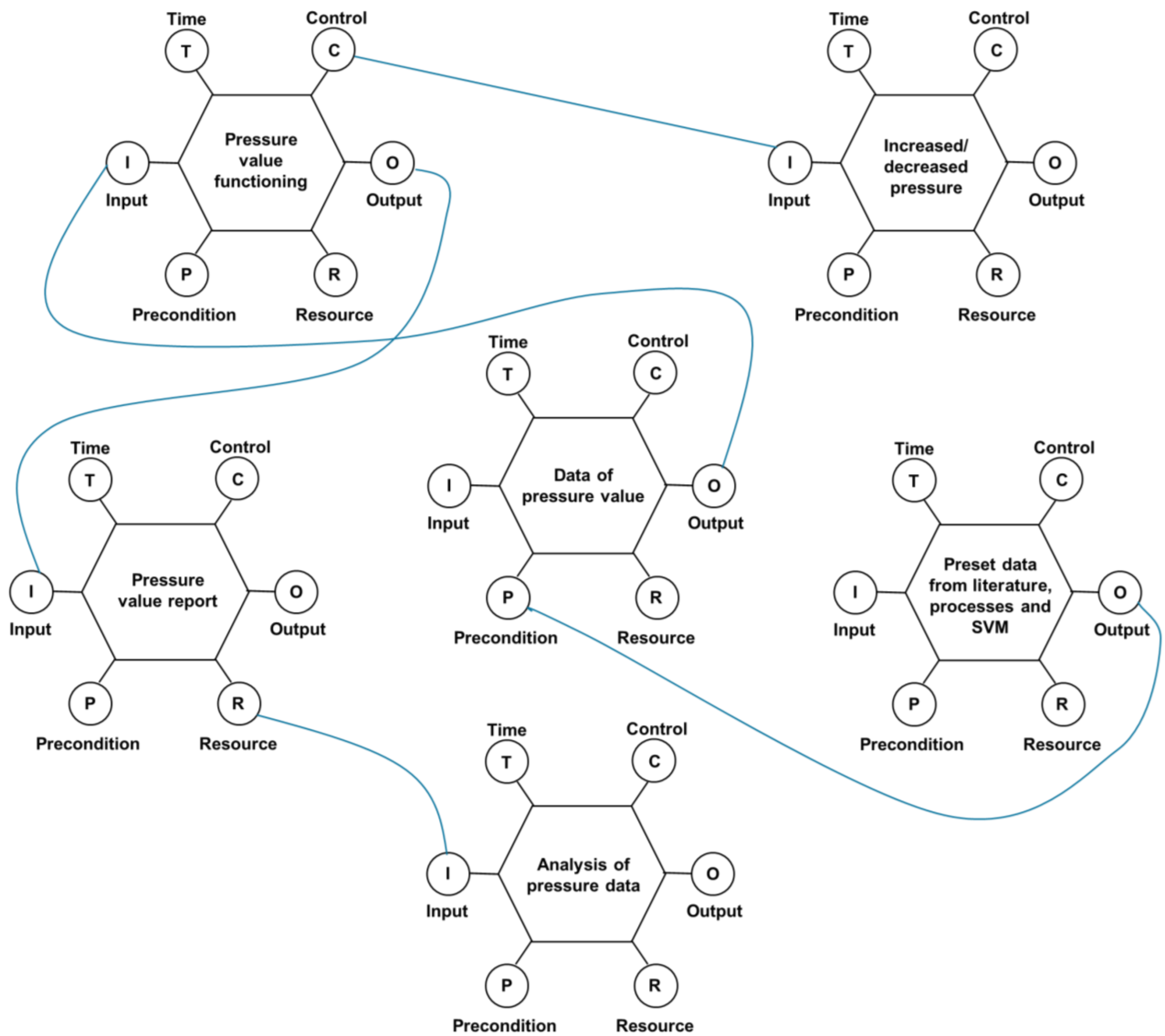

Figure 6. FRAM model for analyzing failure in distillation column due to pressure valve functioning.

