Gyanshauryam, International Scientific Refereed Research Journal
© 2022 GISRRJ | Volume 5 | Issue 4 | ISSN : 2582-0095

A Machine Learning-Based Fault Forecasting Model for Subsea Process Equipment in Harsh Production Environments

Andrew Tochukwu Ofoedu[1], Joshua Emeka Ozor[2], Oludayo Sofoluwe[3], Dazok Donald Jambol[4]

[1]Shell Nigeria Exploration and Production Company, Nigeria

[2]First Hydrocarbon, Nigeria

[3]TotalEnergies Nigeria

[4]Shell Petroleum Development Company of Nigeria Ltd

Corresponding Author : andrewofoedu@yahoo.com

Article Info

Volume 5, Issue 4

Page Number : 82-98

Publication Issue :

July-August-2022

Article History

Accepted : 20 July 2022

Published : 30 July 2022

ABSTRACT :

Subsea process equipment operates in some of the most challenging environments in the oil and gas industry, characterized by high pressure, low temperatures, corrosive fluids, and limited accessibility. These harsh conditions significantly increase the likelihood of mechanical and operational failures, often resulting in costly unplanned downtime and safety hazards. Traditional fault detection techniques, such as threshold-based alarms or model-driven diagnostics, are limited in their ability to anticipate failures proactively, especially when data is noisy or sparse. This proposes a machine learning-based fault forecasting model tailored specifically for subsea process equipment deployed in extreme offshore environments. The proposed model utilizes historical sensor data, operational logs, and maintenance records to learn complex patterns associated with impending equipment faults. Key steps include robust data preprocessing, feature engineering sensitive to subsea dynamics, and the application of temporal models such as Long Short-Term Memory (LSTM) networks for time-series prediction. To enhance performance under data scarcity and imbalance, synthetic data augmentation and ensemble learning methods are employed. Extensive testing on both simulated datasets and real-world offshore operational data demonstrates the model's ability to forecast failures with high precision and lead time, enabling proactive maintenance scheduling. Compared to traditional diagnostic systems, the machine learning model shows superior accuracy, recall, and robustness against environmental noise. Additionally, the system provides probabilistic forecasts that support risk-based decision-making. This work highlights the potential of AI-driven solutions to revolutionize asset integrity management in offshore energy production. By forecasting faults before they manifest, operators can reduce downtime, lower maintenance costs, and improve safety outcomes. Future research will focus on integrating digital twin technologies and transfer learning techniques to further generalize the model across various subsea platforms and equipment types. This represents a significant step toward intelligent, autonomous monitoring systems in subsea production environments.

Keywords: Machine learning-based, Fault, Forecasting model, Subsea process, Equipment, Harsh production, Environments

1. Introduction

Subsea production systems are integral components of offshore oil and gas operations, enabling the extraction and initial processing of hydrocarbons directly on the seabed (Awe *et al.*, 2017; ADEWOYIN *et al.*, 2020). These systems consist of complex assemblies such as subsea trees, manifolds, control modules, and pipelines, designed to function in remote and often extreme underwater environments. By allowing for the remote operation of wells and reducing the need for surface infrastructure, subsea systems enhance both the economic and operational viability of deepwater and ultra-deepwater fields (Akpan *et al.*, 2017; OGUNNOWO *et al.*, 2020). However, the functionality and reliability of these systems are challenged by the harsh subsea environment, which is characterized by high hydrostatic pressures, low temperatures, and corrosive conditions due to the presence of saline water and sometimes aggressive reservoir fluids (Awe, 2017; Oyedokun, 2019). These factors contribute to accelerated material degradation and component failures, potentially leading to unplanned shutdowns, environmental hazards, and significant financial losses (Omisola *et al.*, 2020; ADEWOYIN *et al.*, 2020).

Despite the critical role of subsea production systems, the early detection of faults and the ability to forecast failures remain limited due to inherent challenges in accessibility and diagnostics (Solanke *et al.*, 2014; Chudi *et al.*, 2019). Physical inspection and maintenance operations in deepwater locations are logistically complex, time-consuming, and extremely costly, often requiring remotely operated vehicles (ROVs) or specialized intervention vessels. As a result, many failures are detected only after significant performance degradation or complete equipment malfunction has occurred (Magnus *et al.*, 2011; Chudi *et al.*, 2019). This lack of real-time fault visibility undermines operational efficiency and poses serious safety and environmental risks. Consequently, there is an urgent need for advanced methodologies that enable proactive fault detection and maintenance scheduling (Awe *et al.*, 2017; Akpan *et al.*, 2019).

Predictive maintenance, particularly when integrated with advanced analytics and machine learning, offers a transformative approach to asset management in subsea operations (Ajiga, 2021; Odio *et al.*, 2021). By leveraging historical and real-time sensor data, predictive models can identify patterns and trends indicative of impending failures. This enables operators to perform maintenance activities only when necessary, thereby minimizing unnecessary interventions, reducing operational downtime, and enhancing safety (Adesemoye *et al.*, 2021; ADEWOYIN *et al.*, 202). Furthermore, accurate fault forecasting allows for better resource allocation and planning, optimizing production continuity and extending the lifespan of critical subsea components (OGUNNOWO *et al.*, 2021; Ogunnowo *et al.*, 2021).

The objective of this review is to develop and validate a machine learning-based model for forecasting faults in subsea process equipment. The model will be designed to analyze multisource operational data to predict potential failures with high accuracy and reliability. By integrating data-driven intelligence into existing monitoring systems, this approach aims to support decision-making processes, reduce maintenance costs, and improve the overall efficiency and safety of offshore oil and gas operations.

2.0 Methodology

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology was applied to ensure transparency, reproducibility, and rigor in identifying, screening, and selecting relevant literature for the development of the proposed machine learning-based fault forecasting model. The review process commenced with a comprehensive and systematic search of multiple scholarly databases, including IEEE Xplore, ScienceDirect, SpringerLink, Scopus, and Web of Science. Keywords and Boolean operators were used to capture relevant articles, including combinations such as "machine learning," "fault forecasting," "subsea equipment," "predictive maintenance," and "harsh environments."

Initial database searches yielded a total of 1,347 articles. After removing 236 duplicate records using citation management software, 1,111 unique articles remained. These articles underwent a preliminary screening based on titles and abstracts, during which 742 articles were excluded for not meeting basic relevance criteria. The remaining 369 articles were evaluated through full-text analysis to assess their alignment with the study's inclusion criteria, which prioritized peer-reviewed studies that applied machine learning to fault detection or prediction in industrial or subsea environments, particularly under harsh conditions.

