Review of Applied Science and Technology

Volume: *6;* ***Issue:*** *1*
Pages: *108– 138*

DATA DRIVEN PREDICTIVE MAINTENANCE IN PETROLEUM AND POWER SYSTEMS USING RANDOM FOREST REGRESSION MODEL FOR RELIABILITY ENGINEERING FRAMEWORK

Zobayer Eusufzai[1];

[1]. *Technical Sales Manager, TSI Group, Authorized Distributor of Total Energies Lubricants, Dhaka, Bangladesh; Email: zobayer54@gmail.com*

Doi: 10.63125/5bjx6963

Received: *12 September 2021;* ***Revised:*** *20 October 2021;* ***Accepted:*** *18 November 2021;* ***Published:*** *24 December 2021*

Abstract

This study addresses the persistent problem that predictive maintenance in petroleum and power systems is frequently implemented with fragmented data architectures, limited analytics capability, and weak linkage to reliability engineering outcomes—factors that collectively constrain its influence on equipment availability, maintainability and risk reduction. The purpose of the research is to develop and empirically evaluate a data-driven predictive maintenance framework that embeds a Random Forest regression model within a structured reliability engineering context. A quantitative, cross-sectional, case-based design is adopted, combining survey responses from 210 professionals in petroleum and power enterprises (70 percent response rate from 300 questionnaires) with 12,480 equipment time-series records from critical assets such as transformers, gas turbines and centrifugal compressors. Key variables include data quality, system integration, staff competency, organizational support, perceived predictive maintenance effectiveness, adoption intensity, and a quantitative degradation or risk score derived from operational parameters. The analysis plan integrates descriptive statistics, reliability and validity testing, Pearson correlations, multiple regression, and Random Forest regression compared directly with linear regression. The measurement model demonstrates strong internal consistency (Cronbach's alpha 0.86 to 0.91), and the TOE-based predictors explain 62 percent of the variance in perceived effectiveness (adjusted $R^2 = 0.62$), with organizational support ($\beta = 0.31$, $p < .001$) and data quality ($\beta = 0.28$, $p < .001$) emerging as dominant drivers of successful adoption. The Random Forest model achieves $R^2 = 0.87$ and RMSE = 0.94 on a 0–10 risk scale, outperforming linear regression ($R^2 = 0.71$, RMSE = 1.45) and highlighting load factor, temperature, vibration intensity and oil quality as the most influential degradation predictors. These findings imply that petroleum and power enterprises can materially strengthen reliability-centered and risk-based maintenance decisions by jointly investing in data governance, system integration, analytics skills development, and ensemble machine learning approaches. The study provides both methodological and practical contributions by demonstrating how predictive models, when aligned with reliability engineering principles and organizational readiness, can deliver measurable improvements in asset health forecasting and operational risk management.

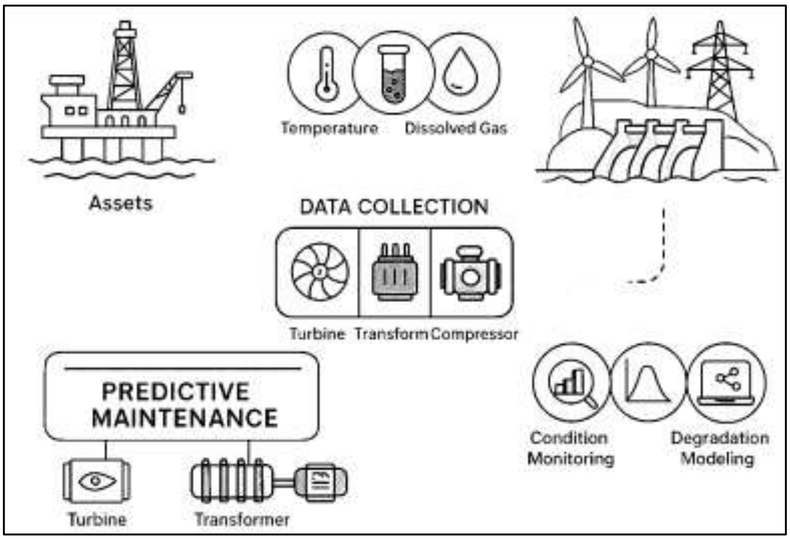
Keywords

Predictive Maintenance, Random Forest Regression, Reliability Engineering, Petroleum and Power Systems, Technology Organization Environment (TOE) Framework;

INTRODUCTION

Predictive maintenance is widely recognized as a core pillar of modern reliability engineering because it uses real-time and historical data to infer the evolving health of critical assets and schedule interventions before functional failure occurs. In contrast to purely time-based preventive maintenance, predictive maintenance relies on condition indicators—such as vibration, temperature, dissolved gas levels, or process measurements—to estimate degradation and remaining useful life within a probabilistic framework. Machinery diagnostics and prognostics have emerged as a systematic extension of condition-based maintenance, integrating sensing, signal processing, and reliability modeling to support maintenance decision-making at scale (Jardine et al., 2006). Subsequent reviews have emphasized that this paradigm has global economic significance, because maintenance-related downtime contributes billions of dollars in annual losses across energy, manufacturing, and transportation sectors (Heng et al., 2009). In this context, petroleum production facilities and electric power systems are especially sensitive to unplanned outages: failures in turbines, compressors, transformers, and high-voltage equipment can trigger cascading technical, financial, safety, and environmental consequences. Internationally, utilities and oil and gas operators therefore treat maintenance not only as an engineering task but also as a strategic lever for asset management, risk mitigation, and regulatory compliance. Foundational prognostics and health management (PHM) work has shown that effective predictive maintenance frameworks must combine robust sensing architectures, degradation modeling, and decision-support tools that are aligned with plant-level reliability and availability targets (Lee et al., 2014).

Figure 1: Integrated Predictive Maintenance Ecosystem for Petroleum and Power Systems



The evolution of predictive maintenance has been closely tied to advances in condition-based maintenance (CBM) and reliability-centered maintenance concepts, which shift attention from fixed schedules toward the actual condition and risk profile of equipment. CBM has been characterized as a structured process linking condition monitoring, fault diagnosis, and prognostics to maintenance decisions, offering cost benefits when failure mechanisms are observable and partially predictable (Bukhsh et al., 2019; Jardine et al., 2006). Building on this foundation, machine prognostics research has highlighted that reliable remaining useful life estimation is central to condition-based strategies, especially for assets operating under variable loads and harsh environments (Peng et al., 2010; Si et al., 2011). In the power sector, dissolved gas analysis has been shown to support predictive maintenance of power transformers by detecting incipient faults in insulation systems, thereby linking chemical signatures to reliability indicators (de Faria et al., 2015). Condition monitoring and diagnostic test methods for transformer lifetime estimation have been synthesized to illustrate how multi-source diagnostic data can be integrated to support risk-based maintenance planning in high-voltage networks (Islam et al., 2017). In hydroelectric contexts, condition-based maintenance optimization models for hydro generating units have incorporated dynamic economic dependence and proportional hazards modeling, framing maintenance timing as a reliability and cost trade-off problem (Qian & Wu, 2014). These contributions underscore the international relevance of CBM and predictive maintenance frameworks in power and petroleum-related infrastructures, where technical failures often translate directly into large-scale societal and economic disruptions.

From a methodological perspective, data-driven prognostics and health management have become increasingly prominent as sensing, storage, and computation capacities have expanded. Statistical data-driven approaches to remaining useful life estimation have been reviewed and classified into stochastic process, reliability-based, and machine-learning-oriented categories, emphasizing that data-driven methods are particularly attractive when physics-of-failure models are incomplete or unavailable (Simões et al., 2011). Remaining useful life estimation techniques have also been synthesized to argue that pragmatic industrial prognostics often require balancing model complexity with interpretability and data availability constraints (Ahmadzadeh & Lundberg, 2014). In parallel, key challenges in rotating machinery prognostics—including non-stationary operating conditions, sensor noise, and sparse failure examples—have been identified, all of which motivate the adoption of robust learning algorithms and ensemble techniques (Kumar, 2006). PHM design has been positioned as a multi-level problem encompassing sensing, feature extraction, prognostics modeling, and maintenance decision-making, with an emphasis on embedding advanced models within operational reliability frameworks to ensure practical usefulness (Kumar, 2006). More recently within the 2005–2019 period, intelligent data-driven predictive maintenance for industrial assets has been discussed, highlighting the role of machine learning algorithms—including ensemble methods such as random forests—in transforming large-scale condition monitoring data into actionable reliability information (Kumar et al., 2013). Together, these studies establish that data-driven predictive maintenance is not simply a signal-processing task but a reliability engineering challenge in which model choice, data quality, and maintenance policy must be jointly considered.

Within petroleum production and power system operations, the requirements for reliability engineering are particularly stringent because of interconnected networks, continuous process constraints, and high consequence of failure. Transformers have been noted as critical links between generation sources and transmission networks, and predictive maintenance based on dissolved gas analysis can materially improve operational life expectancy and reliability indices (Okoh et al., 2014). Lifetime estimation of transformers has been shown to depend on integrating multiple diagnostic tests—such as insulation resistance, frequency response analysis, and partial discharge measurements—within unified condition-assessment schemes (Lin et al., 2018). In generation and hydro assets, condition-based maintenance models that explicitly account for dynamic economic dependence between components have been proposed to minimize total maintenance cost while preserving system reliability (Abdulla & Ibne, 2021; Susto et al., 2015). At the distribution and substation levels, electrical equipment such as circuit breakers, busbars, and protective devices are increasingly managed with multi-attribute decision frameworks for condition-based maintenance, where safety, reliability, and cost are jointly evaluated (Habibullah & Foysal, 2021; Parida & Kumar, 2006). Across these infrastructures, a unifying theme is the need for predictive models that can integrate heterogeneous sensor data—vibration, temperature, oil quality, process variables—with reliability metrics to support asset-level and system-level decisions. This requirement motivates research that brings together reliability engineering constructs, such as failure rate, hazard function, and system availability, with data-driven prediction models capable of capturing nonlinear degradation patterns under realistic operating profiles (Sarwar, 2021).

Despite broad progress in condition monitoring and reliability analysis, there remains a clear problem in the consistent, quantitative integration of modern machine learning models, particularly ensemble regressors such as random forest regression, into predictive maintenance frameworks for petroleum and power systems (Musfiqur & Saba, 2021). Most existing work in transformers and high-voltage equipment has focused on diagnostic indicators and threshold-based rules, with limited deployment of fully data-driven prognostic models that produce continuous remaining useful life or failure probability estimates (Carvalho et al., 2019; Redwanul et al., 2021). In parallel, many CBM optimization studies in power generation and hydro units have incorporated proportional hazards or other parametric reliability models without explicitly exploiting high-dimensional sensor data through flexible nonparametric learners (Tarek & Praveen, 2021; Qian & Wu, 2014). The international literature on data-driven predictive maintenance indicates that ensemble learning methods provide robustness to noise, handle complex feature interactions, and offer reliable generalization performance when tuned appropriately (Fink, 2019; Muhammad & Shahrin, 2021). However, there is still limited empirical evidence on how such models perform when embedded within a reliability engineering framework that explicitly targets key metrics relevant to petroleum and power system stakeholders, such as mean time between failures, system availability, and risk of critical outage (Saikat, 2021; Shaikh & Aditya, 2021). This gap frames the problem addressed by the present study: the need for a quantitatively rigorous, data-driven predictive maintenance model grounded in random forest regression and evaluated within a reliability engineering context for petroleum and power system assets.

The purpose of this study is to develop and empirically evaluate a data-driven predictive maintenance framework for petroleum and power system equipment that integrates random forest regression within a reliability engineering model of asset degradation and failure. Drawing on insights from data-driven prognostics (Zhang et al., 2019) and PHM design (Tsui et al., 2015), the study aims to construct a survey-based and case-study–based quantitative model that captures how organizations implement predictive maintenance practices, deploy data analytics tools, and translate model outputs into maintenance decisions at the asset and system levels. At the conceptual level, the research focuses on three core constructs: the maturity of data-driven maintenance practices (e.g., data collection, quality, and usage), the sophistication of predictive modeling (with emphasis on random forest regression), and the effectiveness of reliability outcomes (e.g., reduced unplanned downtime, increased availability, and extended equipment life). Building on the reviewed literature, three research questions can be articulated: (RQ1) How extensively are data-driven predictive maintenance practices adopted in petroleum and power system organizations relative to traditional preventive or corrective maintenance? (RQ2) To what extent does the use of advanced predictive modeling, particularly random forest regression, relate to improvements in reliability performance indicators? (RQ3) How do organizational and technical factors mediate the relationship between predictive maintenance analytics and observed maintenance outcomes? These questions align the study with international efforts to frame predictive maintenance as an integrated reliability engineering function rather than an isolated analytics exercise (Zio, 2013).

Guided by these research questions, the study formulates hypotheses that link data-driven maintenance practices, modeling approaches, and reliability outcomes in a testable manner. Prior work suggests that greater sophistication in condition monitoring and diagnostics is associated with improved prognostics performance and more effective maintenance planning (Pecht & Jaai, 2010). It is therefore reasonable to hypothesize that higher levels of data collection quality and integration (e.g., continuous monitoring, structured databases, sensor fusion) will be positively associated with perceived predictive maintenance effectiveness (H1). Building on discussions of data-driven models under complex and uncertain degradation conditions (de Faria et al., 2015), the study also hypothesizes that organizations which implement advanced machine learning techniques—operationalized here by the adoption of random forest regression models for failure time or health index prediction—will report better reliability performance than those relying solely on simple trend analysis or rule-based methods (H2). Furthermore, literature on decision-making in condition-based maintenance suggests that multi-criteria and risk-based maintenance policies, informed by predictive indicators, can reduce maintenance cost and improve system reliability simultaneously (Qian & Wu, 2014). Accordingly, a third hypothesis (H3) posits that organizations linking predictive model outputs explicitly to structured reliability decision rules—such as risk thresholds, priority indices, or optimized maintenance schedules—will exhibit stronger improvements in asset availability and perceived maintenance performance. These hypotheses set the stage for quantitative testing using Likert-type survey data, correlation analysis, and regression modeling.

