Published by Avanti Publishers

Global Journal of Energy Technology Research Updates

ISSN (online): 2409-5818

Machine Learning-Based Failure Prediction in Electric Submersible Pumps

Walid Mohamed Mahmud* and Masara A. Aldowro

Department of Petroleum Engineering, Faculty of Engineering, University of Tripoli, Tripoli, Libya

ARTICLE INFO

Article Type: Research Article

Academic Editor: Rongli Gao

Keywords:
Artificial lift
Fault diagnosis
Machine learning
Ensemble models
Predictive maintenance
Artificial intelligence in oilfields
Electric submersible pumps (ESP)

Timeline:
Received: April 04, 2026
Accepted: May 22, 2026
Published: August 01, 2026

Citation: Mahmud WM, Aldowro MA. Machine learning-based failure prediction in electric submersible pumps. Glob J Energ Technol Res Updat. 2026; 13: 11-27.

DOI: https://doi.org/10.15377/2409-5818.2026.13.2

ABSTRACT

Electric Submersible Pumps (ESP) are among the most widely used artificial lift methods in oil production wells. Despite their effectiveness, ESP systems are prone to unexpected failures that lead to significant production losses, costly workover operations, and extended downtime. Traditional maintenance strategies, such as corrective or time-based maintenance, are often insufficient in predicting early signs of ESP degradation. This paper presents a machine learning–based approach for predicting potential ESP failures using historical SCADA operational data. The framework integrates statistical feature engineering via a sliding window, SMOTE-based class balancing, and a comparative evaluation of four algorithms: Random Forest, XGBoost (XGB), Extra Trees, and Logistic Regression. The results demonstrate that tree-based ensemble models achieve exceptional classification accuracy up to 99.96%, significantly outperforming the baseline Logistic Regression model (99.83%) and effectively identifying early indicators of ESP performance degradation with minimal misclassifications. Feature importance evaluation revealed that thermal and hydraulic parameters serve as the primary diagnostic anchors. Integrating this predictive framework into automated systems can successfully transition field operations from reactive to proactive maintenance, optimizing the economic lifecycle of ESP systems.

*Corresponding Author
Email: w.mahmud@uot.edu.ly
Tel: +(218) 091 323 0494

1. Introduction

Artificial lift is a vital part of the life of many oil wells worldwide. Using several artificial lift methods can prolong the life of the wells and increase oil recovery significantly. One of the most widely used artificial lift methods nowadays is the electrical submersible pump (ESP) [1]. This artificial lift method may handle large volumes of hydrocarbons and is applicable under many conditions in both offshore and onshore reservoirs. Even though ESP has been applied extensively for many years, it still suffers from many failures due to electrical, mechanical, and operational problems associated with the ESP downhole assembly [1], often resulting in high operational costs and production downtime. Many case studies have reported severe electrical failures during ESP operations. Electrical failures related to the cable were mainly caused by insulation breakdown due to corrosion, material degradation, abrasion, and overload. Mechanical failures of the ESP assembly have been observed in onshore and offshore fields in more than thirty different countries [2]. Mechanical failures associated with the motor were mainly caused by rotor damage, corrosion, abrasion due to the presence of solids in the produced fluid, overheating due to improper cooling, and motor protector malfunction. Operational failures are usually focused on downhole conditions that may result in failures of the ESP assembly components. These failures can result from severe downhole temperature, sudden change in pressure gradients, introduction of gaseous phase due to reservoir pressure decline, and deposition of scale, asphaltene, or wax on the downhole ESP components.

Traditional ESP systems have been maintained using corrective or preventive approaches, such as scheduled inspections, periodic replacement of worn components, and post-failure repairs. While these methods can mitigate severe failures, they often result in high operational costs, unplanned downtime, and limited ability to detect early signs of malfunction. However, these approaches are often reactive and cannot always detect early signs of failure, highlighting the need for intelligent predictive approaches. Consequently, there is a growing interest in adopting predictive maintenance strategies using AI and machine learning, which can analyze real-time operational data and forecast potential ESP failures before they occur [3, 4]. Monitoring operational parameters is essential for ensuring the reliability of pumping systems [5].

According to Jani *et al.* [6], periodic performance testing is vital to avoid equipment damage, as strength degradation can manifest through increased vibrations and significant pressure drops. Their study emphasizes that integrating sensors to monitor pressure variations and energy consumption facilitates 'smart' process monitoring, which is a fundamental prerequisite for establishing effective maintenance schedules and identifying early signs of degradation. These findings underscore the necessity of developing advanced predictive models, such as the machine learning approaches applied to ESP systems in this study, to transform raw sensor data into actionable maintenance insights.

Building on this shift toward predictive maintenance, several recent studies have explored the use of advanced machine-learning models to enhance the early detection of ESP failures. Similarly, ML-based classification of electrical current and vibration data from ESP wells enables early diagnosis of abnormal operating conditions with high accuracy. These advances show that integrating AI/ML into ESP monitoring can significantly reduce unplanned downtime, operational costs, and production losses [7]. Therefore, understanding ESP failure mechanisms and leveraging AI/ML predictive tools are crucial for enhancing the reliability and efficiency of ESP-equipped wells.

Unlike previous ESP predictive maintenance studies that primarily focused on limited operational parameters or single-model approaches, the present work integrates statistical feature engineering, sliding window analysis, SMOTE balancing, and comparative evaluation of multiple machine learning algorithms using real ESP field data [8]. This framework enables early detection of multiple failure conditions while improving classification robustness and predictive reliability under real operating environments.

In this paper, operational and sensor data from ESP systems were analyzed to identify common failure patterns and root causes. An ensemble-based machine learning framework enhanced with data-balancing techniques was developed to accurately predict ESP failures before their occurrence. The performance of the predictive models in terms of accuracy, reliability, and ability to reduce unplanned downtime was evaluated. In addition, the potential impacts of AI/ML-based predictive maintenance on operational costs and production efficiency were assessed.

2. Related Work

In recent years, machine learning (ML) techniques have gained significant attention for predictive maintenance applications in Electric Submersible Pump (ESP) systems [9]. Traditional monitoring approaches mainly rely on threshold-based alarms and reactive maintenance strategies, which are often insufficient for detecting complex operational patterns that precede ESP failures. In contrast, ML-based approaches can analyze large volumes of operational and sensor data to identify abnormal conditions, predict potential failures, and support proactive maintenance planning.

Saptadi *et al*. [10] investigated the application of K-Nearest Neighbors (KNN) and Decision Tree (DT) algorithms for ESP failure prediction using operational sensor data. Their study incorporated data preprocessing techniques, feature selection, scaling, and Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance problems. Similar improvements in classification performance using SMOTE-based balancing techniques within hybrid predictive frameworks have also been reported in recent studies involving imbalanced industrial datasets [11].

The results demonstrated that KNN achieved higher predictive accuracy compared to DT in detecting early ESP failure patterns with the data distributions before and after SMOTE illustrated in Fig. (**1**) and Fig. (**2**), respectively. However, the study did not address scalability challenges associated with high-dimensional field datasets.

