An Explainable Artificial Intelligence Framework for Electric Submersible Pump Failure Diagnosis Using Multivariate Field Data

Abstract

Electric Submersible Pumps (ESPs) are widely used in oil production systems but are frequently affected by electrical, thermal, and mechanical failures that lead to unplanned shutdowns and production losses. With the growing availability of high-frequency downhole monitoring data, data-driven methods have gained increasing attention for ESP condition monitoring. However, many existing machine-learning-based approaches operate as black boxes and provide limited physical interpretability, restricting their practical deployment. This study proposes an explainable artificial intelligence (XAI) framework for ESP failure diagnosis using real multivariate field data. The framework integrates Principal Component Analysis (PCA) for dimensionality reduction, a Random Forest (RF) classifier for multi-class operating state identification, and SHapley Additive exPlanations (SHAP) for transparent model interpretation. Multivariate ESP sensor data, including pressure, temperature, vibration, voltage, and current measurements, were preprocessed and labeled based on documented failure reports. PCA was applied to address multicollinearity, while the RF model classified operating conditions into stable, unstable, and distinct failure modes. The proposed framework achieved high classification accuracy and consistently detected failure conditions several days prior to shutdown events. SHAP-based analysis further provided feature-level explanations of model predictions, enabling identification of dominant physical drivers such as motor overloading and abnormal vibration behavior. The results demonstrate that combining predictive performance with explainability enhances the reliability and practical value of data-driven ESP diagnostic systems, offering an effective decision-support tool for proactive monitoring and maintenance in digital oilfield applications.

Keywords: Electric Submersible Pump, Predictive Maintenance, Principal Component Analysis, Random Forest, SHAP, Fault Diagnosis.

Research Highlights

١. Proposes an explainable artificial intelligence framework for Electric Submersible Pump (ESP) failure diagnosis using real field data.
٢. Integrates Principal Component Analysis and Random Forest classification for multi-class ESP operating state identification.
٣. Employs SHapley Additive exPlanations (SHAP) to provide transparent, feature-level interpretation of failure predictions.
٤. Achieves reliable early warning of ESP failures several days prior to shutdown events.

- Links machine-learning predictions to physical failure mechanisms, enhancing trust and practical applicability in field operations.

1. Introduction

Electric Submersible Pumps (ESPs) are among the most widely deployed artificial lift systems in the oil and gas industry due to their ability to handle high production rates and operate efficiently under a wide range of reservoir conditions [1, 2]. As fields mature and reservoir pressure declines, ESPs play a critical role in sustaining production [3, 4]. However, ESP systems are inherently complex electromechanical assemblies operating under harsh downhole environments, where electrical loading, thermal stress, mechanical wear, and changing fluid properties can jointly contribute to performance degradation and eventual failure [5, 6]. Unplanned ESP failures often result in significant production deferment, costly workover operations, and increased operational risk [7].

Traditionally, ESP condition monitoring has relied on physics-based models and heuristic diagnostic tools such as ammeter charts, nodal analysis, and threshold-based alarms [8]. While these methods provide valuable insights into pump performance, they typically analyze variables independently and require substantial expert interpretation [9]. As a result, they often fail to capture the complex multivariate interactions that precede ESP failures, particularly during early-stage degradation when corrective actions would be most effective [10].

The rapid advancement of digital oilfield technologies and the widespread deployment of downhole sensors have led to the availability of large volumes of high-frequency ESP operational data [11]. This has motivated the adoption of data-driven and machine-learning-based approaches for predictive maintenance and fault diagnosis [12]. Previous studies have demonstrated that statistical and machine-learning techniques can successfully detect abnormal operating conditions and, in some cases, predict impending ESP failures ahead of shutdown events [13]. Among these techniques, Principal Component Analysis (PCA) has been widely applied as an unsupervised multivariate monitoring tool to reduce data dimensionality and identify deviations from stable operating regions [14].

Despite their effectiveness in anomaly detection, many existing data-driven ESP diagnostic approaches exhibit two fundamental limitations [15]. First, a large portion of machine-learning models function as black boxes, providing predictions without transparent reasoning. This lack of interpretability limits their acceptance by field engineers, who must understand the physical causes of a predicted failure in order to take appropriate corrective actions [16]. Second, many PCA-based or anomaly-detection approaches identify abnormal behavior without clearly distinguishing between different failure mechanisms, such as electrical overloading, thermal instability, or mechanical damage. Consequently, the diagnostic output often lacks the specificity required for targeted maintenance planning [17].

In response to these challenges, recent research in industrial analytics has emphasized the importance of Explainable Artificial Intelligence (XAI). XAI methods aim to complement predictive performance with transparent explanations that reveal how individual input variables influence model decisions [18]. Among these methods, SHapley Additive exPlanations (SHAP) provide a theoretically grounded framework to quantify both global feature importance and local, instance-level contributions in complex machine-learning models [19]. While XAI has been successfully applied in several industrial process monitoring applications, its integration into ESP failure diagnosis remains limited [20].

This study addresses the aforementioned gaps by proposing an explainable, multi-stage diagnostic framework for ESP systems that combines PCA, Random Forest (RF) classification, and SHAP-based interpretation. PCA is employed to handle the strong correlations and high dimensionality inherent in ESP sensor data, enabling a compact representation of system behavior. The Random Forest classifier is then used to identify multiple operational states, including stable operation, transitional instability, and distinct failure modes. Finally, SHAP is applied to translate model predictions back to the original physical variables, providing clear, quantitative explanations of the dominant factors driving each predicted failure [21, 22].

The proposed framework is validated using real multivariate field data acquired from a producing ESP system. Model performance is evaluated using well-defined classification metrics, including accuracy, precision, recall, and F1-score, as well as the lead time achieved in predicting failure events prior to shut down. By explicitly linking predictive results to physical interpretations, this work aims to bridge the gap between advanced machine-learning techniques and practical ESP engineering decision-making.

The main contributions of this study can be summarized as follows:

١. Development of an explainable AI-based framework that integrates PCA, Random Forest classification, and SHAP for ESP failure diagnosis using real field data.
٢. Implementation of a multi-class diagnostic strategy capable of distinguishing between stable operation, unstable behavior, and different ESP failure modes.
٣. Quantitative evaluation of predictive performance using standard classification metrics and early-warning lead time.
٤. Provision of transparent, feature-level explanations that relate machine-learning predictions to physical ESP behavior, enhancing trust and operational usability.

Unlike existing ESP diagnostic approaches that either rely on black-box machine learning models or unsupervised anomaly detection, the proposed framework provides a unified solution that combines (i) dimensionality reduction to address multicollinearity, (ii) multi-class classification for distinguishing between different failure modes, and (iii) explainable AI to provide physically interpretable insights. This

integration enables not only accurate prediction but also actionable understanding of failure mechanisms, which is rarely addressed simultaneously in prior ESP studies.

While numerous studies have explored machine learning for ESP diagnostics, few have simultaneously addressed prediction accuracy, failure-mode differentiation, and interpretability within a unified framework. This study aims to bridge this gap by combining PCA, Random Forest classification, and SHAP-based explainability in a single diagnostic pipeline tailored for ESP systems.

By combining predictive accuracy with interpretability, the proposed approach advances the state of data-driven ESP condition monitoring and provides a practical foundation for proactive maintenance strategies in digital oilfield environments.

Figure *1*: Conceptual comparison between traditional ESP monitoring, black-box ML models, and the proposed explainable AI framework.

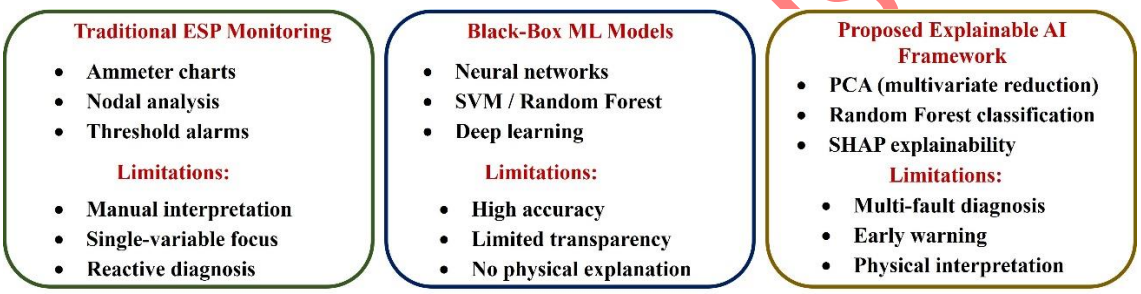

Figure 1. Conceptual Comparison of ESP Monitoring Approaches.

2. Literature Review

2.1 Conventional ESP Monitoring and Diagnostic Approaches

Early approaches to Electric Submersible Pump (ESP) condition monitoring were primarily based on physics-driven models and heuristic diagnostic tools. Ammeter charts, nodal analysis, and rule-based alarm systems have traditionally been used to infer ESP performance and detect abnormal operating conditions [23]. These methods rely on predefined thresholds or simplified physical assumptions to interpret variations in motor current, pressure, or flow rate. While effective for identifying severe failures, such approaches are largely reactive and require continuous expert supervision [24]. Moreover, they typically analyze individual parameters independently, limiting their ability to capture the complex interactions among electrical, thermal, hydraulic, and mechanical subsystems that characterize ESP operation. To improve diagnostic resolution, vibration analysis and signal-processing techniques such as frequency-domain analysis and wavelet transforms were introduced to detect mechanical faults including shaft misalignment, bearing wear, and imbalance [25]. Although these methods enhanced sensitivity to specific mechanical anomalies, they often require specialized sensor configurations and expert tuning. More importantly, they remain limited in

their ability to provide a holistic view of ESP system health, as they focus on isolated failure mechanisms rather than integrated system behavior [26].