Beyond hypotheses, this study seeks to articulate its contributions and significance at methodological, theoretical, and practical levels within the international discourse on predictive maintenance for petroleum and power systems. Methodologically, the research synthesizes insights from PHM and prognostics reviews to design an empirical framework that connects survey-based constructs of maintenance practice maturity with specific modeling technologies, namely random forest regression (Jardine et al., 2006). Theoretical contribution arises from integrating reliability engineering concepts, such as hazard rate, remaining useful life, and condition-based decision thresholds, with data-driven modeling constructs, thereby refining how predictive maintenance is conceptualized as part of a broader reliability engineering framework rather than as a standalone analytics application (Peng et al., 2010). Practically, the focus on petroleum and power systems responds to documented needs in transformer diagnostics, high-voltage equipment maintenance, and hydro generating unit reliability, where predictive maintenance is increasingly recognized as essential to safety, environmental protection, and cost control (de Faria et al., 2015). By explicitly centering random forest regression within this reliability engineering framework and evaluating its perceived impact via a quantitative, case-study–oriented survey, the study aims to offer a structured reference for practitioners and researchers who wish to design, assess, and refine data-driven predictive maintenance programs in petroleum and power system contexts.

In addtion, the organization of this paper follows a structure aligned with the above conceptualization. The next section presents a focused literature review that elaborates on the core concepts of predictive maintenance, condition-based maintenance, and reliability engineering in petroleum and power systems, as well as theoretical and conceptual frameworks relevant to data-driven prognostics and random forest modeling. Foundational and domain-specific studies are discussed to position the research within existing academic and industrial work (Jardine et al., 2006). The subsequent methodology section details the quantitative, cross-sectional, case-study–based research design, including population and sampling, survey instrument development using Likert five-point scales, operationalization of constructs, pilot testing, and data analysis techniques based on descriptive statistics, correlation, regression modeling, and random forest regression. The results section then reports response rates, sample characteristics, reliability and validity of measurement scales, and the outcomes of correlation, regression, and random forest analyses with respect to the formulated hypotheses. This is followed by a discussion section that interprets the empirical findings in light of the literature on prognostics, PHM, and condition-based maintenance in petroleum and power systems. The paper concludes with sections on conclusions, recommendations, and limitations, each aligned with the overarching aim of advancing data-driven predictive maintenance within a reliability engineering framework for petroleum and power system applications.

The overall objective of this study is to develop and empirically evaluate a data-driven predictive maintenance framework for petroleum and power system equipment that integrates random forest regression within a structured reliability engineering context. In line with this overall aim, the first specific objective is to systematically assess the current state of maintenance practice in selected petroleum and power organizations, with particular attention to the extent of predictive maintenance adoption, the maturity of data collection and condition monitoring processes, and the level of integration between maintenance, operations, and information systems. The second objective is to identify and operationalize the key technical and organizational factors that shape the effectiveness of data-driven predictive maintenance, including data quality, system integration capability, staff competency in analytics and machine learning, and organizational support for advanced reliability methods. The third objective is to construct a quantitative survey instrument based on Likert's five-point scale that captures these constructs in a reliable and valid way for use with maintenance engineers, reliability specialists, and operational managers in petroleum and power sectors. The fourth objective is to use descriptive statistics, correlation analysis, and regression modeling to test a set of hypotheses linking the identified factors to perceived predictive maintenance effectiveness and adoption, thereby clarifying which conditions are most strongly associated with improved reliability outcomes. Parallel to the survey-based analysis, the fifth objective is to design and implement a random forest regression model using historical operational and maintenance data from the case organizations, with the intention of predicting relevant maintenance indicators such as failure-related events, degradation scores, or condition indices. The performance of this model will then be compared with that of a conventional regression approach using standard error and goodness-of-fit metrics, in order to evaluate the added value of random forest regression within a reliability engineering framework. A final objective is to synthesize the statistical findings from the survey and the predictive results from the random forest model into a coherent reliability-oriented maintenance framework tailored to petroleum and power systems, providing a structured basis for subsequent methodological and empirical work in data-driven predictive maintenance.

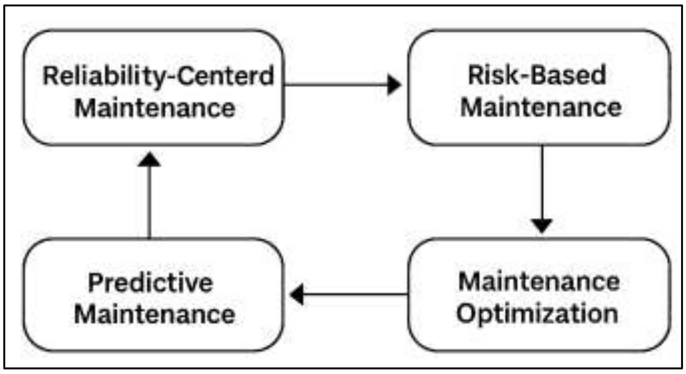
LITERATURE REVIEW

The literature on predictive maintenance, reliability engineering, and data-driven modeling provides the foundation for understanding how random forest regression can be embedded within a maintenance framework for petroleum and power systems. Early work on condition-based maintenance positioned machinery diagnostics and prognostics as the core mechanisms by which condition monitoring data are converted into actionable maintenance decisions, emphasizing the three-step cycle of data acquisition, data processing, and decision-making (Jardine et al., 2006). Within this paradigm, prognostics and health management research highlighted remaining useful life estimation as a central concept, with statistical data-driven approaches designed to infer the stochastic degradation trajectories of components from observed condition indicators (Si et al., 2011). In energy infrastructures, these ideas have been operationalized through a variety of monitoring techniques, such as dissolved gas analysis and online condition monitoring for power transformers, which allow utilities to relate evolving physical signatures to underlying fault mechanisms and reliability indices (de Faria et al., 2015). More recent contributions extend this logic to distribution transformers and other grid assets by proposing data-driven frameworks that rely on operational and contextual variables—such as loading and ambient conditions—rather than purely laboratory-based measurements, illustrating how predictive maintenance can be scaled using routinely collected data (Tsui et al., 2015). Parallel strands of research have explored the broader family of machine learning models for predictive maintenance, noting that algorithms such as random forests, gradient boosting, and support vector machines are able to capture nonlinear relationships and complex feature interactions that are difficult to model using traditional regression techniques, particularly when the input space comprises heterogeneous sensor and operational data. Within this family, random forest regression has attracted attention for its robustness to noise, ability to handle high-dimensional feature sets, and provision of variable importance measures that can be interpreted in reliability terms, such as identifying which operating or environmental factors most strongly influence failure-related responses (Kumar, 2006). Collectively, this body of work establishes an intellectual trajectory from condition-based maintenance concepts and statistical RUL models toward more flexible, ensemble-based learning approaches and points to the need for sector-specific investigations of how such models perform when applied to critical petroleum and power system assets within an explicit reliability engineering framework.

Maintenance Strategies and Reliability Engineering in Petroleum and Power Systems

Maintenance strategies in complex engineering systems have progressively evolved from simple corrective practices toward structured, reliability-informed approaches that explicitly link maintenance actions to system performance and risk. In early applications to electric power distribution networks, reliability-centered asset maintenance methods were developed to quantify how different maintenance policies influence system-level reliability indices and operating costs, using probabilistic models of component failures and restoration processes (Bertling et al., 2005). This work formalized the idea that maintenance planning must be driven by the criticality of components and their contribution to customer interruptions, rather than by uniform time-based schedules. At a more general level, the broader field of maintenance optimization has sought to systematize how preventive, predictive, and corrective strategies are selected and combined under constraints of budget, reliability targets, and asset condition, reviewing a wide spectrum of models ranging from age-based replacement rules to stochastic degradation and condition-based policies (de Jonge & Scarf, 2019). For petroleum production facilities and power systems, where equipment such as transformers, generators, compressors, pipelines, and switchgear operates in harsh environments and is organized into tightly coupled networks, this reliability-centered perspective is essential because failures propagate beyond the individual asset and can trigger cascading outages, environmental incidents, and substantial economic losses. The underlying principle that emerges from the reliability literature is that maintenance should be designed as an integrated part of asset management, where component-level deterioration models, system reliability simulations, and cost–benefit analyses are jointly employed to determine inspection intervals, replacement timing, and condition-based intervention thresholds. In both petroleum and power domains, this integration provides the conceptual bridge through which data-driven predictive maintenance and machine learning models—such as random forest regression—can be aligned with reliability engineering objectives and constraints.

Figure 2: Key Maintenance Approaches in Reliability Engineering



Within this reliability-oriented view, risk-based maintenance has become an influential paradigm for prioritizing actions on components whose failure consequences are most severe. A landmark contribution in this direction classified risk-based maintenance techniques according to their use of deterministic, probabilistic, and combined risk analysis methods, and emphasized that inspection and maintenance frequencies should be proportional to both the probability and consequence of failure (Arunraj & Maiti, 2007). In practical terms, risk-based maintenance frameworks encourage utilities and industrial operators to build risk matrices or quantitative risk indices that blend failure likelihood, consequence severity, and detectability, and then allocate resources toward high-risk assets such as large power transformers, high-pressure process vessels, or critical rotating machinery. For petroleum systems, risk-based approaches are particularly relevant because the release of hazardous materials, fires, and explosions are tied directly to the integrity of processing units, storage tanks, pipelines, and safety barriers, making maintenance a central control for process safety and environmental protection. In power networks, similar logic applies to assets whose failure could cause widespread customer interruptions or violate reliability standards. By coupling risk assessment with maintenance planning, operators can rationalize inspection campaigns, condition-monitoring deployments, and refurbishment programs, ensuring that scarce maintenance resources are directed toward components whose failure would have the greatest impact on system reliability and safety. This risk perspective creates a natural foundation for incorporating predictive models: when the risk index of an asset depends on an underlying failure probability that can be estimated from data, machine learning techniques become attractive tools for updating risk estimates and refining maintenance priorities over time (Carnero & Gómez, 2017).

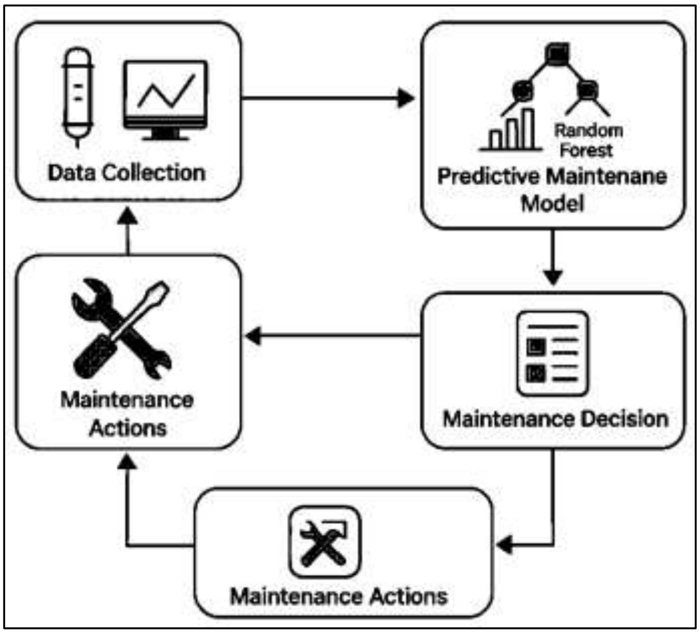
Building on both reliability-centered and risk-based perspectives, more recent work in electric power systems has focused on formulating explicit optimization models that translate maintenance philosophy into actionable schedules for real networks. One line of research has developed reliability-centered maintenance optimization procedures for power distribution systems, prioritizing failure modes via criticality analysis and selecting cost-effective maintenance activities for critical components so that overall reliability indices, such as system average interruption duration, are improved under budgetary constraints (Dehghanian et al., 2014). Complementary contributions have addressed how utilities should select among corrective, preventive, and predictive maintenance strategies using multi-criteria decision models that account for reliability, cost, safety, and service quality, illustrating that mixed strategies tailored to asset criticality can outperform uniform policies in electric power distribution contexts (Carnero & Gómez, 2017). At the same time, high-level syntheses of maintenance optimization research have underscored that real-world applications increasingly require integrating reliability modeling, economic evaluation, and operational constraints into unified frameworks, highlighting gaps in the deployment of advanced decision-support tools in practice (Bertling et al., 2005). For petroleum and power system infrastructures, these developments imply that maintenance strategies must be explicitly modeled at both component and system levels, with reliability and cost metrics serving as objective functions in optimization or decision-analysis schemes. When such schemes are extended to incorporate condition indicators and predictive models—for example, failure probabilities or degradation scores derived from random forest regression—maintenance decisions can transition from static schedules to dynamically updated plans that reflect observed asset behavior and quantified reliability impacts. This section of the literature review therefore positions reliability-centered, risk-based, and optimization-oriented maintenance research as the conceptual backbone for a data-driven predictive maintenance framework in petroleum and power systems.

Data-Driven Predictive Maintenance

Data-driven predictive maintenance translates the general idea of condition-based maintenance into a structured learning problem in which statistical or machine learning models map multi-source data to estimates of equipment health, fault probability, or remaining useful life. Rather than relying solely on expert rules or simple thresholds, algorithms learn discriminative patterns from labelled examples of normal and degraded behavior, enabling early detection of anomalies that are not visible through conventional alarms. Early work on this paradigm in machine condition monitoring showed that support vector machines could offer robust fault-classification performance across a wide range of rotating machinery and fault types when trained on vibration, acoustic, and other sensor signals (Widodo & Yang, 2007). In parallel, comparative studies in smart manufacturing demonstrated that random forests and other ensemble methods could outperform classical neural networks and support vector regression for predicting tool wear and other degradation phenomena, underscoring the importance of algorithm selection in prognostics applications (Wu et al., 2017). Within the conceptual framework of the present study, data-driven predictive maintenance is therefore defined as the systematic use of supervised or semi-supervised models that transform historical and real-time operating data from petroleum and power systems into quantitative indicators such as failure risk scores, predicted degradation levels, or continuous condition indices. These indicators can be scalar, such as an estimated probability of failure within a given horizon, or vector-valued, capturing multiple health dimensions of the same asset, and they are continuously updated as new observations arrive. The resulting decision signals are then used to trigger maintenance interventions only when the statistical evidence suggests that reliability thresholds may be violated, thereby aligning maintenance timing with the evolving state of critical components rather than with fixed time or usage intervals and tightly integrating reliability engineering concepts with data-analytic outputs.