Perspectives: The industries related to environmental informatics could gain major attention if the proposed LSVM +FRAM framework is integrated to their database. The use of framework in this field goes beyond routine operational data by incurporating additional streams such as emissions monitoring, effluent quality, flare-stack combustion, alarm logs, weather conditions, and regulatory thresholds. Our hazard-forecasting model not only acts as a decision-support tool but are integrated to enable proactive risk mitigation. Adding the informatics related to the environmental datasets can serve as a decision-support layer, combining LSVM-based anomaly detection with FRAM-based systemic variability mapping to prioritize critical events (e.g., emission spikes, off-spec discharges) and guide operational adjustments. This integration facilitates early warning, risk tracing, and real-time recommendations, enhancing both resilience and compliance. However, data related to pollutants, emissions, effluxent quality, flare combustion, weather, and regulatory thresholds are confidential and cannot be included in the manuscript.

5. Conclusion

In conclusion, the proposed SVM+FRAM framework provides a robust and effective approach to failure detection and mitigation in distillation columns within refineries. Key findings include:

Enhanced Failure Detection: The integration of the Support Vector Machine (SVM) with the Functional Resonance Analysis Method (FRAM) enables accurate identification of failure events that are often overlooked by traditional fault and event tree analysis.

Improved Diagnostic Capabilities: By leveraging the vectorization of large, complex datasets and applying kernel transformations, the SVM model effectively classifies failure events, providing a reliable decision boundary for operational anomalies.

Comprehensive Failure Analysis: The SVM+FRAM coupled framework allows for the identification of multi-factorial failure scenarios, offering a more nuanced understanding of distillation column failures, including those induced by overpressurization and flooding.

Mitigation Strategy Development: Unlike fault and event tree analysis, which focuses on logical event relationships, the SVM+FRAM framework not only identifies failure frequencies but also facilitates the development of proactive mitigation strategies to prevent catastrophic outcomes.

Operational Decision Support: This model empowers refinery operators by providing a data-driven basis for real-time decision-making, optimizing product quality while minimizing risks related to unforeseen operational failures.

Environmental Informatics Application: The integration of environmental informatics into the SVM+FRAM framework enhances its ability to monitor and predict failure events that may lead to environmental hazards, ensuring that operational safety aligns with environmental sustainability goals. This helps mitigate risks such as chemical spills, emissions, or resource wastage.

In essence, the SVM+FRAM framework offers a significant advancement in resilience engineering, combining predictive analytics with systems-level failure analysis and environmental informatics, making it a promising tool for improving both the reliability, safety, and sustainability of refinery operations.

References

Bartlett, L.M., Hurdle, E.E. and Kelly, E.M. (2009). Integrated system fault diagnostics utilising digraph and fault tree-based approaches. *Reliability Engineering & System Safety*, 94(6), 1107-1115. https://doi.org/10.1016/j.ress.2008.12.005

Besnard, D. and Hollnagel, E. (2014). I want to believe: some myths about the management of industrial safety. *Cognition, Technology & Work*, 16(1), 13-23. https://doi.org/10.1007/s10111-012-0237-4

Castillo-Borja, F., Vàzquez-Romàn, R., Quiroz-Pérez, Efraín., Diaz-Ovalle, C. and Mannan, M.S. (2017). A resilience index for process safety analysis. *Journal of Loss Prevention in the Process Industries*, 50, 184. https://doi.org/10.1016/j.jlp.2017.06.017

De Souza, D.L., Granzotto, M.H., M. de Almeida, G.M. and Oliveira-Lopes, L.C. (2014). Fault detection and diagnosis using support vector machines-a SVC and SVR comparison. *Journal of Safety Engineering*, 3(1), 18-29. https://doi.org/10.5923/j.safety.20140301.03

Erik, H. (2009). The four cornerstones of resilience engineering*, Volume 2: Preparation and Restoration*, CRC Press pp 1-117. https://doi.org/10.1201/9781315244389

Gryllias, K.C. and Antoniadis, I.A. (2012). A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. *Engineering Applications of Artificial Intelligence*, 25(2), 326-344. https://doi.org/10.1016/j.engappai.2011.09.010

Hollnagel, E. (2014). Is safety a subject for science? *Safety Science*, 67, 21-24. https://doi.org/10.1016/j.ssci.2013.07.025

Hollnagel, E. (2017). *FRAM-the functional resonance analysis method: Modelling complex socio-technical systems*. CRC Press, pp 1-160. https://doi.org/10.1201/9781315255071