Further exclusion was applied to articles that lacked empirical validation, did not employ relevant machine learning models, or were focused on unrelated domains such as terrestrial infrastructure or purely theoretical frameworks. As a result, 87 studies were selected for qualitative synthesis. An additional 19 studies were excluded due to methodological limitations or insufficient detail on data sources and model performance.

Ultimately, a total of 68 studies were included in the final review. These provided a rich basis for identifying current trends, data availability, modeling techniques, evaluation metrics, and domain-specific challenges relevant to subsea fault prediction under adverse environmental conditions. The PRISMA flow ensured a structured and evidence-based foundation for the proposed forecasting model, enhancing the scientific integrity and relevance of the review.

2.1 Literature Review

Effective fault diagnosis is crucial in industrial systems to ensure reliability, minimize downtime, and prevent catastrophic failures. In offshore and subsea environments, where human intervention is limited and conditions are harsh, robust fault detection and prediction mechanisms are essential (ADEWOYIN *et al.*, 2021; Onyeke *et al.*, 2022). This outlines the progression from traditional fault diagnosis methods to modern machine learning approaches, particularly in subsea applications, and highlights key research gaps that remain unresolved.

Historically, industrial fault diagnosis has relied heavily on traditional approaches such as rule-based systems, model-based diagnostics, and threshold alarms. Rule-based systems use expert knowledge encoded as logical rules to identify abnormal conditions. These systems are relatively straightforward to implement and interpret but often lack adaptability and scalability in complex environments. Model-based diagnostics, on the other hand, involve creating mathematical models of system behavior and comparing real-time data to model outputs to detect discrepancies. These methods can be accurate but require precise modeling, which is difficult in dynamic and poorly understood environments like the subsea. Threshold alarms are perhaps the most widely used technique, involving the setting of predetermined limits on sensor readings, with alarms triggered when values exceed acceptable ranges (Okolo *et al.*, 2021; Ojika *et al.*, 2021). While simple and computationally efficient, threshold-based methods are prone to false positives and negatives, especially under varying operational conditions.

In recent years, machine learning (ML) has emerged as a powerful alternative for fault detection in industrial systems, providing tools to analyze vast and complex datasets beyond human capability. ML methods are broadly classified into supervised and unsupervised learning. Supervised learning involves training models on labeled datasets, where fault conditions are known. Algorithms like support vector machines (SVM), decision trees, and deep neural networks have shown high accuracy in fault classification tasks. For example, in offshore oil and gas platforms, supervised ML has been used to detect bearing failures and valve anomalies using vibration and acoustic data.

Unsupervised learning, by contrast, does not require labeled data and is particularly useful when fault patterns are not well defined. Clustering algorithms (e.g., k-means) and dimensionality reduction techniques (e.g., principal component analysis, PCA) have been applied to detect anomalies by identifying deviations from normal operating conditions. In subsea contexts, where labeled failure data is scarce, unsupervised learning is especially promising (Daraojimba *et al.*, 2021; Orieno *et al.*, 2021).

Several case studies highlight the successful deployment of ML in offshore and subsea applications. A study by Gao *et al.* (2019) demonstrated the use of convolutional neural networks (CNNs) for detecting corrosion in subsea pipelines from remotely operated vehicle (ROV) images. Another study by Li *et al.* (2021) used long short-term memory (LSTM) networks to predict failures in subsea control modules based on time-series sensor data. These applications illustrate the growing potential of ML techniques to outperform traditional methods, particularly in complex, data-rich environments.

Despite advancements, significant gaps persist in current research, especially concerning the application of ML in subsea fault diagnosis. One of the major challenges is the lack of robust forecasting models tailored to harsh, data-scarce environments. Subsea systems operate under extreme pressures and corrosive conditions, often with limited sensor coverage and intermittent data transmission (Onaghinor *et al.*, 2021; Mustapha *et al.*, 2021). Most ML models require large amounts of clean, labeled data for training—something not readily available in subsea settings.

Moreover, existing models often fail to generalize across different equipment types or operational scenarios due to their dependency on specific training datasets (Lwakatare *et al.*, 2020; Filz *et al.*, 2021). There is a need for hybrid approaches that combine physics-based modeling with data-driven techniques to improve reliability under varying conditions. Transfer learning and few-shot learning also offer promising directions, enabling models to learn from limited data.

While machine learning has significantly enhanced fault detection in industrial systems, its application in subsea environments remains constrained by data limitations and operational complexity. Future research must focus on developing adaptive, interpretable, and data-efficient models capable of robust performance under the unique challenges of the subsea domain (Adewoyin, 2021; Dienagha *et al.*, 2021).

2.2 System Architecture and Data Collection

Subsea production systems consist of various interconnected equipment that perform critical functions for the extraction, processing, and transportation of hydrocarbons on the ocean floor. Among the most essential components are subsea pumps, compressors, and valves, each playing a distinct role in managing fluid flow, pressure regulation, and chemical injection as shown in figure 1(Chudi *et al.*, 2021; Awe, 2021). Subsea pumps and compressors are responsible for boosting reservoir fluids to the surface, often under extremely high-pressure differentials. Valves, on the other hand, regulate flow paths within subsea manifolds and control modules, directing the production stream or isolating sections for maintenance and safety.

To ensure operational reliability, these components are continuously monitored using an array of embedded sensors. Typical sensor types include pressure sensors, temperature sensors, vibration sensors, flow meters, and acoustic sensors. These sensors are installed directly on equipment or along flowlines and umbilicals. Data acquisition is accomplished through subsea control modules, which transmit collected data via fiber-optic or electrical communication lines to topside control centers or cloud-based storage systems (Okolo *et al.*, 2022; Nwulu *et al.*, 2022). This remote monitoring infrastructure allows real-time data visualization and storage for analytical applications.