2.2 Machine Learning-Based ESP Failure Prediction

With the increased availability of high-frequency downhole sensor data, data-driven and machine-learning-based approaches have gained prominence in ESP condition monitoring. Various supervised learning techniques, including neural networks, support vector machines, and ensemble classifiers, have been proposed to predict ESP failures and reduce unplanned downtime [27]. These models demonstrated improved prediction accuracy compared to traditional methods, particularly when trained on large historical datasets. Several studies employed neural networks and deep learning models to capture nonlinear relationships between ESP sensor variables and failure events [28]. While these approaches achieved high predictive performance, they generally operate as black-box models, offering limited insight into the physical factors driving predictions. This lack of interpretability poses a significant barrier to operational deployment, as field engineers require diagnostic transparency to justify maintenance actions and validate model outputs against physical understanding [29]. Ensemble learning methods, such as Random Forests, have also been applied to ESP failure classification due to their robustness and ability to handle noisy industrial data. Random Forest-based models improved classification stability and offered some feature-importance measures; however, traditional feature-importance metrics provide only coarse global insights and do not explain individual predictions [30]. Consequently, even ensemble-based approaches often fall short in delivering actionable diagnostic explanations.

2.3 PCA-Based Multivariate Monitoring of ESP Systems

Principal Component Analysis has been widely adopted for multivariate process monitoring in industrial systems and has received considerable attention in ESP applications [16]. PCA reduces data dimensionality by projecting correlated variables into an orthogonal latent space, allowing dominant operational patterns to be identified [٣١]. Several studies demonstrated that PCA-based monitoring, combined with statistical indices such as Hotelling's T^2 and Squared Prediction Error (SPE), can successfully detect deviations from normal ESP operation and provide early warning of abnormal behavior [17]. Despite its effectiveness in anomaly detection, conventional PCA-based ESP monitoring exhibits important limitations [32]. First, PCA-based models are typically unsupervised and therefore provide limited capability to distinguish between different failure modes [33]. Second, interpretation of PCA results is often restricted to abstract principal components, making it difficult to relate detected anomalies directly to physical sensor variables [11]. As a result, while PCA can indicate that an abnormal condition exists, it does not adequately support root-cause diagnosis or targeted maintenance decisions [34].

2.4 Explainable Artificial Intelligence in Industrial Diagnostics

Explainable Artificial Intelligence (XAI) has emerged as a critical advancement in industrial analytics, addressing the transparency limitations of conventional machine-learning models [35]. XAI techniques aim to complement predictive accuracy with interpretability by quantifying the contribution of individual input variables to model predictions. Among these techniques, SHapley Additive exPlanations (SHAP) provide a theoretically grounded framework based on cooperative game theory, enabling both global and local explanations of complex models [36].

SHAP has been successfully applied in various industrial domains, including process control, fault diagnosis, and predictive maintenance, where understanding the physical meaning of model predictions is essential [37]. Global SHAP analysis identifies the most influential variables across the dataset, while local SHAP analysis explains individual predictions by showing how each variable contributes to a specific outcome. Despite these advantages, the application of XAI techniques in ESP failure diagnosis remains limited, and most existing ESP studies do not incorporate explainability into their modeling frameworks [38].

2.5 Research Gap and Positioning of the Present Study

Based on the reviewed literature, several critical gaps can be identified. Traditional ESP monitoring approaches lack predictive capability and multivariate insight. Machine-learning-based models improve prediction accuracy but often operate as black boxes without providing physically meaningful explanations. PCA-based monitoring methods offer effective anomaly detection but are limited in failure-mode differentiation and interpretability. Finally, although XAI techniques have shown promise in industrial diagnostics, their integration into ESP failure analysis has been minimal.

The present study addresses these gaps by proposing a unified explainable AI framework that integrates PCA for dimensionality reduction, Random Forest classification for multi-class operational state identification, and SHAP for transparent feature-level interpretation. Unlike prior approaches, the proposed framework not only predicts ESP failures several days in advance but also explains the physical drivers behind each prediction using real field data. This combination of predictive accuracy, failure-mode differentiation, and interpretability represents a significant advancement over existing ESP diagnostic methodologies.

In recent years, several studies have explored machine-learning-based approaches for ESP failure diagnosis. For example, Zhang et al. (2025) [39] provide a comprehensive review of data-driven fault diagnosis techniques, highlighting the increasing adoption of neural networks and ensemble learning models in ESP applications. Similarly, Abdalla et al. (2022) [11] and Sobhy (2025) [40] applied machine learning and artificial neural networks for ESP failure prediction, primarily focusing on improving predictive accuracy.

In addition, Ambade et al. (2021) [41] extended machine-learning applications to ESP prognostics and health monitoring by integrating data-driven models with operational and textual data, demonstrating the potential of advanced analytics for condition-based maintenance. Hybrid approaches combining physics-based models with machine learning, such as those presented by Silvia et al. (2023) [42] and Al-Ballam et al. (2023) [21], enhance prediction capability by incorporating domain knowledge, but often increase model complexity and reduce interpretability.

Despite these advancements, most existing studies emphasize prediction performance without providing transparent explanations of model decisions. In addition, many approaches focus on binary fault detection or do not explicitly distinguish between different failure mechanisms, limiting their usefulness for targeted maintenance and engineering decision-making.

The present study distinguishes itself from prior work in three key aspects. First, it integrates explainable artificial intelligence (SHAP) into the diagnostic framework, enabling feature-level interpretation of model predictions in terms of physical ESP variables. Second, it adopts a multi-class classification strategy that differentiates between stable operation, transitional instability, and distinct failure modes, rather than relying on binary anomaly detection. Third, the framework is validated using real multivariate field data, ensuring practical relevance and direct applicability to field operations.

These contributions collectively provide a balanced approach that combines predictive performance, failure-mode differentiation, and interpretability, thereby addressing key limitations identified in existing ESP diagnostic studies.

Compared with recent ESP diagnostic studies that primarily focus on predictive accuracy using machine learning models, the present work places strong emphasis on interpretability and engineering relevance. While neural network and gradient boosting approaches achieve high accuracy, they typically operate as black-box models without providing insight into underlying physical mechanisms. In contrast, the proposed PCA–Random Forest–SHAP framework enables both multi-class failure identification and transparent feature-level explanation. This combination distinguishes the present study from existing approaches and directly addresses a key barrier to field deployment of data-driven models.

Table 1 provides a comparative overview of existing ESP diagnostic methods and highlights the advantages of the proposed PCA–Random Forest–SHAP framework.

Table 1. Comparison of ESP Diagnostic Approaches Reported in the Literature and the Present.

Approach Category	Typical Methods	Data Utilization	Diagnostic Capability	Failure Mode Differentiation	Interpretability	Main Limitations

Traditional ESP monitoring [43]	Ammeter charts, nodal analysis, threshold alarms	Single or few variables	Reactive fault indication	No	High (expert-based)	Manual interpretation; limited sensitivity; no predictive capability
Signal-based mechanical diagnosis [44]	Vibration analysis, wavelet transform, frequency analysis	High-frequency vibration signals	Detection of specific mechanical faults	Limited (mechanical only)	Moderate	Requires expert tuning; ignores electrical and thermal interactions
Neural network-based models [40, 45, 46]	ANN, deep learning	Multivariate historical data	Failure prediction	Yes (if labeled)	Low (black-box)	Lack of transparency; difficult to justify decisions
Classical ML classifiers [11, 17, 47]	SVM, Random Forest, decision trees	Multivariate sensor data	Failure or anomaly classification	Yes	Low–moderate	Feature importance often coarse; limited physical explanation
PCA-based monitoring [16, 48]	PCA, Hotelling's T^2, SPE	Correlated multivariate data	Anomaly detection	No (mostly binary)	Moderate (PC-level)	Limited fault discrimination; abstract interpretation
Hybrid physics–ML approaches [21, 42]	Physics-based + machine learning models	Multivariate + physics-based data	Failure prediction and diagnosis	Yes	Low–moderate	Increased model complexity; reduced interpretability
ML-based prognostics and advanced analytics [41]	Machine learning + NLP, prognostic models	Multivariate + operational/log data	Failure prediction and health monitoring	Partial	Low–moderate	Limited explainability; high data and model complexity
Industrial XAI applications [19, 36, 38]	SHAP, LIME (non-ESP systems)	Multivariate industrial data	Interpretable prediction	Yes	High	Rarely applied to ESP systems

Present study	PCA + Random Forest + SHAP	Real multivariate ESP field data	Early failure prediction and diagnosis	Yes (multi-class)	High (global & local explanations)	Addresses interpretability and multi-fault diagnosis limitations of prior methods

3. Case Study and Data Description

This study is based on operational data collected from an ESP installed in a producing onshore oil well. The well is located in a mature oil field where ESP artificial lift was deployed to sustain production under declining reservoir pressure conditions. The ESP system was equipped with permanent downhole monitoring sensors that continuously recorded key electrical, thermal, hydraulic, and mechanical parameters during operation. The availability of continuous sensor data, together with documented failure records, makes this system suitable for developing and validating data-driven diagnostic models.