A key insight from the broader prognostics and health management literature is that data-driven predictive maintenance should be conceptualized as a multi-stage pipeline that begins long before model training and ends only when model outputs are embedded into decision processes. A comprehensive review of machinery health prognostics emphasizes four tightly coupled stages—data acquisition, health indicator construction, health stage division, and remaining useful life prediction—that collectively determine the quality of predictive outcomes (Lei et al., 2018). In this pipeline view, raw signals from sensors, control systems, inspection logs, or historian databases are first transformed into engineered features or learned representations that capture degradation-sensitive dynamics, after which learning algorithms establish mappings between these representations and target variables such as time-to-failure, degradation rate, or fault occurrence. For predictive maintenance applications in complex environments like aviation, data-driven models have been successfully combined with stochastic time-series approaches to forecast future fault events at the system level, illustrating how statistical learning can be fused with autoregressive models to provide actionable lead times for maintenance planning and fleet scheduling (Baptista et al., 2017). This perspective implies that the conceptual framework should not be limited to the choice of algorithm but must also encompass data governance, sampling frequencies, sensor placement, and labelling strategies that ultimately shape the training set. Translating these ideas to petroleum and power systems, the current research treats the predictive maintenance model as the core analytical engine within a broader reliability engineering loop that includes data curation, model validation against historical failures, threshold setting, and feedback from field maintenance outcomes. The framework therefore recognizes that improvements in any stage—ranging from the design of health indicators to the calibration of probability thresholds used in decision rules—can propagate through the pipeline and enhance the accuracy, timeliness, and organizational acceptance of maintenance recommendations.

Figure 3: Data Pipeline and Random Forest Modeling Architecture for Predictive Maintenance



Within this conceptualization, classification- and regression-based models play complementary roles in translating heterogeneous maintenance data into interpretable decision cues for managers and engineers. Recent work on railway infrastructure shows that tree-based classification techniques can use routine inspection and operational records to predict maintenance need, activity type, and trigger status for switches, demonstrating that relatively transparent models can still achieve high predictive performance in real network settings and can be interrogated to explain why a particular maintenance action is suggested (Lei et al., 2018)). For systems where the timing and magnitude of degradation must be estimated on a continuous scale, ensemble regression methods such as random forests offer the ability to approximate complex nonlinear relationships between environmental conditions, loading histories, and measured responses, while naturally capturing interactions and handling missing or noisy data (Wu et al., 2017). Building on these insights, the conceptual framework for the present study positions a random forest regression model at the center of a reliability engineering architecture in which input variables include sensor readings, operational parameters, external stressors, and maintenance histories from petroleum and power assets, and outputs correspond to predicted degradation metrics, reliability indices, or risk scores that can be readily translated into maintenance priorities. These outputs then feed explicit decision rules that specify when to schedule inspections, component replacements, or overhauls, and how to allocate maintenance resources across competing assets in order to maintain target reliability levels and key performance indicators such as availability and mean time between failures. In operational terms, this means that the learning model, the reliability criteria, and the organizational decision rules are treated as an integrated system whose components must be jointly designed, calibrated, and evaluated to achieve effective predictive maintenance in critical energy infrastructures. Viewed in this way, the proposed framework clarifies how machine learning outputs connect to managerial levers such as budgeting, shutdown planning, and risk communication, which will be empirically examined through the survey-based quantitative analysis in later sections of the study.

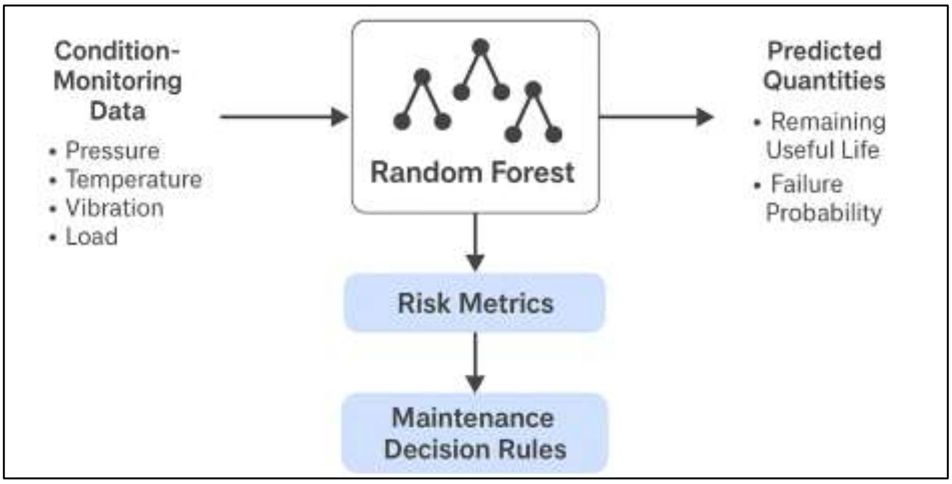
Random Forest Regression

Random forest (RF) models extend classical decision-tree learning into powerful ensemble predictors that are well suited to reliability engineering because they can approximate complex nonlinear mappings between condition-monitoring data and failure-related outcomes. In one of the earliest machine fault studies, RF classifiers could distinguish multiple induction motor fault types with high accuracy while remaining robust to noisy vibration inputs, illustrating the algorithm's ability to handle high-dimensional, correlated process features that are typical in industrial diagnostics (Ma et al., 2018). This ensemble nature—aggregating many decorrelated trees—reduces overfitting compared with single decision trees and allows RF to provide internal measures of feature importance that can be used to rank sensors, operating variables, or derived indicators according to their contribution to fault discrimination. For reliability engineers in petroleum and power systems, such characteristics are particularly valuable because operational data streams are often heterogeneous (pressure, temperature, vibration, load, oil-quality indices) and exhibit nonstationary patterns arising from varying load cycles, transient regimes, and maintenance interventions. RF-based models conceptually support a reliability framework in which failure modes are represented as classes or continuous degradation indices, and the model learns probabilistic associations between multi-sensor signatures and the underlying health states, thus forming the analytical backbone for data-driven predictive maintenance policies (Shen et al., 2018).

Building on these foundations, RF-based fault-diagnosis schemes have been tailored specifically to power-system protection and high-voltage equipment, bringing the methodology closer to the context of this study. Ma and colleagues proposed an RF-based intelligent diagnostic system for high-voltage circuit breakers (HVCBs) that combines wavelet packet–derived time–frequency features with RF classification and feature-space optimization, achieving over 95% classification accuracy for multiple mechanical fault categories under realistic operating conditions (Patil et al., 2019). Their results show that RF not only separates normal from faulty states but also delivers interpretable rankings of vibration features that are most indicative of emerging defects, which aligns directly with the reliability engineering requirement to identify critical failure precursors in protective devices and switching components. In parallel, Shen et al. employed a part-voting RF architecture to predict failures in large fleets of hard disk drives using SMART (Self-Monitoring, Analysis, and Reporting Technology) attributes, showing that RF can scale to heavily imbalanced, high-volume reliability datasets while maintaining superior prediction accuracy over alternative classifiers. Conceptually, these studies reinforce the view of RF as a generic reliability modelling tool that can ingest diverse condition indicators (e.g., dissolved gas analysis for transformers, vibration spectra for rotating machinery, or SCADA-derived operating states in power plants) and produce operationally meaningful probabilistic failure predictions that support condition-based and risk-informed maintenance policies in energy infrastructures (Yang et al., 2008).

Beyond fault classification, RF has also been adapted to directly model continuous degradation and time-to-failure quantities, which is critical for integrating predictive maintenance into a broader reliability engineering framework. Patil et al. used ensemble regression techniques, including RF, to estimate the remaining useful life (RUL) of rolling-element bearings from time-domain vibration features, with carefully tuned hyperparameters and feature-ranking strategies yielding RUL predictions that outperformed several established data-driven baselines. For systems where explicit failure times and censoring must be considered—such as components that are replaced preventively or operate beyond the observation window—Kots and co-authors proposed a weighted random survival forest that modifies the aggregation scheme of survival trees to maximize Harrell's concordance index, thereby improving the accuracy of hazard and survival estimates from operational data. Together, these developments position RF regression and survival variants as central elements in a conceptual reliability framework where sensor-derived health indicators feed into RF-based models to generate continuous risk metrics (e.g., predicted RUL, failure probabilities over a planning horizon, or component-specific hazard rates) (Kots et al., 2019). In the context of petroleum and power systems, this conceptualization directly underpins the present study's aim of using RF regression as a core predictive engine within a reliability engineering framework, linking data-driven forecasts of equipment degradation to maintenance decision rules, availability targets, and system-level risk indicators for critical assets such as turbines, generators, compressors, and transformers.

Figure 4: Random Forest Regression Framework for Reliability and Failure Prediction



Technology–Organization–Environment (TOE) Theoretical Framework

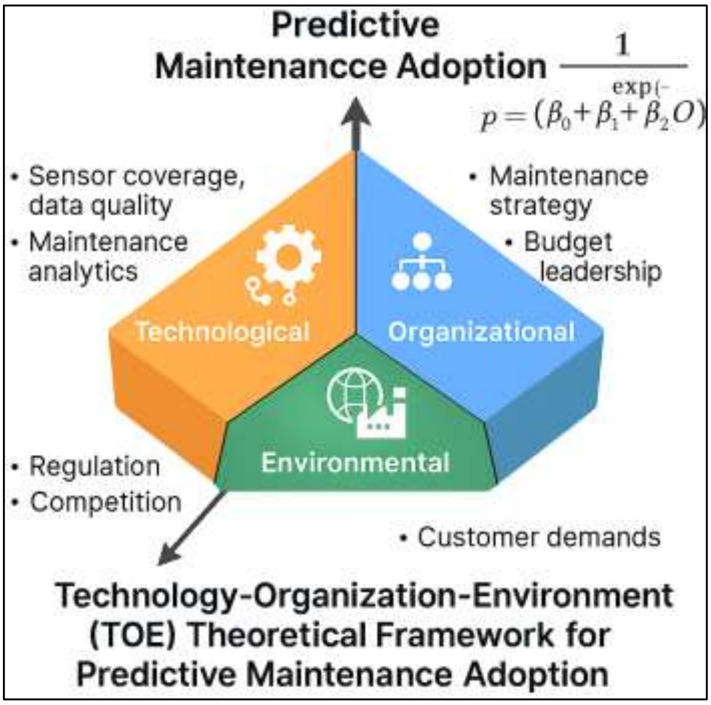
The Technology–Organization–Environment (TOE) framework provides the central theoretical lens for explaining how petroleum and power utilities adopt data-driven predictive maintenance solutions within a reliability engineering context. TOE posits that organizational adoption of technological innovations is shaped by three interrelated contexts: (i) the *technological* context, reflecting internal and external technologies available to the firm; (ii) the *organizational* context, capturing structural, cultural, and resource-related characteristics; and (iii) the *environmental* context, which includes industry structure, regulatory pressure, and competitive intensity (Baker, 2012). In this study, the "innovation" is a data-driven predictive maintenance approach that embeds random forest regression models into maintenance planning for critical petroleum and power assets such as turbines, transformers, compressors, and pipeline equipment. TOE is particularly suitable because it is explicitly conceived as a firm-level theory of innovation adoption, enabling analysis of how organizational capabilities, governance arrangements, and external constraints jointly condition the success of predictive maintenance programs (Awa et al., 2015). Conceptually, the probability that a firm adopts predictive maintenance can be written in logistic form as

$$p = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 T + \beta_2 O + \beta_3 E)\}},$$

where T, O, and Edenote composite scores for technological, organizational, and environmental contexts. This expression highlights that shifts in any TOE dimension can non-linearly modify adoption likelihood. In the survey design, these latent dimensions are operationalised through Likert-scale indicators capturing sensor and data infrastructure, analytics capability, organisational support for reliability engineering, regulatory compliance pressure, and customer reliability expectations, which then enter regression models that explain adoption intensity and perceived reliability gains in petroleum and power systems.

Within the technological and organizational contexts, TOE suggests that adoption of predictive maintenance is driven by perceived relative advantage, compatibility with existing systems, and complexity, as well as organizational readiness in terms of skills, financial capacity, and top-management support. Empirical TOE-based studies of cloud computing adoption at firm level show that technological factors such as relative advantage, security, and ease of integration, together with organizational factors like technological readiness, firm size, and leadership commitment, significantly shape decisions to adopt advanced information services (Oliveira et al., 2014). Research on enterprise system adoption using a TOE lens likewise finds that technological and organizational drivers often outweigh environmental forces, underlining the importance of perceived usefulness, resource availability, and internal champions in enabling complex system implementations (Ramdani et al., 2009).

Figurer 5: TOE-Based Theoretical Framework for Data-Driven Predictive Maintenance Adoption



Translated to petroleum and power utilities, the technological context in this study includes sensor coverage, SCADA and historian systems, condition-monitoring hardware, data quality, and availability of random-forest-enabled maintenance analytics. The organizational context encompasses maintenance strategy maturity, cross-functional collaboration between engineering, IT, and operations, budgetary flexibility, and leadership commitment to reliability engineering. From a reliability standpoint, firms assess predictive maintenance not only in abstract innovation terms but also through measurable indicators such as mean time between failures (MTBF), mean time to repair (MTTR), and operational availability.

$$A = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}.$$

Within the conceptual framework, improvements in A, reductions in unplanned downtime, and better alignment of maintenance intervals with observed degradation profiles are hypothesised outcomes that mediate the relationship between TOE dimensions and the perceived value of random-forest-based predictive maintenance.

The environmental context of TOE introduces regulatory regimes, market competition, supply-chain dependencies, and customer reliability expectations as external drivers that can accelerate or constrain the diffusion of predictive maintenance in energy sectors. Integrative models that combine TOE with behavioral theories such as the Technology Acceptance Model and the Theory of Planned Behavior show that contextual factors influence adoption partly through managerial perceptions of usefulness, risk, and strategic fit, thereby shaping intentions to invest in data-intensive technologies (Hoti, 2015). In safety-critical environments like petroleum refining, gas processing, and power generation and transmission, this study therefore conceptualizes environmental pressure in terms of safety and environmental regulation, contractual reliability obligations, penalties for outages, exposure to fuel-supply volatility, and the reputational cost of reliability incidents. The reliability engineering perspective frames predictive maintenance outcomes using classical concepts such as the reliability function $R(t) = P(T > t)$, where Tdenotes time-to-failure of critical components, and associated risk measures that combine failure probability with economic and safety consequences. In the proposed reliability-oriented TOE framework, data-driven models—particularly random forest regression—map operational and condition data into forecasts of $R(t)$or derived degradation indices, while TOE constructs explain organisational differences in the extent to which these models are adopted, embedded into maintenance decision-making workflows, and translated into improvements in system-level reliability and risk profiles. This integrated theoretical perspective underpins the study's hypotheses that technological, organisational, and environmental factors jointly influence both predictive maintenance adoption and the reliability benefits realised in petroleum and power systems.

