Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

DATA-DRIVEN PREDICTIVE MAINTENANCE IN GAS PLANTS: FAILURE PREDICTION WITH SUPPORT VECTOR MACHINES

Chukwuemeka Johnson Nwafor and Adewale Samuel Ogunleye
University of Port Harcourt, Department of Mechanical Engineering, Nigeria.
DOI: https://doi.org/10.5281/zenodo.17192420

Abstract
This study optimized data-driven predictive maintenance in gas plant using machine learning based Support Vector Machines (SVM) in predicting failures. The application software was developed to improve the operational efficiency and reliability of turbo-compressors in gas injection plants. The Gas injection plant produced below maximum capacity due to failure problems of the Turbo-compressors, these affected the targeted oil production negatively. The unavailability and unreliable gas plant led to revenue losses. Machine learning techniques using Support Vector Machines (SVM), was employed to develop the failure predictive application software. According to the findings, the Efficient Linear SVM model detected failures with a 99.5% true positive rate and classified non-failure events with a 99.9% classification precision. Although it showed a 0.5% false negative rate, the Boosted Trees model obtained a 99.5% true positive rate (TPR) for failure detection, underscoring the need for additional optimization and integration with ensemble approaches to reduce operational risks. Additionally, the SVM model demonstrated a minimum false negative occurrence and 99.9% classification precision for non-failure events. The outcomes of this study yielded a highly effective, computationally efficient machine learning-based application software capable of reliably predicting turbo-compressor failures. The study concluded that the developed application software is a powerful tool for predicting failures in gas injection plants, supporting decision-making processes, and enhancing operational safety. Recommendations for future works included refining existing models, exploring additional feature engineering techniques, and evaluating the robustness of the models under varying operational conditions.

***Keywords**: Predictive Maintenance, Support Vector Machines, Gas Plants, Failure Prediction, Machine Learning*

1. Introduction

In the very competitive oil and gas sector, increasing production efficiency is essential to remain in business. In gas injection plants, turbo-compressors are essential components that have a direct impact on operational availability and reliability. Significant production losses result from these systems' frequent failures, nevertheless, particularly in light of Nigeria's security issues, pipeline vandalism, and falling oil prices.

Predicting failures is crucial for maximizing planned maintenance and reducing unscheduled downtime. However, because modern turbo-compressor systems integrate several components under the supervision of sophisticated computerized controls. Recurrent turbo-compressor failures have resulted in decreased production capacity at the gas injection plant leading to financial losses.

This study addresses this issue by creating application software based on machine learning using support vector machines to improve the precision of failure predictions, computational efficiencies and usability by maintenance staff.

The accuracy, computational efficiency and practical usefulness of current failure prediction models are limited. Hybrid models as indicated by Anil et al (2015) could improve predictions, but their application in the oil and gas sector remains limited.

The paper seeks to develop an application software based on machine learning using SVM that can accurately predict gas injection plant failures. Data will be drawn from the maintenance department for analysis and develop the application software. The gas injection plant's historical failure data was used to validate the software and an easy-to-use software interface was developed for the maintenance staff.

By reducing production losses through precise failure prediction and proactive maintenance, the outcome of this study will be of immense benefit to the oil and gas sector and enhance revenue streams to promote economic stability in oil-dependent nation like Nigeria. The study also lays the groundwork for further predictive maintenance research in other crucial industries like manufacturing and power generation.

2. Literature Review

Recent advancements have integrated artificial intelligence (AI) into failure prediction. Anirbid and James (2022) documented AI's transformative potential in oil and gas, while Arash et al. (2021) demonstrated how hybrid models combining Artificial Neural Networks (ANNs) and Genetic Algorithms could optimize gas injection processes. Similarly, Steve et al. (2018) applied machine learning to diesel engine prognostics, though they emphasized the need for improved data quality to enhance predictive accuracy.

Failure theory remains an emerging field of study that has gained significance in recent decades due to increasing industrial complexity (Shokufe et al., 2018). As machinery systems grow more sophisticated—with higher costs and performance demands—the consequences of failure have become substantially more severe, underscoring the critical need for accurate lifespan prediction of equipment components (Anirbid & James, 2022). This is particularly relevant in gas injection plants, which are deployed to enhance reservoir pressure for improved oil recovery (James, 2015).

According to Dragomir et al. (2022), prognostics and remaining useful life (RUL) prediction are critical enablers for cost-effective condition-based maintenance of industrial assets. However, they noted prognostic approaches must account for variable operating conditions and load profiles which accelerate degradation. Their review analyzed recent advances in datadriven prognostics under variable operating conditions. Based on the literature, physicsinformed machine learning and adaptive modelling techniques showed promise for enhancing robustness and accuracy by incorporating operating context (Dragomir et al., 2022). However further validation on real-world assets with diverse usage profiles was needed. Overall, their review highlighted challenges and opportunities to advance prognostics to support maintenance decision-making for systems facing variable operating environments.

Zeeshan et al. (2021) and Joerg et al. (2021) systematically reviewed ML applications in oil and gas and industrial maintenance, respectively, serving as key references for failure prediction methodologies. Kadir et al. (2020) achieved notable success in aircraft equipment failure prediction using hybrid data preparation, while Michail et al. (2020) validated datadriven fault detection in maritime systems. Andrea et al. (2014) further tailored ConditionBased Maintenance (CBM) to naval gas turbines, emphasizing real-time monitoring.

Oyedepo et al. (2014) conducted a comprehensive 10-year evaluation (2001-2010) of gas turbine operations in Nigeria, revealing critical availability-performance gaps. Their analysis demonstrated only 64.3% operational capacity utilization against industry benchmarks of 95%, with generation losses amounting to approximately $251 million. The researchers advocated for enhanced operator training, optimized spare parts management, and improved maintenance protocols as key corrective measures. These findings established a baseline for subsequent studies on African energy infrastructure reliability. Recent scholars have advanced various methodologies for failure prediction: Fernando and Gilberto (2009) established the correlation between component reliability, maintenance policies, and overall system availability in gas turbine operations. Anil et al. (2015) developed Markov models for urea synthesis systems, demonstrating their efficacy in calculating Mean Time Between Failures (MTBF) despite noted limitations in handling complex differential equations. Zhiqiang et al. (2018) pioneered dynamic Bayesian networks for multi-state systems, incorporating condition-based maintenance (CBM) parameters that improved failure prediction accuracy by 22% in validation studies. Contemporary research emphasizes integrated analytical approaches: Federick et al. (2021) introduced a hybrid reliability model for rolling stock that reduced unplanned downtime by 18% through combined physical and data-driven analytics. Sangje et al. (2016) demonstrated superior performance of Markov-regression hybrid models in compressor failure prediction, achieving 92% accuracy in offshore operational tests. Basheer et al. (2023) innovated with accumulative artificial neural networks, combining FCN and ACN architectures to predict mechanical component degradation with <5% error margins.

Nanda et al. (2017) trained their model using vibration data collected from a laboratory centrifugal pump setup under normal operation, cavitation, impeller damage, and bearing wear conditions. They reported over 98% classification accuracy in detecting these major fault categories. The SOM feature extraction combined with the pattern recognition capabilities of SVM enabled high-precision diagnostics of equipment conditions (Nanda et al., 2017). However, the researchers acknowledged that real-world applications could prove more complex. Factors like varying pump operating points and limited available sensor data may impact model performance outside controlled lab environments. They recommended further research under field conditions to evaluate the robustness of the approach for commercial pump monitoring. Nonetheless, their hybrid SOM-SVM methodology demonstrates a promising advancement in utilizing machine learning techniques for predictive maintenance of centrifugal pump systems (Nanda et al., 2017).

Providing a comprehensive overview, Lei et al. (2020) reviewed the landscape of machine learning techniques applied to machine fault diagnosis. The authors systematically categorized the literature based on the diagnosis task including fault detection, fault classification, fault identification, and fault prognosis. For each category, key methods were benchmarked and comparative strengths and limitations were analyzed.

A major finding was that hybrid approaches coupling physics-based feature extraction with machine learning models consistently achieved state-of-the-art accuracy across all diagnosis subtasks. However, challenges remain around model generalization, handling noisy sensor data, and integrating domain knowledge. The authors mapped out opportunities where machine learning, especially hybrid techniques, could significantly advance fault diagnosis. However, realizing real-world implementations would require focused efforts to address the identified constraints. This review highlighted the promise while pinpointing areas needing further innovation.

Djeziri et al. (2016) focused their research on using artificial neural networks for equipment condition monitoring and fault diagnosis. They argued machine learning methods like neural networks can learn complex equipment fault signatures from sensor data without requiring extensive analytical modelling. In their methodology, they employed a radial basis function (RBF) network for bearing fault detection based on vibration measurements (Djeziri et al., 2016). The RBF architecture can approximate nonlinear functions and classify patterns through supervised training.

In their implementation, Djeziri et al. (2016) trained the RBF network on experimental vibration data collected from a rotating machinery test rig under normal conditions and with artificially induced bearing faults. The RBF model reliably detected faults like inner race defects, ball defects, and outer race defects with over 95% accuracy.

3. Materials and Methodology

The context in which the present case study was carried out was in the Gas Injection plant in Nigeria. The maintenance or failure data were collected from the Maintenance department of the plant. A comprehensive site visit was conducted at the Gas Injection Plant, involving indepth discussions with Field Engineers, Maintenance Managers, Supervisors, and Technicians. Relevant documents were meticulously reviewed to identify specific dates and times associated with equipment failures.

The reviewed documents from the Gas Plant encompassed: Operating Logs, Correspondence, Inspection/Surveillance Records, Maintenance Records, Minutes of Technical Meetings, Computer Process Data, Procedural and Instructional Documentation, Vendor Manuals, Drawings and Specifications, Equipment History Records, Design Basis Information, Trend Charts and Graphs, Facility Parameter Readings, Operational Safety Requirements and Work orders.