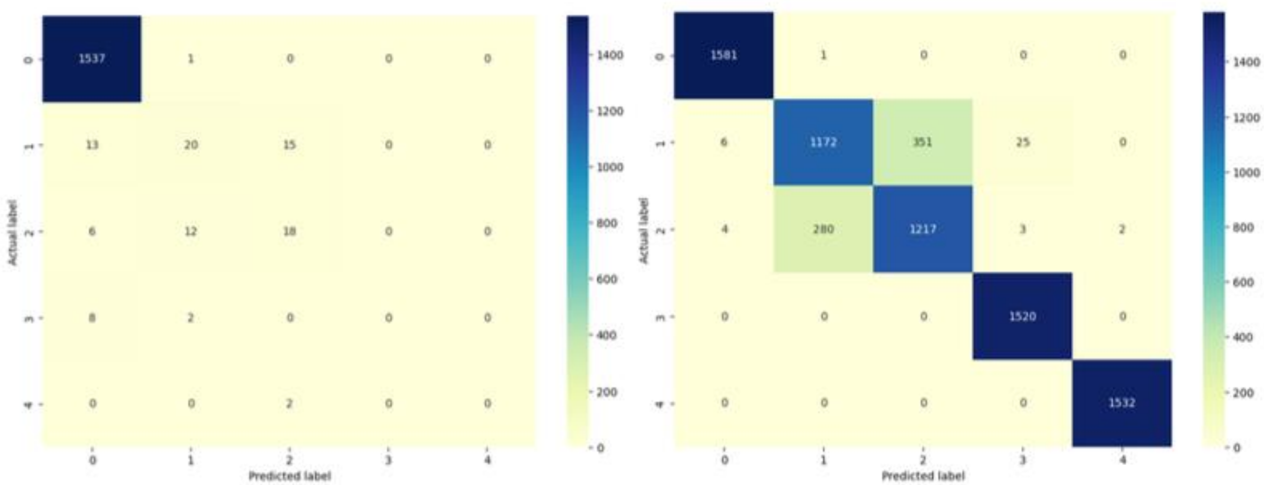

Figure 1: KNN - Data distribution before & after SMOTE.

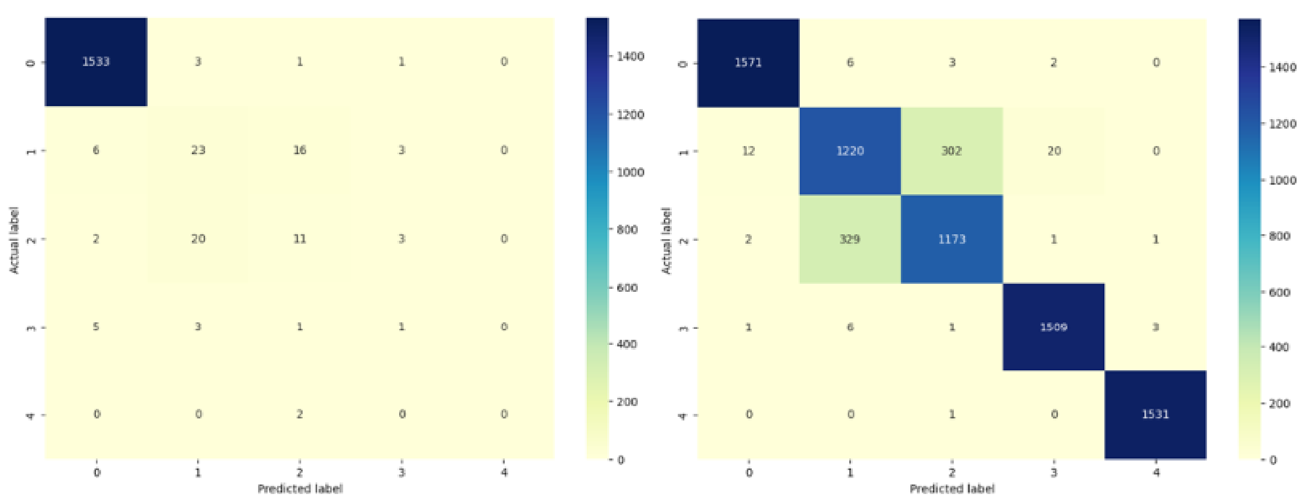

Figure 2: Decision Tree - Data distribution before & after SMOTE.

Batallas *et al.* [8] focused on integrating machine learning–based predictive models with operational maintenance practices to improve ESP reliability and reduce unplanned shutdowns. Real operational field data, including electrical, thermal, and operational parameters, were utilized for model development. Their findings demonstrated the practical applicability of ML techniques for supporting maintenance decision-making in real ESP field environments. Nevertheless, the study mainly relied on relatively simple classification approaches, leaving opportunities for further enhancement through advanced feature engineering and ensemble learning techniques.

Brasil *et al.* [7] applied multiple machine learning models, including Decision Tree (DT), Support Vector Machine (SVM), KNN, and Neural Networks, for diagnosing abnormal operating conditions in ESP systems. The models were trained using electrical current and operational sensor data collected from ESP installations. Their study demonstrated the effectiveness of ML techniques in detecting abnormal conditions at early stages and supporting preventive maintenance actions. However, the work primarily focused on fault diagnosis rather than long-term predictive maintenance and failure forecasting.

Christopher *et al.* [3] developed a machine learning–based predictive maintenance framework for ESP systems using historical and real-time operational data. Multiple algorithms [12], including Linear Regression, Random Forest, Gradient Boosting, and XGBoost, were evaluated for forecasting ESP failures. The results showed that ensemble-based models achieved superior predictive performance, highlighting the effectiveness of data-driven techniques for early fault detection. Despite the promising results, the study did not incorporate physics-based knowledge of ESP systems, which may affect model robustness under varying operating conditions.

Similar findings regarding the effectiveness of ensemble learning approaches for ESP predictive maintenance were also reported in recent studies employing trend-based labeling and time-aware predictive frameworks [13].

Jani *et al.* [14] highlighted the importance of monitoring pressure variation, cavitation behavior, and hydraulic performance in pumping systems to improve operational reliability and maintenance planning. Their studies demonstrated that abnormal flow conditions and cavitation phenomena can significantly reduce pump efficiency and contribute to operational instability and mechanical degradation. These findings support the importance of continuous monitoring and intelligent diagnostic approaches in industrial pumping applications.

Although previous studies demonstrated the effectiveness of machine learning techniques for ESP predictive maintenance, several limitations remain. Many studies relied on individual classifiers, simple feature engineering approaches, or limited operational parameters. In addition, challenges related to class imbalance, model generalization, computational efficiency, and real-time field applicability are still insufficiently addressed.

Previous studies have also highlighted that comparative evaluation of multiple machine learning models is essential for improving predictive maintenance performance and selecting computationally efficient frameworks for industrial applications [15].

Therefore, the present study proposes an ensemble-based machine learning framework integrated with statistical feature engineering, sliding window analysis, and SMOTE balancing techniques using real ESP operational field data. The proposed framework aims to improve predictive accuracy, enhance model robustness, and support practical implementation for real-world ESP predictive maintenance applications.