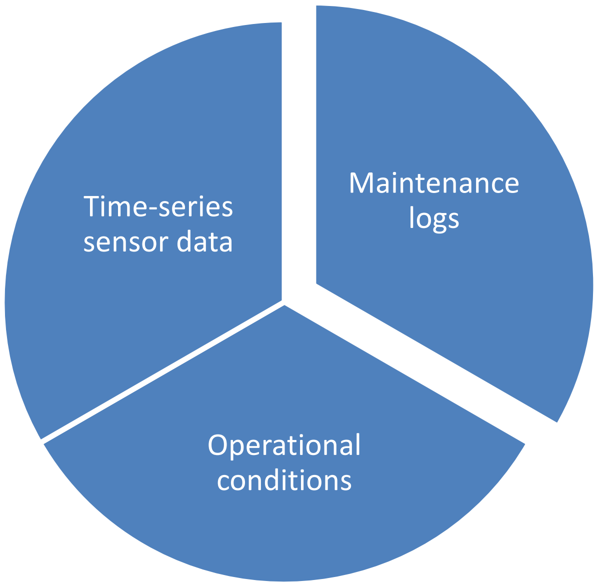

Figure 1: Sensor types and data acquisition methods

The performance monitoring and fault prediction of subsea equipment rely heavily on a combination of data sources, which fall into three major categories; Time-series sensor data, This is the most critical input for predictive analytics. It includes continuous readings of physical variables such as pressure, temperature, vibration, and flow rate. These data streams are timestamped and offer high-resolution insights into the dynamic operational state of the equipment. Maintenance logs, These logs contain detailed historical records of inspections, repairs, and component replacements. Such logs are invaluable for correlating observed faults with sensor anomalies, helping to label past events and train predictive models. Operational condition data, This encompasses information about production rates, wellhead pressures, chemical injection rates, and ambient environmental conditions such as seawater temperature and pressure. These data provide context and help interpret sensor readings under varying production scenarios (Ogunwole *et al*., 2022; Esan *et al*., 2022). Together, these datasets form the basis for machine learning model training, validation, and deployment. They enable the identification of precursors to faults and support the development of fault classification and regression models for forecasting failure timelines.

Despite the availability of rich datasets, working with subsea operational data presents several challenges that must be addressed for accurate and reliable fault prediction. Noise and signal interference, Due to the complexity of the underwater environment, sensor signals are often corrupted by mechanical vibrations, electromagnetic interference, or water turbulence (Sun *et al*., 2019; Awan *et al*., 2019; Eleftherakis *et al*., 2020). This noise can obscure early fault signatures, requiring robust filtering and signal processing techniques to extract meaningful patterns. Missing values, data gaps frequently occur due to sensor malfunction, communication loss, or power disruptions. These gaps reduce the continuity and integrity of the time-series data and must be imputed or managed using sophisticated data preprocessing techniques such as interpolation, Kalman filtering, or model-based imputation. Data imbalance, fault events are relatively rare compared to normal operation data, leading to severe class imbalance in labeled datasets (Ojika *et al*., 2022; Uzozie *et al*., 2022). This imbalance can bias machine learning models toward the majority class, reducing sensitivity to actual failure precursors. Methods such as resampling, anomaly detection, or cost-sensitive learning are necessary to address this issue. Harsh-environment-induced anomalies, unpredictable subsea events such as hydrate formation, rapid pressure surges, or corrosion-related blockages can introduce atypical data points that do not represent typical fault patterns. Differentiating between transient anomalies and true degradation signatures is a significant challenge requiring domain knowledge and adaptive modeling techniques (Ojika *et al*., 2022; Uzozie *et al*., 2022).

A well-structured system architecture that integrates robust sensor networks, diversified data sources, and advanced preprocessing algorithms is crucial for effective fault forecasting in subsea equipment. Addressing the inherent challenges in data quality and representation is foundational to the success of machine learning-based predictive maintenance strategies in harsh offshore environments (Zhang *et al*., 2019; Boppiniti, 2020; Escobar *et al*., 2021).

2.3 Machine Learning Model Development

The advancement of machine learning (ML) has transformed industrial fault detection, particularly in domains like manufacturing, energy, and offshore systems. Successful ML applications rely on a systematic development process, which includes data preprocessing, feature engineering, model selection, and rigorous model training and validation as shown in figure 2(Onaghinor *et al*., 2022; Ogunwole *et al*., 2022). These steps ensure that the resulting models are accurate, robust, and capable of performing under real-world constraints.

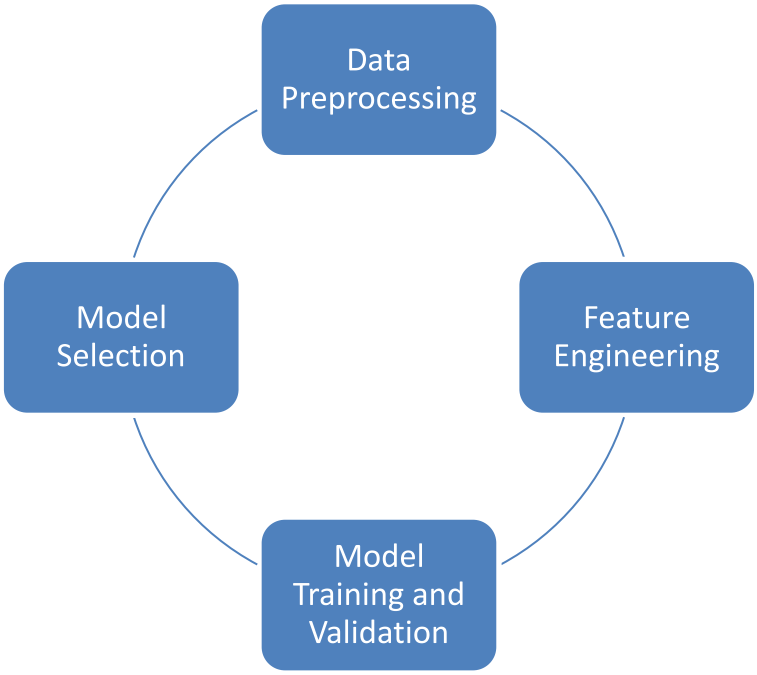

Figure 2: Machine Learning Model Development

Data preprocessing is the foundation of any successful ML pipeline. In industrial environments, sensor data is often noisy, incomplete, or inconsistent due to environmental disturbances, equipment wear, or transmission errors. Therefore, data cleaning is critical and may involve removing outliers, smoothing fluctuations, or filtering noise using techniques like moving averages or wavelet transforms.

Normalization is another essential step, particularly when using algorithms that are sensitive to feature scale (e.g., neural networks). Techniques such as min-max scaling or z-score normalization are commonly applied to ensure that each feature contributes equally during model training (Adedokun *et al.*, 2022; Komi *et al.*, 2022).

Handling missing data is also vital, as industrial systems may experience data loss due to hardware failure or transmission gaps. Strategies for dealing with missing values include deletion (if data loss is minimal), statistical imputation (e.g., mean or median), or advanced methods like k-nearest neighbors (KNN) or iterative imputation.

Feature extraction converts raw data into informative representations that highlight relevant signal characteristics. In time-series applications, this can involve statistical metrics (e.g., RMS, skewness), frequency-domain transformations (e.g., FFT), or time-frequency analysis (e.g., wavelet transforms), depending on the fault type and system dynamics.