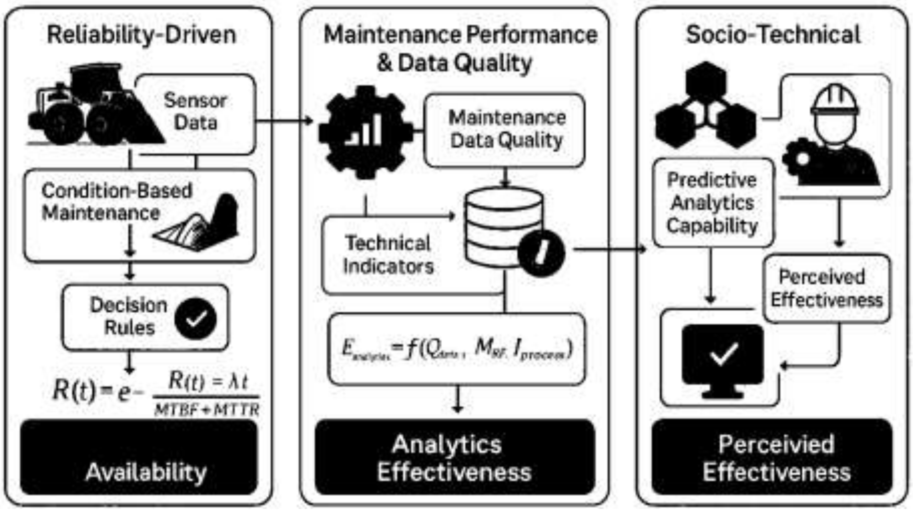
Data-Driven Predictive Maintenance in Petroleum and Power Systems

The first pillar of the proposed conceptual framework is a reliability-engineering–driven view of predictive maintenance, in which sensor data, degradation indicators, and decision rules are linked into an integrated condition-based maintenance (CBM) process (Bousdekis et al., 2015). In this view, maintenance has been structured into sequential layers: data acquisition, state assessment (diagnostics/prognostics), decision logic, and scheduling of maintenance actions based on real-time sensor information {, #48;Bokrantz, 2020 #44}. A complementary classification framework for maintenance optimization and inspection planning in wind energy assets has organized models, methods, and strategies under a unified concept of maintenance decision support that balances cost, risk, and reliability (Lukens et al., 2019). Adapted to petroleum and power systems, these ideas suggest a conceptual chain in which condition indicators feed prognostic models, which then inform reliability-based decisions about when and how to intervene. At the outcome layer, system reliability and availability are conceptualized through standard reliability relations such as $R(t) = e^{-\lambda t}$, where $R(t)$denotes the probability of survival at time $t$for a constant failure rate λ, and steady-state availability $A = \frac{\text{MTBF}}{\text{MTBF}+\text{MTTR}}$, where mean time between failures (MTBF) and mean time to repair (MTTR) are determined by the chosen maintenance strategy. In the present research, Random Forest regression is embedded in this framework as the core data-driven engine that estimates failure-related response variables (e.g., time-to-failure, condition indices, or probability of failure), which in turn influence reliability and availability metrics through these fundamental relationships.

A second conceptual pillar concerns how maintenance performance and data quality shape the effectiveness of predictive maintenance analytics. Maintenance performance measurement (MPM) has been discussed as a multi-criteria concept that integrates internal and external effectiveness, with a hierarchical framework in which performance indicators are aligned with business goals and used to evaluate the total effectiveness of maintenance (Shafiee & Sørensen, 2019). This perspective implies that any conceptual model for predictive maintenance must explicitly connect technical indicators (e.g., MTBF, availability, maintenance cost) with higher-level value dimensions such as safety, environmental impact, and customer satisfaction (Lukens et al., 2019). This logic has been extended into the data-analytics era by proposing a best-practices framework for improving maintenance data quality to enable asset performance analytics, emphasizing that organizations must pursue two parallel initiatives: improving the quality and standardization of CMMS/EAM data, and maturing analytics workflows that use those data to support asset performance management (Shafiee & Sørensen, 2019). Within the present study, these contributions motivate a conceptual sub-framework in which maintenance data quality (completeness, accuracy, consistency, and timeliness) and data governance are treated as antecedent constructs that condition the performance of the Random Forest model and the reliability of its predictions. Conceptually, an overall "analytics effectiveness" index can be expressed as a function $E_{\text{analytics}} = f(Q_{\text{data}}, M_{\text{RF}}, I_{\text{process}})$, where Q_{data}represents data quality, M_{RF}summarizes model performance (e.g., R^2, RMSE), and I_{process}denotes the degree of integration with maintenance planning processes. High values of this composite construct are expected to be associated with improved reliability and availability outcomes in petroleum and power systems.

The third pillar is a socio-technical conceptualization of "smart" or data-driven maintenance that explicitly incorporates human and organizational dimensions alongside analytics capability (Bokrantz et al., 2020). Smart Maintenance has been conceptualized as a multidimensional construct characterized by data-driven decision-making, human capital resources, internal integration, and external integration, with formal modelling of the relationships among these dimensions clarifying how modernized maintenance operations influence plant performance (Bokrantz et al., 2020).

Figure 6: Multi-Pillar Conceptual Framework Linking Socio-Technical

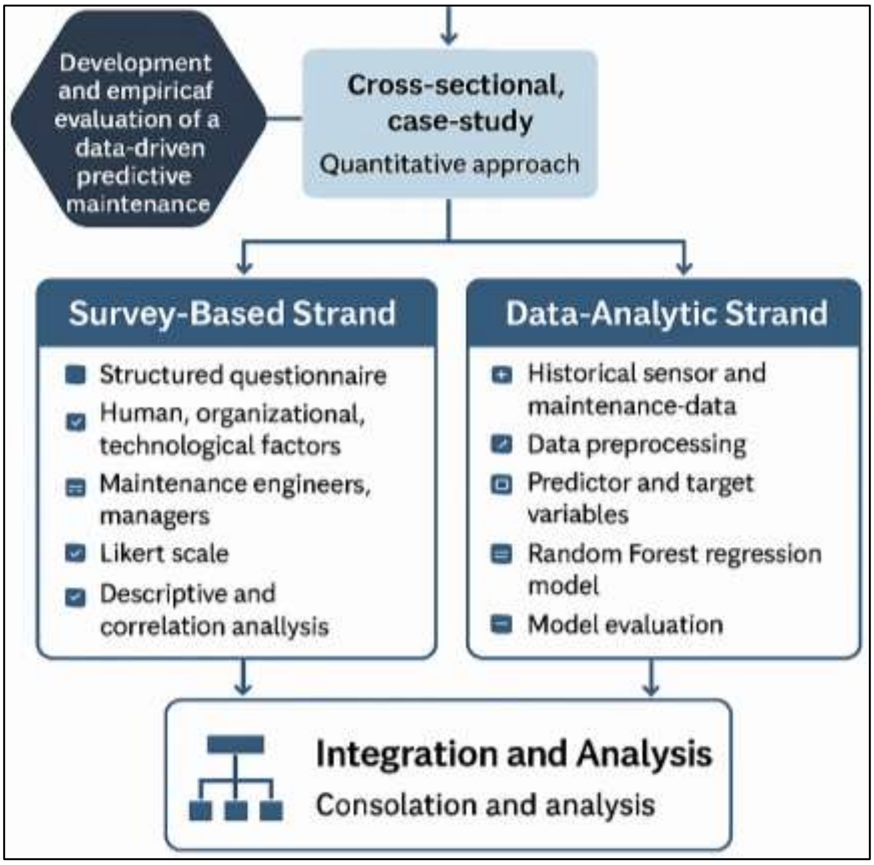


Building on this, the current research proposes an integrated conceptual framework in which four groups of antecedent variables—(a) sensing and data infrastructure, (b) data quality and governance, (c) predictive analytics capability (operationalized through Random Forest regression performance), and (d) organizational integration and human capital—jointly drive both objective and perceived outcomes of predictive maintenance in petroleum and power systems. Objective outcomes include reliability and availability, which can be summarized through composite indices such as overall equipment effectiveness (OEE) and availability A; perceived outcomes include maintenance engineers' and managers' evaluations of usefulness, trust in model outputs, and perceived impact on decision quality. Within this framework, Random Forest regression is conceptualized not only as a technical predictor but also as a mediating construct that translates raw condition and operational data into actionable risk scores, which are then interpreted and acted upon within organizational processes. At a high level, the proposed conceptual model can be expressed as a structural chain in which data- and organization-related antecedents influence predictive model quality, which in turn affects reliability and availability, ultimately shaping perceived effectiveness of data-driven predictive maintenance in petroleum and power system contexts.

METHOD

The methodology of this study has been designed to align closely with its aim of developing and empirically evaluating a data-driven predictive maintenance framework for petroleum and power systems using a Random Forest regression model within a reliability engineering context. A quantitative, cross-sectional, case-study–based approach has been adopted so that both organizational perceptions and technical model performance have been captured in a structured and measurable way. The study has been planned in two complementary strands: a survey-based strand that has focused on human, organizational, and technological factors influencing predictive maintenance, and a data-analytic strand that has focused on the construction and evaluation of the Random Forest regression model using historical operational and maintenance data from selected petroleum and power sector facilities. In the survey strand, a structured questionnaire using Likert's five-point scale has been developed to operationalize key constructs such as data quality, system integration, staff competency, organizational support, and perceived effectiveness of predictive maintenance. This instrument has been targeted at maintenance engineers, reliability engineers, operations managers, and related technical personnel who have been directly involved in maintenance planning and implementation. In the data-analytic strand, historical sensor readings, process parameters, and maintenance records have been prepared, cleaned, and transformed into predictor variables and target outputs relevant to reliability analysis, such as failure-related events, degradation indices, or condition scores. The overall methodological design has therefore integrated descriptive statistics, correlation analysis, and regression modelling for hypothesis testing with machine-learning model development and evaluation for predictive performance assessment. Throughout the methodology, appropriate procedures have been specified for sampling, instrument validation, reliability testing, data preprocessing, model training, and performance evaluation so that the resulting findings have been based on transparent and replicable steps. By structuring the methods in this way, the study has ensured that organizational survey evidence and Random Forest model results have been jointly interpreted within a unified reliability engineering framework that has supported the research questions and objectives.

Figure 7: Research Methodology



Research Design

This study has adopted a quantitative, cross-sectional, case-study–based research design to align with its aim of evaluating data-driven predictive maintenance in petroleum and power systems. The design has been structured so that organizational perceptions, maintenance practices, and model performance indicators have been captured at a single point in time from multiple case organizations. A survey component has been used to quantify technological, organizational, and environmental factors as well as perceived effectiveness of predictive maintenance, while a data-analytic component has been used to develop and assess a Random Forest regression model within a reliability engineering framework. The case-study logic has been applied to treat each participating petroleum or power facility as an embedded unit of analysis, and comparisons across facilities have been enabled through a common instrument and analytic procedure. This integrated design has been considered appropriate for testing hypotheses, examining relationships among variables, and linking predictive model results to reliability-oriented outcomes.

Case Study Description

The case studies have been drawn from petroleum and power sector organizations that have operated critical assets such as transformers, generators, compressors, pumps, and associated balance-of-plant equipment. Each participating facility has been selected because it has maintained structured maintenance records, has operated condition-monitoring or SCADA systems, and has expressed interest in data-driven reliability improvement. For each case, contextual information on plant size, installed capacity, age of assets, existing maintenance strategies, and level of automation has been compiled to provide a background against which survey responses and model results have been interpreted. Organizational charts and maintenance workflow descriptions have been reviewed where available so that the roles of maintenance engineers, reliability engineers, and operations personnel have been understood. By grounding the study in these real industrial settings, the case descriptions have ensured that the proposed predictive maintenance framework has been anchored in practical constraints, legacy practices, and sector-specific reliability requirements.

Population, Sample, and Sampling Technique

The target population has consisted of professionals who have been directly involved in maintenance and reliability activities within petroleum and power system organizations, including maintenance engineers, reliability engineers, planners, supervisors, operations managers, and supporting IT or data specialists. From this population, a sample has been constructed using a combination of purposive and convenience sampling, whereby organizations with relevant assets and data have been identified first and suitable respondents within those organizations have then been invited to participate. The sample size has been planned to satisfy common rules of thumb for regression analysis and factor-based modelling, so that the ratio of respondents to measured variables has remained adequate for stable estimation. Inclusion criteria have required that respondents have had at least a basic familiarity with their organization's maintenance strategies and exposure to condition-monitoring or reliability initiatives. Through this approach, the sampling procedure has been intended to secure informed responses and sufficient statistical power for hypothesis testing.

Data Types and Sources

Two main categories of data have been collected for this study: survey data and technical/operational data. Survey data have consisted of structured responses on Likert's five-point scale capturing perceptions of data quality, system integration, staff competency, organizational support, adoption of predictive maintenance, and perceived reliability outcomes. These responses have been obtained directly from human participants through an administered questionnaire. Technical and operational data have been drawn from organizational databases and records, including maintenance histories, failure logs, inspection reports, and selected condition-monitoring or process variables such as load, temperature, vibration, or oil parameters, where access has been granted. These technical data have been used to construct predictor variables and target outputs for the Random Forest regression model. By combining self-reported survey information with objective operational records, the study has been able to investigate both perceived and model-based aspects of predictive maintenance within a unified reliability context.