3. Methodology

This study is based on the Sharara Oil Field, a major onshore oil field asset located in the Murzuq Basin, southwestern Libya. The field represents a significant joint venture that manages its vital infrastructure and production activities. The field comprises several productive blocks as field's geological structure characterized by complex reservoir distribution, directly influences the operational conditions of ESPs deployed across the field. To evaluate the performance and predict failures of the ESPs, representative vertical and horizontal producer wells were selected that range in total depth from 4,556 to 6,208 ft. Fig. (**3**) and (**4**), respectively, show well schematic for the ESP completion for one of the vertical and horizontal wells utilized. Tables **1** and **2** show ESP System Specifications and downhole configuration for those wells.

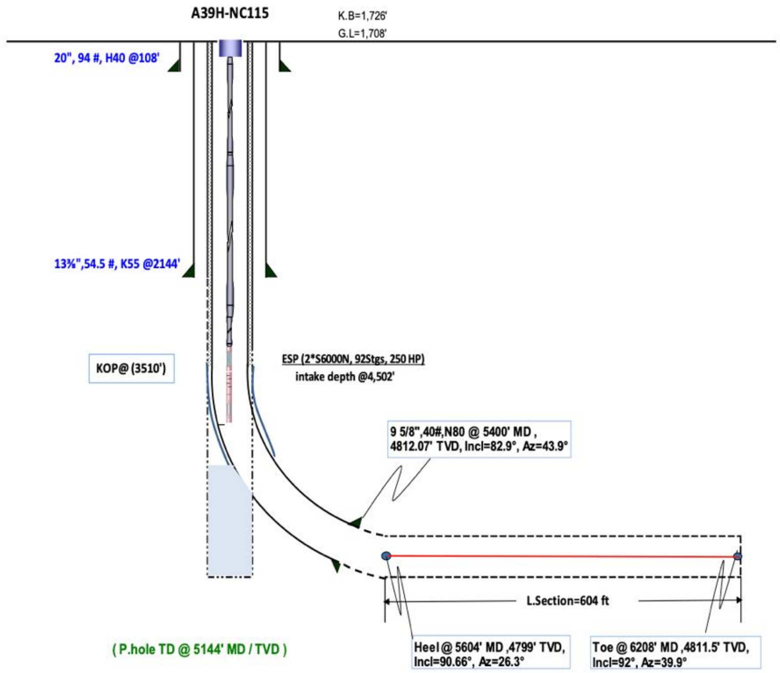

Figure 3: Well schematic for a horizontal well showing the ESP completion.

Table 1: ESP system specifications for a horizontal well.

Parameter	Value
Pump Type	S6000N
No. of Stages	92
Pump Intake Depth	4,502 ft (MD)
Motor Power	250 HP (RA-S)

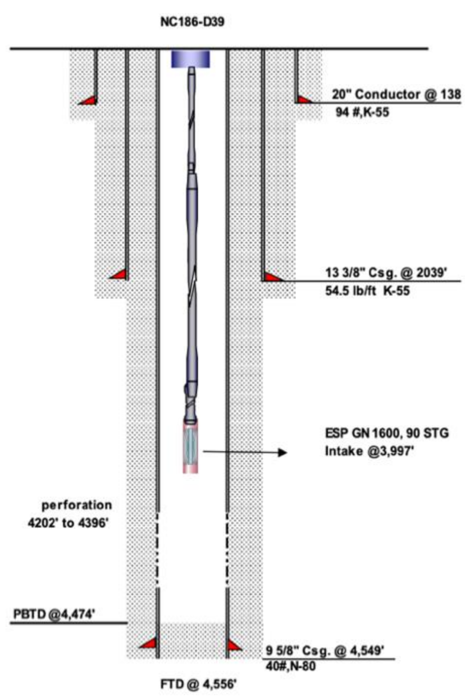

Figure 4: Well schematic for a vertical well showing the downhole configuration.

Table 2: ESP system specifications for a vertical well.

Parameter	Value
Pump Type	GN 1600
No. of Stages	90
Pump Intake Depth	3,997 ft
Motor Power	52 HP (DA-UT)

The success of the predictive maintenance model depends on selecting parameters that directly reflect the downhole conditions of the ESP. Data used were obtained from the SCADA system, specifically focusing on several vertical and horizontal wells producing from different formations to ensure a diverse range of failure scenarios.

For each well, eight key operational parameters were monitored in real-time. These features provide the necessary physical indicators for both normal operations and the onset of mechanical or electrical degradation. Table **3** provides description of input operational parameters.

Table 3: Description of input operational parameters.

No.	Parameter Name	Unit	Description
1	DH Discharge Temperature	°C	The temperature of the fluid as it leaves the pump stages.
2	DH Vibration	in/s	Measures mechanical stability: high levels indicate wear or imbalance.
3	DH Differential Pressure	psi	The difference between discharge and intake pressure; indicates pump efficiency.
4	DH Discharge Pressure	psi	The pressure at which the pump delivers fluid to the tubing.
5	DH Intake Pressure	psi	The pressure of the fluid entering the pump from the reservoir.
6	DH Intake Temperature	°C	The temperature of the reservoir fluid at the pump suction point.
7	DH Motor Temperature	°C	The internal temperature of the ESP motor; critical for preventing burnout.
8	Active Current Leakage	Amps	Measures electrical insulation health; a primary indicator for 'Motor Grounded' and 'Cable Grounded' failures.

Raw data collected from the SCADA system often contained noise or missing values due to sensor communication gaps. The total raw dataset utilized for this analysis comprised 130,354 per-minute records integrated from three different wells in the Sharara Oil Field: Well 1 (12,678 records), Well 2 (67,603 records), and Well 3 (50,073 records). This substantial dataset includes 99,312 records of normal operation and 31,042 records representing various failure transitions. To improve the model's accuracy, a multi-step pre-processing workflow was applied. All records with null values or non-logical sensor readings were identified and removed.

During this stage, several sensors reported constant zero values due to compatibility mismatches between specific downhole sensor models and the surface switchboard units. Since these zeros did not represent physical downhole conditions, they were treated as noise and moved to maintain a high-quality dataset. Instead of using raw instantaneous values, a 5-minute sliding window was applied to each of the eight parameters. This transformation effectively condensed the high-frequency per-minute data into a final dataset of 26,075 representative samples, capturing the underlying physical patterns while significantly reducing noise.

For every window, three statistical features were calculated: mean, standard deviation, and slope. The mean represents the average behavior of the sensor, the standard deviation captures the stability of the downhole conditions, and the slope measures the rate of change over time, which is critical for identifying gradual degradation trends. These 26,075 samples were assigned specific labels (0, 1, 2, or 3) to represent the operational health of the ESP based on a detailed manual inspection of sensor trends over a 30-day period prior to shut down.

Label 0 (Normal) represented stable operation, while Labels 1, 2, and 3 represented the gradual transition toward failure (pre-failure).

Well 1 (Electrical Fault - Motor Grounded): Labels were assigned based on identifying the onset of internal insulation degradation. The manual tracing prioritized Active Current Leakage and Motor Temperature to detect the transition leading to the motor being grounded.