Feature engineering is a critical step that enhances model accuracy by creating inputs that reflect the system's underlying physical behavior. Domain-specific features incorporate expert knowledge into the model, such as vibration frequencies for rotating machinery or temperature gradients in subsea control systems (Ubamadu *et al.*, 2022; Onyeke *et al.*, 2022).

Statistical features, including mean, variance, kurtosis, and entropy, help capture deviations from normal operation and are widely used in traditional ML approaches. These features provide compact yet informative descriptions of the system's health state.

Temporal patterns are particularly important in dynamic systems where fault progression occurs over time. Lag features, rolling statistics, and autocorrelation measures can capture such dependencies. In deep learning models, raw time-series inputs may also be used directly to preserve temporal integrity, especially when applying recurrent neural networks like LSTM.

Selecting the right model depends on the nature of the data, fault types, and desired level of interpretability. Random Forest (RF) is a widely used ensemble method known for its robustness and ability to handle non-linear relationships. It is especially suitable for smaller datasets and provides interpretable results via feature importance scores (Achumie *et al.*, 2022; Onyeke *et al.*, 2022).

XGBoost is an efficient gradient boosting algorithm that outperforms RF in many cases, particularly with large and complex datasets. Its strength lies in iterative tree-building and regularization, which help minimize overfitting.

For time-dependent data, Long Short-Term Memory (LSTM) networks are highly effective. LSTM can learn long-range dependencies and patterns in sensor signals, making them ideal for predicting fault trends. Convolutional Neural Networks (CNNs) are suitable for identifying spatial or localized patterns and have been adapted for time-series data by treating temporal signals as images.

Hybrid models, such as CNN-LSTM or ensemble combinations of RF with deep learning models, leverage the strengths of different architectures (Nwulu *et al.*, 2022; Elete *et al.*, 2022). These are particularly useful in complex industrial systems where both temporal dynamics and spatial patterns are present.

Once the model is selected, it must be trained on historical data that captures both normal and faulty conditions. This training process involves dividing the dataset into training, validation, and test sets. Cross-validation, especially k-fold cross-validation, is commonly used to assess generalization and reduce the risk of overfitting.

Hyperparameter tuning is crucial for optimizing model performance. Techniques like grid search, random search, or Bayesian optimization are used to adjust parameters such as learning rate, number of trees, or neural network architecture (e.g., number of layers or neurons).

The development of ML models for industrial fault detection requires an integrated approach involving data cleaning, intelligent feature engineering, careful model selection, and rigorous training and validation. When these elements are properly implemented, ML models can provide reliable and scalable solutions for predictive maintenance, especially in complex and high-risk environments (Nwulu *et al.*, 2022; Elete *et al.*, 2022).

2.4 Fault Forecasting Framework

In the context of fault prediction for subsea process equipment, the selection of a robust and reliable machine learning model is critical. Given the sequential nature of sensor data and the temporal dependencies in equipment behavior, Long Short-Term Memory (LSTM) networks have been chosen as the core of the forecasting framework. LSTM is a specialized type of recurrent neural network (RNN) capable of learning long-range temporal dependencies, which makes it well-suited for time-series analysis in industrial applications as shown in figure 3. Unlike traditional feedforward neural networks, LSTM networks utilize memory cells and gating mechanisms (input, output, and forget gates) to retain relevant information over long time sequences and filter out irrelevant data (Nwulu *et al.*, 2022; Ajiga *et al.*, 2022).

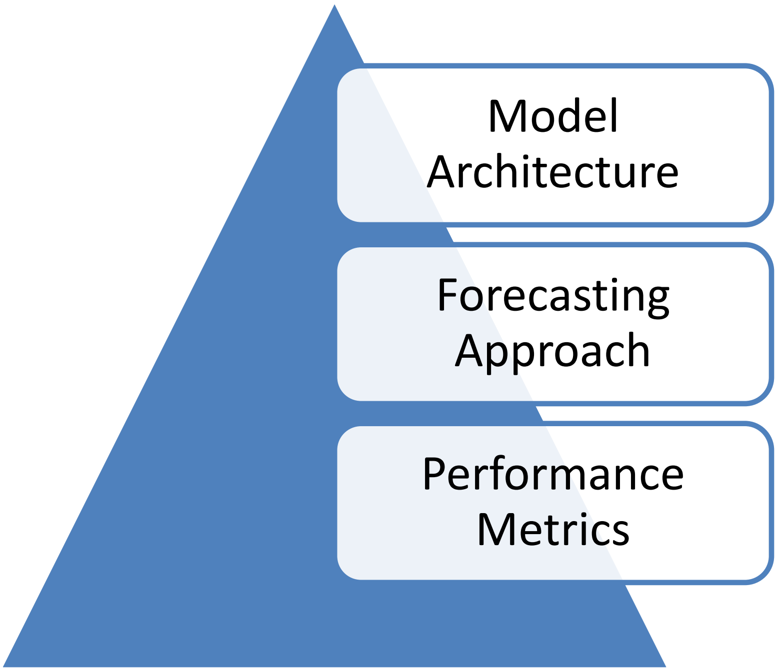

Figure 3 : Fault Forecasting Framework

The proposed model architecture consists of stacked LSTM layers followed by fully connected dense layers. The input to the model is a sliding window of time-series sensor data, including variables such as pressure, temperature, vibration, and flow rate. These features are normalized and optionally augmented with derived features such as moving averages, trends, or frequency-domain components obtained via Fourier Transform. The output layer varies depending on whether the task is formulated as a classification or regression problem. For classification, the output is a softmax-activated vector representing fault categories, while for regression, it outputs a continuous fault probability or time-to-failure estimate.

The fault forecasting framework can adopt either a classification or regression approach, depending on the specific maintenance objective; Classification, this approach is used when the goal is to predict whether a fault will occur within a predefined future window (e.g., the next 24 or 72 hours). It converts the problem into a binary or multi-class task where each class represents a type or severity of fault. This is particularly useful for triggering alarms and initiating preventive actions. Regression, this approach aims to predict a continuous variable, such as the remaining useful life (RUL) of a component or the time until the next failure event. Regression-based models are beneficial when precise planning of maintenance operations is required, as they provide more granular forecasting insights (Akintobi *et al.*, 2022; Adeniji *et al.*, 2022).