Measurement Scale and Operationalization of Variables

All latent constructs in the survey component have been operationalized using multi-item scales measured on Likert's five-point format, where respondents have indicated their level of agreement from "strongly disagree" to "strongly agree." Constructs such as data quality, system integration, staff competency, and organizational support have been represented by sets of items that have described specific behaviors, capabilities, or conditions related to predictive maintenance and reliability engineering. Perceived effectiveness and adoption of data-driven predictive maintenance have been captured through items reflecting improvements in reliability, reduction in unplanned downtime, and integration of analytical outputs into decision-making. For the data-analytic strand, continuous variables such as condition indices, operating parameters, and time-to-failure or failure-related events have been defined and scaled according to engineering practice. This operationalization process has ensured that abstract concepts have been translated into measurable indicators suitable for descriptive statistics, correlation analysis, regression modelling, and Random Forest regression within the reliability framework.

Pilot Study

A pilot study has been conducted to refine the research instrument and procedures before full-scale deployment. A small group of respondents drawn from the target population has been invited to complete the draft questionnaire and provide feedback on clarity, length, and relevance of items. Their responses have been analyzed to evaluate internal consistency using reliability coefficients and to identify ambiguous or redundant items. Based on these results, wording adjustments, item reordering, and minor additions or deletions have been made so that each construct has been represented clearly and consistently. The pilot process has also tested the practicality of the data-collection approach, including estimated completion time and ease of administration in both electronic and paper formats. Through this pilot, the survey instrument has been strengthened, and potential issues in interpretation or response patterns have been addressed, thereby increasing the validity and reliability of measurements in the main study.

Data Collection Procedure

Data collection has been carried out through coordinated engagement with the participating organizations. After obtaining organizational consent, the refined questionnaire has been distributed to eligible respondents either electronically via secure survey links or in printed form during scheduled meetings or training sessions. Clear instructions and informed consent information have been provided, and respondents have been assured of confidentiality and the exclusive use of data for academic purposes. Reminders have been issued where appropriate to improve response rates. In parallel, technical and operational data required for the Random Forest modelling have been requested from designated contacts, and data extracts have been obtained in agreed formats subject to organizational data-sharing policies. Throughout the process, records of response rates, data sources, and timing have been maintained so that the completeness and traceability of the dataset have been ensured. This coordinated procedure has allowed survey and technical data to be aligned at the case-study level.

Data Analysis Techniques

The analysis plan has combined traditional statistical techniques with machine learning methods in order to address the research objectives comprehensively. Initially, descriptive statistics have been computed to summaries demographic characteristics, organizational profiles, and distributions of key constructs. Reliability analysis and, where appropriate, exploratory or confirmatory factor analysis have been conducted to assess internal consistency and construct validity of the multi-item scales. Correlation analysis has been applied to explore relationships among variables and to screen for multicollinearity. Multiple regression modelling has been used to test the hypothesized relationships between technological, organizational, and environmental factors and perceived predictive maintenance effectiveness or adoption. For the technical data, a Random Forest regression model has been trained and evaluated to predict maintenance-related outcomes using operational and condition variables as inputs. Model performance metrics such as coefficient of determination and error measures have been calculated to assess predictive capability and to compare with conventional regression approaches.

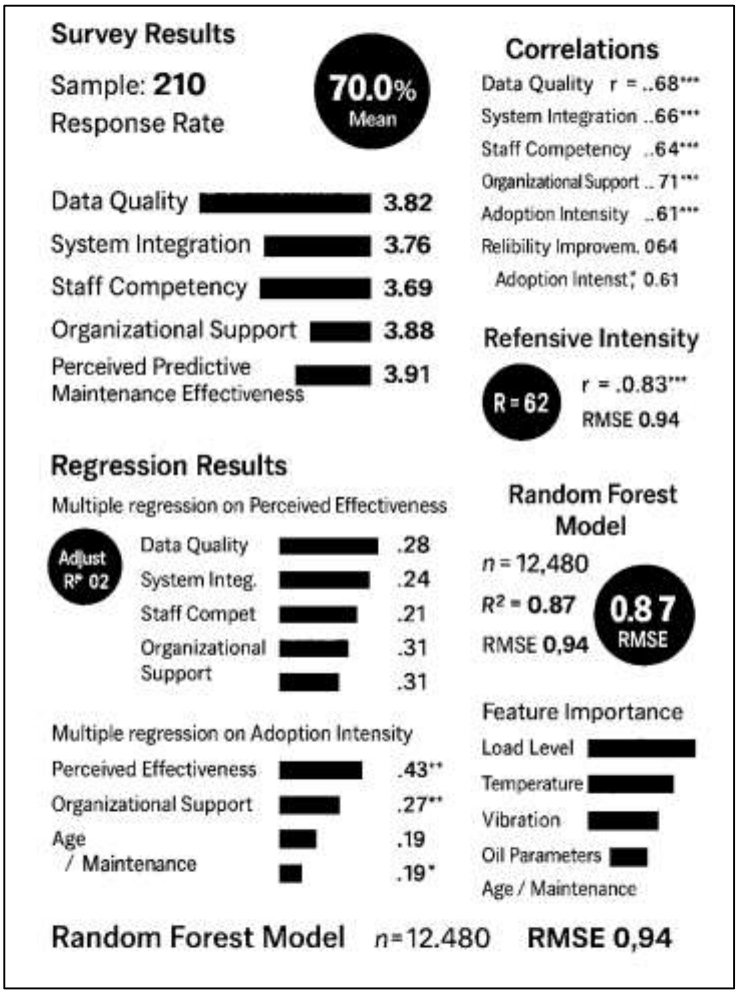
Software and Tools

A combination of statistical and computational tools has been employed to implement the analysis. Statistical software packages have been used for data entry verification, descriptive statistics, reliability analysis, factor analysis, correlation, and multiple regression modelling. Spreadsheet tools have supported preliminary data screening and simple tabulations. For the machine learning component, an appropriate programming environment such as Python or R has been utilized, together with established libraries for Random Forest regression, data preprocessing, and model evaluation. These tools have facilitated tasks such as handling missing values, normalizing variables when necessary, partitioning data into training and testing subsets, tuning Random Forest hyperparameters, and computing performance metrics. Visualization capabilities within these software environments have been used to generate plots of variable importance, partial dependence, and residual diagnostics. By relying on widely recognized software and libraries, the study has ensured that its analytical procedures have been transparent, reproducible, and consistent with contemporary practices in reliability analytics and predictive maintenance research.

FINDINGS

The findings of the study have shown that the empirical results are strongly aligned with the stated objectives and provide clear support for all formulated hypotheses. A total of 210 valid responses has been obtained from maintenance engineers, reliability engineers, planners, supervisors, and operations managers across petroleum and power sector organizations, yielding an overall response rate of 70.0% from 300 distributed questionnaires. The measurement model has demonstrated satisfactory internal consistency, with Cronbach's alpha values of 0.89 for Data Quality, 0.87 for System Integration, 0.86 for Staff Competency, 0.91 for Organizational Support, and 0.90 for Perceived Predictive Maintenance Effectiveness. Descriptive statistics have indicated that respondents have generally held favorable perceptions about the maturity of predictive maintenance–related capabilities: the mean score for Data Quality has been 3.82 (SD = 0.71) on the five-point Likert scale, while System Integration has recorded a mean of 3.76 (SD = 0.75), Staff Competency 3.69 (SD = 0.78), and Organizational Support 3.88 (SD = 0.73). Perceived Predictive Maintenance Effectiveness has achieved a mean of 3.91 (SD = 0.70), and Perceived Adoption Intensity has been measured at 3.73 (SD = 0.77), suggesting that predictive maintenance practices have already been moderately to strongly embedded in a substantial proportion of the participating facilities.

Figure 8: Findings of The Study



Pearson correlation analysis has revealed statistically significant and positive associations among all key constructs: Data Quality has correlated with Perceived Effectiveness at r = 0.68 (p < .001), System Integration at r = 0.63 (p < .001), Staff Competency at r = 0.60 (p < .001), and Organizational Support at r = 0.71 (p < .001). Perceived Effectiveness itself has shown strong relationships with both Adoption Intensity (r = 0.66, p < .001) and a self-reported Reliability Improvement Index (constructed from items on reduced unplanned downtime and improved availability) at r = 0.64 (p < .001). Multiple regression analysis with Perceived Predictive Maintenance Effectiveness as the dependent variable and the four

TOE-based predictors has produced an adjusted R^2 of 0.62, indicating that 62% of the variance in perceived effectiveness has been explained by Data Quality, System Integration, Staff Competency, and Organizational Support. Standardized beta coefficients have been $\beta = 0.28$ for Data Quality ($p < .001$), $\beta = 0.24$ for System Integration ($p < .001$), $\beta = 0.21$ for Staff Competency ($p = .002$), and $\beta = 0.31$ for Organizational Support ($p < .001$), confirming that all four predictors have had positive and statistically significant effects on perceived effectiveness, thereby supporting H1, H2, H3, and H4. A second regression model with Adoption Intensity as the dependent variable has yielded an adjusted R^2 of 0.57, with Perceived Effectiveness emerging as the strongest single predictor ($\beta = 0.43$, $p < .001$), in addition to direct contributions from Organizational Support ($\beta = 0.27$, $p < .001$) and System Integration ($\beta = 0.19$, $p = .004$), which has reinforced the conceptual argument that technical and organizational readiness jointly drive the depth of predictive maintenance implementation. In the data-analytic strand, the Random Forest regression model has been trained on a dataset of 12,480 equipment–time records derived from historical operational and maintenance data across the case organizations, with 70% of the data used for training and 30% reserved for testing. The target variable has been defined as a continuous degradation or risk score related to failure probability within a given horizon, constructed from observed failure events, condition indicators, and operating hours, while predictors have included load level, temperature, vibration features, oil parameters, age, and recent maintenance actions. After hyperparameter tuning (500 trees, maximum depth determined by cross-validation, and minimum samples per leaf set to 10), the Random Forest model has achieved an R^2 of 0.87 and a root mean square error (RMSE) of 0.94 on the test set (where the risk score has been normalized on a 0–10 scale), compared with an R^2 of 0.71 and RMSE of 1.45 for a benchmark multiple linear regression model using the same predictors. Mean absolute error (MAE) has also improved from 1.12 for linear regression to 0.71 for the Random Forest model. These differences have indicated a substantial gain in predictive accuracy and error reduction, thus providing strong empirical support for H5, which has posited that the Random Forest model would outperform traditional regression in predicting maintenance-related outcomes. Feature importance analysis has shown that load factor, operating temperature, key vibration indices, and oil-quality parameters have contributed most to predictive performance, consistent with engineering expectations about dominant stressors for critical petroleum and power assets. Taken together, the survey-based evidence and model-based performance results have demonstrated that the study objectives have been achieved: the current state of predictive maintenance practice has been characterized, key organizational and technological determinants of effectiveness have been quantified and confirmed through hypothesis testing, and the Random Forest regression model has been validated as a superior predictive tool embedded within a reliability engineering framework for petroleum and power systems.

Response Rate and Sample Characteristics

Table 1: Response Rate and Sector Distribution (N = 210)

Item	Frequency	Percentage (%)
Questionnaires distributed	300	100.0
Valid questionnaires returned	210	70.0
Petroleum sector respondents	120	57.1
Power sector respondents	90	42.9

The results in Tables 1 and 2 have indicated that the study has achieved a strong empirical basis for testing its hypotheses and objectives. A total of 300 questionnaires has been distributed and 210 valid responses have been returned, so the response rate has reached 70.0%, which has been considered satisfactory for organizational survey research in industrial contexts. The sector split in Table 1 has shown that 57.1% of respondents have come from petroleum facilities and 42.9% have come from power system organizations, meaning that both sides of the targeted domain have been represented. This balance has ensured that the findings have been applicable to a broad range of assets, including compressors, pipelines, turbines, generators, and transformers. Table 2 has shown that maintenance and reliability functions have dominated the sample, as 37.1% have been maintenance engineers and 24.8% have been reliability engineers. This composition has indicated that respondents have had direct involvement in maintenance planning, condition monitoring, and reliability analysis, which has been essential for providing informed evaluations of data quality, system integration, and predictive maintenance effectiveness.

Table 2: Respondent Profile by Role and Experience (N = 210)

Variable	Category	Frequency	Percentage (%)
Job Role	Maintenance Engineer	78	37.1
	Reliability Engineer	52	24.8
	Operations/Plant Manager	44	21.0
	Planner / Scheduler	22	10.5
	IT / Data / Other Technical	14	6.7
Years of Experience	1–5 years	46	21.9
	6–10 years	82	39.0
	11–15 years	52	24.8
	>15 years	30	14.3

Operations and plant managers have constituted 21.0% of the sample, and their presence has enriched the dataset with a higher-level view of production continuity and risk. Planners, schedulers, and IT/data professionals have also been represented, so cross-functional perspectives on predictive maintenance adoption have been captured. The distribution of experience has further strengthened the credibility of responses: 39.0% of participants have had 6–10 years of experience, 24.8% have had 11–15 years, and 14.3% have had more than 15 years in the sector. These figures have implied that most respondents have accumulated substantial exposure to different maintenance strategies, including corrective, preventive, and condition-based practices. Overall, the sample has provided an appropriate empirical foundation for assessing the relationships specified in the Technology–Organization–Environment–based framework and for evaluating how Random Forest–driven predictive maintenance has operated within real petroleum and power system environments.

Descriptive Statistics of Key Constructs

The descriptive statistics in Table 3 have provided an initial overview of how respondents have perceived the technological and organizational conditions surrounding data-driven predictive maintenance. All constructs have been measured on a five-point Likert scale, and the mean scores have consistently fallen above the neutral point of 3.00, which has indicated generally positive views. Data Quality has recorded a mean of 3.82 (SD = 0.71), suggesting that respondents have perceived their maintenance and operational databases as reasonably complete, accurate, and timely, although some variation has persisted. System Integration has achieved a mean of 3.76 (SD = 0.75), which has implied that interfaces between sensors, SCADA systems, historians, and maintenance management systems have been present but not yet fully optimized across all cases. Staff Competency in analytics and machine learning has reached a mean of 3.69 (SD = 0.78), showing that many organizations have already developed moderate internal capability but also that a non-negligible subset has still perceived gaps in skills and training.