The desk-based research involved referencing engineering journals, technical papers, and standard texts pertinent to this study. Key references included the Engineering Design Handbook, Handbook on

Reliability, Operations and Maintenance manuals of the Gas Injection Plant, and Root Cause Analysis guidance documents. The Machine Learning application software will be developed using data collected from the Gas Injection Plant. The findings from both field visits and desk research will form the foundation for developing a predictive failure application software. Upon a thorough review and preliminary analysis of the plant data, about thirty-five (35) significant causes of failures in the Gas Plant were identified.

3.1.1 Purpose and Significance of the Materials and Methods Used

The primary purpose of the materials and methods outlined here is to establish a robust and reliable framework for developing a machine learning-based application software capable of predicting failures in a gas injection plant. This section underscores the rationale behind selecting specific materials and the significance of each methodological step.

3.1.2 Purpose of the Material 1. **Hardware Components:**

Sensors and Monitoring Systems: Essential for collecting real-time operational data and environmental parameters. These devices ensure accurate and continuous data acquisition, which is crucial for training and validating the predictive Application software.

Computational Resources: High-performance computing hardware, such as GPUs and multi-core processors, are necessary for handling large datasets and complex computations involved in model training and evaluation.

2. **Software Tools and Libraries:**

Programming Languages (e.g., Python): Chosen for its versatility and extensive libraries, Python provides the necessary tools for data manipulation, model development, and evaluation.

Machine Learning Libraries (e.g., TensorFlow, Scikit-Learn): These libraries offer prebuilt functions and algorithms that simplify the implementation of machine learning models, facilitating efficient experimentation and optimization.

3. **Data Sources and Databases:**

Historical Failure Data: Historical data on past failures is vital for identifying patterns and correlations that the machine learning model can learn from to predict future failures. **Operational Parameters and Environmental Conditions:** These datasets provide the contextual information needed to understand the factors influencing system performance and potential failure modes.

3.1.3 Significance of the Method

1. **Data Collection:**

Collecting high-quality and relevant data is foundational to the success of any machine learning project. The methods used to gather and preprocess data ensure that the dataset is comprehensive, accurate, and suitable for training predictive models.

2. **Feature Selection and Engineering:**

Identifying and engineering the right features is crucial for enhancing the Application software's ability to learn and make accurate predictions. This step transforms raw data into meaningful inputs that capture the underlying patterns related to system failures.

3. **Model Selection:**

Choosing the appropriate machine learning algorithm is critical for balancing model complexity, interpretability, and performance. The selected models are evaluated based on their ability to accurately predict failures while being computationally efficient and scalable.

4. **Model Training and Hyperparameter Tuning:**

The training process and hyperparameter tuning are essential for optimizing model performance. These methods ensure that the model generalizes well to new data, minimizing both underfitting and overfitting.

5. **Model Evaluation:**

Rigorous evaluation using relevant metrics and validation techniques provides confidence in the model's predictive capabilities. This step verifies that the model performs well not only on training data but also on unseen data, ensuring its reliability in real-world applications.

6. **Implementation and Deployment:**

The methods for implementing and deploying the Application software are significant for integrating it into the operational environment of a gas injection plant. This ensures that the software can provide real-time predictions and actionable insights, thereby enhancing plant safety and efficiency.

In summary, the materials and methods used in this paper are meticulously selected and applied to achieve a high-performing, reliable, and ethically sound predictive application software. Each component and methodological step contributes to the overarching goal of enhancing the safety and operational efficiency of gas injection plants through advanced machine learning techniques.

3.2 Materials

The materials utilized in this study are integral to the development and validation of the machine learning-based application software for predicting failures in a typical gas injection plant. The materials can be categorized into three main groups: hardware components, software tools, and data sources.

3.2.1 Hardware Components

The hardware components play a critical role in the acquisition, processing, and storage of data necessary for model development.

1. **Sensors and Monitoring Systems:**

Pressure Sensors: These sensors measure the pressure within the gas injection system, providing crucial data that can indicate potential failure points due to pressure anomalies. **Temperature Sensors:** Monitoring temperature variations is essential as extreme temperatures can lead to equipment failures. These sensors help track the thermal conditions within the system.

Flow Meters: These devices measure the flow rate of gas being injected, offering insights into the operational efficiency and potential blockages or leaks in the system.

Vibration Sensors: Installed on key machinery, vibration sensors detect unusual vibrations that often precede mechanical failures.

2. **Data Acquisition Systems:**

Data Loggers: Data loggers are used to continuously record sensor readings over time. They store large volumes of data, which is later retrieved for analysis.

SCADA Systems (Supervisory Control and Data Acquisition): SCADA systems provide a comprehensive interface for monitoring and controlling industrial processes. They collect real-time data from sensors and actuators, facilitating centralized data collection.

3. **Computational Resources:**

High-Performance Servers: These servers are equipped with powerful CPUs and large memory capacities to handle the extensive computations required for data processing and model training.

Graphics Processing Units (GPUs): GPUs are utilized for their parallel processing capabilities, significantly accelerating the training of complex machine learning models. **Storage Solutions:** High-capacity storage devices, such as SSDs and cloud storage, ensure efficient data management and quick access to large datasets.

3.2.2 Software Tools and Libraries

The software tools and libraries employed in this study provide the necessary infrastructure for data processing, model development, and deployment.

1. **Programming Languages:**

Python: Chosen for its simplicity and extensive ecosystem, Python is the primary programming language used in this study. It offers a wide range of libraries and frameworks for data science and machine learning.

2. **Data Processing and Analysis Tools:**

Pandas: This library is used for data manipulation and analysis. It provides data structures and functions needed to clean, transform, and analyze large datasets efficiently.

NumPy: NumPy supports numerical operations on large arrays and matrices, which are fundamental in data preprocessing and feature engineering.

3. **Machine Learning Libraries:**

Scikit-Learn: A versatile library that offers simple and efficient tools for data mining and data analysis. It is used for implementing various machine learning algorithms and evaluation metrics.

TensorFlow: This open-source library developed by Google is used for building and training deep learning models. TensorFlow's flexibility and scalability make it suitable for complex neural network architectures.

Keras: An API running on top of TensorFlow, Keras simplifies the process of building and training neural networks, providing a user-friendly interface.

4. **Visualization Tools:**

Matplotlib and Seaborn: These libraries are used to create informative and attractive visualizations. They help in understanding data distributions, feature correlations, and model performance.

3.4 Feature Selection and Engineering

3.4.1 Identification of Relevant Features

Effective feature selection is critical for the success of any machine learning model. In this research, identifying relevant features involves selecting the most informative and significant variables from the dataset. This process ensures that the model can learn meaningful patterns and relationships that contribute to accurate failure predictions. The relevant features for this study are categorized into three main groups: operational parameters, environmental factors, and historical failure indicators.

3.5 Model Selection

3.5.1 Overview of Machine Learning Algorithms Considered

In this study, a variety of machine learning algorithms were considered for predicting failures in a gas injection plant. The goal was to explore both supervised and unsupervised learning techniques to identify the most suitable model for the task. The choice of algorithms was guided by their ability to handle the complexity and nature of the dataset, as well as their track record in similar predictive maintenance applications.

Support Vector Machines (SVM):

An algorithm that finds the hyperplane that best separates the data into classes. It is effective in high-dimensional spaces and when the number of dimensions exceeds the number of samples. SVM is particularly useful for binary classification tasks.

In Linear SVM for a binary classification problem, the goal is to find the hyperplane that maximizes the margin between the two classes. The decision function is as in Equation (3.1).

$$f(x) = w.x + b \tag{3.1}$$

where:

w is the weight vector.

b is the bias term.

x is the input feature vector.

The optimization problem is in Equation (3.2).

$$min \frac{1}{2} ||w||^2 \tag{3.2}$$

Subject to Equation (3.3).

$$y_i(w.x_i + b) \geq 1 \tag{3.3}$$

where:

$y_i \in \{-1,1\}$ are the class labels.

x_i are the input feature vectors.

In Kernel SVM for non-linear decision boundaries, SVM uses kernel functions to map the input features into a higher dimension space where a linear separator can be found. The decision function is as in Equation (3.4).

n

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

$$(x) = \sum_{i=1} a_i y_i K(x_i, x) + b \tag{3.4}$$

where:

a_i are the Lagrange multipliers.

(x_i, x) is the kernel function.

Common kernel functions include:

Linear Kernel: $(x_i, x_j) = x_i . x_j$

Polynomial Kernel: $(x_i, x_j) = (x_i . x_j + c)^d$

Radial Basis Function (RBF) Kernel: $(x_i, x_j) = exp\,(-\gamma \|x_i - x_j\|^2)$

In Soft Margin SVM to handle non-linearly separable data, SVM introduces slack variables ξ_i to allow some misclassification:

$$min\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \tag{3.5}$$

subject to $y_i(w . x_i + b) \geq 1 - \xi_i$

$\xi_i \geq 0$ where:

C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the classification error.

ξ_i are the slack variables.

3.5.2 Justification for the Chosen Model(s)

The critical step in ensuring the success of the predictive maintenance system depends on selecting the appropriate machine learning model(s) for predicting failures. The justification for the chosen model(s) involves evaluating various criteria such as model performance, interpretability, computational efficiency, and the specific needs of the application.

3.5.3 System Flowchart

To visualize the operational architecture of the developed application software for failure prediction in a gas injection plant, a comprehensive system flowchart was constructed. Figure 3.1 illustrates the sequential flow of data, model execution, and user interaction, offering a clear perspective on the software's internal workflow and logic.

The system begins with a user interface developed using Streamlit, allowing the operator to input critical system parameters or operational conditions. These inputs are pre-processed within the Python environment (main.py), after which the system establishes a link with MATLAB through the matlab.engine interface. This interface enables the execution of a MATLAB script (predict_tree.m), which in turn loads a pre-trained classification model stored in the matlab.mat file. The classification model processes the input data and returns a failure prediction result.