Well 2 (Mechanical Fault): Labels were assigned based on abnormal physical behaviors, such as significant fluctuations in DH Vibration and Discharge Pressure, indicating mechanical wear or instability within the pump-motor assembly.

Well 3 (Electrical Fault - Cable Grounded): This well represents an external electrical failure. The labeling focused on analyzing specific Current Leakage patterns to distinguish cable grounding from internal motor faults before the final shutdown. The complete sequence of this data preparation and the subsequent machine learning pipeline is illustrated in Fig. (**5**).

Figure 5: Systematic flowchart of the proposed methodology.

The predictive diagnostic system was developed using the Python programming language within the Anaconda environment. To address the challenge of imbalanced fault classes, the SMOTE (Synthetic Minority Over-sampling Technique) was implemented, ensuring the models receive a balanced representation of all operational states. Four high-performance machine learning algorithms were selected for comparative analysis: Random Forest Classifier (RF), XGBoost Classifier (XGB), Extra Trees Classifier (ET) and Logistic Regression (LR). The implementation utilized the Scikit-Learn, XGBoost, and Imbalanced-learn libraries. Each model was trained and validated using a stratified 80/20 split. To ensure the reliability of the results and prevent overfitting, a 5-fold stratified cross-validation was also employed during the training phase. Crucially, the SMOTE technique was applied exclusively to the training sets to address class imbalance, ensuring that the models were evaluated on original, non-synthetic testing data to maintain the validity of the diagnostic results.

The sensor data from the wells were loaded and integrated into a unified data frame. This step ensures that the training process encompasses a wide range of operational conditions and different failure transitions, providing the models with a comprehensive dataset for robust failure prediction. The process involved cleaning the data by removing missing values and redundant columns, such as 'Unnamed' indices, and converting all sensor readings into a numerical format. Finally, the dataset was partitioned into features (X) and the target output (Y) to facilitate the machine learning implementation. Random Forest (RF) classifier was implemented with 100 estimators. This ensemble learning method was chosen for its high accuracy in handling complex nonlinear relationships within the ESP sensor data. The XGBoost model, an optimized gradient boosting library, was utilized to enhance predictive performance. It is particularly effective in identifying subtle patterns preceding mechanical failures. The Extra Trees (ET) classifier was employed to further test the robustness of the diagnostic system. This model introduces additional randomness to the tree-building process, which helps in reducing variance. As a baseline linear model, Logistic Regression (LR) was implemented to provide a clear interpretation of the feature coefficients and to compare linear vs. non-linear model performance. For all models, hyperparameters such as the number of estimators (e.g., 100 for RF) and maximum depth were optimized through a systematic grid search to ensure the best possible generalization on the Sharara field dataset. After balancing the dataset, the selected models were subjected to the training phase using the fit method on the resampled training data (Xres, Yres). This process allows the algorithms to learn the underlying patterns associated with different ESP operational states. Subsequently, the predict function was applied to the unseen testing set (Xtest) to generate diagnostic predictions (preds). This separation between training and testing ensures an unbiased evaluation of the models' ability to generalize to new sensor data.

To evaluate the effectiveness of the implemented diagnostic models, the performance was measured using the Classification Report and the Confusion Matrix. The primary metric utilized is Accuracy, which quantifies the overall proportion of correct predictions across the four operational states (Healthy, Motor Grounded, Mechanical Failure, and Cable Grounded). Furthermore, the implementation included the calculation of Precision, Recall, and F1-score for each specific fault category to ensure the system’s reliability in detecting critical ESP failures without bias, especially after applying the SMOTE balancing technique.

Additionally, the ROC-AUC (Area Under the Receiver Operating Characteristic curve) was calculated to measure the models' ability to distinguish between different failure modes, providing a more comprehensive evaluation of the diagnostic performance.

4. Results and Discussion

Analysis of the results obtained from four machine learning "algorithms RF, XGB, ET, and LR are presented". These evaluations are supported by diagnostic audit logs for specific wells to ensure a comprehensive assessment of each model's predictive reliability across various ESP failure modes.

The Random Forest (RF) model demonstrated a strong ability to capture complex variations in sensor readings. The confusion matrix (Fig. **6**) indicates that only 2 instances of 'Motor Grounded' (Class 1) were misclassified as 'Mechanical Failure' (Class 2). This slight overlap is physically justified as severe electrical insulation breakdown can trigger mechanical vibrations and torque fluctuations that mimic mechanical wear signatures, leading to these localized boundary errors. This performance is attributed to the model's ability to capture complex variations in sensor readings, which exhibit distinct physical patterns during failure events. as detailed in the classification report presented in Table **4**.

Table 4: Classification report for random forest model.

Support	F1-Score	Recall	Precision	Class
3973	1.00	1.00	1.00	Healthy
48	0.979	0.958	1.00	Motor Grounded
976	0.999	1.00	0.998	Mechanical Failure
218	1.00	1.00	1.00	Cable Grounded
5215	**0.9996**			Overall Accuracy

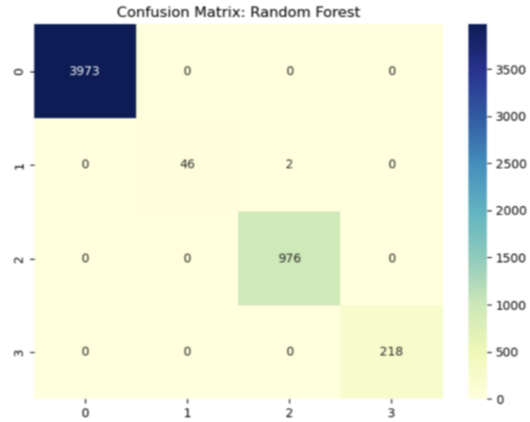

Figure 6: Confusion matrix for RF classifier.

To understand the logic behind these predictions, the Feature Importance was analyzed, as shown in Fig. (**7**). This ranking illustrates the relative contribution of each sensor to the diagnostic process. The RF model relies most heavily on DH Intake Temp MEAN, followed closely by DH Discharge Temperature MEAN and DH Motor Temp MEAN as the primary diagnostic anchors. This strategic focus on thermal parameters indicates that the model prioritizes thermodynamic shifts within the ESP system to distinguish between various failure modes. By assigning significant weight to these temperature-based indicators, the algorithm effectively captures the underlying physical patterns that precede both electrical and mechanical degradation.

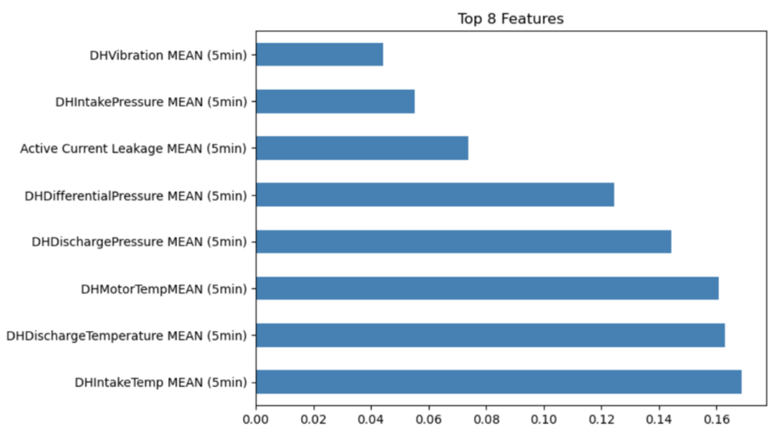

Figure 7: Feature Importance for the RF model.