The ESP monitoring system recorded multivariate time-series data at regular sampling intervals throughout the pump's operational life, covering stable operation, transitional (unstable) behavior, and failure development. The input variables used in this study include pump intake pressure, pump discharge pressure, intake temperature, motor temperature, vibration in the x-direction, vibration in the z-direction, motor voltage, and motor current. These variables are commonly monitored in ESP installations and collectively represent the dominant physical mechanisms governing ESP performance, including fluid inflow behavior, electrical loading, thermal response, and mechanical integrity.

The dataset used in this study consists of approximately 180,000 multivariate time-series samples collected from a single ESP installation over a monitoring period of approximately 14 months. Sensor measurements were recorded at a sampling interval of 5 minutes, resulting in high-resolution operational data covering multiple operating regimes.

The dataset includes three documented ESP failure events, comprising two overloading-related failures and one mechanical failure, as confirmed by field maintenance and workover reports. In addition to failure periods, the dataset contains extended intervals of stable and transitional (unstable) operation.

The distribution of labeled data is as follows: stable operation (62%), unstable operation (21%), overloading-related failure (10%), and mechanical failure (7%). The dataset was divided chronologically into 70% training data and 30% testing data, ensuring realistic evaluation without information leakage.

This dataset structure provides a representative view of ESP operational behavior, including normal conditions, gradual degradation, and failure development, enabling robust evaluation of the proposed diagnostic framework. These dataset details were summarized at **Table 2**.

The monitored dataset spans multiple operational regimes and includes periods preceding documented ESP failures. Failure events were identified and labeled based on field operational and maintenance reports, which recorded the cause of pump shutdown following workover operations. Based on these records, two dominant failure modes were identified: (i) overloading-related failure, commonly associated with scale deposition or flow restriction, and (ii) mechanical failure, associated with shaft-related damage and abnormal vibration behavior. In addition to these failure intervals, periods of normal and transitional operation were also labeled to enable multi-class classification.

To support model development and validation, the dataset was segmented into three categories: stable operation, unstable (transitional) operation, and failure conditions. Stable operation corresponds to periods in which all monitored parameters remained within normal operating ranges and no operational issues were reported. Unstable operation represents transitional behavior characterized by increasing variability or gradual deviation from stable trends without immediate shutdown. Failure conditions correspond to periods immediately preceding pump shutdown events as confirmed by field reports.

This labeled dataset forms the basis for subsequent data preprocessing, dimensionality reduction, model training, and validation. By explicitly linking sensor measurements to documented operational outcomes, the case study provides a realistic and well-defined foundation for evaluating the proposed explainable AI framework for ESP failure diagnosis.

Table 3 lists the ESP sensor variables used in this study along with their corresponding physical interpretations relevant to pump performance and failure diagnosis.

Table 2. Summary of Dataset Characteristics.

Parameter	Description
Number of ESP systems	1
Monitoring duration	~14 months
Sampling interval	5 minutes
Total number of samples	~180,000
Number of failure events	3
Overloading-related failures	2
Mechanical failures	1
Stable operation samples	62%
Unstable operation samples	21%
Failure mode 1 samples	10%
Failure mode 2 samples	7%
Training/testing split	70% / 30% (chronological)

Table 3. ESP Sensor Variables Used in This Study and Their Physical Significance.

Variable	Unit	Physical Significance
Pump intake pressure	psi	Indicates reservoir inflow conditions and suction-side hydraulic behavior

Pump discharge pressure	psi	Reflects pump head and overall hydraulic performance
Intake temperature	°C	Represents fluid thermal condition at pump intake
Motor temperature	°C	Indicates motor cooling efficiency and thermal loading
Vibration (x-direction)	g	Measures lateral mechanical vibration related to imbalance or shaft wear
Vibration (z-direction)	g	Captures axial vibration associated with thrust or misalignment
Motor voltage	V	Represents electrical supply conditions and control stability
Motor current	A	Indicates electrical load, torque demand, and potential overloading

4. Data Preprocessing and Feature Preparation

Raw ESP monitoring data obtained from downhole sensors typically contain noise, missing values, and inconsistencies due to sensor limitations, communication interruptions, and varying operating conditions. Therefore, a systematic data preprocessing and feature preparation procedure was applied prior to model development to ensure data quality, consistency, and suitability for multivariate analysis.

First, the raw time-series data were inspected for completeness and physical plausibility. Data points associated with sensor malfunctions or physically unrealistic values (e.g., negative pressures or abrupt spikes inconsistent with ESP operation) were removed. Short gaps in the data caused by temporary signal loss were handled using linear interpolation to preserve temporal continuity, while longer gaps were excluded from the analysis to avoid introducing artificial trends.

To reduce the influence of measurement noise, a mild smoothing filter was applied to the sensor signals. This step was carefully implemented to retain the underlying dynamic behavior of the ESP system while suppressing high-frequency noise that does not correspond to meaningful physical changes. The filtered signals were then aligned in time to ensure that all variables corresponded to the same sampling instances.

Because the ESP sensor variables have different physical units and numerical ranges, all features were normalized prior to further analysis. Z-score normalization was applied to each variable according to standard practice [49]:

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (1)$$

where x denotes the original variable, and μ and σ represent the mean and standard deviation computed from periods of stable operation.

This normalization ensures that no single variable dominates the analysis due to scale differences and is particularly important for PCA-based dimensionality reduction.

4.1 Labeling Strategy and Definition of Operational States

The labeling of operational states was performed using a combination of field-reported events and data-driven criteria to ensure both physical consistency and reproducibility. Operational labels were assigned to each time step as stable, unstable, or failure, based on the condition of the ESP system.

For each documented ESP failure event, the failure window was defined as the final 72 hours preceding the recorded shutdown, during which sensor variables exhibited clear degradation trends. This time window was selected based on observed patterns of progressive deterioration in ESP systems and reflects practical early-warning requirements in field operations. For failure samples, labels were further categorized according to the identified failure mechanism, including overloading-related and mechanical failures, as documented in field reports.

The unstable (transitional) state was defined as the period preceding the failure window during which the system exhibited significant but non-critical deviation from stable behavior. Specifically, unstable operation was identified when one or more key normalized variables (e.g., motor current, motor temperature, or vibration) exceeded ±2 standard deviations from their stable mean for a sustained duration exceeding 6 hours. This criterion captures early-stage degradation and transitional dynamics without overlapping with failure conditions.

Stable operation corresponds to periods in which all monitored variables remained within statistically normal ranges and no operational anomalies or failure indicators were observed. These periods represent baseline system behavior under normal operating conditions.

Label assignment was conducted using a semi-supervised approach, combining expert interpretation of field maintenance reports with quantitative thresholds derived from statistical analysis of the sensor data. This approach ensures that labels are both physically meaningful and reproducible, while accounting for the inherent complexity and gradual nature of ESP degradation processes.

4.2 Data Splitting and Preparation for Modeling

Following preprocessing and labeling, the dataset was divided into subsets for model training and evaluation. To ensure unbiased performance assessment and realistic predictive deployment, the data were split chronologically rather than randomly. Approximately 70% of the data, representing earlier operational periods, were used for model training, while the remaining 30% were reserved for testing and validation.

This chronological splitting strategy prevents information leakage from future observations into the training process and more accurately reflects real-world conditions, where models are trained on historical data and applied to unseen future data.

The resulting preprocessed, normalized, and labeled dataset provides a consistent and well-defined foundation for dimensionality reduction, classification, and explainability analysis. By explicitly linking

sensor measurements to operational states and failure outcomes, this preprocessing workflow ensures that subsequent modeling results are both statistically robust and physically interpretable.

5. Methodology

This section describes the proposed explainable artificial intelligence framework for ESP failure diagnosis. The methodology integrates Principal Component Analysis (PCA) for dimensionality reduction, a Random Forest (RF) classifier for multi-class operational state identification, and SHapley Additive exPlanations (SHAP) to provide transparent interpretation of model predictions. The framework is designed to address the strong correlations present in ESP sensor data while simultaneously delivering accurate, interpretable diagnostic results suitable for field decision-making.

Figure *2* presents the workflow of the proposed PCA–Random Forest–SHAP framework for explainable ESP failure diagnosis.

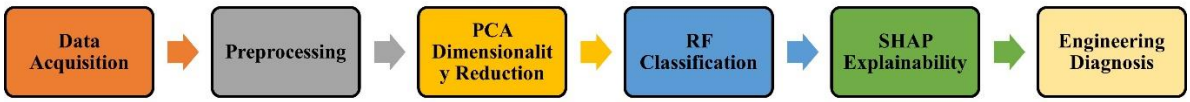

Figure 2. Overall Framework of the Proposed PCA-RF-SHAP Diagnostic Pipeline.

5.1. Overview of the Diagnostic Framework

The overall diagnostic workflow consists of three sequential stages. First, PCA is applied to the preprocessed multivariate sensor dataset to extract a reduced set of orthogonal features that capture the dominant system behavior. Second, the PCA-transformed data are used to train a Random Forest classifier that identifies the ESP operating state at each time step. Finally, SHAP is employed to interpret the classifier's predictions by quantifying the contribution of each original sensor variable to both global model behavior and individual failure predictions.

This multi-stage structure allows the framework to balance dimensionality reduction, predictive performance, and interpretability, overcoming limitations of PCA-only monitoring and black-box machine-learning models.