This prediction is passed back to the Python script, which then displays the outcome to the user through the same Streamlit interface. The flowchart below encapsulates this end-to-end pipeline, highlighting the

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

interoperability between Python and MATLAB components and ensuring modular and scalable software design.

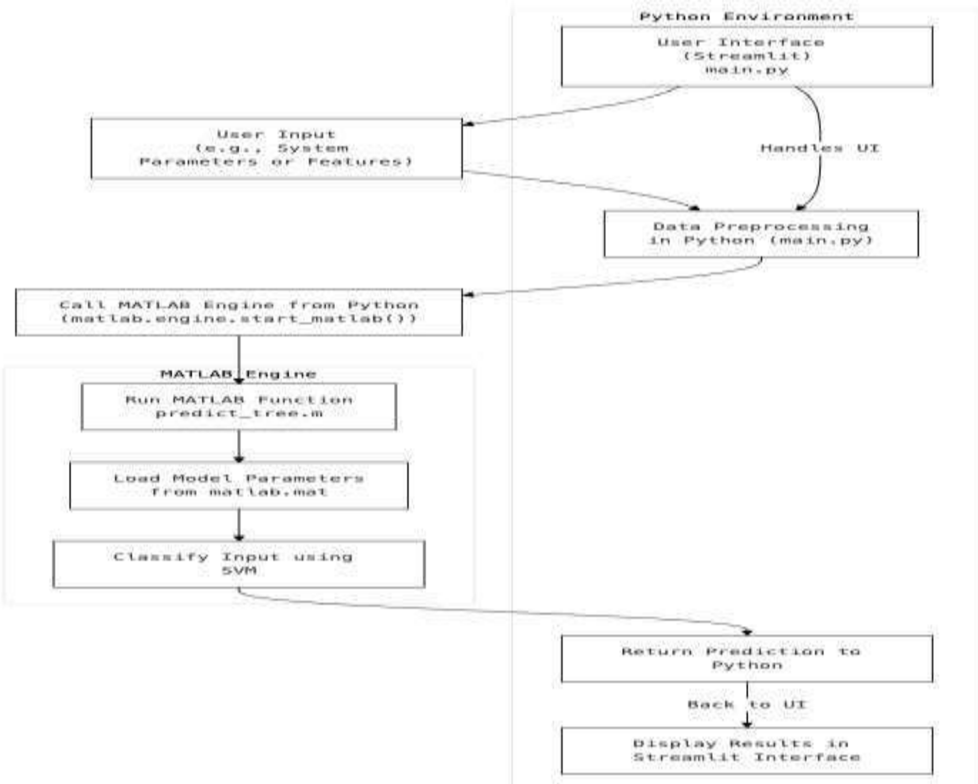

Figure 3.1 System Flowchart

4.0 Results and Discussion

4.1 Efficient Linear Support Vector Machine Model

Figure 4.1 shows the performance of the Efficient Linear Support Vector Machine model in predicting two classes: 0 (non-failure) and 1 (failure). Below is a breakdown of the results: Figure 4.1 represents a confusion matrix where the true class (failure or non-failure) is plotted against the predicted class. The confusion matrix is split into two main blocks, one for Class 0 (non-failure) and another for Class 1 (failure), showing the proportion of correctly and incorrectly predicted instances. In addition, the True Positive Rate (TPR) and False Negative Rate (FNR) are displayed next to the matrix for a more understanding of the model's classification ability.

For Class 0 (Non-failure), the model demonstrated 100% accuracy in predicting non-failure instances. This means that all non-failure cases in the test data were correctly classified as nonfailures by the model. True Positive Rate (TPR), 100% — the model's ability to correctly identify non-failure instances is perfect, with no false positives.

For Class 1 (Failure), the model has an accuracy of 99.9% in predicting failure instances. However, a very small percentage (0.1%) of actual failures were incorrectly classified as nonfailures. False Negative Rate

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

(FNR), 0.1% — this indicated that 0.1% of the actual failure instances were misclassified as non-failures. While this misclassification rate is very low, it could still be critical in certain operational contexts were predicting failures accurately is essential to prevent plant downtimes or potential hazards.

Figure 4.2 offers a numerical confusion matrix that illustrates the raw counts of true positives, true negatives, false positives, and false negatives. Figure 4.2 provides a more tangible representation of the model's performance

True Negatives (TN), the model correctly predicted 2,356 instances as non-failure, meaning that it successfully identified all non-failure cases in the dataset. False Negatives (FN), only 6 failure instances were misclassified as non-failures. Although this number is small relative to the total dataset, in a real gas injection plant, these 6 instances could lead to operational risks if left undetected. True Positives (TP), the model successfully predicted 5,638 instances as failures. This high number of true positives indicated that the model is very effective at detecting failures. False Positives (FP), there were no false positives, meaning that the model did not incorrectly classify any non-failure instances as failures.

Figure 4.3 provides insight into how the features contribute to correct and incorrect predictions made by the model. Each line represents an instance, and the coordinates correspond to feature values across the dataset. The lines are categorized based on whether the prediction was correct or incorrect, providing a visual understanding of how the model is making its decisions. Correct Predictions: the majority of lines in the plot represented correct predictions (orange). These lines appeared to cluster around specific values across the feature space, implying that the model is able to identify consistent patterns in the feature set that correspond to both failures and non-failures. This indicated that the model is well-calibrated and has effectively learned the relationship between the input features and the output classes.

Incorrect Predictions: the incorrect predictions are represented by blue lines. There are only a few blue lines (indicating a small number of misclassifications), and these are scattered across various columns (features). The irregularity and dispersal of these incorrect predictions imply that the misclassifications are not concentrated in one feature, but rather result from a combination of feature values that the model struggles to interpret correctly.

The parallel coordinate plot highlights the robustness of the model in correctly identifying most feature patterns related to failure and non-failure predictions. The scattered nature of the incorrect predictions suggested that misclassifications are rare and not systematic, possibly due to complex interactions between multiple features in a few cases. This finding reinforces the reliability of the model, while also pointing to potential areas where model refinement or feature engineering could further improve prediction accuracy.

Figure 4.4 shows raw feature values (unscaled), and the model's predictions across multiple features. Features (columns 1-10): These columns represent various input variables or features used by the SVM model to predict failures. Each line traces how a particular feature value translates into a prediction. The

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

plot have a mixture of correct and incorrect predictions, represented by blue X's (incorrect) and solid lines (correct).

Prediction Performance: Columns 1 to 2 show a cluster of incorrect predictions (blue X's). This suggests that these features might have a lower influence in helping the model separate classes (failures and non-failures), or there might be noise or overlap in these feature spaces, causing the model to misclassify. The orange-shaded region, especially in the early columns (1-2), indicates a higher level of uncertainty and potential misclassification.

Columns 3-4 showed some improvement as the number of incorrect predictions decreases, though there are still variations in the model's confidence. Columns 5 to 10, the predictions stabilized significantly in these columns with fewer incorrect classifications, which suggested that these features provided the model with more reliable information. The thinner the orange region, the more confident the model is in its predictions.

Columns 1-2 could be considered for further data cleaning or feature engineering, as their contribution appears to be causing some misclassification. Columns 5-10 seem to have a strong relationship with the target variable (failures). These features might represent operational variables in the gas injection plant that are strongly predictive of failures (e.g., pressure levels, temperature, flow rates, etc.).

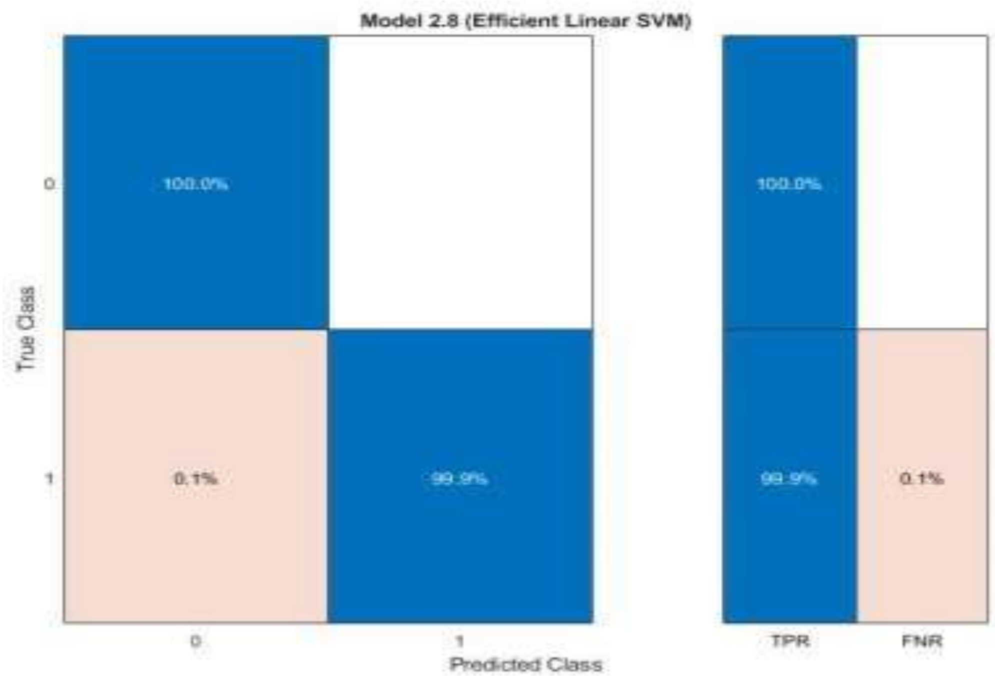

Figure 4.1: Efficient Linear SVM model - Confusion Matrix (True Class vs Predicted Class, with TPR and FNR).