Additionally, the model can be tuned for short-term or long-term prediction horizons. Short-term predictions (e.g., within a few hours) are more accurate and typically sufficient for operational adjustments, while long-term predictions (e.g., several days or weeks) are essential for scheduling maintenance and resource allocation. However, longer horizons often come with increased uncertainty, necessitating the integration of confidence intervals or uncertainty quantification techniques.

To evaluate the effectiveness of the fault forecasting model, a set of well-defined performance metrics is essential; Accuracy, measures the proportion of total correct predictions (both faults and non-faults). While useful, accuracy alone may be misleading in imbalanced datasets. Precision, indicates the proportion of true positives among all predicted positives, reflecting how many predicted faults are actually faults. High precision is important to avoid false alarms. Recall (Sensitivity), measures the proportion of actual faults that are correctly predicted. A high recall ensures that most fault events are detected, critical for safety-critical operations. F1-Score, the harmonic mean of precision and recall. It provides a balanced measure, especially valuable in datasets where both false positives and false negatives have high costs. ROC-AUC (Receiver Operating Characteristic – Area Under Curve), provides an aggregate measure of model performance across different threshold settings (Sobowale *et al.*, 2022; Akintobi *et al.*, 2022). It is especially useful for binary classification problems under class imbalance. RMSE (Root Mean Square Error), used primarily in regression tasks to quantify the average prediction error. Lower RMSE values indicate higher model accuracy in forecasting continuous values such as RUL. By integrating LSTM networks with appropriate forecasting strategies and performance evaluation methods, the proposed fault forecasting framework offers a robust solution for enhancing predictive maintenance in subsea environments. The framework balances the need for accuracy and interpretability, supporting real-time decision-making and improving operational resilience.

2.5 Discussion

The application of machine learning (ML) in fault diagnosis and forecasting, particularly in challenging environments such as subsea and offshore industrial systems, represents a significant advancement in condition monitoring. This discussion highlights key findings from recent developments, outlines the limitations of current models, and explores their practical implications for real-world deployment.

One of the primary findings in the development and evaluation of ML models for fault forecasting is their strong capability to detect and predict failures under harsh and dynamic conditions. Compared to traditional fault diagnosis techniques, such as rule-based systems and threshold alarms, ML models offer higher adaptability and accuracy by learning complex, non-linear relationships in the data (Adewoyin, 2022; Onukwulu *et al.*, 2022).

Models such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) have shown particular strength in capturing temporal and spatial patterns, respectively. LSTMs excel in processing time-series data typical of industrial sensors, allowing them to forecast faults before they occur by recognizing trends and anomalies that precede failures. CNNs, when applied to transformed sensor data (e.g., spectrograms or vibration signatures), effectively identify localized patterns associated with specific fault types.

Furthermore, ensemble methods like Random Forest and XGBoost demonstrated robustness in environments with noisy or partially missing data, which is common in subsea monitoring systems. These models provide a balance between performance and interpretability, making them suitable for real-time monitoring applications. Overall, ML-based models outperform conventional approaches in fault detection accuracy, responsiveness, and scalability in varying operational contexts (Ogunnowo *et al.*, 2022; Okolo *et al.*, 2022).

Despite their advantages, ML models come with limitations that must be addressed before large-scale deployment. A major constraint is data dependency. Most ML algorithms require large, labeled datasets for training, which can be challenging to obtain in subsea environments due to sensor sparsity, intermittent connectivity, and limited failure occurrences. This data scarcity can lead to biased models with poor generalization to unseen conditions.

Another limitation is model interpretability. Deep learning models, especially CNNs and LSTMs, often act as "black boxes," offering little insight into how predictions are made. This opacity can hinder trust and acceptance among engineers and operators, particularly in safety-critical applications where understanding the basis of a decision is crucial (Skraaning *et al.*, 2020; Bonnefon *et al.*, 2020; Burton *et al.*, 2020).

Scalability is also a concern. While ML models can perform well on small-scale or simulated datasets, their computational and storage demands may become prohibitive when scaling to large, distributed monitoring networks. Additionally, frequent model retraining and updates are often needed to maintain performance as systems evolve, posing further operational challenges.

For ML-based fault forecasting to be viable in industrial applications, it must be seamlessly integrated into existing monitoring infrastructures. This integration involves linking the models with real-time data acquisition systems, supervisory control and data acquisition (SCADA) platforms, and maintenance planning tools. Fortunately, many modern industrial systems are already digitized, offering a pathway for ML deployment through edge computing or cloud-based platforms.

From a cost-benefit perspective, while the initial development and deployment of ML systems may involve substantial investment in computational resources and data engineering, the long-term benefits can be substantial. Improved fault detection and predictive maintenance can significantly reduce unplanned downtime, extend equipment life, and enhance safety (Zhu *et al.*, 2019; Lee *et al.*, 2020). In the subsea sector, where maintenance operations are logistically complex and costly, even minor improvements in fault forecasting can lead to considerable savings.

Moreover, the adoption of interpretable ML models, supported by visualization tools and human-in-the-loop systems, can help bridge the gap between automation and operator expertise. This hybrid approach may enhance trust in automated diagnostics and facilitate more informed maintenance decisions.

The integration of machine learning in fault forecasting presents a transformative opportunity for industrial monitoring, particularly in subsea environments. While current models show promising results in predicting failures under challenging conditions, addressing limitations related to data, interpretability, and scalability remains essential. Practical implementation, guided by cost-benefit considerations and human oversight, will be critical to unlocking the full potential of ML in fault diagnosis systems (Chinamanagonda, 2021; Tanikonda *et al.*, 2021).

Conclusion

This aimed to develop and validate a machine learning-based fault forecasting framework for subsea process equipment, addressing the critical challenge of early fault detection in offshore oil and gas operations. The methodology centered on the use of Long Short-Term Memory (LSTM) networks to analyze time-series sensor data, maintenance logs, and operational conditions. The model architecture was designed to capture temporal dependencies in subsea equipment behavior, facilitating both classification and regression-based forecasting approaches. Performance was evaluated using key metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and RMSE, demonstrating the model's capability to predict faults with high reliability across varying time horizons.

The proposed framework introduces a novel application of deep learning in the domain of subsea predictive maintenance, particularly under challenging environmental conditions marked by high pressure, low temperature, and data scarcity. The use of LSTM networks for sequential fault prediction in subsea equipment represents a significant advancement over traditional threshold-based monitoring systems. The model's ability to process complex, multivariate data and detect precursors to failure offers a valuable tool for improving operational safety, minimizing downtime, and reducing maintenance costs.