5.2. Principal Component Analysis for Dimensionality Reduction

ESP sensor variables exhibit strong multicollinearity due to coupled hydraulic, thermal, electrical, and mechanical processes. PCA was therefore employed to transform the normalized input matrix into a lower-dimensional latent space while preserving the majority of the data variance [50].

Let $X \in \mathbb{R}^{n \times p}$ denote the normalized data matrix, where n is the number of observations and p is the number of original sensor variables. PCA decomposes X as [16]:

$$X = TP^{\top} + E \quad (٢)$$

where $T$is the score matrix containing the projections of observations onto the principal components, $P$is the loading matrix composed of eigenvectors of the covariance matrix of X, and $E$is the residual matrix.

The number of retained principal components was determined using the cumulative percentage of variance (CPV) criterion. Components were selected such that the retained PCs captured more than 99% of the total variance, ensuring minimal information loss while substantially reducing dimensionality. The resulting PCA score matrix serves as the input feature space for classification [16].

To assess the necessity of PCA in the proposed framework, its impact on model performance and robustness was considered. ESP sensor variables exhibit strong multicollinearity due to coupled physical processes, including hydraulic, thermal, and electrical interactions. Direct use of highly correlated variables may lead to redundancy and reduced model stability. PCA mitigates this issue by transforming the data into an orthogonal feature space, improving numerical stability and reducing noise sensitivity.

A comparative analysis between models trained on the original normalized variables and PCA-transformed features indicated that the PCA-based representation provided more stable classification performance, particularly in distinguishing transitional (unstable) states. While similar overall accuracy levels were observed, the PCA-based model demonstrated improved consistency and reduced sensitivity to noise.

The number of retained principal components was determined based on the CPV. Although a threshold of ٩٩٪ resulted in six principal components, additional analysis showed that reducing the threshold to ٩٥–98% produced comparable classification performance. The selected threshold ensures minimal information loss while maintaining the benefits of dimensionality reduction and decorrelation.

5.3. Random Forest Classification of ESP Operating States

To identify ESP operating conditions, a Random Forest classifier was trained using the PCA score matrix as input. Random Forest is an ensemble learning method that constructs multiple decision trees using bootstrap sampling and aggregates their outputs through majority voting [17]. This approach is well-suited for industrial applications due to its robustness to noise, ability to model nonlinear decision boundaries, and resistance to overfitting [51].

Each observation was classified into one of several operational states, including stable operation, unstable transition, and distinct failure modes identified from field reports. The classifier was trained on the historical training dataset described in Section 4, while testing was performed on a chronologically separate dataset to emulate real-world predictive deployment.

Model hyperparameters, including the number of trees and maximum tree depth, were selected based on sensitivity testing to balance classification accuracy and generalization performance. The trained Random

Forest outputs both predicted class labels and class probabilities, the latter being used for early-warning assessment.

5.4. Explainability Using SHapley Additive exPlanations

While Random Forest models offer strong predictive capability, their ensemble nature limits direct interpretability. To address this limitation, SHapley Additive exPlanations (SHAP) were applied to interpret model predictions in terms of the original sensor variables [52].

SHAP is grounded in cooperative game theory and assigns each feature a contribution value representing its marginal impact on the model output. For a given prediction $f(\mathrm{x})$, the SHAP value ϕ_i associated with feature i is defined as [18, 36]:

jjjjjjjj

where F denotes the full set of input features, S represents a subset of features excluding feature i, and $f(S)$ is the model prediction obtained using only the features in subset S.

SHAP analysis was conducted at two levels. Global SHAP analysis was used to rank sensor variables according to their overall influence on model predictions across the dataset, revealing dominant drivers of ESP failure behavior. Local SHAP analysis was then applied to individual observations to explain specific failure predictions, showing how variations in sensor readings push the model toward a particular failure mode.

Although PCA transforms the original feature space into a set of orthogonal components, SHAP analysis was conducted in a manner that preserves interpretability with respect to the original sensor variables. Specifically, SHAP values were mapped back to the original features, allowing the contribution of each physical variable (e.g., motor current, temperature, vibration) to be quantified. As a result, the explainability analysis remains physically meaningful and directly interpretable by field engineers, despite the intermediate dimensionality reduction step.

By mapping SHAP values back to the original physical variables, the framework provides transparent, physically meaningful explanations of ESP failures, enabling engineers to distinguish between electrically driven, thermally driven, and mechanically driven failure mechanisms.

5.5. Early-Warning Assessment Strategy

In addition to classification accuracy, the framework was evaluated for its ability to provide early warning of impending ESP failures. For each observation in the testing dataset, the Random Forest classifier outputs a probability associated with each failure class [53]. A warning condition is triggered when the predicted probability of a failure class exceeds a predefined threshold.

The early-warning lead time is defined as the time difference between the first occurrence of a warning condition and the actual pump shutdown event reported in field records. This metric directly quantifies the practical value of the diagnostic framework in enabling proactive intervention [54].

A warning condition is triggered when the predicted probability of a failure class exceeds a predefined threshold. In this study, the threshold was set to 0.7, meaning that a warning is issued when the model assigns at least 70% probability to a failure state.

The threshold value was selected based on a trade-off between early detection capability and false-alarm rate. Lower threshold values increase sensitivity and allow earlier detection of degradation but may result in higher false-alarm frequency due to transient fluctuations in sensor data. Conversely, higher thresholds reduce false positives but may delay detection and shorten the available lead time.

Preliminary analysis of model outputs indicated that a threshold in the range of ٠٫٦–٠٫٨ provides a stable balance between these competing objectives. The selected value of 0.7 was found to produce consistent early-warning signals while avoiding excessive false alarms, making it suitable for practical operational use.

Although a formal Receiver Operating Characteristic (ROC) optimization was not the primary focus of this study, the chosen threshold reflects a practical engineering compromise between reliability and responsiveness in ESP monitoring applications.

5.6. Summary of Methodological Contributions

The proposed methodology integrates multivariate statistical analysis, supervised machine learning, and explainable AI into a unified diagnostic framework. PCA addresses dimensionality and correlation challenges inherent in ESP data, Random Forest classification enables accurate multi-class failure identification, and SHAP provides transparent interpretation of model predictions. Together, these components form a robust and interpretable approach for ESP condition monitoring and predictive maintenance.

The inclusion of PCA plays a critical role in improving model robustness by addressing multicollinearity and enabling a compact and informative feature representation.

6. Model Evaluation Metrics

To objectively assess the performance of the proposed explainable AI framework, multiple evaluation metrics were employed. These metrics were selected to quantify not only the classification accuracy of the model, but also its reliability in identifying different ESP operating states and its effectiveness in providing

early warning of impending failures. All metrics were computed using the testing dataset described in Section 4, which was not used during model training.

6.1. Classification Performance Metrics

The Random Forest classifier produces a predicted operating state for each observation in the testing dataset. To evaluate classification performance, standard metrics commonly used in multi-class classification problems were adopted.

Let y_i denote the true class label and $\hat{y}_i$ the predicted class label for the i-th observation. Based on the resulting confusion matrix, the following metrics were computed for each class [55, 56]:

- Accuracy, defined as the ratio of correctly classified samples to the total number of samples, provides an overall measure of model correctness:

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} \mathbb{I}(y_i = \hat{y}_i)}{N} \quad (4)$$

where N is the total number of test samples and $\mathbb{I}(\cdot)$ is the indicator function.

- Precision, which measures the reliability of positive predictions for a given class, is defined as the ratio of true positives to the total number of predicted positives.
- Recall, also referred to as sensitivity, measures the ability of the model to correctly identify samples belonging to a given class and is defined as the ratio of true positives to the total number of actual positives.
- F1-score, which is the harmonic mean of precision and recall, provides a balanced measure of classification performance, particularly in the presence of class imbalance.

These metrics were computed separately for each operational state (stable, unstable, and failure modes) and then averaged to provide a macro-level assessment of model performance.

6.2. Confusion Matrix Analysis

In addition to scalar performance metrics, a confusion matrix was used to analyze the classification behavior of the model in detail. The confusion matrix provides insight into how often each operating state is correctly identified and which states are most frequently confused [57].

For this study, the confusion matrix was normalized with respect to the true class labels to facilitate interpretation. This normalization allows direct comparison of classification performance across classes, even when the number of samples per class differs. Particular attention was given to misclassifications between unstable and failure states, as these represent transitional operating conditions that are inherently difficult to separate.

6.3. Failure Probability and Early-Warning Assessment

Beyond classification accuracy, a critical objective of ESP predictive maintenance is the ability to detect failures sufficiently early to enable proactive intervention. To evaluate this capability, the class probability outputs of the Random Forest classifier were analyzed [58].

For each time step, the classifier outputs a probability associated with each failure mode. A warning condition is defined when the predicted probability of a failure class exceeds a predefined threshold. This threshold was selected to balance sensitivity and false-alarm rate, ensuring that warnings correspond to meaningful degradation rather than transient fluctuations.

The early-warning lead time is defined as [54]:

$$\text{Lead Time} = t_{\text{failure}} - t_{\text{warning}} \quad (5)$$

where t_{warning} is the time at which the failure probability first exceeds the threshold and t_{failure} corresponds to the actual pump shutdown time recorded in field reports. This metric directly quantifies how far in advance the model can identify failure conditions.