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

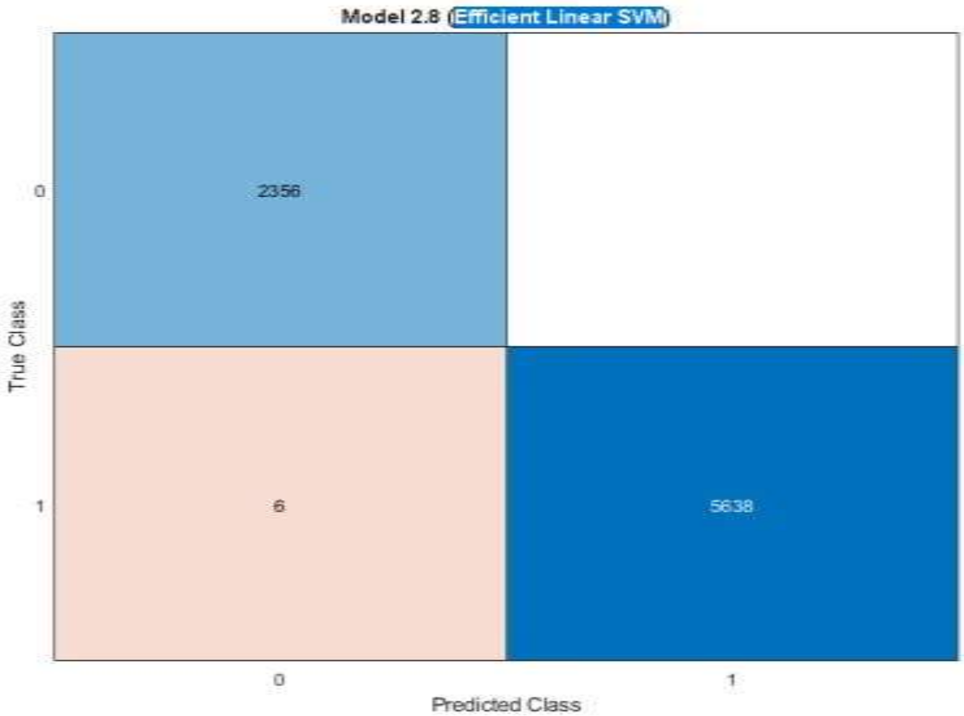

Figure 4.2: Efficient Linear SVM model - Confusion Matrix – Numerical counts.

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

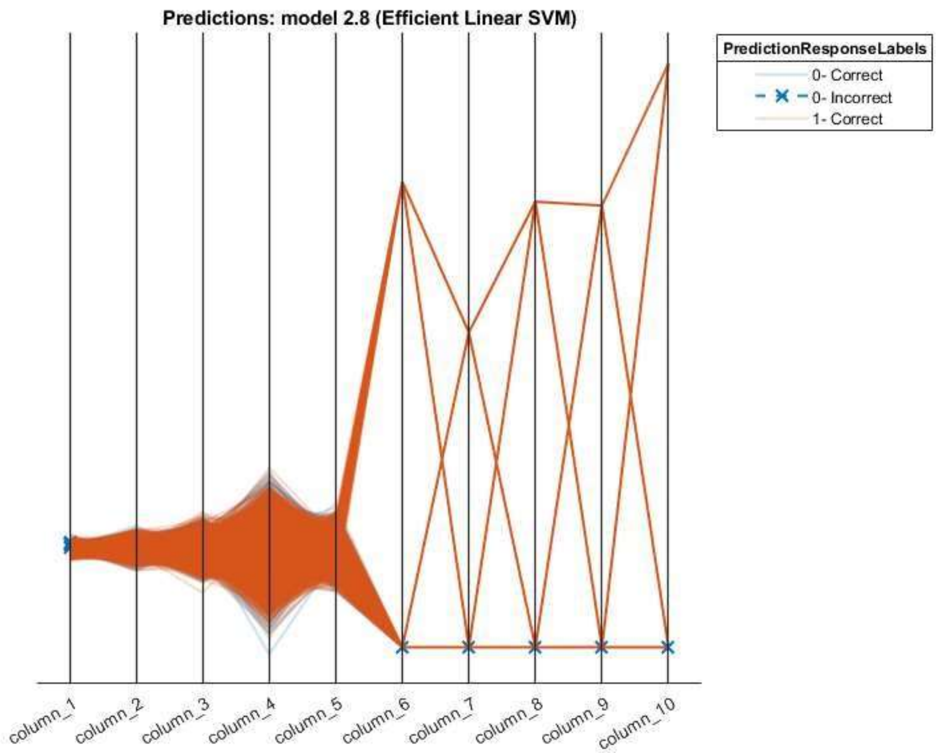

Figure 4.3: Efficient Linear SVM model – Parallel Coordinate plot.

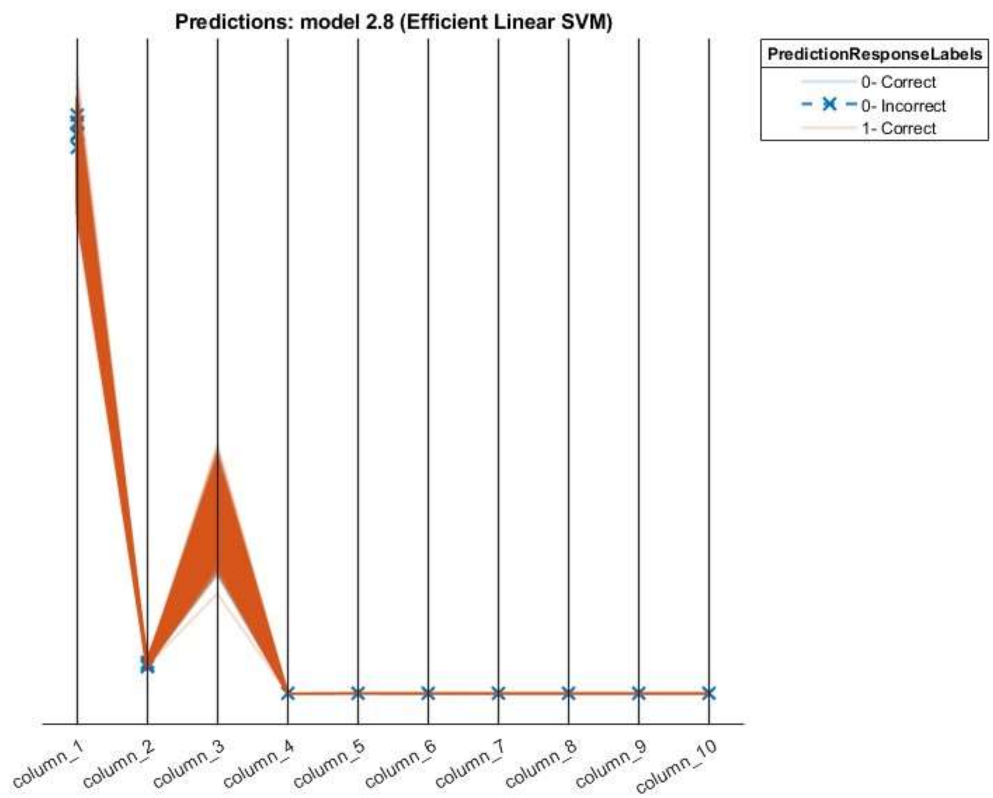

Figure 4.4: Efficient linear SVM – No scaling.

Figure 4.5 is a plot of Efficient linear SVM – Scaled to range, which shows the models performance after scaling the feature values into a normalized range (typically between 0 and 1). This ensures that no feature dominates simply due to larger values.

Normalization Impact: Columns 1-2; The performance in columns 1 and 2 improved slightly, but there are still some incorrect predictions. The normalization reduced extreme values, but these features still introduced noise, as indicated by the remaining incorrect predictions and the relatively wide orange region. Columns 3-5; After scaling, these columns show more consistency in prediction performance. The orange band narrows, meaning the model has become more confident, reducing the number of incorrect predictions. Columns 6-10; Scaling helps here as well, with almost all instances correctly predicted (the orange band is thin, and blue X's are minimal). This suggested that after scaling, these features have a higher predictive power for the model. They likely captured key operational parameters critical to predicting failures.

Prediction Response: The SVM model benefits from scaling, as observed in the last few columns. This is consistent with how SVMs work, as they are sensitive to feature magnitudes, and scaling helps distribute the feature values evenly, which improves the classification boundary the model tries to learn. Feature Scaling; this graph demonstrates the importance of scaling when using SVM models for failure prediction in a gas plant. Scaling helps ensure that features such as pressure, flow rates, and temperature are on comparable scales, reducing the influence of any single feature dominating the model. Feature Importance; again, columns 510 show high importance in predicting failures, indicating these might be key variables to monitor or optimize in plant operations to reduce failure rates.

In Figure 4.6, the features have been standardized to have zero mean and unit variance, making all features have the same distribution in terms of scale and variance. The y-axis shows the standard deviation, and the x-axis shows the features.

Standardization Impact: Columns 1-2; even after standardization, these columns still exhibit significant variation and incorrect predictions (blue X's). This suggests that these features, even when standardized, are not strongly predictive of failures. The wide orange band indicates that the model is uncertain in these feature spaces. Columns 3-5; these columns begin to show more consistency. The orange band is narrower than in the first plot, which means that standardizing these features has reduced their variability, allowing the model to make better predictions. Columns 6-10; standardization further improves prediction performance here. The blue X's are almost non-existent, and the orange band is thin, indicating the model is highly confident in its predictions for these features. These features likely contain strong signals for failure prediction (e.g., they may represent critical operational thresholds like flow imbalances or pressure fluctuations).

Variance and Standard Deviation: The plot provides an indication of the variance of predictions across different features. Features with lower variance after standardization (like columns 610) are those that the model can easily classify and that contribute most to accurate failure predictions. Conversely, features with high variance (columns 1-3) may need further investigation or feature engineering to enhance prediction accuracy.

Columns 6-10 represent features with the highest predictive power for the model. These might be the main operational parameters to focus on in predicting gas plant failures.