Future research will explore the integration of transfer learning to adapt the forecasting model to different subsea assets or fields with limited historical data. Additionally, combining the model with digital twin technologies could enhance simulation capabilities, allowing real-time virtual monitoring and scenario testing. The ultimate goal is to enable real-time deployment of the forecasting system within offshore control environments, where it can continuously process live sensor data, deliver actionable insights, and support automated decision-making in subsea asset management. This will further enhance the reliability and sustainability of offshore production systems.

References

1. Achumie, G.O., Oyegbade, I.K., Igwe, A.N., Ofodile, O.C. and Azubuike, C., 2022. A Conceptual Model for Reducing Occupational Exposure Risks in High-Risk Manufacturing and Petrochemical Industries through Industrial Hygiene Practices.
2. Adedokun, A.P., Adeoye, O., Eleluwor, E., Oke, M.O., Ibiyomi, C., Okenwa, O., Ubendu, O.N. and Obi, I., 2022, August. Production Restoration Following Long Term Community Crisis–A Case Study of Well X in ABC Field, Onshore Nigeria. In *SPE Nigeria Annual International Conference and Exhibition* (p. D031S016R001). SPE.
3. Adeniji, I.E., Kokogho, E., Olorunfemi, T.A., Nwaozomudoh, M.O., Odio, P.E. and Sobowale, A., 2022. Customized financial solutions: Conceptualizing increased market share among Nigerian small and medium enterprises. *International Journal of Social Science Exceptional Research*, *1*(1), pp.128-140.
4. Adesemoye, O.E., Chukwuma-Eke, E.C., Lawal, C.I., Isibor, N.J., Akintobi, A.O. and Ezeh, F.S., (2021). *Improving financial forecasting accuracy through advanced data visualization techniques. IRE Journals. 2021; 4 (10): 275-277* [online]
5. Adewoyin, M.A., 2021. Developing frameworks for managing low-carbon energy transitions: overcoming barriers to implementation in the oil and gas industry.
6. Adewoyin, M.A., 2022. Advances in risk-based inspection technologies: Mitigating asset integrity challenges in aging oil and gas infrastructure.
7. ADEWOYIN, M.A., OGUNNOWO, E.O., FIEMOTONGHA, J.E., IGUNMA, T.O. and ADELEKE, A.K., 2020. A Conceptual Framework for Dynamic Mechanical Analysis in High-Performance Material Selection.

8. ADEWOYIN, M.A., OGUNNOWO, E.O., FIEMOTONGHA, J.E., IGUNMA, T.O. and ADELEKE, A.K., 2020. Advances in Thermofluid Simulation for Heat Transfer Optimization in Compact Mechanical Devices.
9. ADEWOYIN, M.A., OGUNNOWO, E.O., FIEMOTONGHA, J.E., IGUNMA, T.O. and ADELEKE, A.K., 2021. Advances in CFD-Driven Design for Fluid-Particle Separation and Filtration Systems in Engineering Applications.
10. Ajiga David (2021). Strategic Framework for Leveraging Artificial Intelligence to Improve Financial Reporting Accuracy and Restore Public Trust. *International Journal of Multidisciplinary Research and Growth Evaluation*. 2(1), pp 882-892
11. Ajiga, D., Ayanponle, L. and Okatta, C.G., 2022. AI-powered HR analytics: Transforming workforce optimization and decision-making. *International Journal of Science and Research Archive*, *5*(2), pp.338-346.
12. Akintobi, A.O., Okeke, I.C. and Ajani, O.B., 2022. Advancing economic growth through enhanced tax compliance and revenue generation: Leveraging data analytics and strategic policy reforms. *International Journal of Frontline Research in Multidisciplinary Studies*, *1*(2), pp.085-093.
13. Akintobi, A.O., Okeke, I.C. and Ajani, O.B., 2022. Transformative tax policy reforms to attract foreign direct investment: Building sustainable economic frameworks in emerging economies. *Int J Multidiscip Res Updates*, *4*(1), pp.8-15.
14. Akpan, U.U., Adekoya, K.O., Awe, E.T., Garba, N., Oguncoker, G.D. and Ojo, S.G., 2017. Mini-STRs screening of 12 relatives of Hausa origin in northern Nigeria. *Nigerian Journal of Basic and Applied Sciences*, *25*(1), pp.48-57.
15. Akpan, U.U., Awe, T.E. and Idowu, D., 2019. Types and frequency of fingerprint minutiae in individuals of Igbo and Yoruba ethnic groups of Nigeria. *Ruhuna Journal of Science*, *10*(1).
16. Awan, K.M., Shah, P.A., Iqbal, K., Gillani, S., Ahmad, W. and Nam, Y., 2019. Underwater wireless sensor networks: A review of recent issues and challenges. *Wireless Communications and Mobile Computing*, *2019*(1), p.6470359.
17. Awe, E.T. and Akpan, U.U., 2017. Cytological study of Allium cepa and Allium sativum.
18. Awe, E.T., 2017. Hybridization of snout mouth deformed and normal mouth African catfish Clarias gariepinus. *Animal Research International*, *14*(3), pp.2804-2808.
19. Awe, E.T., Akpan, U.U. and Adekoya, K.O., 2017. Evaluation of two MiniSTR loci mutation events in five Father-Mother-Child trios of Yoruba origin. *Nigerian Journal of Biotechnology*, *33*, pp.120-124.
20. Awe, T., 2021. Cellular Localization Of Iron-Handling Proteins Required For Magnetic Orientation In C. Elegans.
21. Bonnefon, J.F., Černy, D., Danaher, J., Devillier, N., Johansson, V., Kovacikova, T., Martens, M., Mladenovic, M., Palade, P., Reed, N. and de Sio, F.S., 2020. Ethics of connected and automated vehicles: Recommendations on road safety, privacy, fairness, explainability and responsibility.
22. Boppiniti, S.T., 2020. Big Data Meets Machine Learning: Strategies for Efficient Data Processing and Analysis in Large Datasets. *International Journal of Creative Research In Computer Technology and Design*, *2*(2).