In addition to the overall performance metrics, a diagnostic audit log was generated to evaluate the Random Forest model's accuracy on specific real-world wells. This stage of testing is crucial to verify that the model maintains its predictive power when applied to individual well datasets with varying operational conditions. As shown in Table **5**, the model was tested across selected wells representing different failure categories, ensuring robust validation against field-specific data.

Table 5: Diagnostic audit log for random forest on selected wells.

Cases	Predicted State	Actual State	Well
46	Motor Grounded	Motor Grounded	1
2	Mechanical Failure	Motor Grounded	1
218	Cable Grounded	Cable Grounded	3
976	Mechanical Failure	Mechanical Failure	2

Regarding the minor misclassifications observed in Well 1, where two instances of 'Motor Grounded' were predicted as 'Mechanical Failure', this is attributed to the physical nature of severe electrical grounding failures. Such faults can induce rapid thermal expansions and torque fluctuations that mimic the mechanical instability and pressure signatures characteristic of pump-motor wear. Since the RF model relies significantly on temperature-based features specifically DH Intake Temp and DH Discharge Temp—these overlapping physical patterns led to a localized boundary misinterpretation where electrical disturbances manifested as mechanical anomalies. However, identifying 46 out of 48 electrical fault instances confirms the model's high reliability in distinguishing complex failure modes under real field conditions.

The performance of the XGBoost (XGB) model, as detailed in Table **6**, is primarily attributed to its gradient boosting architecture. This approach effectively captures complex variations in sensor readings, although it faced slight difficulty where early-stage failure signatures overlapped with stable operational conditions. as illustrated in the confusion matrix in Fig. (**8**).

Table 6: Classification report for XGB model.

Support	F1-Score	Recall	Precision	Class
3973	1.000	1.000	0.999	Healthy
48	0.989	0.979	1.000	Motor Grounded
976	0.999	0.999	1.000	Mechanical Failure
218	1.000	1.000	1.000	Cable Grounded
5215	**0.9996**			Overall Accuracy

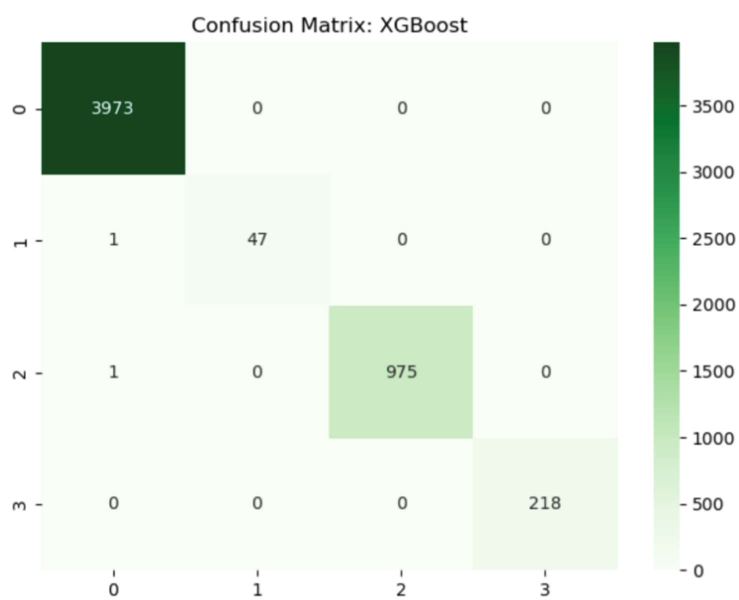

Figure 8: Confusion matrix for XGBoost (XGB) classifier.

To delve into the decision-making logic of the XGBoost algorithm, the Feature Importance was extracted as shown in Fig. (**9**). XGBoost demonstrates a sophisticated prioritization of data, where DH Discharge Pressure MEAN acts as the primary diagnostic anchor, followed by DH Intake Temp MEAN and DH Discharge Temperature MEAN. This strategic focus indicates that the model is exceptionally sensitive to hydraulic shifts and thermodynamic variations within the well. By assigning the highest weight to discharge pressure, the model ensures that even the most subtle pressure fluctuations—which are often the earliest precursors to pump failure—are captured, providing a robust defense against misclassification in complex downhole environments.

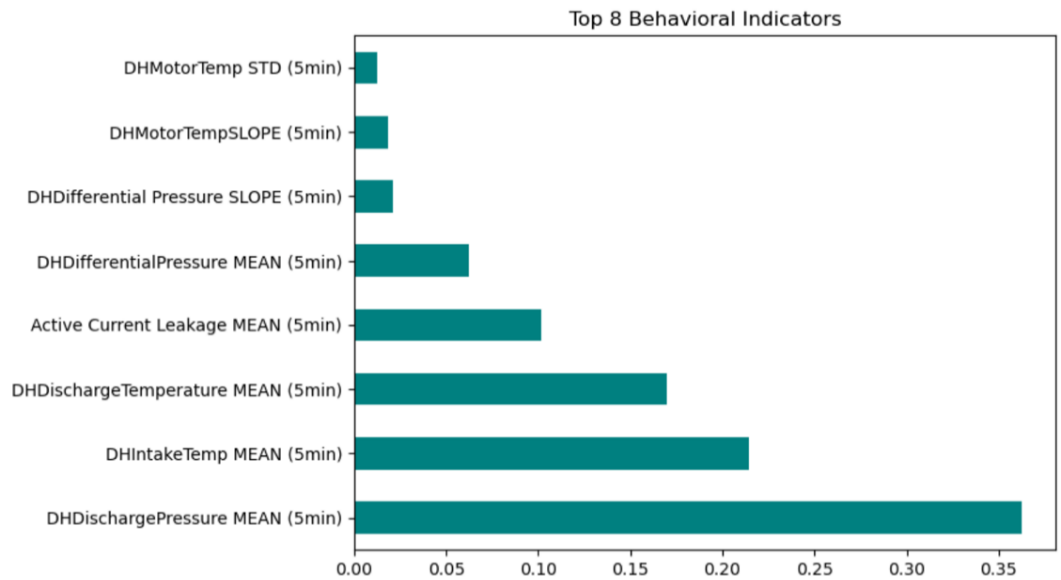

Figure 9: Feature Importance for the XGB model.

To further validate these findings, the performance of the XGB model was also assessed through a diagnostic audit log applied to specific well datasets. This evaluation ensures that the model maintains consistent predictive capabilities across different well environments and failure modes. As presented in Table **7**, the audit covers the diagnostic outcomes for three wells.

Table 7: Diagnostic audit log for XGB on selected wells.