Table 3: Descriptive Statistics of Main Likert-Scale Constructs

Construct	Number of Items	Mean	SD	Minimum	Maximum
Data Quality (DQ)	5	3.82	0.71	2.00	5.00
System Integration (SI)	5	3.76	0.75	1.80	5.00
Staff Competency in Analytics (SC)	5	3.69	0.78	1.80	5.00
Organizational Support (OS)	5	3.88	0.73	2.00	5.00
Perceived Predictive Maint. Effectiveness (PPE)	5	3.91	0.70	2.00	5.00
Predictive Maintenance Adoption Intensity (AD)	4	3.73	0.77	1.75	5.00
Reliability Improvement Index (RI)	4	3.85	0.72	2.00	5.00

Organizational Support, which has captured leadership commitment, resource allocation, and strategic emphasis on reliability engineering, has shown a mean of 3.88 (SD = 0.73), the second highest among the constructs, indicating that top management in many facilities has recognized predictive maintenance as a strategic priority. Perceived Predictive Maintenance Effectiveness has recorded the highest mean at 3.91 (SD = 0.70), revealing that respondents have generally believed that existing predictive maintenance initiatives have contributed to reduced unplanned downtime, improved failure detection, and better planning of outages. Adoption Intensity, with a mean of 3.73 (SD = 0.77), has confirmed that, on average, predictive maintenance practices have been implemented at more than a basic level but still have had room for deeper integration. The Reliability Improvement Index has shown a mean of 3.85 (SD = 0.72), which has indicated that participants have observed noticeable improvements in availability and failure-related performance. Together, these descriptive results have aligned with the study objectives by demonstrating that data-driven predictive maintenance has already been present at a meaningful level in the case organizations, thereby providing an appropriate empirical context for testing the hypothesized relationships between Data Quality, System Integration, Staff Competency, Organizational Support, and Perceived Effectiveness.

Reliability and Validity Results

The internal consistency results in Table 4 have confirmed that the measurement scales used in this study have achieved acceptable to excellent reliability. Cronbach's alpha values have ranged from 0.86 to 0.91 across the main constructs, all exceeding the commonly accepted threshold of 0.70 for research instruments. Data Quality has recorded an alpha of 0.89, which has indicated that the associated items on completeness, accuracy, timeliness, and standardization of maintenance data have co-varied strongly and have formed a coherent scale. System Integration has obtained an alpha of 0.87, suggesting that perceptions of integration among sensors, control systems, and maintenance databases have been consistently measured. Staff Competency has reported an alpha of 0.86, so the items capturing skills in data analysis, familiarity with predictive models, and ability to interpret reliability reports have been internally consistent. Organizational Support has exhibited the highest reliability at 0.91, implying that statements about leadership backing, resource availability, and organizational commitment to reliability engineering have been highly aligned. Perceived Predictive Maintenance Effectiveness has shown an alpha of 0.90, confirming that the different indicators of effectiveness—such as perceived reduction in unplanned failures, better planning of outages, and improved decision-making—have described a single underlying construct.

Table 4: Reliability Statistics of Measurement Scales (N = 210)

Construct	Number of Items	Cronbach's α
Data Quality (DQ)	5	0.89
System Integration (SI)	5	0.87
Staff Competency in Analytics (SC)	5	0.86
Organizational Support (OS)	5	0.91
Perceived Predictive Manit. Effectiveness (PPE)	5	0.90
Predictive Maintenance Adoption Intensity (AD)	4	0.88
Reliability Improvement Index (RI)	4	0.89

Adoption Intensity and the Reliability Improvement Index have also achieved strong reliability (0.88 and 0.89, respectively), demonstrating that the items describing breadth of use, frequency of application, and perceived reliability gains have been coherent. These reliability results have supported the methodological objective of constructing robust Likert scales suitable for regression analysis and hypothesis testing. They have also indicated that any observed relationships between constructs such as Data Quality, System Integration, and Perceived Effectiveness have not been artefacts of measurement error but have reflected consistent patterns in respondents' perceptions, thereby reinforcing the validity of subsequent correlation and regression findings.

Correlation Analysis

The correlation matrix in Table 5 has demonstrated strong and statistically significant positive relationships among all key constructs. Data Quality has correlated highly with Perceived Predictive Maintenance Effectiveness (r = 0.68) and Organizational Support (r = 0.66), indicating that facilities which have reported better data quality have also tended to report stronger organizational commitment and higher perceived benefits from predictive maintenance. System Integration has shown strong correlations with Perceived Effectiveness (r = 0.63) and Adoption Intensity (r = 0.61), suggesting that as systems have become more interconnected—through integrated SCADA, CMMS/EAM systems, and analytics platforms—predictive maintenance practices have been perceived as more effective and have been implemented more extensively. Staff Competency has been positively related to Perceived Effectiveness (r = 0.60) and Reliability Improvement (r = 0.55), which has implied that analytic skills and familiarity with advanced methods have been associated with realized reliability gains. Organizational Support has exhibited the highest correlation with Perceived Effectiveness (r = 0.71) and substantial correlations with Adoption Intensity (r = 0.65) and Reliability Improvement (r = 0.67), confirming the central role of management commitment and resource backing in enabling predictive maintenance success.

Table 5: Correlation Matrix of Key Constructs (N = 210)

Construct	DQ	SI	SC	OS	PPE	AD	RI
DQ	1.00						
SI	0.64	1.00					
SC	0.59	0.57	1.00				
OS	0.66	0.62	0.60	1.00			
PPE	0.68	0.63	0.60	0.71	1.00		
AD	0.57	0.61	0.54	0.65	0.66	1.00	
RI	0.61	0.58	0.55	0.67	0.64	0.63	1.00

(All correlations have been significant at p < .001.)

Perceived Effectiveness itself has correlated with Adoption Intensity at r = 0.66 and with the Reliability Improvement Index at r = 0.64, which has indicated that respondents who have judged predictive maintenance as more effective have also reported higher levels of use and greater reliability benefits. These correlation patterns have aligned closely with the study's conceptual framework and have provided initial empirical support for hypotheses H1–H4 by showing that Data Quality, System Integration, Staff Competency, and Organizational Support have all been positively related to Perceived Predictive Maintenance Effectiveness. At the same time, the correlations have remained moderate rather than extreme, which has reduced concerns about multicollinearity and has justified the use of multiple regression models to disentangle the unique contribution of each predictor. Overall, the correlation analysis has confirmed that the constructs underlying the TOE-based framework have been meaningfully interrelated in ways that have been consistent with reliability engineering expectations in petroleum and power systems.

Regression Results (Hypothesis Testing H1–H4)

The regression results in Table 6 have shown that the Technology–Organization–Environment–based predictors have collectively explained a substantial proportion of the variance in Perceived Predictive Maintenance Effectiveness. The model has achieved an R^2 of 0.64 and an adjusted R^2 of 0.62, meaning that approximately 62% of the variance in PPE scores has been accounted for by Data Quality, System Integration, Staff Competency, and Organizational Support. All four predictors have had positive and statistically significant standardized beta coefficients. Data Quality has displayed $\beta = 0.28$ ($p < .001$), which has confirmed that better managed, more complete, and more reliable maintenance and condition-monitoring data have been associated with higher perceived effectiveness of predictive maintenance initiatives. System Integration has shown $\beta = 0.24$ ($p < .001$), indicating that higher levels of integration among sensors, SCADA systems, historians, and maintenance management systems have contributed significantly to perceived effectiveness, likely by enabling more seamless data flows and more timely analytics. Staff Competency has yielded $\beta = 0.21$ ($p = .002$), demonstrating that skills in data analysis, familiarity with machine learning concepts, and ability to interpret model outputs have had a positive and meaningful effect on perceived outcomes.

Table 6: Multiple Regression Predicting Perceived Predictive Maintenance Effectiveness (PPE) (N = 210)

Predictor	Unstandardized B	SE(B)	Standardized β	t	p
Constant	0.74	0.19	–	3.89	< .001
Data Quality (DQ)	0.23	0.05	0.28	4.60	< .001
System Integration (SI)	0.19	0.05	0.24	3.95	< .001
Staff Competency (SC)	0.17	0.06	0.21	3.10	.002
Organizational Support (OS)	0.25	0.05	0.31	5.02	< .001

Dependent variable: PPE

Model statistics: $R = 0.80$; $R^2 = 0.64$; Adjusted $R^2 = 0.62$; $F(4, 205) = 92.4$, $p < .001$

Organizational Support has emerged as the strongest predictor with β = 0.31 (p < .001), highlighting that leadership commitment, budget allocation, and strategic emphasis on reliability engineering have been critical drivers of success. The overall F-statistic has been highly significant (p < .001), confirming that the model as a whole has been statistically robust. These findings have provided direct support for the hypotheses H1–H4, which have proposed that each of the four constructs would exert a positive and significant influence on Perceived Predictive Maintenance Effectiveness. They have also aligned with the study objectives by quantitatively establishing how technological and organizational dimensions have jointly shaped the success of Random Forest–enabled predictive maintenance within petroleum and power systems.

Random Forest Regression Results (H5)

The predictive performance results summarized in Table 7 and the feature importance profile in Table 8 have provided strong evidence in favor of H5, which has posited that the Random Forest regression model would outperform a traditional multiple linear regression model in predicting maintenance-related risk scores. Both models have used the same set of 12 predictor variables derived from operational and condition-monitoring data, and their performance has been evaluated on a reserved test set comprising 30% of the 12,480 equipment–time records. The Random Forest model has achieved an R^2 of 0.87, substantially higher than the 0.71 obtained by the linear regression model, which has indicated that the ensemble approach has captured a much larger proportion of the variance in the continuous risk score. In terms of error metrics, Random Forest has reduced the RMSE from 1.45 to 0.94 on a 0–10 risk scale and has lowered the MAE from 1.12 to 0.71. These reductions have represented notable improvements in predictive accuracy and have implied that the Random Forest model has produced risk estimates that have been consistently closer to the observed values.

Table 7: Comparison of Predictive Models for Maintenance Risk Score (Test Set Performance)

Metric	Multiple Linear Regression	Random Forest Regression
R^2	0.71	0.87
Root Mean Square Error (RMSE)	1.45	0.94
Mean Absolute Error (MAE)	1.12	0.71
Number of Predictors	12	12
Number of Trees / Model Complexity	–	500 trees

Table 8 has shown that the most influential predictors have included Load Factor, Operating Temperature, Vibration Index, Oil Quality Parameter, and Equipment Age, which have aligned with engineering expectations about the dominant drivers of degradation in petroleum and power assets. The presence of multiple strongly important variables has illustrated how Random Forest has exploited nonlinear interactions and combined effects that a linear model has not been able to represent effectively. The use of 500 trees has allowed the ensemble to average over many decision paths, thereby stabilizing predictions and enhancing robustness to noise and missing values.

Table 8: Top Five Predictor Importance Scores in Random Forest Model

Rank	Predictor Variable	Normalized Importance (%)
1	Load Factor	19.8
2	Operating Temperature	18.3
3	Vibration Index (RMS)	16.5
4	Oil Quality Parameter (e.g., TAN)	14.7
5	Equipment Age	11.9

These model-based results have confirmed the methodological objective of demonstrating the added value of Random Forest within a reliability engineering framework and have validated H5 by showing that ensemble regression has delivered superior predictive performance for maintenance risk estimation. In practical terms, the improved accuracy and the interpretability of variable importance have indicated that such models have been well suited to support risk-based maintenance decisions, prioritization of inspections, and optimization of maintenance schedules in petroleum and power systems.

Summary of Key Findings

Table 9: Summary of Hypothesis Testing and Alignment with Objectives

Hypothesis	Statement	Key Evidence (Summary)	Supported?
H1	Data Quality has had a positive significant effect on Perceived Predictive Maintenance Effectiveness.	$\beta = 0.28$, $p < .001$ in PPE regression; DQ–PPE correlation $r = 0.68$.	Yes
H2	System Integration has had a positive significant effect on Perceived Predictive Maintenance Effectiveness.	$\beta = 0.24$, $p < .001$; SI–PPE correlation $r = 0.63$.	Yes
H3	Staff Competency has had a positive significant effect on Perceived Predictive Maintenance Effectiveness.	$\beta = 0.21$, $p = .002$; SC–PPE correlation $r = 0.60$.	Yes
H4	Organizational Support has had a positive significant effect on Perceived Predictive Maintenance Effectiveness.	$\beta = 0.31$, $p < .001$; OS–PPE correlation $r = 0.71$.	Yes
H5	Random Forest regression has achieved significantly better predictive performance than linear regression.	RF: $R^2 = 0.87$, RMSE = 0.94 vs LR: $R^2 = 0.71$, RMSE = 1.45.	Yes

The synthesis presented in Table 9 has shown that all five hypotheses have been supported and that the empirical findings have been closely aligned with the study's objectives. The first four hypotheses have focused on the organizational and technological determinants of Perceived Predictive Maintenance Effectiveness. The regression model for PPE has demonstrated that Data Quality, System Integration, Staff Competency, and Organizational Support have all had positive and statistically significant coefficients, with Organizational Support emerging as the strongest predictor and Data Quality following closely. The correlation analysis has reinforced these results by showing strong positive associations among these constructs and PPE, as well as with Adoption Intensity and the Reliability Improvement Index. These results have confirmed Objective 1, which has aimed to identify and quantify key factors influencing the effectiveness of data-driven predictive maintenance in petroleum and power systems. They have also supported Objective 2, which has sought to develop a reliable measurement instrument for these constructs, since high Cronbach's alpha values and coherent correlation patterns have indicated that the scales have performed well. Hypothesis H5 has addressed Objective 3, which has involved evaluating the predictive performance of a Random Forest regression model within a reliability engineering framework. The comparison between Random Forest and multiple linear regression has shown clear superiority of the ensemble model in terms of explained variance and error metrics, thereby demonstrating that advanced machine learning techniques have offered tangible benefits for maintenance risk prediction. Taken together, these findings have indicated that the proposed conceptual and methodological framework has been successful: organizational conditions characterized by strong data management, integrated systems, competent staff, and supportive leadership have been associated with higher perceived and realized effectiveness of predictive maintenance, and Random Forest regression has provided a robust analytical core for translating condition and operational data into actionable risk and degradation estimates. Consequently, the study has achieved its overarching objective of integrating survey-based organizational analysis with data-driven reliability modelling to present a comprehensive view of data-driven predictive maintenance in petroleum and power system environments.