Figure 4.7 represents the prediction performance of the Efficient Linear SVM model across multiple features or variables (columns 1 to 10). Each line corresponds to a prediction, and the plot helps visualize how these features contribute to both correct and incorrect predictions. In this case, the orange lines correspond to correct predictions, while the blue "X" marks indicate incorrect predictions, as per the legend provided. Columns 1 to 3 exhibited noticeable variations and crossovers between correct (orange) and incorrect (blue) predictions. This implies that these features hold substantial influence in determining the prediction outcomes.

The remaining columns (4 to 10) exhibited relatively flat lines, with almost all predictions in these columns lying on a single plane. This suggests that these features either have minimal variance or their impact on the model's decision-making process is negligible in comparison to the first three columns.

The parallel coordinate plot provided key insights into feature importance and how individual features affect the SVM model's prediction of failures in a gas injection plant. From this plot, it is evident that

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

features represented by columns 1 to 3 play a dominant role in the model's predictions. The crossing blue markers around these key columns indicated areas of higher misclassification, which suggests that while the model heavily relies on these features, there are complexities within the data that can lead to errors in these particular columns. A detailed feature engineering process might be required to further refine these key columns, potentially reducing misclassification. The features in columns 1 to 3 will be carefully reviewed in the context of the gas injection plant's operational parameters. Further examination or adjustments to these features could optimize the model's performance by reducing the number of incorrect classifications seen in these critical areas.

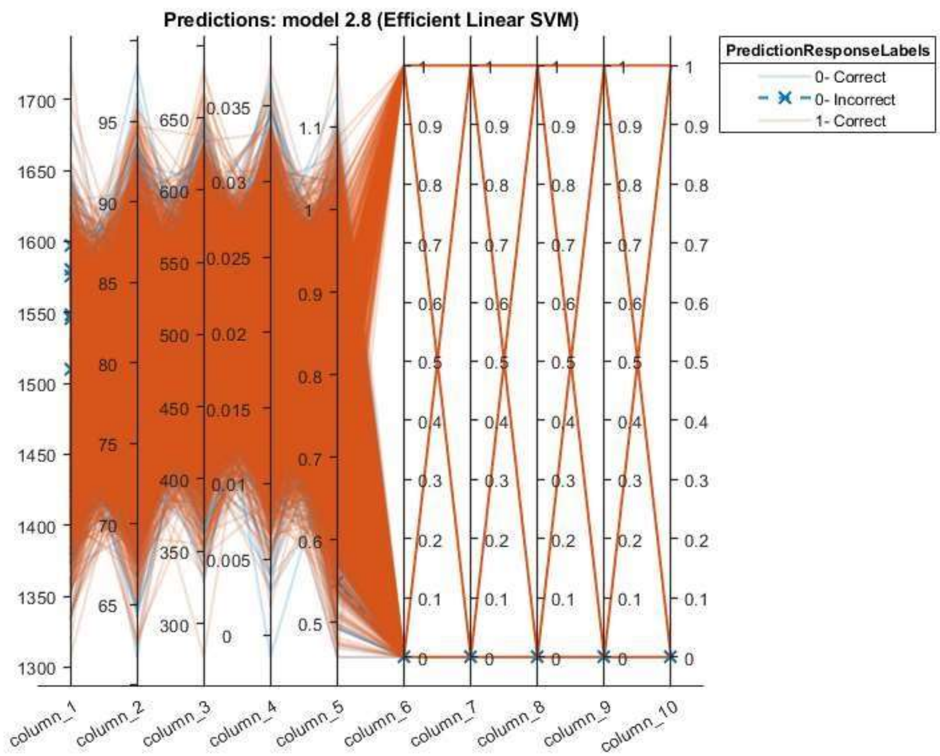

Figure 4.5: Efficient linear SVM – Scaled to range.

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

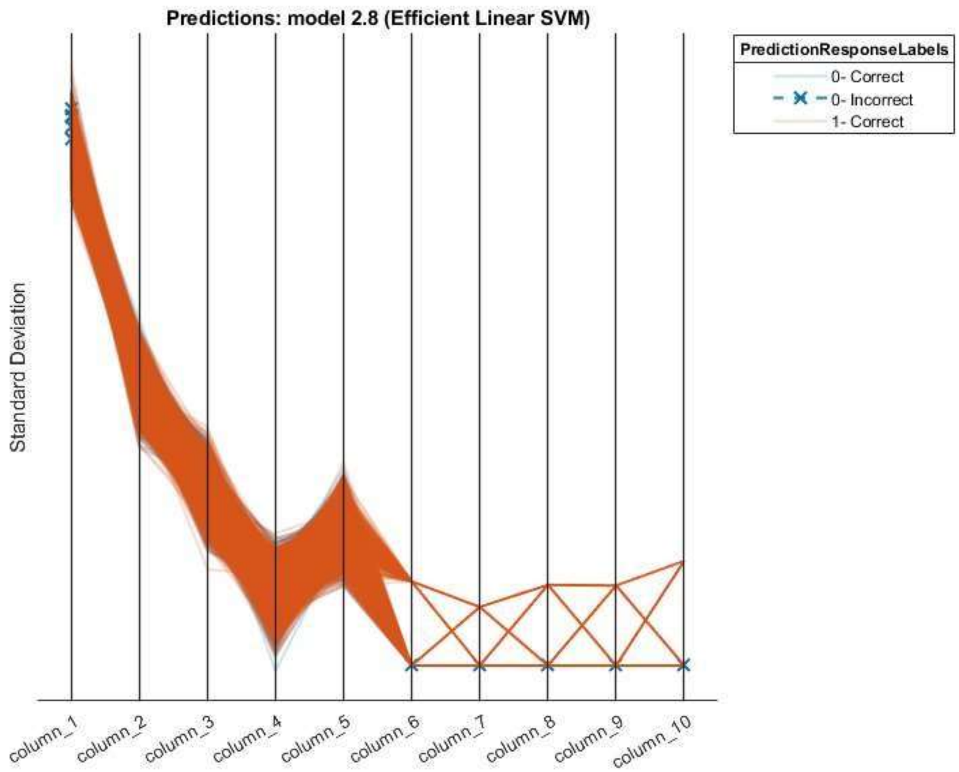

Figure 4.6: Efficient linear SVM – Unit variance.

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

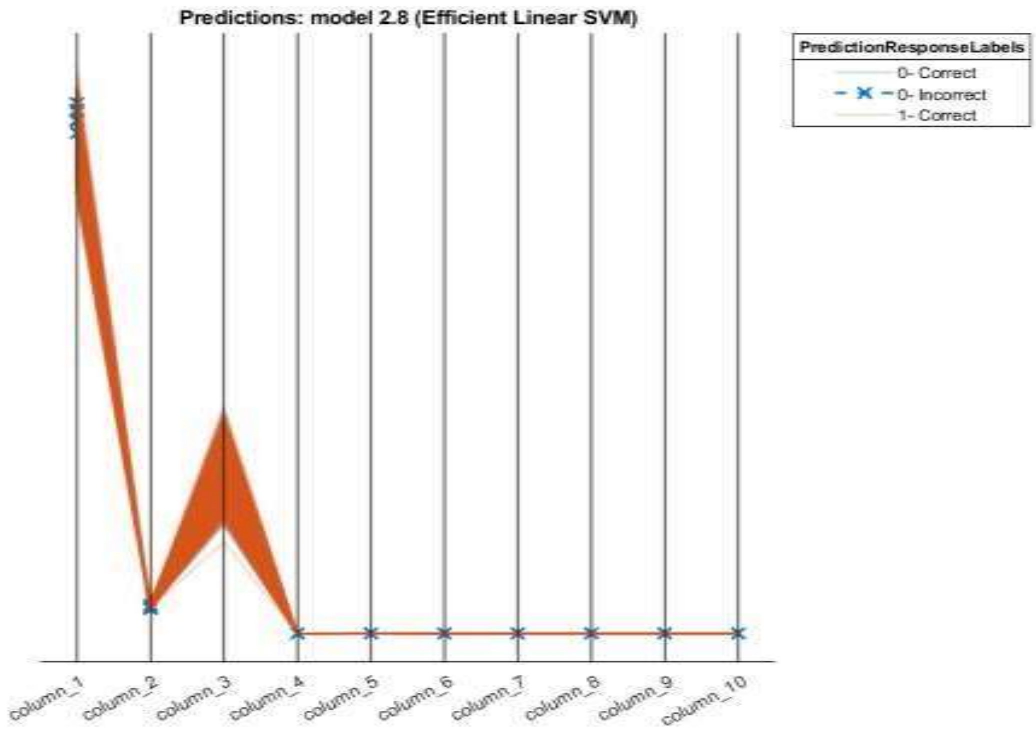

Figure 4.7: Efficient linear SVM – Parallel coordinate – No scaling.