Cases	Predicted State	Actual State	Well ID
47	Motor Grounded	Motor Grounded	1
1	Healthy	Motor Grounded	1
218	Cable Grounded	Cable Grounded	3
975	Mechanical Failure	Mechanical Failure	2
1	Healthy	Mechanical Failure	2

The diagnostic performance of the XGB model reveals subtle insights into the physical complexity of ESP failures. In (Well 1), the single misclassification of 'Motor Grounded' as 'Healthy' suggests that the initial insulation breakdown had not yet generated a thermal or electrical signature strong enough to be distinguished from normal operation by the gradient boosting algorithm.

Furthermore, the instance in (Well 2) where a 'Mechanical Failure' was predicted as 'Healthy' highlights a critical diagnostic challenge; in the incipient stages of mechanical wear, sensor deviations may remain within the statistical thresholds of normal operation. These results demonstrate that while XGBoost is highly sensitive, the physical overlap between early-stage mechanical degradation and stable operational signatures can occasionally mask the fault's presence in certain wells like (Well 2).

The classification performance of the Extra Trees (ET) model, as summarized in Table **8**, reveals a high level of diagnostic precision. The algorithm maintained perfect precision for the 'Motor Grounded' category, with the Confusion Matrix (Fig. **10**) confirming that only two instances were misclassified as 'Mechanical Failure'. This performance highlights the model's effectiveness in isolating complex failure signatures within the ESP sensor data, even when dealing with the high-dimensional noise inherent in the Sharara field datasets.

Table 8：**Classification report for extra trees model.**

Support	F1-Score	Recall	Precision	Class
3973	1.00	1.00	1.00	Healthy
48	0.979	0.958	1.00	Motor Grounded
976	0.999	1.00	0.998	Mechanical Failure
218	1.00	1.00	1.00	Cable Grounded
5215	**0.9996**			Overall Accuracy

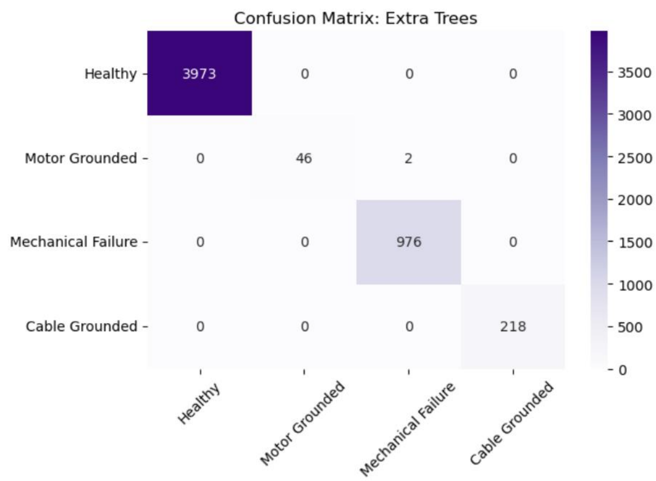

Figure 10: Confusion matrix for Extra Trees (ET) classifier.

The relative contribution of each sensor parameter to the ET model's decision-making process is illustrated in Fig. (**11**). The analysis identifies DH Intake Temp MEAN as the most influential feature, followed by DH Discharge Temperature MEAN and DH Motor Temp MEAN. This strong reliance on thermal differentials highlights the critical role of temperature variations in diagnosing ESP failure modes. Conversely, parameters such as Active Current Leakage exhibit minimal predictive importance in this model, indicating that thermodynamic readings are the primary drivers for fault identification within the ET algorithm.

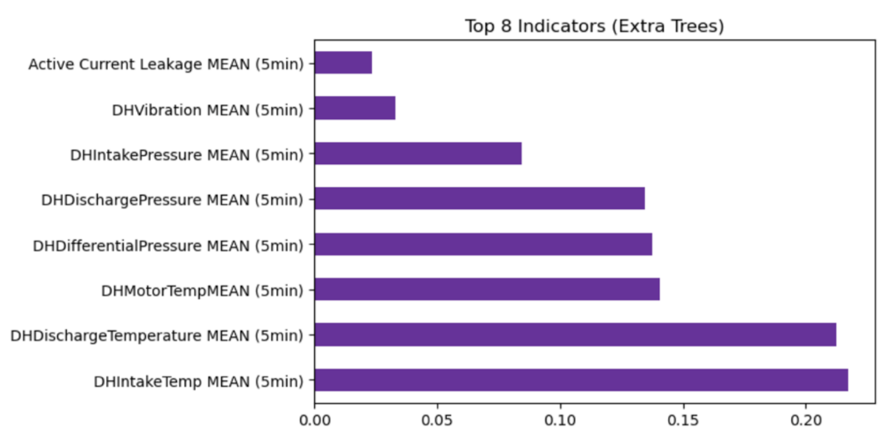

Figure 11: Feature Importance for the ET model.

To establish a deeper level of confidence in the ET model's diagnostic accuracy, its performance was further scrutinized through a diagnostic audit log targeting specific operational wells. This step is essential to confirm that the algorithm remains robust when faced with diverse well conditions and different failure categories. As detailed in Table **9**, this audit focuses on the classification results obtained from a sample of three wells.

Table 9: Diagnostic audit log for extra trees on selected wells.

Cases	Predicted State	Actual State	Well
46	Motor Grounded	Motor Grounded	1
2	Mechanical Failure	Motor Grounded	1
218	Cable Grounded	Cable Grounded	3
976	Mechanical Failure	Mechanical Failure	2

The audit log results confirm that the ET model provided high-fidelity diagnostics. The misclassification in (Well 1), where two Motor Grounded instances were tagged as Mechanical Failure, highlights a known physical complexity. This misclassification may be related to overlapping thermal and mechanical behavior during severe fault conditions. Since the ET algorithm relies heavily on thermal features like Intake and Discharge temperatures, this specific physical overlap led to a localized boundary error. Nonetheless, achieving a correct identification rate of 96% for electrical faults in this well underscores the model's integrity.

The classification performance of the Logistic Regression (LR) model, as summarized in Table **10**, demonstrates a high level of diagnostic precision. While the model successfully classified the 'Motor Grounded' state with perfect scores, the Confusion Matrix (Fig. **12**) reveals a total of 9 misclassifications in other categories. These discrepancies, where healthy instances were flagged as failures and some actual failures were missed, suggest that the linear decision boundaries of the LR algorithm are more sensitive to noise compared to ensemble-based methods. This highlights the sensitivity limits of linear models when dealing with the complex, non-linear degradation patterns typical of ESP failures.

Table 10: Classification report for logistic regression model.