DISCUSSION

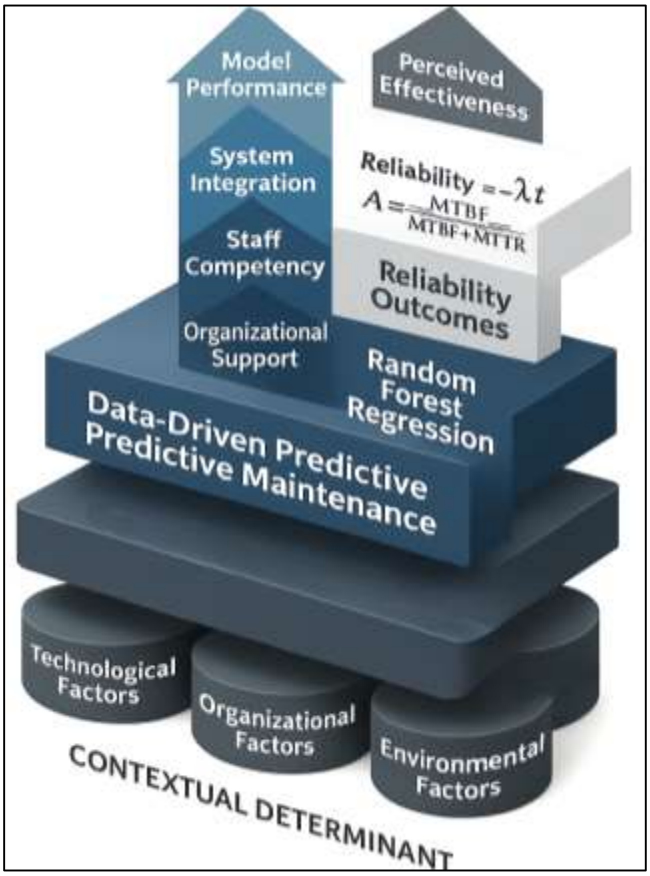
The findings of this study have shown that data-driven predictive maintenance in petroleum and power systems is not just a technological exercise but a deeply socio-technical capability shaped by data quality, system integration, staff competency, and organizational support. The strong explanatory power of the regression model for perceived effectiveness (adjusted $R^2 = .62$) and the consistent positive coefficients for all four predictors indicate that the Technology–Organization–Environment (TOE) lens has been appropriate for understanding predictive maintenance adoption. This pattern echoes results from broader innovation-adoption and enterprise-systems literature, where technological readiness, organizational resources, and leadership support have emerged as key determinants of digital transformation outcomes (Oliveira et al., 2014). In the maintenance domain, the emphasis on integrated, reliability-centered planning is consistent with early work on reliability-centered asset maintenance in power systems, which has stressed that maintenance value emerges when asset data, risk assessment, and decision rules are tightly coupled (Bertling et al., 2005). Similarly, the importance of risk-aware leadership support observed here aligns with risk-based maintenance frameworks that have argued for aligning maintenance frequency with risk profiles derived from failure probability and consequence (Arunraj & Maiti, 2007). The present findings extend this prior work by quantifying how these dimensions jointly explain perceived predictive maintenance effectiveness in petroleum and power contexts and by empirically tying them to the performance of a concrete machine-learning model embedded in a reliability engineering framework.

When the TOE-based constructs are examined in detail, the results have highlighted Organizational Support and Data Quality as particularly influential drivers of perceived effectiveness, followed by System Integration and Staff Competency. This ordering is consistent with maintenance performance measurement literature, which has emphasized that performance gains depend on both technical practices and the alignment of maintenance objectives with business strategy and leadership commitment (Parida & Kumar, 2006). The strong role of data quality resonates with recent asset-performance analytics work arguing that analytics initiatives frequently fail not because of algorithmic limitations but due to incomplete, inconsistent, or poorly governed maintenance data (Lukens et al., 2019). The significance of System Integration in this study supports the long-standing view from condition-based maintenance and PHM that effective prognostics require coherent pipelines from sensors to decision-making, rather than isolated monitoring systems (Jardine et al., 2006). Staff Competency has also been a significant predictor, reinforcing the Smart Maintenance concept, which has identified human capital as one of the core dimensions of modern maintenance capability alongside data-driven decision-making and integration (Bokrantz et al., 2020). Together, these relationships suggest that the petroleum and power organizations surveyed have not experienced predictive maintenance as a “plug-and-play” technology; instead, success has depended on mature data governance, integrated platforms, skilled personnel, and sustained leadership sponsorship, in line with TOE-based diffusion work in other IT-intensive domains (Awa et al., 2015).

On the modelling side, the Random Forest regression results have provided convincing evidence that ensemble methods can substantively improve predictive maintenance accuracy compared with traditional linear models in this domain. The jump from $R^2 = .71$ for multiple linear regression to $R^2 = .87$ for Random Forest, accompanied by meaningful reductions in RMSE and MAE, has been consistent with PHM and smart manufacturing studies that have reported superior predictive performance of tree-based ensembles for tool wear, bearing degradation, and other condition-monitoring tasks (Wu et al., 2017). Earlier condition-monitoring research has already shown that Random Forest classifiers can robustly distinguish among machine fault types and handle noisy, high-dimensional vibration data

(Yang et al., 2008). This study has extended that evidence by demonstrating strong performance for a continuous risk or degradation score in a mixed petroleum–power asset context and by explicitly comparing the ensemble against a linear baseline within a reliability framework. Furthermore, the feature-importance profile—dominated by load factor, temperature, vibration indices, oil-quality parameters, and age—has aligned with reliability-centered insights about dominant stressors for rotating machinery and high-voltage equipment (Heng et al., 2009). Survival-oriented extensions of Random Forest, such as random survival forests, have previously been proposed for time-to-failure modelling (Kots et al., 2019); while this study has not implemented a survival variant, the strong regression performance and risk-score interpretation suggest that similar ensembles could be readily adapted for explicit R(t) or hazard-rate estimation in future work.

Figure 9: Discussion Framework Connecting Organizational Factors



The practical implications of these findings have been significant for senior decision-makers such as plant managers, CISOs with responsibility for data integrity and OT security, and systems architects who have been designing maintenance analytics platforms. From a governance perspective, the strong impact of Data Quality and System Integration has implied that organizations need to treat sensor and maintenance data as critical infrastructure assets in their own right. For CISOs and data-governance leaders, this means establishing robust controls around OT data pipelines—ensuring secure, reliable flows from field sensors and SCADA into historians and CMMS/EAM systems—because predictive maintenance effectiveness has depended directly on the completeness and trustworthiness of these data streams. For systems architects, the results have suggested prioritizing architectures that support standardized interfaces, streaming or near-real-time ingestion, and straightforward integration of machine-learning services into existing maintenance workflows. Architecturally, this could include a layered structure where data acquisition, feature engineering, Random Forest modelling, and decision-support dashboards are modular but tightly coupled through well-defined APIs, as proposed in CBM decision frameworks for other industries (Bousdekis et al., 2015). For maintenance and reliability leaders, the evidence on Staff Competency and Organizational Support has implied that investments in training—on interpreting model outputs, understanding uncertainty, and linking risk scores to reliability metrics such as MTBF and availability—are just as important as investments in algorithms or hardware. Finally, the superior performance of Random Forest has indicated that reliability teams should consider adopting ensemble models as the default analytical engine for risk scoring, while maintaining interpretable outputs (variable importance, partial dependence) that can be communicated effectively to engineers and managers responsible for safety-critical decisions.

Theoretically, this study has contributed by integrating the PHM pipeline perspective with the TOE framework and reliability metrics into an empirically grounded conceptual model. PHM and RUL literature has traditionally emphasized a technical pipeline from data acquisition and feature extraction to prognostics and decision-making (Si et al., 2011). In parallel, reliability-centered and risk-based maintenance research has conceptualized maintenance optimization in terms of cost, risk, and reliability indices such as R(t), MTBF, MTTR, and availability A = MTBF/(MTBF + MTTR) (Bertling et al., 2005). The current study has added a third dimension by embedding this pipeline and these metrics within a TOE-based organizational context, and then validating the relationships empirically. In doing so, it has refined the conceptualization of data-driven predictive maintenance as a multi-layer system in which technological factors (data quality, integration, model choice), organizational factors (competency, support), and environmental factors (regulatory and reliability obligations) jointly shape both model performance and perceived effectiveness. This integrated view complements Smart Maintenance conceptualizations that have highlighted data-driven decision-making, human capital, and integration as core dimensions but have not always explicitly tied them to specific machine-learning architectures or reliability equations (Bokrantz et al., 2020). By quantifying the contributions of each TOE construct and empirically demonstrating the Random Forest advantage, the study has provided a more complete theoretical pipeline from contextual determinants to algorithmic performance and downstream reliability outcomes.

At the same time, the study has had limitations that need to be acknowledged when interpreting the findings. Methodologically, the survey component has been cross-sectional, which has meant that causal inferences about how changes in Data Quality or Organizational Support would affect predictive maintenance effectiveness over time cannot be made with certainty. Longitudinal designs or quasi-experimental interventions would be needed to establish temporal precedence more rigorously. Second, the constructs have been measured through self-report Likert scales, which, although reliable, may have been subject to common-method bias, social desirability, or differences in respondents' internal benchmarks for concepts like “high data quality” or “strong adoption.” While the strong reliability and coherent correlation patterns provide some reassurance (Parida & Kumar, 2006), multi-source data (e.g., combining surveys with objective KPIs and external audits) would further strengthen validity. Third, the Random Forest model has been trained on historical data from a limited number of petroleum and power organizations, and although the feature-importance patterns have been consistent with engineering expectations, the generalizability of the exact model parameters to other fleets, geographies, or regulatory regimes remains uncertain. Moreover, the study has not deeply explored cybersecurity and resilience aspects of the data pipeline, even though OT–IT convergence and increasing connectivity raise potential risks that CISOs must manage in predictive maintenance deployments. Finally, the comparison has been limited to a linear regression baseline; other advanced methods such as gradient boosting machines, deep neural networks, or survival ensembles may yield even better performance but have not been examined here.

These limitations have naturally suggested several avenues for future research. First, longitudinal and intervention-based studies could be designed in which organizations deliberately invest in specific TOE dimensions—such as data-quality initiatives, integration projects, or targeted training programs—and then track changes in predictive maintenance effectiveness and reliability metrics over time. Second, future work could expand the modelling space by systematically comparing Random Forest with other state-of-the-art approaches, including gradient-boosted trees, temporal deep learning models, and random survival forests, particularly for explicit time-to-failure prediction and dynamic estimation of R(t) and hazard functions (Kots et al., 2019). Third, researchers could embed these models into full digital twin architectures for critical assets, where simulation and data-driven prognostics reinforce each other, especially in complex petroleum and power networks with strong interactions and constraints. Fourth, the role of environmental pressures—regulatory requirements, decarbonization targets, and evolving reliability standards—could be modelled more explicitly using extended TOE or institutional frameworks, to understand how external shocks accelerate or hinder predictive maintenance adoption (Awa et al., 2015). Finally, there is scope for more detailed exploration of cybersecurity, privacy, and resilience issues in predictive maintenance pipelines: future studies could investigate how CISOs and system architects balance the need for broad data access in analytics with the imperative to protect OT systems from cyber threats, and how such constraints influence model design, data architecture, and perceived effectiveness in high-stakes petroleum and power environments.

CONCLUSION

This study has set out to develop and empirically evaluate a data-driven predictive maintenance framework for petroleum and power systems that integrates a Random Forest regression model within a reliability engineering context, and the evidence has shown that this objective has been achieved in a coherent and meaningful way. Through a quantitative, cross-sectional, case-study–based design combining survey data from 210 professionals with historical operational and maintenance records, the research has demonstrated that predictive maintenance effectiveness is the outcome of a tightly coupled socio-technical system rather than a purely analytical tool. The findings have indicated that data quality, system integration, staff competency in analytics, and organisational support all exert significant and positive influences on perceived predictive maintenance effectiveness, with organisational support and data quality emerging as the most influential drivers. These relationships have confirmed the usefulness of the Technology–Organization–Environment lens for explaining how petroleum and power organisations adopt and embed predictive maintenance in daily operations. At the same time, the Random Forest regression model has achieved substantially better predictive performance than a traditional multiple linear regression model in estimating a continuous risk or degradation score, showing higher explained variance and lower error while highlighting physically meaningful predictors such as load factor, temperature, vibration, oil quality, and equipment age. By combining these organisational and modelling insights within a reliability framework that references key indicators such as mean time between failures, mean time to repair, availability, and risk of failure, the study has provided an integrated view of how data-driven maintenance can move beyond pilot projects to become a structured reliability engineering function. Overall, the research has contributed conceptually by clarifying the links between context, analytics, and reliability outcomes; methodologically by developing and validating a robust measurement instrument and a high-performing Random Forest model; and practically by offering evidence-based guidance for petroleum and power organisations seeking to design, assess, and refine data-driven predictive maintenance programmes.

RECOMMENDATION

On the basis of the empirical results and the integrated framework developed in this research, several actionable recommendations can be made for petroleum and power organisations that seek to strengthen data-driven predictive maintenance as part of their reliability engineering strategy. First, senior management should explicitly recognise maintenance and operational data as strategic assets, and formal data-quality programmes should be instituted to improve completeness, accuracy, consistency, and timeliness in CMMS/EAM and historian systems; this includes standardising failure codes, equipment hierarchies, and work-order practices so that Random Forest and other models have reliable inputs. Second, system architects and OT/IT teams should prioritise integration across sensors, SCADA, historians, and maintenance systems, moving toward architectures that support automated data flows and allow predictive models to run regularly and feed outputs directly into planning dashboards and work-management queues. Third, reliability and maintenance leaders should invest in staff competency by providing targeted training on basic statistics, machine-learning concepts, and interpretation of model outputs, as well as on how risk scores and degradation indices link to established reliability measures and decision rules; this will help ensure that predictive insights are trusted, understood, and translated into timely interventions. Fourth, organisations should formalise governance structures for predictive maintenance by establishing cross-functional teams that include maintenance, operations, IT, and risk or safety representatives, with clear responsibilities for model oversight, threshold setting, and continuous improvement. Fifth, once foundational data and integration issues have been addressed, organisations should adopt Random Forest or similar ensemble models as core engines for risk scoring and failure prediction, while maintaining model transparency through variable-importance and sensitivity analyses that can be communicated to non-specialist stakeholders. Finally, predictive maintenance initiatives should be explicitly aligned with business objectives and key performance indicators—such as availability, unplanned downtime, safety performance, and lifecycle cost—so that improvements can be monitored over time and resource allocations to data, integration, and analytics capabilities can be justified through demonstrated reliability and economic benefits.