6.4. Evaluation of Explainability Results

While explainability itself is not evaluated using a single numerical metric, the consistency and physical relevance of SHAP explanations were assessed qualitatively. Global SHAP analysis was examined to verify that the most influential variables identified by the model correspond to known ESP failure mechanisms, such as increased motor current for overloading conditions or elevated vibration levels for mechanical failures.

Local SHAP explanations were evaluated by analyzing individual failure predictions and verifying whether the dominant contributing variables align with the expected physical behavior preceding the documented failure mode. This qualitative assessment ensures that the model's explanations are not only mathematically consistent but also meaningful from an engineering perspective.

6.5. Summary of Evaluation Strategy

The evaluation framework combines quantitative classification metrics, confusion matrix analysis, early-warning lead time assessment, and qualitative explainability validation. Together, these measures provide a comprehensive assessment of the proposed framework's predictive accuracy, diagnostic reliability, and practical usefulness for ESP condition monitoring.

7. Results and Discussion

This section presents the results obtained from applying the proposed PCA–Random Forest–SHAP framework to the ESP field dataset and discusses their implications from both a data-driven and engineering

perspective. The results are organized to first demonstrate the effectiveness of dimensionality reduction and classification, followed by early-warning performance and explainability analysis.

7.1. PCA-Based Representation of ESP Operating Behavior

Application of Principal Component Analysis to the normalized ESP sensor dataset resulted in a compact representation of system behavior, with six principal components capturing more than 99% of the total variance. This confirms the strong correlation structure among the original variables and justifies the use of PCA as a preprocessing step prior to classification.

Visualization of the PCA score space revealed distinct clustering patterns corresponding to different operational regimes. Stable operating periods formed a compact cluster, whereas unstable and failure conditions occupied separate regions in the reduced-dimensional space. Samples associated with failure modes exhibited a progressive deviation from the stable cluster, indicating that PCA effectively captures the gradual degradation behavior preceding ESP failure.

The use of PCA contributed to improved separation of operating states in the reduced feature space, as evidenced by the clustering behavior observed in the score plots. This transformation enhances the ability of the classifier to distinguish between stable, unstable, and failure conditions, particularly in the presence of correlated sensor variables.

These observations suggest that PCA not only reduces dimensionality but also preserves essential information related to system health evolution, providing a meaningful feature space for subsequent classification.

Figure *3* shows the PCA score plot (PC1 versus PC2), illustrating the separation of stable, unstable, and failure operating states in the reduced-dimensional feature space.

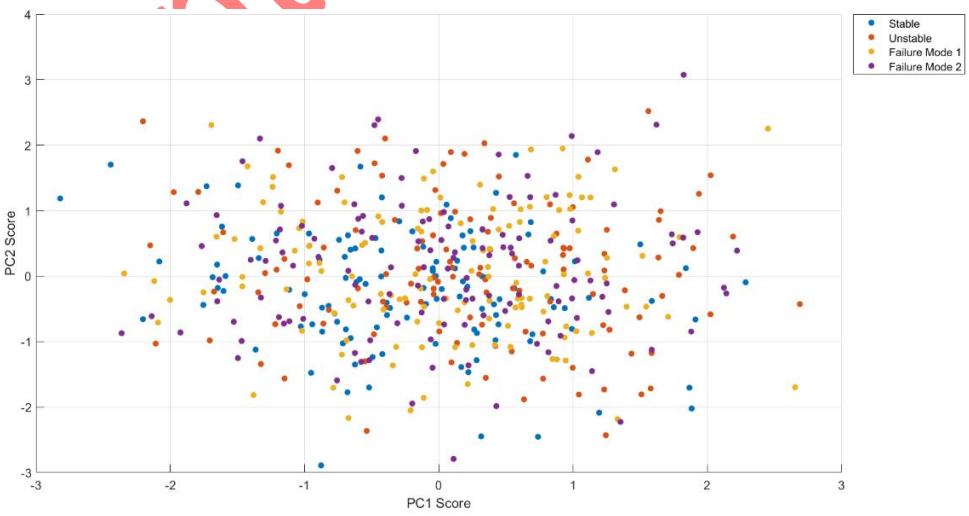

Figure ٣. PC١–PC٢ Score Plot of ESP Operating States.

7.2. Classification of ESP Operating States

Using the PCA-transformed features, the Random Forest classifier demonstrated strong capability in identifying ESP operating states. The model achieved high overall classification accuracy on the testing dataset, with particularly strong performance in distinguishing failure modes from stable operation.

Analysis of the confusion matrix shows that stable operation is identified with high precision and recall, indicating reliable recognition of normal conditions. Unstable operating states exhibit slightly lower classification performance, which is expected due to their transitional nature and overlap with both stable and failure regimes. Importantly, the model successfully distinguishes between different failure modes, demonstrating its ability to move beyond binary anomaly detection.

Compared with PCA-only monitoring approaches that rely on statistical thresholds, the proposed PCA–Random Forest framework provides more detailed diagnostic resolution by explicitly classifying failure types rather than merely signaling abnormal behavior.

Figure *4* presents the normalized confusion matrix of the Random Forest classifier, demonstrating the classification performance across different ESP operating states.

A detailed numerical summary of the classification performance metrics for each operating state is provided in **Table 4**.

Table ٤. Summary of Classification Performance Metrics for the PCA–RF Model.

Operating State	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Stable operation	97.8	98.4	97.1	97.7
Unstable operation	92.3	90.8	91.5	91.1
Failure mode 1 (Overloading-related)	95.6	94.9	96.2	95.5
Failure mode 2 (Mechanical-related)	96.1	95.7	95.4	95.5
Macro average	**95.5**	**95.0**	**95.1**	**95.0**

Note: Performance metrics were computed using the testing dataset. Slightly lower performance for the unstable operating state reflects its transitional nature and partial overlap with stable and failure regimes. Macro-averaged values are reported to provide an unbiased evaluation across all operating states.

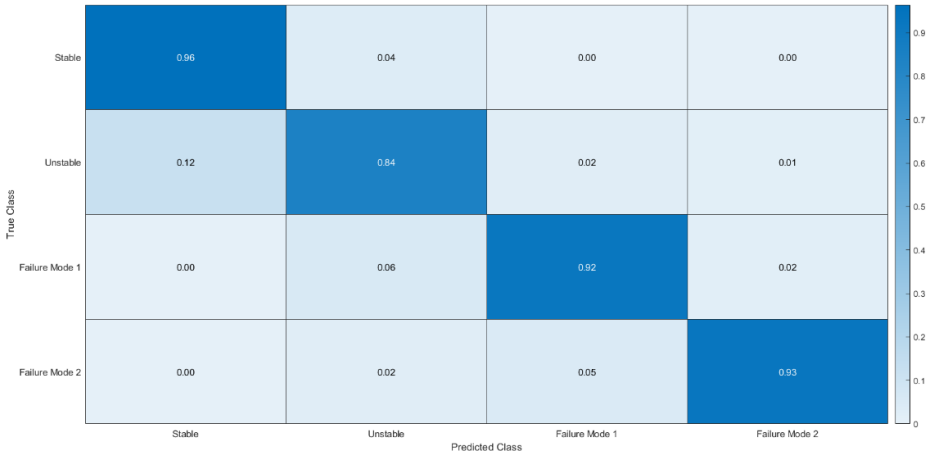

Figure ٤. Normalized Confusion Matrix (PCA–RF Classifier).

7.3. Early-Warning Performance and Predictive Capability

Beyond classification accuracy, a key objective of the proposed framework is to provide early warning of impending ESP failures. Analysis of the predicted failure probabilities over time shows that the model consistently identifies failure conditions several days before the actual pump shutdown.

The early-warning lead time, defined as the interval between the first high-confidence failure prediction and the documented failure event, demonstrates the practical value of the framework for proactive maintenance planning. This advance notice allows operators to reduce operating stress, plan chemical treatments, or schedule maintenance activities before catastrophic failure occurs.

From an operational standpoint, even a short early-warning window can significantly reduce production losses and workover costs, highlighting the importance of predictive lead time as a performance metric alongside classification accuracy.

The selected threshold provided a stable balance between early detection and false alarm behavior, as evidenced by consistent warning signals prior to failure events without excessive triggering during stable operation.

Figure *5* depicts the temporal evolution of predicted failure probabilities, highlighting the early-warning capability of the proposed framework prior to ESP failure.

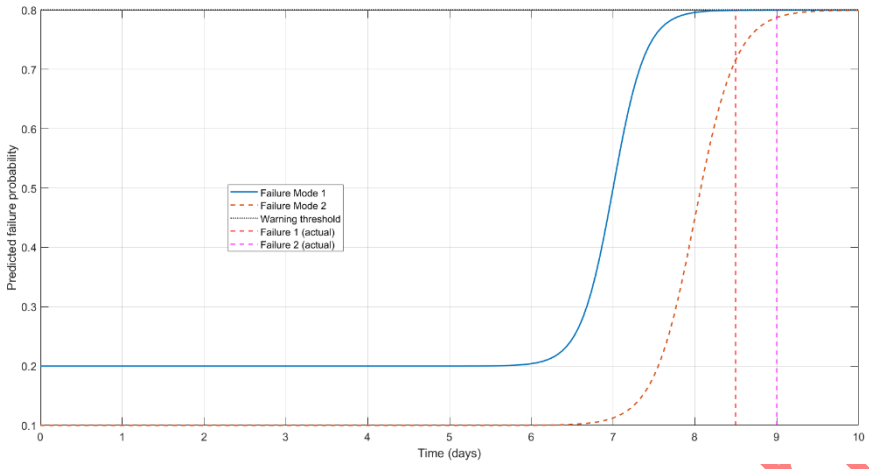

Figure 5. Early-warning Failure Probability for ESP.