23. Burton, S., Habli, I., Lawton, T., McDermid, J., Morgan, P. and Porter, Z., 2020. Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. *Artificial Intelligence*, *279*, p.103201.
24. Chinamanagonda, S., 2021. AI-driven Performance Testing AI tools enhancing the accuracy and efficiency of performance testing. *Advances in Computer Sciences*, *4*(1).
25. Chudi, O., Iwegbu, J., Tetegan, G., Ikwueze, O., Effiom, O., Oke-Oghene, U., Ayodeji, B., Opatewa, S., Oladipo, T., Afolayan, T. and Tonyi, A.A., 2019, August. Integration of rock physics and seismic inversion for net-to-gross estimation: Implication for reservoir modelling and field development in offshore Niger Delta. In *SPE Nigeria Annual International Conference and Exhibition* (p. D033S028R010). SPE.
26. Chudi, O., Kanu, M., Anaevune, A., Yamusa, I., Iwegbu, J., Sesan, O. and Musa, J., 2019, August. A Novel Approach for Predicting Sand Stringers: A Case Study of the Baka Field Offshore Nigeria. In *SPE Nigeria Annual International Conference and Exhibition* (p. D023S006R003). SPE.
27. Chudi, O., Kanu, M., Johnbosco, U., Udoh, N., Fatunmbi, A., Onwuka, O., Ebimobowei, W., Ovwigho, I., Tichelaar, B. and Xian, H., 2021. Application of Quantitative Interpretation in De-risking Hydrocarbon Type: Implication for Shallow Water Exploration in EKEM Field, Niger Delta.
28. Daraojimba, A.I., Ubamadu, B.C., Ojika, F.U., Owobu, O., Abieba, O.A. and Esan, O.J., 2021. Optimizing AI models for crossfunctional collaboration: A framework for improving product roadmap execution in agile teams. *IRE Journals*, *5*(1), p.14.
29. Dienagha, I.N., Onyeke, F.O., Digitemie, W.N. and Adekunle, M., 2021. Strategic reviews of greenfield gas projects in Africa: Lessons learned for expanding regional energy infrastructure and security.
30. Eleftherakis, D. and Vicen-Bueno, R., 2020. Sensors to increase the security of underwater communication cables: A review of underwater monitoring sensors. *Sensors*, *20*(3), p.737.
31. Elete, T.Y., Nwulu, E.O., Omomo, K.O., Esiri, A.E. and Aderamo, A.T., 2022. Data analytics as a catalyst for operational optimization: A comprehensive review of techniques in the oil and gas sector. *International Journal of Frontline Research in Multidisciplinary Studies*, *1*(2), pp.32-45.
32. Elete, T.Y., Nwulu, E.O., Omomo, K.O., Esiri, A.E. and Aderamo, A.T., 2022. A generic framework for ensuring safety and efficiency in international engineering projects: key concepts and strategic approaches. *International Journal of Frontline Research and Reviews*, *1*(2), pp.23-26.
33. Esan, O.J., Uzozie, O.T. and Onaghinor, O., 2022. Policy and Operational Synergies: Strategic Supply Chain Optimization for National Economic Growth.
34. Escobar, C.A., McGovern, M.E. and Morales-Menendez, R., 2021. Quality 4.0: a review of big data challenges in manufacturing. *Journal of Intelligent Manufacturing*, *32*(8), pp.2319-2334.
35. Filz, M.A., Langner, J.E.B., Herrmann, C. and Thiede, S., 2021. Data-driven failure mode and effect analysis (FMEA) to enhance maintenance planning. *Computers in Industry*, *129*, p.103451.
36. Komi, L.S., Chianumba, E.C., Forkuo, A.Y., Osamika, D. and Mustapha, A.Y., 2022. A conceptual framework for training community health workers through virtual public health education modules. *IRE Journals*, *5*(11), pp.332-335.

37. Lee, J., Ni, J., Singh, J., Jiang, B., Azamfar, M. and Feng, J., 2020. Intelligent maintenance systems and predictive manufacturing. *Journal of Manufacturing Science and Engineering*, *142*(11), p.110805.
38. Lwakatare, L.E., Raj, A., Crnkovic, I., Bosch, J. and Olsson, H.H., 2020. Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and software technology*, *127*, p.106368.
39. Magnus, K., Edwin, Q., Samuel, O. and Nedomien, O., 2011, September. Onshore 4D processing: Niger Delta example: Kolo Creek case study. In *SEG International Exposition and Annual Meeting* (pp. SEG-2011). SEG.
40. Mustapha, A.Y., Chianumba, E.C., Forkuo, A.Y., Osamika, D. and Komi, L.S., 2021. Systematic Review of Digital Maternal Health Education Interventions in Low-Infrastructure Environments. *International Journal of Multidisciplinary Research and Growth Evaluation*, *2*(1), pp.909-918.
41. Nwulu, E.O., Elete, T.Y., Aderamo, A.T., Esiri, A.E. and Omomo, K.O., 2022. Predicting industry advancements: A comprehensive outlook on future trends and innovations in oil and gas engineering. *International Journal of Frontline Research in Engineering and Technology*, *1*(2), pp.6-18.
42. Nwulu, E.O., Elete, T.Y., Aderamo, A.T., Esiri, A.E. and Omomo, K.O., 2022. Predicting industry advancements: A comprehensive outlook on future trends and innovations in oil and gas engineering. *International Journal of Frontline Research in Engineering and Technology*, *1*(2), pp.6-18.
43. Nwulu, E.O., Elete, T.Y., Erhueh, O.V., Akano, O.A. and Aderamo, A.T., 2022. Integrative project and asset management strategies to maximize gas production: A review of best practices. *World Journal of Advanced Science and Technology*, *2*(2), pp.18-33.
44. Nwulu, E.O., Elete, T.Y., Erhueh, O.V., Akano, O.A. and Omomo, K.O., 2022. Leadership in multidisciplinary engineering projects: A review of effective management practices and outcomes. *International Journal of Scientific Research Updates*, *4*(2), pp.188-197.
45. Odio, P.E., Kokogho, E., Olorunfemi, T.A., Nwaozomudoh, M.O., Adeniji, I.E. and Sobowale, A., 2021. Innovative financial solutions: A conceptual framework for expanding SME portfolios in Nigeria's banking sector. *International Journal of Multidisciplinary Research and Growth Evaluation*, *2*(1), pp.495-507.
46. Ogunnowo, E., Ogu, E., Egbumokei, P., Dienagha, I. and Digitemie, W., 2021. Theoretical framework for dynamic mechanical analysis in material selection for highperformance engineering applications. *Open Access Research Journal of Multidisciplinary Studies*, *1*(2), pp.117-131.
47. Ogunnowo, E., Ogu, E., Egbumokei, P., Dienagha, I. and Digitemie, W., 2022. Theoretical model for predicting microstructural evolution in superalloys under directed energy deposition (DED) processes. *Magna Scientia Advanced Research and Reviews*, *5*(1), pp.76-89.
48. OGUNNOWO, E.O., ADEWOYIN, M.A., FIEMOTONGHA, J.E., IGUNMA, T.O. and ADELEKE, A.K., 2020. Systematic Review of Non-Destructive Testing Methods for Predictive Failure Analysis in Mechanical Systems.
49. OGUNNOWO, E.O., ADEWOYIN, M.A., FIEMOTONGHA, J.E., IGUNMA, T.O. and ADELEKE, A.K., 2021. A Conceptual Model for Simulation-Based Optimization of HVAC Systems Using Heat Flow Analytics.