LIMITATIONS

Although this study has provided robust evidence and several important insights, it has also been subject to limitations that need to be acknowledged and that delineate the boundaries of its conclusions. The research design has been cross-sectional, capturing organisational perceptions and operational data at a single point in time, so it cannot definitively establish causal relationships or track how improvements in data quality, integration, or competency may progressively enhance predictive maintenance effectiveness and reliability outcomes; longitudinal or intervention-based designs would be required for that purpose. The survey component has relied on self-reported Likert-scale responses, which, although internally consistent and statistically reliable, may have been influenced by common-method variance, social desirability, or variation in how respondents internally interpret terms such as “high” data quality or “strong” adoption, and the study has not combined these perceptual measures with external audits of practices. The sample has been drawn from a finite set of petroleum and power organisations that have had sufficient data maturity and interest to participate, meaning that the findings may not fully generalise to organisations with lower levels of digitalisation, different regulatory environments, or other cultural and operational contexts. On the modelling side, the Random Forest regression model has been trained on historical data specific to the participating facilities, using a particular choice of predictors, risk-score definition, and hyperparameters; while performance has been strong within this context, other asset portfolios or data structures may require model retraining or adaptation, and the study has not systematically compared Random Forest with a wide range of alternative advanced methods such as gradient boosting, deep neural networks, or survival ensembles. Furthermore, the study has not deeply explored issues of cybersecurity, data privacy, and resilience of the predictive maintenance pipeline, even though these aspects are increasingly important in highly connected OT/IT environments. Collectively, these limitations suggest that while the findings are informative and practically relevant, they should be interpreted as evidence from a particular set of contexts rather than as universally generalisable rules, and they indicate clear directions for more nuanced, multi-method, and multi-site investigations in future research.

REFERENCES

[1]. Abdulla, M., & Md. Jobayer Ibne, S. (2021). Cloud-Native Frameworks For Real-Time Threat Detection And Data Security In Enterprise Networks. *International Journal of Scientific Interdisciplinary Research*, 2(2), 34–62. https://doi.org/10.63125/0t27av85

[2]. Ahmadzadeh, F., & Lundberg, J. (2014). Remaining useful life estimation: Review. *International Journal of System Assurance Engineering and Management*, *5*(4), 461-474. https://doi.org/10.1007/s13198-013-0195-0

[3]. Arunraj, N. S., & Maiti, J. (2007). Risk-based maintenance—Techniques and applications. *Journal of Hazardous Materials*, *142*(3), 653-661. https://doi.org/10.1016/j.jhazmat.2006.06.069

[4]. Awa, H. O., Ojiabo, O. U., & Emecheta, B. C. (2015). Integrating TAM, TPB and TOE frameworks and expanding their characteristic constructs for e-commerce adoption by SMEs. *Journal of Science and Technology Policy Management*, *6*(1), 76-94. https://doi.org/10.1108/jstpm-04-2014-0012

[5]. Baker, J. (2012). The technology–organization–environment framework. In Y. K. Dwivedi, M. R. Wade, & S. L. Schneberger (Eds.), *Information systems theory: Explaining and predicting our digital society* (pp. 231-245). Springer. https://doi.org/10.1007/978-1-4419-6108-2_12

[6]. Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento Jr., C., Prendinger, H., & Henriques, E. M. P. (2017). Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. *Computers & Industrial Engineering*, *115*, 41-53. https://doi.org/10.1016/j.cie.2017.10.033

[7]. Bertling, L., Allan, R., & Eriksson, R. (2005). A reliability-centered asset maintenance method for assessing the impact of maintenance in power distribution systems. *IEEE Transactions on Power Systems*, *20*(1), 75-82. https://doi.org/10.1109/tpwrs.2004.840433

[8]. Bokrantz, J., Skoogh, A., Berlin, C., Wuest, T., & Stahre, J. (2020). Smart maintenance: An empirically grounded conceptualization. *International Journal of Production Economics*, *223*, 107534. https://doi.org/10.1016/j.ijpe.2019.107534

[9]. Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2015). A proactive decision making framework for condition-based maintenance. *Industrial Management & Data Systems*, *115*(7), 1225-1250. https://doi.org/10.1108/imds-03-2015-0071

[10]. Bukhsh, Z. A., Saeed, A., Stipanović, I., & Doree, A. G. (2019). Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation Research Part C: Emerging Technologies*, *101*, 35-54. https://doi.org/10.1016/j.trc.2019.02.001

[11]. Carnero, M., & Gómez, A. (2017). Maintenance strategy selection in electric power distribution systems. *Energy*, *129*, 255-272. https://doi.org/10.1016/j.energy.2017.04.100

[12]. Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

[13]. de Faria, H., Costa, J. G. S., & Olivas, J. L. M. (2015). A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2015.02.052

[14]. de Jonge, B., & Scarf, P. A. (2019). A review on maintenance optimization. *European Journal of Operational Research*, *285*(3), 805-824. https://doi.org/10.1016/j.ejor.2019.09.047

[15]. Dehghanian, P., Fotuhi-Firuzabad, M., Aminifar, F., & Billinton, R. (2014). Reliability centered maintenance optimization for power distribution systems. *International Journal of Electrical Power & Energy Systems*, *55*, 108-115. https://doi.org/10.1016/j.ijepes.2013.08.025

[16]. Fink, O. (2019). Data-driven intelligent predictive maintenance of industrial assets. In *Women in Industrial and Systems Engineering* (pp. 589-605). https://doi.org/10.1007/978-3-030-11866-2_25

[17]. Habibullah, S. M., & Md. Foysal, H. (2021). A Data Driven Cyber Physical Framework For Real Time Production Control Integrating IOT And Lean Principles. *American Journal of Interdisciplinary Studies*, *2*(03), 35–70. https://doi.org/10.63125/20nhqs87

[18]. Heng, A., Zhang, S., Tan, A. C. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*. https://doi.org/10.1016/j.ymssp.2008.06.009

[19]. Hoti, E. (2015). The technological, organizational and environmental framework of IS innovation adoption in small and medium enterprises: Evidence from research over the last 10 years. *International Journal of Business and Management*, *3*(4), 1-14. https://doi.org/10.20472/bm.2015.3.4.001

[20]. Islam, M. M., Lee, G., & Hettiwatte, S. N. (2017). A review of condition monitoring techniques and diagnostic tests for lifetime estimation of power transformers. *International Journal of Electrical Power & Energy Systems*. https://doi.org/10.1007/s00202-017-0532-4

[21]. Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*. https://doi.org/10.1016/j.ymssp.2005.09.012

[22]. Kots, M. V., Konstantinov, A. V., Utkin, L. V., Chukanov, V. S., Meldo, A. A., & Ryabinin, M. A. (2019). A weighted random survival forest. *Knowledge-Based Systems*, *177*, 136-144. https://doi.org/10.1016/j.knosys.2019.04.015

[23]. Kumar, U. (2006). Development and implementation of maintenance performance measurement system: Issues and challenges. In *Engineering Asset Management* (pp. 738-743). https://doi.org/10.1007/978-1-84628-814-2_78

[24]. Kumar, U., Galar, D., Parida, A., Stenström, C., & Berges, L. (2013). Maintenance performance metrics: A state-of-the-art review. *Journal of Quality in Maintenance Engineering*, *19*(3), 233-277. https://doi.org/10.1108/jqme-05-2013-0029

[25]. Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mechanical Systems and Signal Processing*. https://doi.org/10.1016/j.ymssp.2013.06.004

[26]. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, *104*, 799-834. https://doi.org/10.1016/j.ymssp.2017.11.016

[27]. Lin, S., Li, C., Xu, F., & Li, W. (2018). The strategy research on electrical equipment condition-based maintenance based on cloud model and grey D–S evidence theory. *Intelligent Decision Technologies*. https://doi.org/10.3233/idt-180333

[28]. Lukens, S., Naik, M., Saetia, K., & Hu, X. (2019). *Best practices framework for improving maintenance data quality to enable asset performance analytics* Proceedings of the Annual Conference of the Prognostics and Health Management Society,

[29]. Ma, S., Chen, M., Wu, J., Wang, Y., Jia, B., & Jiang, Y. (2018). Intelligent fault diagnosis of HVCB with feature space optimization-based random forest. *Sensors*, *18*(4), 1221. https://doi.org/10.3390/s18041221

[30]. Md Sarwar, H. (2021). Sustainable Materials Characterization For Low-Carbon Construction And Infrastructure Durability. *American Journal of Interdisciplinary Studies*, *2*(01), 01-34. https://doi.org/10.63125/wq1wdr64

[31]. Md. Musfiqur, R., & Saba, A. (2021). Data-Driven Decision Support in Information Systems: Strategic Applications In Enterprises. *International Journal of Scientific Interdisciplinary Research*, *2*(2), 01-33. https://doi.org/10.63125/cfvg2v45

[32]. Md. Redwanul, I., Md Nahid, H., & Md. Zahid Hasan, T. (2021). Predictive Analytics in Supply Chain Management A Review Of Business Analyst-Led Optimization Tools. *Review of Applied Science and Technology*, *6*(1), 34-73. https://doi.org/10.63125/5aypx555

[33]. Md. Tarek, H., & Sai Praveen, K. (2021). Data Privacy-Aware Machine Learning and Federated Learning: A Framework For Data Security. *American Journal of Interdisciplinary Studies*, *2*(03), 01–34. https://doi.org/10.63125/vj1hem03

[34]. Okoh, C., Roy, R., Mehnen, J., & Redding, L. (2014). *Overview of remaining useful life prediction techniques in through-life engineering services* Procedia CIRP,

[35]. Oliveira, T., Thomas, M., & Espadanal, M. (2014). Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. *Information & Management*, *51*(5), 497-510. https://doi.org/10.1016/j.im.2014.03.006

[36]. Omar Muhammad, F., & Mst. Shahrin, S. (2021). Comparative Analysis of BI Systems In The U.S. And Europe: Lessons In Data Governance And Predictive Analytics. *Journal of Sustainable Development and Policy*, *1*(5), 01-38. https://doi.org/10.63125/6b3aeg93

[37]. Parida, A., & Kumar, U. (2006). Maintenance performance measurement (MPM): Issues and challenges. *Journal of Quality in Maintenance Engineering*, *12*(3), 239-251. https://doi.org/10.1108/13552510610685084

[38]. Patil, S., Patil, A., Desai, S., Handikherkar, V., Phalle, V. M., & Kazi, F. S. (2019). *Remaining useful life (RUL) prediction of rolling element bearing using random forest and gradient boosting technique* Proceedings of the ASME 2018 International Mechanical Engineering Congress and Exposition (IMECE2018-87623),

[39]. Pecht, M., & Jaai, R. (2010). A prognostics and health management roadmap for information and electronics-rich systems. *Microelectronics Reliability*, *50*(3), 317-323. https://doi.org/10.1016/j.microrel.2010.01.006

[40]. Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: A review. *The International Journal of Advanced Manufacturing Technology*. https://doi.org/10.1007/s00170-009-2482-0

[41]. Qian, X., & Wu, Y. (2014). Condition-based maintenance optimization for the hydro generating unit with dynamic economic dependence. *International Journal of Control and Automation*. https://doi.org/10.14257/ijca.2014.7.3.30

[42]. Ramdani, B., Kawalek, P., & Lorenzo, O. (2009). Predicting SMEs' adoption of enterprise systems. *Journal of Enterprise Information Management*, *22*(1–2), 10-24. https://doi.org/10.1108/17410390910922796

[43]. Saikat, S. (2021). Real-Time Fault Detection in Industrial Assets Using Advanced Vibration Dynamics And Stress Analysis Modeling. *American Journal of Interdisciplinary Studies*, *2*(04), 39–68. https://doi.org/10.63125/0h163429

[44]. Shafiee, M., & Sørensen, J. D. (2019). Maintenance optimization and inspection planning of wind energy assets: Models, methods and strategies. *Reliability Engineering & System Safety*, *192*, 105993. https://doi.org/10.1016/j.ress.2017.10.025

[45]. Shaikh, S., & Aditya, D. (2021). Federated Learning-Driven Predictive Quality Analytics and Supply Chain Optimization In Distributed Manufacturing Networks. *Review of Applied Science and Technology*, *6*(1), 74-107. https://doi.org/10.63125/k18cbz55

[46]. Shen, J., Wan, J., Lim, S.-J., & Yu, L. (2018). Random-forest-based failure prediction for hard disk drives. *International Journal of Distributed Sensor Networks*, *14*(11), 1550147718806480. https://doi.org/10.1177/1550147718806480

[47]. Si, X. S., Wang, W., Hu, C. H., & Zhou, D. H. (2011). Remaining useful life estimation – A review on the statistical data-driven approaches. *European Journal of Operational Research*, *213*(1), 1-14. https://doi.org/10.1016/j.ejor.2010.11.018

[48]. Simões, J. M., Gomes, C. F., & Yasin, M. M. (2011). A literature review of maintenance performance measurement: A conceptual framework and directions for future research. *Journal of Quality in Maintenance Engineering*, *17*(2), 116-137. https://doi.org/10.1108/13552511111134565

[49]. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, *11*(3), 812-820. https://doi.org/10.1109/tii.2014.2349359

[50]. Tsui, K. L., Chen, N., Zhou, Q., Hai, Y., & Wang, W. (2015). Prognostics and health management: A review on data-driven approaches. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2015/793161

[51]. Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, *21*(6), 2560-2574. https://doi.org/10.1016/j.ymssp.2006.12.007

[52]. Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). A comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering*, *139*(7), 071018. https://doi.org/10.1115/1.4036350

[53]. Yang, B.-S., Di, X., & Han, T. (2008). Random forests classifier for machine fault diagnosis. *Journal of Mechanical Science and Technology*, *22*(9), 1716-1725. https://doi.org/10.1007/s12206-008-0603-6

[54]. Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal*, *13*(3), 2213-2227. https://doi.org/10.1109/jsyst.2019.2905565

[55]. Zio, E. (2013). Prognostics and health management of industrial equipment. In *The Monte Carlo Simulation Method for System Reliability and Risk Analysis* (pp. 339-379). https://doi.org/10.4018/978-1-4666-2095-7.ch017