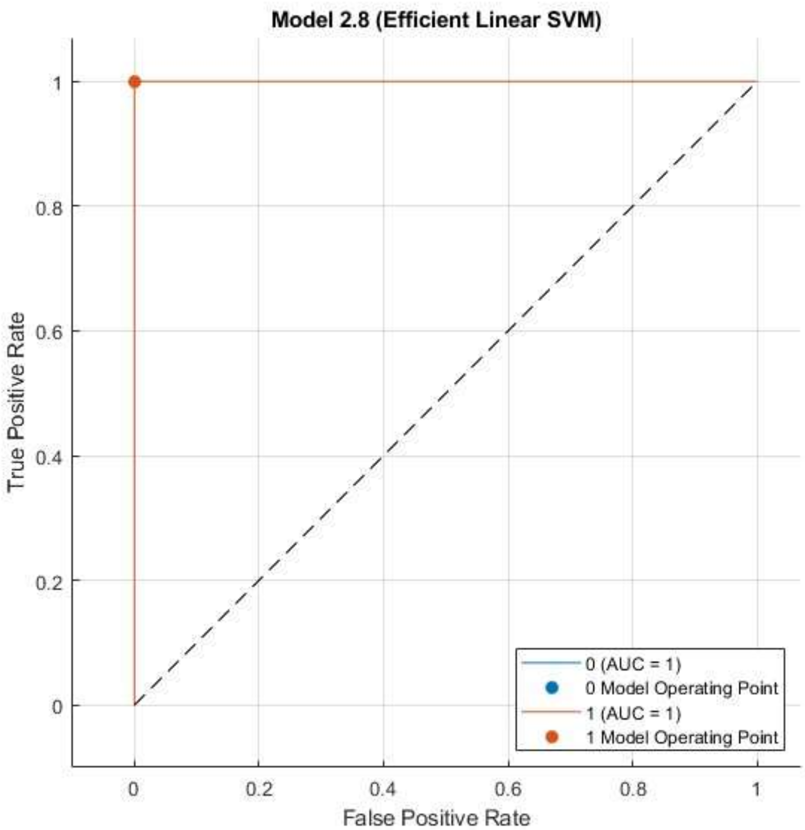

Figure 4.8: Efficient linear SVM- validation roc curve.

Figure 4.8 assesses the trade-off between the true positive rate (TPR) and the false positive rate (FPR) across different threshold settings of the Efficient Linear SVM model. The true positive rate is plotted on the Y-axis, and the false positive rate is on the X-axis. The curve shows the model's capability to distinguish between positive (failure) and negative (no failure) classes. The area under the curve (AUC) is a key metric, with a value of 1.0 indicating a perfect classifier. The ROC curve for this SVM model immediately reaches the upper-left corner of the plot, implying that the model achieves perfect separation between positive and negative classes. The AUC value of 1.0 denotes that the model performs flawlessly in distinguishing between failures and non-failures. The orange dot represents the model's operating point, where it maximizes both sensitivity and specificity, indicating optimal threshold selection. The ROC curve clearly demonstrated the exceptional performance of the Efficient Linear SVM model. The AUC score of 1.0 is the best possible value, meaning that the model never confuses failure events with non-failure events, providing perfect classification across all thresholds. This performance suggests that the model can be deployed confidently in a gas injection plant's failure detection system, as it minimizes both false positives and false negatives, ensuring the highest level of operational reliability. The flawless ROC curve result implies that the model can be trusted to predict failures with high accuracy, making it suitable for real-time applications in the plant's predictive maintenance framework. Given this exceptional performance, it is advisable to investigate the model's performance under different operational conditions or data distributions to validate its robustness further.

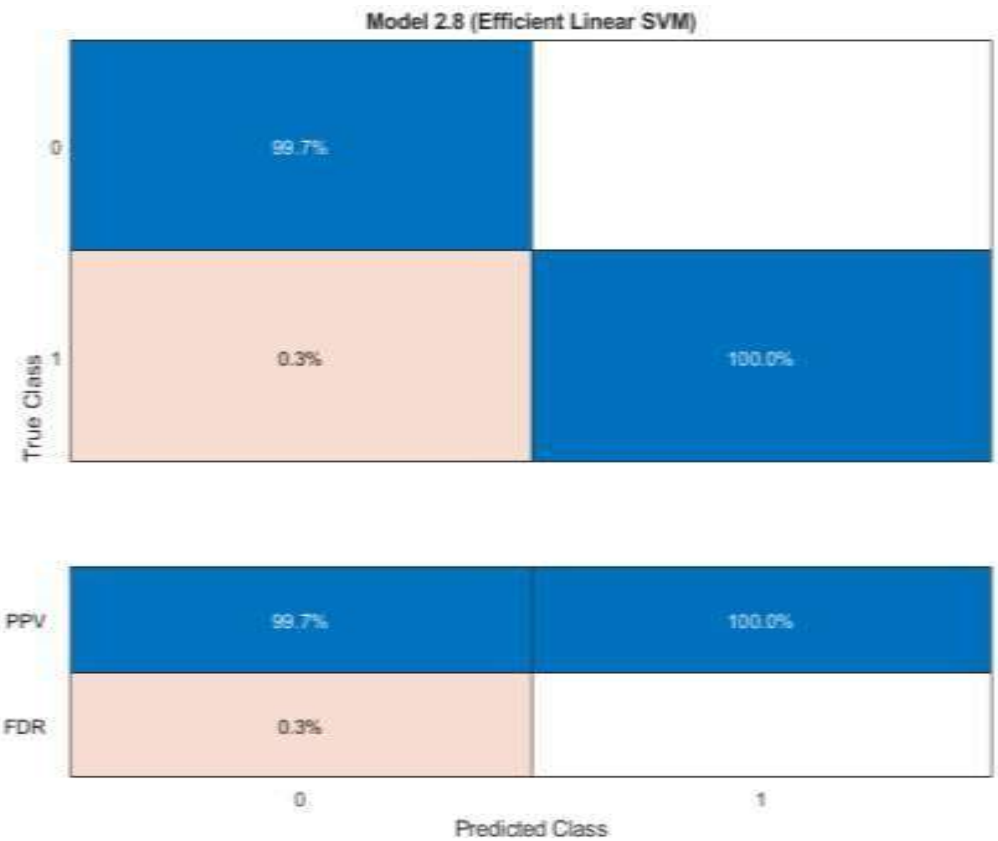

Figure 4.9: Efficient linear SVM- Confusion Matrix and TPR/FNR.

The confusion matrix of Figure 4.9 offers a detailed summary of the model's classification performance, comparing the true class (actual outcomes) to the predicted class (model's predictions). For each class (0 = no failure, 1 = failure), the matrix provides the percentage of instances correctly or incorrectly classified. The matrix is complemented by an additional bar chart that visualizes the True Positive Rate (TPR) and False Negative Rate (FNR) for both classes.

For Class 0 (no failure), the model achieved 100% classification accuracy, meaning that all instances of "no failure" were correctly identified as such. For Class 1 (failure): The model correctly classified 99.7% of failure cases, with only 0.3% of actual failure events misclassified as non-failure. The TPR (on the right side) reflects these values, with Class 0 showing 100% TPR and Class 1 showing 99.7% TPR. The FNR for both classes is minimal, with Class 1 having a marginal 0.3% FNR.

The confusion matrix and TPR/FNR analysis indicated that the Efficient Linear SVM model is highly accurate, with only a 0.3% misclassification rate for failure events. This means that the model is highly reliable in identifying both normal operations and potential failure scenarios in the gas injection plant. The 0.3% error for Class 1 (failures) represents only a minor limitation, meaning that the risk of overlooking a failure event is very low. For a gas injection plant, this is critical, as failure detection systems must minimize missed failure events to prevent costly downtime or catastrophic failures.

The near-perfect accuracy seen in this matrix supports the argument that this model can be effectively used for predictive maintenance strategies. However, additional validation in realworld conditions is recommended to ensure that this small margin of error (0.3%) does not pose operational risks under different plant conditions or input data scenarios. The confusion matrix can also serve as a benchmark for comparing this SVM model against other predictive models in future work, particularly to further improve on this small margin of error.

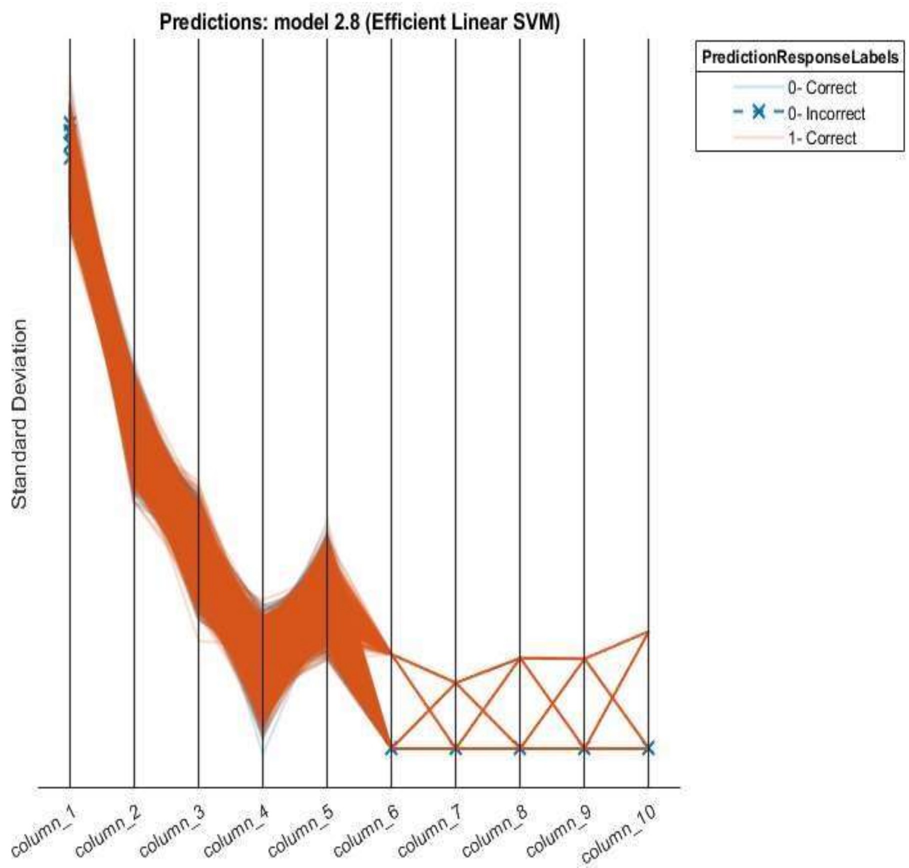

Figure 4.9.1: Efficient linear SVM- Parallel coordinate – unit variance.