Holt, B.R. and Morari, M. (1985). Design of resilient processing plants-VI. The effect of right-half-plane zeros on dynamic resilience. *Chemical Engineering Science*, 40(1), 59-74. https://doi.org/10.1016/0009-2509(85)85047-8

Hong, W.C. and Pai, P.F. (2006). Predicting engine reliability by support vector machines. *The International Journal of Advanced Manufacturing Technology*, 28(1-2), 154-161. https://doi.org/10.1007/s00170-004-2340-z

Jia, S., Qian, X. and Yuan, X. (2017). Optimal design for dividing wall column using support vector machine and particle swarm optimization. *Chemical Engineering Research and Design*, 125, 422-432. https://doi.org/10.1016/j.cherd.2017.07.028

Kumar, A., Singh, S.K., Samanta, B. and Bhattcherjee A. (2025). Risk assessment in sociotechnical systems based on functional resonance analysis method and hierarchical fuzzy inference tree. *Scientific Reports*. 15(1), 23827. https://doi.org/10.1038/s41598-025-10063-5

Madni, A.M. and Jackson, S. (2009). Towards a conceptual framework for resilience engineering. *IEEE Systems Journal*, 3(2), 181-191. https://doi.org/10.1109/JSYST.2009.2017397

Naik, A.S., Yin, S., Ding, S.X. and Zhang, P. (2010). Recursive identification algorithms to design fault detection systems. *Journal of Process Control*, 20(8), 957-965. https://doi.org/10.1016/j.jprocont.2010.06.018

Pani AK, Soofastaei A. (2025). Designing Intelligence: Harnessing Soft Sensors and Advanced Analytics in Petroleum Refining for Industry 4.0. *Advanced Analytics for Industry 4.0.* CRC Press, pp. 83-116. https://doi.org/10.1201/9781003186823

Patriarca, R., Di Gravio, G., Woltjer, R., Costantino, F., Praetorius, G., Ferreira, P. and Hollnagel, E. (2020). Framing the FRAM: A literature review on the functional resonance analysis method. *Safety Science*, 129, 104827. https://doi.org/10.1016/j.ssci.2020.104827

Qi, R. *Fault Diagnosis and Fault Severity Prediction Based on Computational Intelligence Techniques for Industrial Processes* (Ph.D. dissertation, Newcastle University. UK. https://theses.ncl.ac.uk/jspui/bitstream/10443/6297/1/QiR2024.pdf

Sarbayev, M., Yang, M. and Wang, H. (2019). Risk assessment of process systems by mapping fault tree into artificial neural network. *Journal of Loss Prevention in the Process Industries*, 60, 203-212. https://doi.org/10.1016/j.jlp.2019.05.006

Saghir, H., Ahmad, I., Kano, M., Caliskan, H. and Hong, H. (2024). Prediction and optimisation of gasoline quality in petroleum refining: The use of machine learning model as a surrogate in optimisation framework. *CAAI Transactions on Intelligence Technology*, 9(5), 1185-1198. https://doi.org/10.1049/cit2.12324

Shirali, G., Motamedzade, M., Mohammadfam, I., Ebrahimipour, V. and Moghimbeigi, A. (2016). Assessment of resilience engineering factors based on system properties in a process industry. *Cognition, Technology & Work*, 18(1), 19. https://doi.org/10.1007/s10111-015-0343-1

Suthaharan, S. (2016). *Machine Learning Models and Algorithms for Big Data Classification*, Springer, pp 1-359. https://doi.org/10.1007/978-1-4899-7641-3

Wang, H., Zheng, H., Zhang, Z. and Wang, G. (2024). A deep learning-based acoustic signal analysis method for monitoring the distillation columns' potential faults. *Applied Sciences*, 14(16), 7026. https://doi.org/10.3390/app14167026

Wan Jusoh, W.N., Omar, M.B., Sami, A., Bingi, K. and Ibrahim, R. (2025). Crude oil yield estimation: Recent advances and technological progress in the oil refining industry. *Sensors*, 25(17), 5511. https://doi.org/10.3390/s25175511

Woods, D.D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering & System Safety*, 141, 5-9. https://doi.org/10.1016/j.ress.2015.03.018

Xu, J., Xu, C., Zou, B., Tang, Y.Y., Peng, J.T. and You, X.G. (2019). New incremental learning algorithm with support vector machines. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(11), 2230. https://doi.org/10.1109/TSMC.2018.2791511