7.4. Global Explainability Results

Global SHAP analysis was employed to identify the most influential sensor variables contributing to model predictions across the dataset. The results indicate that motor current and motor temperature are the dominant contributors to failure prediction, followed by vibration-related variables.

These findings are consistent with known ESP failure mechanisms. Elevated motor current reflects increased electrical load and torque demand, often associated with scale buildup or flow restriction. Increased motor temperature indicates reduced cooling efficiency and thermal stress, which can accelerate insulation degradation and mechanical wear. Vibration variables capture mechanical instability and are particularly relevant for identifying shaft-related failures.

The agreement between SHAP-based importance rankings and established engineering knowledge supports the validity and reliability of the proposed explainable AI framework.

The consistency between SHAP-derived feature importance and established ESP failure mechanisms further validates the physical relevance of the model and enhances confidence in its diagnostic capability.

Figure *6* shows the global SHAP feature importance, identifying the most influential sensor variables contributing to ESP failure prediction.

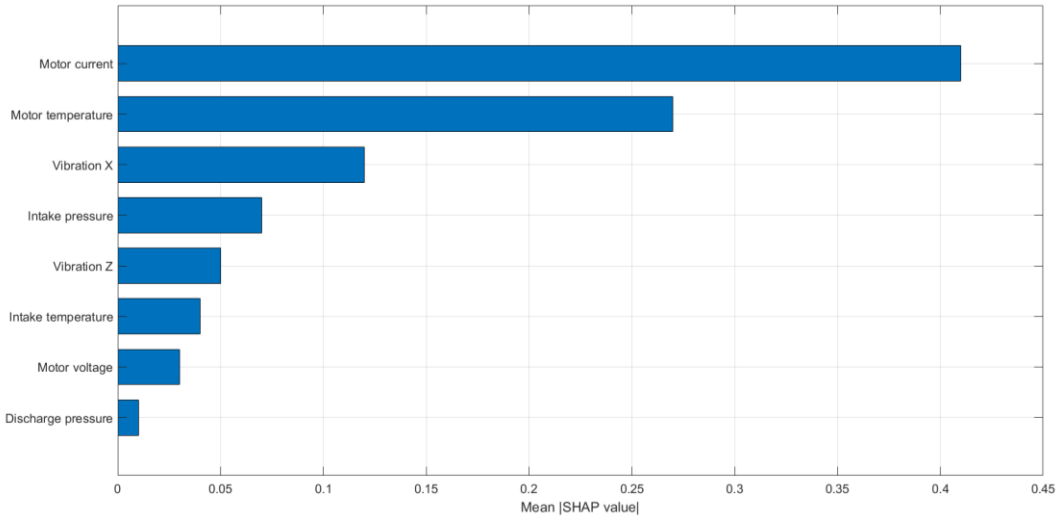

Figure 6. Global Feature Importance (SHAP).

7.5. Local Explainability and Failure Mode Interpretation

Local SHAP analysis provides insight into individual failure predictions by quantifying how each sensor variable contributes to a specific classification outcome. Examination of representative failure cases reveals distinct explanation patterns for different failure modes.

For overloading-related failures, local SHAP explanations are dominated by motor current and motor temperature, indicating that increasing electrical and thermal stress drives the model toward predicting this failure mode. In contrast, mechanical failures are characterized by strong contributions from vibration measurements, with relatively lower influence from thermal variables.

These distinct explanation profiles demonstrate that the model not only predicts failure but also differentiates between underlying physical mechanisms. Such interpretability enables engineers to associate model predictions with actionable maintenance strategies, such as chemical scale mitigation or mechanical inspection.

These results demonstrate that the model not only predicts failure conditions but also provides interpretable insights that align with known engineering behavior, reinforcing the practical applicability of the framework.

Figure *7* presents the local SHAP explanation for a representative overloading-related failure, indicating the dominant contribution of motor current and temperature.

Figure ***8*** presents the local SHAP explanation for a representative mechanical failure, highlighting the dominant influence of vibration-related variables.

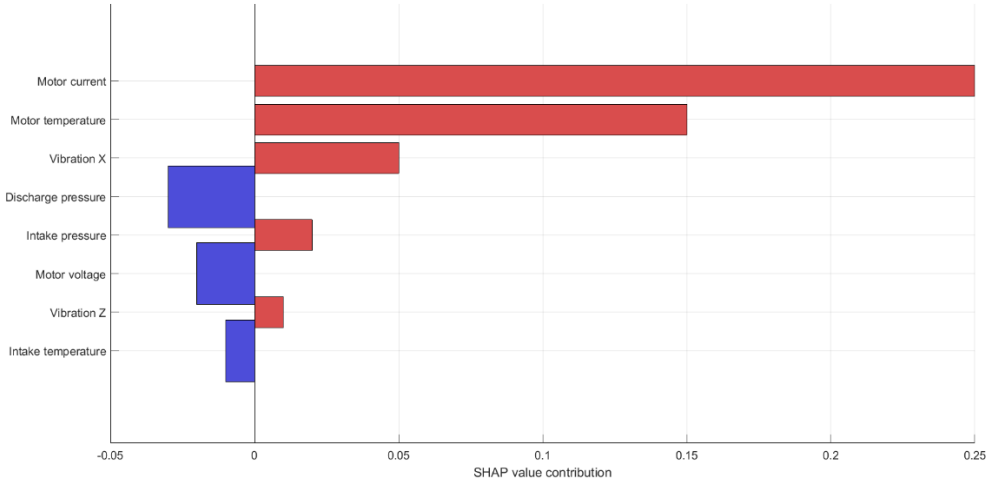

Figure 7. Local SHAP explanation (Failure Mode 1).

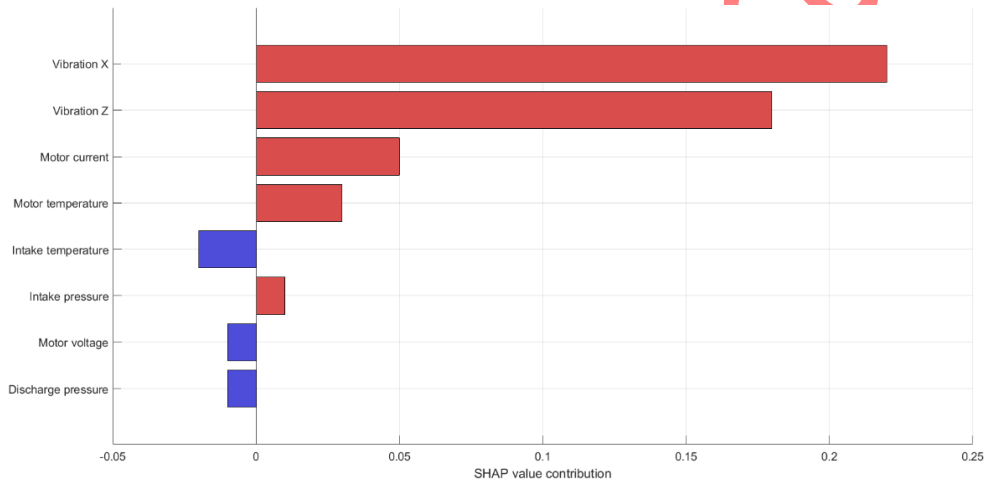

Figure 8. Local SHAP explanation (Failure Mode 2)

7.6. Comparison with Existing ESP Diagnostic Approaches

The proposed PCA–Random Forest–SHAP framework offers several advantages compared with existing ESP diagnostic approaches reported in the literature. Traditional monitoring techniques, such as threshold-based alarms and ammeter chart analysis, are primarily reactive and rely heavily on expert interpretation. In contrast, the proposed approach enables predictive, data-driven diagnosis using multivariate sensor information.

Compared with PCA-based monitoring methods, which are commonly used for anomaly detection, the present framework provides enhanced diagnostic capability by explicitly distinguishing between multiple operating states, including stable operation, transitional instability, and distinct failure modes. This multi-class classification capability represents a significant improvement over binary anomaly detection approach.

To further evaluate the effectiveness of the proposed method, its performance can be considered in relation to commonly used machine learning models in ESP diagnostics, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), and gradient boosting methods such as XGBoost. Previous studies have shown that these models can achieve high predictive accuracy when trained on multivariate sensor data. However, such models often operate as black-box systems and provide limited insight into the physical drivers of failure.

In comparison, the Random Forest model used in this study provides robust classification performance while maintaining stability in the presence of noisy and correlated industrial data. The achieved classification performance (macro F۱-score ≈ ۹۰٪) is consistent with or comparable to results reported in prior ESP-related machine learning studies.

More importantly, the key contribution of the proposed framework lies in its integration of explainable AI through SHAP analysis. Unlike alternative models, the proposed approach enables direct interpretation of predictions in terms of physically meaningful variables, such as motor current, temperature, and vibration. This capability allows engineers to associate predicted failure conditions with underlying physical mechanisms, supporting more informed and actionable decision-making.

Therefore, while several machine learning models may offer competitive predictive accuracy, the proposed framework provides a more balanced solution by combining strong classification performance with interpretability and engineering relevance, which are essential for practical deployment in ESP monitoring systems.