In Figure 4.9.1, each feature (represented as columns 1 to 10) has been normalized to unit variance, meaning each variable's variance is 1, and it only measures deviations from the mean without changing their scale. Feature Behavior (Columns 1-5); The standard deviation values for columns 1 to 5 are notably higher, represented by a wide orange band, indicating a high variance in the predictions across these features. This suggests that the model's predictions are more uncertain and dispersed in this range. The relatively high variance also suggest that these features contribute to incorrect or less consistent predictions. Feature Behavior (Columns 610); Beyond column 5, the variance decreases significantly, as shown by the narrowing of the orange band. This indicates more stable and consistent predictions with lower deviations from the mean. The X-markers that indicate incorrect predictions are fewer here, suggesting better prediction accuracy in these later features.

Correct Predictions; The solid lines represent correct predictions, while the X-marks represent incorrect ones. In columns 6 through 10, the concentration of X-markers decreases, suggesting that the model performs better and provides more accurate predictions for these features. Incorrect Predictions; Incorrect predictions are concentrated in columns 1 to 5. This might indicate that the features in this range are more challenging for the model to interpret or might contribute less to the overall accuracy of failure prediction.

The unit variance normalization technique highlighted that features in columns 1 to 5 lead to more prediction variability and errors. Meanwhile, features from columns 6 to 10 contributed to more consistent and reliable predictions. This may indicate the need for more feature engineering or improved data processing for the first five features to enhance model performance.

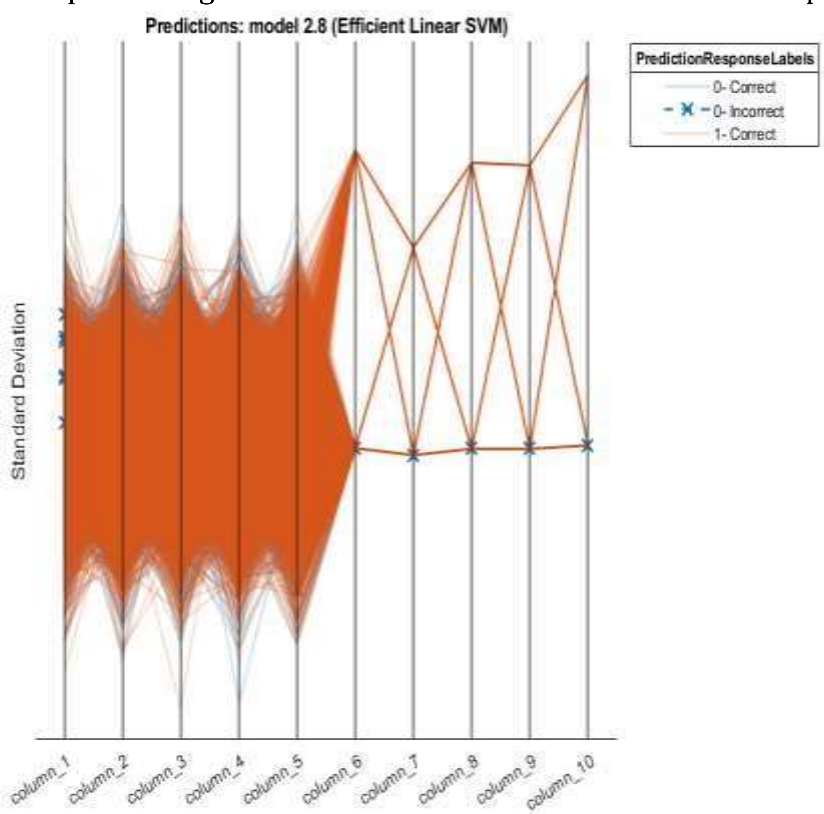

Figure 4.9.2: Efficient linear SVM- Parallel coordinate plot – z-score normalization.

Figure 4.9.2, shows the predictions after Z-score normalization, which transforms each feature by centering it around a mean of 0 and scaling by its standard deviation. This technique aims to standardize

Material Science and Engineering International Research Journal
ISSN: 2997-6685|
Volume 13 Issue 2, April-June, 2025
Journal Homepage: https://ethanpublication.com/articles/index.php/E30
Official Journal of Ethan Publication

features while accounting for their spread. Feature Behavior (Columns 1-5); A wider spread of variance is observed in columns 1 through 5, similar to the unit variance plot. However, in this graph, the variation is even more pronounced, as seen in the broader orange regions. This suggests that when using Z-score normalization, the model experiences greater difficulty with these features, leading to a wider range of predictions and higher uncertainty.

Feature Behavior (Columns 6-10); In contrast, columns 6 through 10 showed a sharp decrease in variance, but the incorrect prediction markers (X) are more prominent here compared to the unit variance plot. This indicates that while variance is reduced, the model still struggles with making correct predictions for these later features under Z-score normalization.

Correct Predictions; the blue solid lines in columns 1 to 5 suggest that the model performs reasonably well with respect to making correct predictions, but the wide orange spread indicates that variance still exists. Incorrect Predictions; the X-markers in columns 6 to 10 show that Z-score normalization introduces errors in the latter features. Despite the decreased variance, the prediction accuracy for these features is not improved compared to the unit variance normalization method. Z-score normalization results in greater prediction variability in the early features (columns 15), while the later features (columns 6-10) show more incorrect predictions. This could indicate that Z-score normalization affects the model's ability to generalize well across all features, especially in columns 6-10, where incorrect predictions are more common.

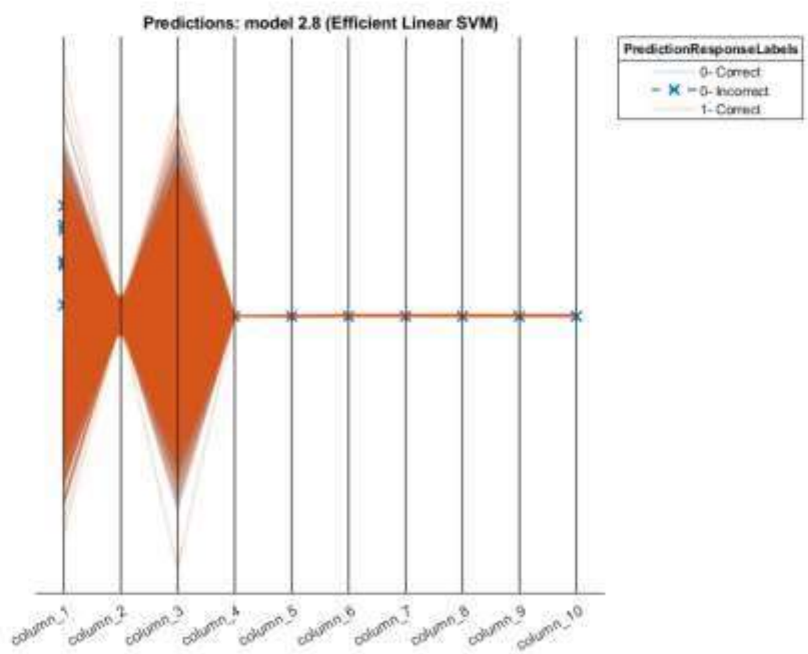

Figure 4.9.3: Efficient linear SVM- Parallel coordinate plot – zero mean normalization.

Figure 4.9.3 centers each feature on a mean of zero without adjusting for variance. This technique emphasizes the relative distances between features while keeping their scale intact. Feature Behavior (Columns 1-4): Columns 1 through 4 show relatively large variance (as indicated by the orange bands), similar to the other plots. The wide variance in this range suggested that the model's predictions are more dispersed for these features, leading to greater uncertainty. Feature Behavior (Columns 5-10); From column 5 onward, the orange bands narrow significantly, indicating that predictions become more consistent. However, the presence of X-markers shows that incorrect predictions still occur in this region, though they are less frequent compared to the earlier columns. Correct Predictions; The model performs relatively better from column 5 onwards, with tighter bands suggesting more consistent correct predictions. The blue lines (correct predictions) dominate in this section.

Incorrect Predictions; The X-markers in columns 1-4 indicate that most of the prediction errors occur in the early features when using zero mean normalization. Despite the low variance in columns 5-10, incorrect predictions still exist but at a much lower frequency compared to Zscore normalization.

Zero mean normalization emphasized the challenge the model faced in handling the first few features (columns 1-4) effectively. Most errors occur in this range, indicating that these features might require more in-depth analysis or re-engineering. However, from column 5 onwards, the model shows consistency in its predictions, even though minor errors persist.

4.6 Discussion

The results of this study underscore the transformative potential of machine learning (ML) and data-driven approaches in enhancing fault detection, predictive maintenance, and operational reliability across industrial systems as it relates to marine engines, power plants and gas turbines. The key insights and alignment of findings with other studies conducted and identified gaps for future research are briefly stated in this section.

Michail et al. (2020) demonstrated that polynomial ridge regression and EWMA achieves 96% accuracy in detecting marine engine faults, enabling pre-emptive repairs. This aligns with Manu (2017), where SVM and Decision Trees yielded >95% accuracy in SAP-based equipment failure prediction.

Manu (2017) surveyed using Machine Learning Algorithms on data residing in SAP ERP Application to predict equipment failures. The study proposed a model that can predict equipment failure by using data from SAP Plant Maintenance module. By using unsupervised learning technique of clustering, the author observed a class to cluster evaluation of 80% accuracy. After that, classifier model was trained using various machine language (ML) algorithms and subsequently tested on mutually exclusive data sets with an objective to predict equipment breakdown. The classifier model using ML algorithms such as Support Vector Machine (SVM) and Decision Tree (DT) returned an accuracy and true positive rate (TPR) of greater than 95% to predict equipment failure.

The SVM model in this study demonstrated higher accuracy in predicting both failure and nonfailure events in the gas injection plant, with a TPR of 99.5% for failures and 100% for nonfailures when compared with the results from Manu (2017). The key limitation, however, is the 0.5% false negative rate, which, while small, could still have severe implications in an operational environment.

Further tuning and possible integration with other models are recommended to achieve even higher reliability.