Support	F1-Score	Recall	Precision	Class
3973	0.999	0.998	0.999	Healthy
48	1.00	1.00	1.00	Motor Grounded
976	0.997	0.999	0.995	Mechanical Failure
218	0.993	0.995	0.991	Cable Grounded
5215	**0.9983**			Overall Accuracy

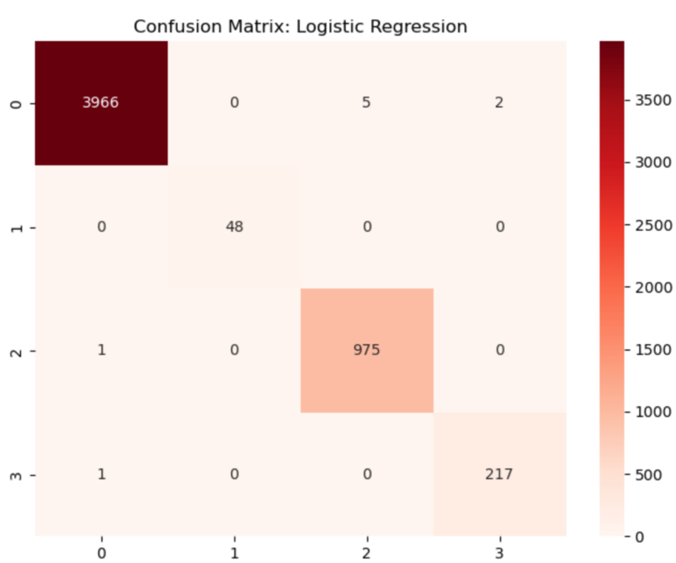

Figure 12: Confusion matrix for Logistic Regression (LR) classifier.

The relative importance of the input features for the Logistic Regression (LR) model is illustrated in Fig. (**13**). The analysis of the model's coefficients reveals that 'DHintake Temp MEAN' is the most influential predictor, followed by 'DHDischargeTemperature MEAN' and 'Active Current Leakage MEAN'. Interestingly, the LR model also assigns high significance to statistical variations, such as the standard deviations of motor and intake temperatures (DHMotorTemp STD and DHintakeTemp STD). This indicates that the linear classifier relies on both the absolute thermal trends and the stability of the sensor readings to establish its decision boundaries between operational and failure states.

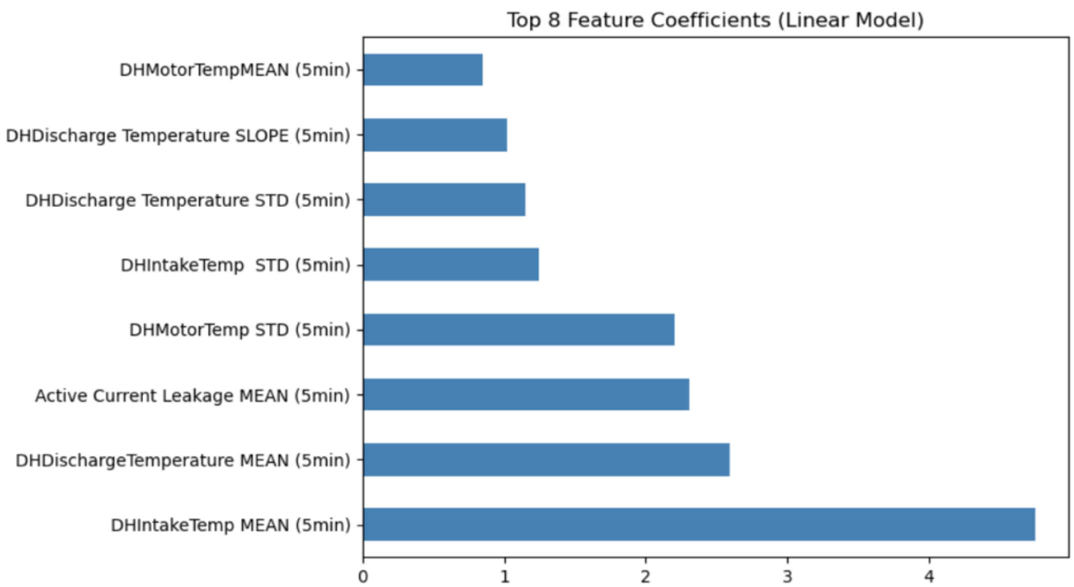

Figure 13: Feature Importance for the LR model.

To further validate the practical reliability of the Logistic Regression (LR) model, a diagnostic audit was conducted using real-time operational data. This step is essential to evaluate how the linear classifier performs when subjected to field-specific sensor noise and actual failure events. The following audit log, presented in Table **10**, provides a detailed breakdown of the model's predictive accuracy across these three diverse well environments. Table **11** illustrates the diagnostic audit log for logistic regression on selected wells.

Table 11: Diagnostic audit log for logistic regression on selected wells.

Cases	Predicted State	Actual State	Well ID
48	Motor Grounded	Motor Grounded	1
3996	Healthy	Healthy	General (Field Data)
5	Mechanical Failure	Healthy	General (Field Data)
2	Cable Grounded	Healthy	General (Field Data)
975	Mechanical Failure	Mechanical Failure	2
1	Healthy	Mechanical Failure	2
217	Cable Grounded	Cable Grounded	3
1	Healthy	Cable Grounded	3

The audit log results indicate that while the Logistic Regression (LR) model maintained high reliability for the electrical fault category (Well 1), it exhibited significant limitations in other scenarios. Specifically, the model produced 7 False Positives in the healthy data and missed critical transitions in (Well 2) and (Well 3). These discrepancies suggest that the linear boundary of the LR algorithm struggles to distinguish between stable operation and subtle mechanical anomalies or cable-related fluctuations. While the model is mathematically efficient, these 9 omissions highlight the sensitivity limits of linear models when dealing with the complex, non-linear degradation patterns typical of ESP failures.

The test results demonstrated a remarkable performance across the four models, with the tree-based ensembles (RF, XGB, and ET) showing a clear superior robustness by achieving an accuracy of 99.96% with only 2 minor errors each. This reflects the clarity of the physical patterns within the data, which confirms the reliability of the results obtained. However, the slightly lower accuracy of the LR model (99.83%) and its 9 misclassifications provide a deeper insight into the data's nature; it confirms that while the patterns are distinct, they are not strictly linear.

Table 12: Overall performance comparison of evaluated models.

Model	Overall Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score
Random Forest (RF)	99.96%	0.999	0.989	0.994
XGBoost (XGB)	99.96%	0.999	0.994	0.997
Extra Trees (ET)	99.96%	0.999	0.989	0.994
Logistic Regression (LR)	99.83%	0.996	0.998	0.997

The performance metrics in Table **12** underscore the exceptional capability of the evaluated models, particularly the tree-based ensembles (ET, RF, and XGB), which achieved accuracies up to 99.96% and ROC-AUC scores exceeding 0.99. These results were consistently validated across all 5 folds of the cross-validation process with a minimal standard deviation (less than 0.005). Such statistical stability confirms that the diagnostic framework has successfully captured the authentic physical signatures of ESP failures, effectively eliminating concerns regarding data overfitting and ensuring the reliability of the diagnostic results.

Unlike Logistic Regression, which struggled with linear decision boundaries and higher error rates, the ensemble models proved superior in capturing the complex, non-linear interactions between thermodynamic and mechanical parameters (temperature, pressure, and vibration). This flexibility is critical in real field environments, as evidenced in Well 2, where these models precisely identified early mechanical wear that the linear baseline failed to detect. Moreover, the integration of Feature Importance analysis provides a transparent physical interpretation of each variable's contribution to the failure onset.

Consequently, tree-based frameworks are the recommended choice for real-time ESP failure detection. To bridge the gap between model predictions and field reality, the Well Test Data summarized in Table **13** was utilized as a benchmarking reference, providing the necessary operational context to ground these findings in actual field conditions.