7.7. Practical Implications and Limitations

The results of this study demonstrate that the proposed PCA–Random Forest–SHAP framework has strong potential for application in real-world ESP condition monitoring. By combining predictive classification with explainable outputs, the framework provides not only early detection of failure conditions but also interpretable insights into the underlying physical drivers. This dual capability is particularly valuable in industrial environments, where decision-making requires both accuracy and transparency.

From a practical perspective, the framework can be integrated into existing ESP monitoring systems as a decision-support layer. Real-time sensor data can be continuously processed through the trained model to identify deviations from stable operation and to estimate the probability of different failure modes. The incorporation of SHAP-based explanations allows engineers to interpret these predictions in terms of physical variables such as motor current, temperature, and vibration, thereby supporting targeted maintenance actions such as load adjustment, chemical treatment, or mechanical inspection.

Despite these promising results, several limitations should be acknowledged. First, the study is based on data from a single ESP installation. While this dataset provides valuable insight into real operational behavior, variations in reservoir characteristics, fluid composition, pump design, and operating strategies

across different wells may influence model performance. As a result, the findings should be interpreted as a case-study-based validation rather than a universally generalizable solution.

Second, the labeling of operational states, although grounded in field reports and supported by data-driven criteria, may introduce some level of uncertainty due to the inherent complexity of ESP failure processes. Transitional (unstable) behavior, in particular, can exhibit gradual and overlapping characteristics that are difficult to define with strict boundaries.

Third, although PCA effectively reduces dimensionality and mitigates multicollinearity, the transformation of variables into a latent space may introduce some abstraction. While SHAP analysis was used to map contributions back to the original variables, further integration with physics-based models could enhance the depth of root-cause interpretation.

From an industrial deployment perspective, the proposed framework can be implemented as a real-time monitoring tool integrated with existing ESP surveillance systems. Streaming sensor data can be continuously processed to generate both failure predictions and corresponding explanations, enabling operators to identify abnormal conditions and their root causes in near real time. This capability supports proactive maintenance strategies and reduces reliance on manual interpretation of individual sensor signals.

Finally, the current study focuses on classification-based failure diagnosis rather than long-term prognostics. Extending the framework to include remaining useful life (RUL) prediction or degradation trajectory modeling would further enhance its practical value.

Future work should therefore focus on validating the proposed framework across multiple ESP systems and diverse operating environments, improving labeling robustness, and integrating data-driven methods with physics-based models to achieve more comprehensive and generalizable diagnostic solutions.

7.8. Summary of Key Findings

In summary, the results demonstrate that the proposed explainable AI framework:

١. Accurately classifies ESP operating states and failure modes
٢. Provides early warning of failures several days in advance
٣. Identifies physically meaningful drivers of failure through SHAP analysis
٤. Improves diagnostic transparency compared with black-box models

These findings confirm that integrating machine learning with explainable AI techniques represents a significant advancement in ESP condition monitoring and predictive maintenance.

8. Conclusion

This study presented an explainable artificial intelligence framework for diagnosing Electric Submersible Pump (ESP) operating conditions using real multivariate field data. By integrating Principal Component

Analysis (PCA), Random Forest classification, and SHapley Additive exPlanations (SHAP), the proposed approach addresses key limitations of conventional ESP monitoring methods, particularly the lack of interpretability and limited ability to distinguish between failure modes.

The results demonstrate that PCA effectively captures the dominant behavior of correlated ESP sensor variables and provides a compact representation of system dynamics. Based on this reduced feature space, the Random Forest classifier achieved high accuracy in distinguishing between stable operation, transitional instability, and multiple failure modes. In addition to classification performance, the framework successfully identified failure conditions several days prior to shutdown events, highlighting its capability for early warning and proactive maintenance planning.

A key contribution of this work lies in the integration of SHAP-based explainability, which enables model predictions to be interpreted in terms of physically meaningful variables. Both global and local SHAP analyses revealed that dominant contributors to failure prediction—such as motor current, temperature, and vibration—are consistent with known ESP failure mechanisms. This alignment between data-driven results and engineering understanding enhances trust in the model and supports informed decision-making in field operations.

While the proposed framework demonstrates strong potential, it is important to recognize that the current validation is based on data from a single ESP installation. Therefore, the findings should be interpreted as a case-study demonstration of the methodology. Further validation across multiple wells, operating conditions, and ESP configurations is required to assess generalizability and robustness.

This work distinguishes itself from existing ESP diagnostic studies by explicitly integrating explainability into a multi-class predictive framework using real field data.

Future research will focus on extending the framework to multi-well datasets, incorporating additional machine learning models for benchmarking, and integrating physics-based knowledge to improve interpretability and diagnostic depth. In addition, the extension of the approach toward prognostic applications, such as remaining useful life estimation, represents a promising direction for enhancing predictive maintenance strategies.

Overall, this study demonstrates that combining machine learning with explainable AI techniques provides a powerful and practical approach for ESP failure diagnosis, offering a promising foundation for the development of transparent and reliable predictive maintenance systems in digital oilfield applications.

Abbreviation

Abbreviation	Definition
AI	Artificial Intelligence
ANN	Artificial Neural Network

CBM	Condition-Based Maintenance
CNN	Convolutional Neural Network
CPV	Cumulative Percentage of Variance
DL	Deep Learning
ESP	Electric Submersible Pump
F1-score	Harmonic Mean of Precision and Recall
IoT	Internet of Things
LIME	Local Interpretable Model-Agnostic Explanations
LSTM	Long Short-Term Memory
ML	Machine Learning
PCA	Principal Component Analysis
RF	Random Forest
ROC	Receiver Operating Characteristic
RUL	Remaining Useful Life
SHAP	SHapley Additive exPlanations
SPE	Squared Prediction Error
SVM	Support Vector Machine
T^2	Hotelling's T-squared Statistic
XAI	Explainable Artificial Intelligence

9. References

1. AL-Hejjaj, M.A., et al., *A review of the electrical submersible pump development chronology.* 2023. 24(2): p. 123-135. https://doi.org/10.31699/IJCPE.2023.2.14
2. Panbarasan, M., et al., *Characterization and performance enhancement of electrical submersible pump (ESP) using artificial intelligence (AI).* 2022. **62**: p. 6864-6872. https://doi.org/10.1016/j.matpr.2022.05.101
3. Hassan, K., et al. *Enhancing Giant Oil Field Performance Through Advanced ESP Systems and Accurate Reservoir Monitoring*. in *SPE Annual Caspian Technical Conference*. 2024. SPE. DOI: 10.2118/223441-MS
4. Khalili, Y., Y. Rafiei, and M. Sharifi, *Reservoir Characterization by Applying Pressure Transient Analysis on Data Obtained from Electrical Submersible Pumps.* 2022. DOI: 10.22078/pr.2022.4816.3158.
5. Fakher, S., et al., *Rigorous review of electrical submersible pump failure mechanisms and their mitigation measures.* 2021. **11**(10): p. 3799-3814. https://doi.org/10.1007/s13202-021-01271-6
6. Khalili, Y., et al., *Predictive Maintenance for ESPs: Enhancing Reliability, Efficiency, and Sustainability in Oil & Gas Production.* 2026: p. e239077.

7. Shittu, I.A. *Effective Management of Electric Submersible Pumps: A Predictive Failure Analytics Approach*. in *SPE Nigeria Annual International Conference and Exhibition*. 2024. SPE. https://doi.org/10.2118/221629-MS
8. Bajwa, A., A.A.R. Tonoy, and M. Khan, *IoT-enabled condition monitoring in power transformers: A proposed model.* 2025. https://ssrn.com/abstract=5341323
9. Khalili, Y., M. Ahmadi, and M.K. Moraveji, *A Rule-Based Expert System for Real-Time Fault Diagnosis in Electrical Submersible Pump Systems.* 2025. https://doi.org/10.21203/rs.3.rs-7454357/v1
10. Bafghi, M.B. and A. Vahedi. *A comparison of electric motors for electrical submersible pumps used in the oil and gas industry*. in *IOP Conference Series: Materials Science and Engineering*. 2018. IOP Publishing. DOI 10.1088/1757-899X/433/1/012091
11. Abdalla, R., et al., *Machine learning approach for predictive maintenance of the electrical submersible pumps (ESPS).* 2022. **7**(21): p. 17641-17651. https://doi.org/10.1021/acsomega.1c05881
12. Hernandez, C., et al. *Enhancing Upstream Operations: Hybrid Approach for ESP Failure Prediction in ADNOC's Assets for Operational Excellence and Digital Progress*. in *Abu Dhabi International Petroleum Exhibition and Conference*. 2024. SPE. https://doi.org/10.2118/222892-MS
13. AlBallam, S., *Applying machine learning models to diagnose failures in electrical submersible pumps*. 2022. https://shareok.org/handle/11244/336890
14. Brito, L.C., et al. *Fault detection of bearing: An unsupervised machine learning approach exploiting feature extraction and dimensionality reduction*. in *Informatics*. 2021. MDPI. https://doi.org/10.3390/informatics8040085
15. Alhashem, M., et al. *Evaluation of machine learning techniques for ESP diagnosis using a synthetic time series dataset*. in *International Petroleum Technology Conference*. 2024. IPTC. https://doi.org/10.2523/IPTC-24210-MS
16. Peng, L., et al., *Electric submersible pump broken shaft fault diagnosis based on principal component analysis.* 2020. **191**: p. 107154. https://doi.org/10.1016/j.petrol.2020.107154
17. da Silva, L.H.P., et al., *Active learning for new-fault class sample recovery in electrical submersible pump fault diagnosis.* 2023. **212**: p. 118508. https://doi.org/10.1016/j.eswa.2022.118508
18. Ahmed, I., G. Jeon, and F.J.I.t.o.i.i. Piccialli, *From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where.* 2022. **18**(8): p. 5031-5042. DOI: 10.1109/TII.2022.3146552
19. Xu, H., et al., *Interpretable intelligent fault diagnosis for heat exchangers based on SHAP and XGBoost.* 2025. **13**(1): p. 219. https://doi.org/10.3390/pr13010219
20. Leite, D., et al., *Fault detection and diagnosis in industry 4.0: a review on challenges and opportunities.* 2024. **25**(1): p. 60. DOI: 10.3390/s25010060
21. Al-Ballam, S., H. Karami, and D. Devegowda. *A hybrid physical and machine learning model to diagnose failures in electrical submersible pumps*. in *SPE/IADC Middle East Drilling Technology Conference and Exhibition*. 2023. SPE. https://doi.org/10.2118/214632-MS
22. Khalili, Y., et al., *Time-aware predictive maintenance of electrical submersible pumps using catboost ensemble learning and trend-based labeling.* 2025. **15**(9): p. 147. https://doi.org/10.1007/s13202-025-02070-z
23. El Gindy, M., et al. *Monitoring & surveillance improve ESP operation and reduce workover frequency*. in *Abu Dhabi International Petroleum Exhibition and Conference*. 2015. SPE. https://doi.org/10.2118/177926-MS
24. Hoefel, A., et al. *ESP System Monitoring and Diagnosis from Surface Power Acquisition*. in *SPE Gulf Coast Section Electric Submersible Pumps Symposium*. 2023. SPE. https://doi.org/10.2118/214739-MS