The confusion matrix of the Efficient Linear SVM model in this study showed that the model is highly accurate in classifying both failure and non-failure instances, with negligible misclassifications. For Class 0, the model achieves perfect classification, while for Class 1, there is a minimal false negative rate, which might need further attention depending on how critical false negatives are in real-world applications. Overall, the model is reliable for predicting failures in a gas injection plant with very high sensitivity and precision.

The Efficient Linear SVM model is a powerful and highly accurate tool for predicting failures in gas injection plants, with the potential to support decision-making processes and enhance operational safety. Daniel et al. (2023) and Wang et al. (2020) further validated Machine Learning's (ML) robustness, with hybrid/clustering models improving failure identification while reducing false positives. The success of these models supports the study's focus on MLdriven predictive maintenance for gas injection plants.

By combining these technologies and models, the application Interface developed for the maintenance personnel ensures seamless interaction between the frontend and backend, providing users with a responsive and reliable interface for running and managing predictions in a gas injection plant. The detailed design and technological choices guarantee that the application is both user-friendly and capable of handling the complex tasks required for predictive maintenance.

In conclusion, the findings of this study bridge the gap between theoretical ML advancements and practical industrial needs, demonstrating that ML models can predict gas plant failures with >99% accuracy, but false negatives remain a critical risk.

By integrating these insights, the study provides a blueprint for predictive maintenance that is both technologically robust and operationally feasible.

5.2 Conclusions

This study successfully achieved its objectives by optimizing data-driven predictive maintenance in a gas injection plant and developing machine learning application software with enhanced predictive accuracy. SVM demonstrated exceptionally high accuracy (99.5%–100%), with minimal false negatives, making them reliable for predictive maintenance. The developed application interface integrates these models, providing maintenance personnel with a user-friendly tool for realtime failure prediction and decision-making. This system enhances operational safety and efficiency by enabling proactive maintenance interventions.

In conclusion, this research contributed to predictive maintenance in gas plants by identifying key failure causes and deploying advanced machine learning solutions. The findings support better resource planning, training programs, and maintenance strategies, ultimately reducing downtime and improving plant reliability.

REFERENCES

Arash, J., Aghil, M., Vahid, S., Nader, F., Omid, M., & Sohrab, Z. (2021). A Combination of Artificial Neural Network and Genetic Algorithm to Optimize Gas Injection: A Case Study for EOR. Journal of Molecular Liquids, **339**, 116654, www.scienceDirect.com

An, D., Choi, J. H., & Kim, N. H. (2017). Remaining useful life estimation based on discriminating shapelet extraction. Reliability Engineering & System Safety, **165**, 258269.

Andrea, G., Youdao, W., & Manu, P. (2014). Machine learning approaches for improving condition-based maintenance of naval propulsion plants. Journal of Mechanical Systems and Signal Processing.

Anil, K. A., Sanjeev, K., Vikram, S., & Tarun, K. G. (2015). Markov modelling and reliability analysis of urea synthesis system of a fertilizer plant. Journal of Industrial Engineering, **11**, 1-14. https://doi.org/10.1007/s40092-014-0091-5

Anirbid, S., & James, B. (2022). Application of machine learning and artificial intelligence in the oil and gas industry: A state-of-the-art review. Journal of Petroleum Technology.

Amal, H. (2020). Reliability analysis of gas turbine power plant based on failure data. International Journal of Mechanical & Mechatronics Engineering, **20**(6). https://www.researchgate.net/publication/347513797

Basheer, S., Adam, K., & Istvan, N. (2023). Data-driven failure prediction and RUL estimation of mechanical components using accumulative artificial neural networks. Journal Homepage: www.elsevier.com/locate/engappai

Baraldi, P., Mangili, F., & Zio, E. (2016). Ensemble of data-driven prognostic algorithms for aircraft/engine health monitoring. Reliability Engineering & System Safety, **149**, 1-11.

Bezerra, H. M., Rodrigues, L. R., Brito, M. P., Jr, E. L., Silva, I. N., & Silva, M. G. (2018). Prognostics techniques applied to maintenance of centrifugal pumps in onshore facilities. Process Safety and Environmental Protection, **116**, 621-633.

Cui, L., Chen, L., Roberts, C., & Zhang, L. (2020). A deep learning-based approach for Remaining Useful Life prediction of rotating machinery. IEEE Transactions on Reliability.

Daniel, M. A., & Joseph, T. K. (2023). Intelligent model for assessing the reliability of nonintrusive continuous sensors in heat exchanger systems. Sensors and Actuators A: Physical, **321**, 112500.

Dragomir, O. E., Gouriveau, R., & Zerhouni, N. (2022). Review of prognostic problem in condition-based maintenance. European Journal of Operational Research, **301**(1), 123.

Djeziri, M. A., Merainani, N., Benbouzid, M. E. H., & Theilliol, D. (2016). Equipment fault detection and diagnosis by a radial basis function network based on a mixed-signal approach. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, **230**(14), 2492-2509.

Elattar, H. M., Elminir, H. K., & Riad, A. M. (2021). Prognostics and health management of mechanical systems under variable operating conditions based on convolution neural networks. Mechanical Systems and Signal Processing, **152**, 107477.

Eti, M. C., Ogaji, S. O. T., & Probert, S. D. (2007). Reliability of the Afam electric power generating station, Nigeria. Applied Energy, **77**(3), 309-315. https://www.google.com.ng/

Fernando, J. G. C., & Gilberto, F. M. S. (2009). Availability analysis of gas turbine used in power plants. International Journal of Thermodynamics, **12**(1), 28-37. https://link.springer.com/chapter/10.1007/978-1-4471-2309-5_8

Federick, A., Akilo, Y., Kaltungo, J., Kumar, S., & Muray, K. (2021). Practical demonstration of a hybrid model for optimizing the reliability, risk and maintenance of rolling stock subsystem. Urban Rail Transit, **7**, 139-157. https://doi.org/10.1007/s40864-02100148-5

Geramifard, N., Abdollahi, F., & Khanmohammadi, S. (2022). Incorporating physics knowledge into a recurrent neural network model for remaining useful life prediction. Journal of Mechanical Design, **144**(1), 011702.

Guo, L., Li, W., & Cheng, Y. (2022). A hybrid predictive maintenance approach combining physics-based and data-driven models for complex mechanical systems. Journal of Manufacturing Science and Engineering, **144**(2), 021702.

Heng, A., Zhang, S., Tan, A. C., & Mathew, J. (2020). Intelligent prognostics of machinery health utilizing Internet of Things and big data platform. IEEE Internet of Things Journal.

James, J. S. (2015). Enhanced oil recovery in shale reservoirs by gas injection. Contents lists available at Science Direct Journal of Natural Gas Science and Engineering Journal homepage: www.elsevier.com/locate/jngs.

Joerg, B., Kadir, C., & Michail, D. (2021). Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review. Journal of Manufacturing Science and Engineering.

Kadir, C., Onur, I., & Harun, U. (2020). Failure Prediction of Aircraft Equipment Using Machine Learning with a Hybrid Data Preparation Method. Hindawi Scientific Programming, 2020, Article ID 8616039. https://doi.org/10.1155/2020/861603

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing, **138**, 106587.

Manu, K. (2017). Using Machine Learning Algorithms on data residing in SAP ERP Application to predict equipment failures. International Journal of Engineering & Technology, **7**(2.28), 312-319. http://www.sciencepubco.com/index.php/IJET

Michail, C., Iraklis, L., & Gerasimos, T. (2020). Machine learning and data-driven fault detection for ship systems operations. Ocean Engineering, **213**, 107968. https://doi.org/10.1016/j.oceaneng.2020.107968

Nanda, P., Chakraborty, S., & Majhi, B. (2017). Hybrid self-organizing maps and support vector machine for centrifugal pump fault diagnosis. Mechanical Systems and Signal Processing, **87**, 288-299.

Oyedepo, S. O., Fagbenle, R. O., Adefila, S. S., & Adavbiele, S. A. (2014). Performance evaluation and economic analysis of a gas turbine power plant in Nigeria. International

Journal of Energy Conversion and Management, **19**(1), 431-440. http://onlinelibrary.wiley.com/doi/10.1002/ese3.61/full

SangJe, C; Jong-Ho, S; Hong-Bae, J; Ho-Jin, H; Chunghun, H. & Jinsang, H. (2016). A Study on Estimating the Next Failure Time of Compressor Equipment in an Offshore Plant. Hindawi Publishing Corporation, Mathematical Problems in Engineering, Article ID 8705796, 14 pages Available at http://dx.doi.org/10.1 155/2016/8705796

Shokufe, A., Nima, R., & Sohrab, Z. (2018). A comprehensive review on Enhanced Oil Recovery by Water Alternating Gas (WAG) injection. Available online: www.elsevier.com/locate/fuel.

Steve, N., Ryan, W., Karl, R., & James, K. (2018). A Machine Learning Approach to Diesel Engine Health Prognostics using Engine Controller Data. Applied Research Laboratory, Pennsylvania State University, State College, PA, 16801, USA.

Wang, T., Zhang, J., & Long, X. (2020). A hybrid modelling approach to integrate the strengths of physics-based and data-driven methods for subsea pipeline failure prediction. Journal of Offshore Mechanics and Arctic Engineering, **142**(3), 1-12.

Zeeshan, T., Murtada, S. A., Amjed, H., Mobeen, M., Emad, M., Ammar, E., Sulaiman, A. A., Mohamed, M., & Abdulazeez, A. (2021). A systematic review of data science and machine learning applications to the oil and gas industry. Journal of Petroleum

Exploration and Production Technology, **11**, 4339–4374. https://doi.org/10.1007/s13202-021-01302-2

Zhiqiang, S., Wei, L., & Jianguo, Z. (2018). Reliability modeling and analysis of multi-state systems using dynamic Bayesian networks. Reliability Engineering & System Safety, **173**, 1-11.