Table 13: Summary of field production profiles from well test records.

Well	Avg. Liquid Rate (Sbbl/d)	Avg. Water Cut (%)	Avg. Gas Rate (Mscf/d)
1	6109.40	1.50	321.18
2	1144.50	9.75	58.05
3	2031.45	0.42	112.76

The analysis reveals significant variations in fluid composition and flow rates across the field. While parameters such as Water Cut (WC) and Gas Rate were not directly utilized as training features in the machine learning models, they serve as critical Performance Indicators (PIs) for evaluating model robustness. For instance, (Well 2) exhibits a significantly lower production rate and a distinct gas profile compared to the high flow (Well 1). The fact that the tree-based models (ET, XGB, and RF) maintained high diagnostic accuracy across these diverse scenarios demonstrates their adaptability to different reservoir behaviors and fluid dynamics.

This benchmarking confirms that the proposed machine learning approach effectively bridges the gap between periodic manual well tests and continuous real-time monitoring, offering a more responsive early warning system for ESP failures. The technical performance analysis, combined with the benchmarking against actual field well test data, demonstrates that tree-based ensemble models are particularly effective at capturing the non-linear dynamics of ESP operations, even under varying production profiles and water cut levels.

4.1. Practical Implications and SCADA Integration

The proposed machine learning framework is designed for seamless integration with existing Supervisory Control and Data Acquisition (SCADA) systems used in oilfield operations. A key advantage of the tree-based models (RF, ET, and XGB) is their low computational latency, which allows for real-time inference on standard industrial servers without the need for specialized high-performance computing infrastructure. By deploying these models as automated diagnostic layers, operators can receive instantaneous alerts regarding early-stage degradation in Motor and Mechanical components. This transition from reactive to proactive maintenance not only reduces downtime but also minimizes the risks of catastrophic pump failure, thereby optimizing the overall economic lifecycle of ESP-lifted wells.

5. Conclusions

A robust machine learning-based framework for the automated diagnosis of ESP failures was developed and validated through rigorous testing and benchmarking against actual field data. The following conclusions are established:

- **Superiority of Ensemble Learning:** Tree-based ensemble models (ET, XGB, and RF) consistently outperformed traditional linear methods. While all ensemble models achieved an exceptional accuracy of 99.96%, they demonstrated a superior ability to handle the non-linear complexity of downhole sensor data, recording only 2 minor misclassifications compared to 9 for the Logistic Regression model.
- **Capturing System Physics:** Unlike linear models, the ensemble-based approach effectively captures the complex interactions between operational parameters. This ensures a more faithful representation of the ESP system during transitional phases between normal operation and failure, particularly in identifying incipient mechanical wear in (Well 2).
- **Robustness Across Field Variations:** The benchmarking results prove the models' capability to maintain high precision across diverse production profiles and varying water cut levels (from 1.5% in Well 1 to nearly 10% in Well 2). This validates the framework's readiness for real-world oilfield environments where conditions are dynamic.
- **Diagnostic Interpretability:** The use of Feature Importance analysis adds a critical layer of transparency. By identifying Intake Temperature, Discharge Pressure, and Motor Temperature as primary failure indicators, the system provides insights that align with established thermodynamic and mechanical principles.
- **Operational Impact:** Transitioning to a continuous, automated diagnostic system can significantly reduce Non-Productive Time (NPT) and prevent catastrophic failures. By providing early detection of anomalies, the system enables proactive maintenance, optimizes the workover process, and ensures the ESP operates closer to its Best Efficiency Point (BEP), ultimately extending the equipment's lifecycle and reducing the need for frequent site visits in remote locations.

Conflict of Interest

The authors serve on the editorial board of the Global Journal of Energy Technology Research Updates. The peer-review process was managed by an independent editor, and the author has no further competing interests to declare.

Funding

The study received no financial support.

References

[1] Mahmud WM, Jaluta AM. Decade of electrical submersible pumps failures: case study on performance and optimization. J Appl Sci Environ Manag. 2025; 29(3): 747-54. http://doi.org/10.4314/jasem.v29i3.8

[2] Mahmud WM. Integrated evaluation of electric submersible pump failures under diverse field operating conditions. J Energy Power Technol. 2026; 8(1): 006. http://doi.org/10.21926/jept.2601006

[3] Christopher EF, Igbinere SA, Ohenhen I. A machine learning-based predictive maintenance for electric submersible pumps in petroleum production. NIPES Adv Eng Des Technol. 2025; 7(2): 37-42. http://doi.org/10.5281/zenodo.16637356

[4] Abdalla R, Samara H, Perozo N, Paz Carvajal C, Jaeger P. Machine learning approach for predictive maintenance of the electrical submersible pumps (ESPs). ACS Omega. 2022; 7(21): 17641-51. http://doi.org/10.1021/acsomega.1c05881

[5] Peng L, Han G, Sui X, Pagou AL, Zhu L, Shu J. Predictive approach to perform fault detection in electrical submersible pump systems. ACS Omega. 2021; 6(11): 7554-63. http://doi.org/10.1021/acsomega.0c05808

[6] Jani B, Chetan C, Baria D, Barjod U, Chauhan J, Bangadiya J. Performance investigation on double acting reciprocating pump. Int J Innov Res Technol. 2019; 6(4): 14-7.

[7] Brasil J, Maitelli C, Nascimento J, Chiavone-Filho O, Galvão E. Diagnosis of operating conditions of the electrical submersible pump via machine learning. Sensors (Basel). 2023; 23(1): 279. http://doi.org/10.3390/s23010279

[8] Batallas E, Enriquez L, Hill C, Reyes L, Lavin M, Loaiza R, *et al*. Enhancing ESP reliability: synergizing novel machine learning prediction model with operational practices for failure reduction. In: ADIPEC; 2025 Nov; Abu Dhabi, United Arab Emirates. http://doi.org/10.2118/230049-MS

[9] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015; 349(6245): 255-60. http://doi.org/10.1126/science.aaa8415

[10] Saptadi S, Widodo A, Athaillah M, Ayyasyi M. Implementation of machine learning methods in predicting failures in electrical submersible pump machines. Multidiscip Sci J. 2024; 7(3): 2025137. http://doi.org/10.31893/multiscience.2025137

[11] Buabeng A, Simons A, Frempong NK, Ziggah YY. A novel hybrid predictive maintenance model based on clustering, SMOTE and multi-layer perceptron neural network optimised with grey wolf algorithm. SN Appl Sci. 2021; 3: 593. https://doi.org/10.1007/s42452-021-04598-1

[12] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16); 2016. p. 785-94. http://doi.org/10.1145/2939672.2939785

[13] Guedes GS, Silva GC, Rossi RG. Time-aware predictive maintenance of electrical submersible pumps using CatBoost ensemble learning and trend-based labeling. Eng Appl Artif Intell. 2023; 123: 106206.

[14] Jani DB. An overview on causes and prevention of cavitation in centrifugal pumps. Int J Inventory Control Manag. 2019; 9(1): 1-10.

[15] Adhikari S, Majumdar J. Cost-optimised machine learning model comparison for predictive maintenance. In: 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI); 2020. p. 1-6.