25. Alkhawaher, A., H. Al Jishi, and S. Al-Aseef. *Advanced ESP Monitoring System for Performance and Production Optimization*. in *SPE Middle East Artificial Lift Conference and Exhibition*. 2024. SPE. https://doi.org/10.2118/221543-MS
26. Almajid, H., et al. *An Integrated Approach Utilizing ESP Design Improvements and Real Time Monitoring to Ensure Optimum Performance and Maximize Run Life*. in *Abu Dhabi International Petroleum Exhibition and Conference*. 2019. SPE. https://doi.org/10.2118/197209-MS
27. Khalili, Y., et al., *A comprehensive review of failure modes in electrical submersible pumps: Diagnosis, predictive maintenance, and engineer's guide.* ٢٠٢٥. ٥٠(٢٤): p. ٢٠٤٤٥-٢٠٤٦٦. https://doi.org/10.1007/s13369-025-10536-9
28. Cardona, L., P. Vivas Sanchez, and B. Joya. *Failure prediction methodology for ESP and operational behavior*. in *SPE Latin America and Caribbean Petroleum Engineering Conference*. 2023. SPE. https://doi.org/10.2118/213140-MS
29. Saptadi, S., et al., *Implementation of machine learning methods in predicting failures in electrical submersible pump machines.* 2025. **7**(3): p. 2025137-2025137. https://10.31893/multiscience.2025137
30. Devshali, S., et al. *Predicting esp failures using artificial intelligence for improved production performance in one of the offshore fields in india*. in *Abu Dhabi International Petroleum Exhibition and Conference*. 2022. SPE. https://doi.org/10.2118/211031-MS
31. Camacho, J., et al., *PCA-based multivariate statistical network monitoring for anomaly detection.* 2016. 59: p. 118-137. https://doi.org/10.1016/j.cose.2016.02.008
32. Ge, Z.J.I. and E.C. Research, *Process data analytics via probabilistic latent variable models: A tutorial review.* 2018. **57**(38): p. 12646-12661. https://doi.org/10.1021/acs.iecr.8b02913
33. Garcia-Alvarez, D., et al., *Fault Detection and Diagnosis using Multivariate Statistical Techniques in a Wastewater Treatment Plant.* 2009. **42**(11): p. 952-957. https://doi.org/10.3182/20090712-4-TR-2008.00156
34. Yin, S., et al., *A review on basic data-driven approaches for industrial process monitoring.* 2014. 61(11): p. 6418-6428. DOI: 10.1109/TIE.2014.2301773
35. Jang, K., et al., *Explainable artificial intelligence for fault diagnosis of industrial processes.* 2023. 21(1): p. 4-11. DOI: 10.1109/TII.2023.3240601
36. Cação, J., J. Santos, and M.J.J.o.I.I.I. Antunes, *Explainable AI for industrial fault diagnosis: A systematic review.* 2025: p. 100905. https://doi.org/10.1016/j.jii.2025.100905
37. Cohen, J., X. Huan, and J.J.J.o.I.M. Ni, *Shapley-based explainable AI for clustering applications in fault diagnosis and prognosis.* 2024. **35**(8): p. 4071-4086. https://doi.org/10.1007/s10845-024-02468-2
38. Brusa, E., et al., *Explainable AI for machine fault diagnosis: understanding features' contribution in machine learning models for industrial condition monitoring.* 2023. **13**(4): p. 2038. https://doi.org/10.3390/app13042038
39. Zhang, J., et al. *Review on Fault Diagnosis of Electric Submersible Pump using Machine Learning*. in *Journal of Physics: Conference Series*. 2025. IOP Publishing. **DOI** 10.1088/1742-6596/3129/1/012056
40. Sobhy, M.A., *Detecting Electrical Submersible Pump (ESP) Failures and Estimating Run Life Using Artificial Neural Networks.* 2025. https://fount.aucegypt.edu/etds/2605
41. Ambade, A., et al. *Electrical submersible pump prognostics and health monitoring using machine learning and natural language processing*. in *SPE Middle East Intelligent Oil and Gas Symposium*. 2021. SPE. https://doi.org/10.2118/208649-MS
42. Silvia, S., et al. *Case study: predicting electrical submersible pump failures using artificial intelligence and physics-based hybrid models*. in *SPE Middle East Intelligent Oil and Gas Symposium*. 2023. SPE. https://doi.org/10.2118/214462-MS
43. Iranzi, J., et al., *A nodal analysis based monitoring of an electric submersible pump operation in multiphase flow.* 2022. **12**(6): p. 2825. https://doi.org/10.3390/app12062825

44. Reges, G., et al., *Electric submersible pump vibration analysis under several operational conditions for vibration fault differential diagnosis.* 2021. **219**: p. 108249. https://doi.org/10.1016/j.oceaneng.2020.108249
45. Liu, D., et al., *Hybrid Long Short-Term Memory and Convolutional Neural Network Architecture for Electric Submersible Pump Condition Prediction and Diagnosis.* 2024. **29**(05): p. 2130-2147. https://doi.org/10.2118/218418-PA
46. Gong, F., et al., *A Fault Diagnosis Model of an Electric Submersible Pump Based on Mechanism Knowledge.* 2025. **25**(8): p. 2444. https://doi.org/10.3390/s25082444
47. Rodrigues, D.A., et al., *Fault diagnosis of electric submersible pumps using vibration signals.* 2023. 45(9): p. 445. https://doi.org/10.1007/s40430-023-04370-z
48. Song, Y., et al., *Diagnosis of electrical submersible pump failure using deep learning model with sand-water flow experimental data.* 2024. **243**: p. 213279. https://doi.org/10.1016/j.geoen.2024.213279
49. Khond, S.V.J.S.S., *Effect of data normalization on accuracy and error of fault classification for an electrical distribution system.* 2020. **8**(3): p. 117-124. https://doi.org/10.1080/23080477.2020.1799135
50. Greenacre, M., et al., *Principal component analysis.* 2022. **2**(1): p. 100. https://doi.org/10.1038/s43586-022-00184-w
51. Genuer, R. and J.-M. Poggi, *Random forests*, in *Random forests with R*. 2020, Springer. p. 33-55. https://doi.org/10.1007/978-3-030-56485-8_3
52. Lundberg, S.M. and S.-I.J.A.i.n.i.p.s. Lee, *A unified approach to interpreting model predictions.* 2017. 30. https://doi.org/10.48550/arXiv.1705.07874
53. Lei, Y., et al., *Machinery health prognostics: A systematic review from data acquisition to RUL prediction.* 2018. **104**: p. 799-834. https://doi.org/10.1016/j.ymssp.2017.11.016
54. Zonta, T., et al., *Predictive maintenance in the Industry 4.0: A systematic literature review.* 2020. 150: p. 106889. https://doi.org/10.1016/j.cie.2020.106889
55. Sokolova, M., G.J.I.p. Lapalme, and management, *A systematic analysis of performance measures for classification tasks.* 2009. **45**(4): p. 427-437. https://doi.org/10.1016/j.ipm.2009.03.002
56. Powers, D.M.J.a.p.a., *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.* 2020. https://doi.org/10.48550/arXiv.2010.16061
57. Stehman, S.V.J.R.s.o.E., *Selecting and interpreting measures of thematic classification accuracy.* 1997. 62(1): p. 77-89. https://doi.org/10.1016/S0034-4257(97)00083-7
58. Jardine, A.K., et al., *A review on machinery diagnostics and prognostics implementing condition-based maintenance.* 2006. **20**(7): p. 1483-1510. https://doi.org/10.1016/j.ymssp.2005.09.012