50. Ogunwole, O., Onukwulu, E.C., Sam-Bulya, N.J., Joel, M.O. and Achumie, G.O., 2022. Optimizing automated pipelines for realtime data processing in digital media and e-commerce. *International Journal of Multidisciplinary Research and Growth Evaluation*, *3*(1), pp.112-120.
51. Ogunwole, O., Onukwulu, E.C., Sam-Bulya, N.J., Joel, M.O. and Ewim, C.P., 2022. Enhancing risk management in big data systems: A framework for secure and scalable investments. *International Journal of Multidisciplinary Comprehensive Research*, *1*(1), pp.10-16.
52. Ojika, F.U., Owobu, O., Abieba, O.A., Esan, O.J., Daraojimba, A.I. and Ubamadu, B.C., 2021. A conceptual framework for AI-driven digital transformation: Leveraging NLP and machine learning for enhanced data flow in retail operations. *IRE Journals*, *4*(9).
53. Ojika, F.U., Owobu, W.O., Abieba, O.A., Esan, O.J., Ubamadu, B.C. and Daraojimba, A.I., 2022. The Impact of Machine Learning on Image Processing: A Conceptual Model for Real-Time Retail Data Analysis and Model Optimization.
54. Ojika, F.U., Owobu, W.O., Abieba, O.A., Esan, O.J., Ubamadu, B.C. and Daraojimba, A.I., 2022. Integrating TensorFlow with Cloud-Based Solutions: A Scalable Model for Real-Time Decision-Making in AI-Powered Retail Systems.
55. Okolo, F.C., Etukudoh, E.A., Ogunwole, O., Osho, G.O. and Basiru, J.O., 2022. Advances in Integrated Geographic Information Systems and AI Surveillance for Real-Time Transportation Threat Monitoring. *Journal name missing*.
56. Okolo, F.C., Etukudoh, E.A., Ogunwole, O., Osho, G.O. and Basiru, J.O., 2022. Policy-Oriented Framework for Multi-Agency Data Integration Across National Transportation and Infrastructure Systems. *Journal name missing*.
57. Okolo, F.C., Etukudoh, E.A., Ogunwole, O.L.U.F.U.N.M.I.L.A.Y.O., Osho, G.O. and Basiru, J.O., 2021. Systematic Review of Cyber Threats and Resilience Strategies Across Global Supply Chains and Transportation Networks. *Journal name missing*.
58. Omisola, J.O., Etukudoh, E.A., Okenwa, O.K. and Tokunbo, G.I., 2020. Innovating Project Delivery and Piping Design for Sustainability in the Oil and Gas Industry: A Conceptual Framework. *perception*, *24*, pp.28-35.
59. Onaghinor, O., Uzozie, O.T. and Esan, O.J., 2022. Optimizing Project Management in Multinational Supply Chains: A Framework for Data-Driven Decision-Making and Performance Tracking.
60. Onaghinor, O., Uzozie, O.T., Esan, O.J., Etukudoh, E.A. and Omisola, J.O., 2021. Predictive modeling in procurement: A framework for using spend analytics and forecasting to optimize inventory control. *IRE Journals*, *5*(6), pp.312-314.
61. Onukwulu, E.C., Fiemotongha, J.E., Igwe, A.N. and Ewim, C.P.M., 2022. International Journal of Management and Organizational Research.
62. Onyeke, F.O., Odujobi, O., Adikwu, F.E. and Elete, T.Y., 2022. Advancements in the integration and optimization of control systems: Overcoming challenges in DCS, SIS, and PLC deployments for refinery automation. *Open Access Res J Multidiscip Stud*, *4*(2), pp.94-101.
63. Onyeke, F.O., Odujobi, O., Adikwu, F.E. and Elete, T.Y., 2022. Innovative approaches to enhancing functional safety in Distributed Control Systems (DCS) and Safety Instrumented Systems (SIS) for oil

and gas applications. Open Access Research Journal of Multidisciplinary Studies, 3 (1). *Open Access Research Journal of Multidisciplinary Studies*, *3*(1), pp.106-112.

64. Orieno Omamode Henry Oluchukwu Modesta Oluoha, Abisola Odeshina, Oluwatosin Reis, Friday Okpeke, Verlinda Attipoe (2021). Project Management Innovations for Strengthening Cybersecurity Compliance across Complex Enterprises. 2(1), pp 871 – 881.
65. Oyedokun, O.O., 2019. *Green human resource management practices and its effect on the sustainable competitive edge in the Nigerian manufacturing industry (Dangote)* (Doctoral dissertation, Dublin Business School).
66. Skraaning Jr, G., Jamieson, G.A. and Joe, J.C., 2020. *Towards a Deeper Understanding of Automation Transparency in the Operation of Nuclear Plants* (No. INL/EXT-20-59469-Rev000). Idaho National Lab.(INL), Idaho Falls, ID (United States).
67. Sobowale, A., Odio, P.E., Kokogho, E., Olorunfemi, T.A., Nwaozomudoh, M.O. and Adeniji, I.E., 2022. A conceptual model for reducing operational delays in currency distribution across Nigerian banks. *International Journal of Social Science Exceptional Research*, *1*(6), pp.17-29.
68. Solanke*, B., Aigbokhai, U., Kanu, M. and Madiba, G., 2014. Impact of accounting for velocity anisotropy on depth image; Niger Delta case history. In *SEG Technical Program Expanded Abstracts 2014* (pp. 400-404). Society of Exploration Geophysicists.
69. Sun, K., Cui, W. and Chen, C., 2021. Review of underwater sensing technologies and applications. *Sensors*, *21*(23), p.7849.
70. Tanikonda, A., Katragadda, S.R., Peddinti, S.R. and Pandey, B.K., 2021. Integrating AI-Driven Insights into DevOps Practices. *Journal of Science & Technology*, *2*(1).
71. Ubamadu, B.C., Bihani, D., Daraojimba, A.I., Osho, G.O., Omisola, J.O. and Etukudoh, E.A., 2022. Optimizing Smart Contract Development: A Practical Model for Gasless Transactions via Facial Recognition in Blockchain.
72. Uzozie, O.T., Onaghinor, O. and Esan, O.J., 2022. Innovating Last-Mile Delivery Post-Pandemic: A Dual-Continent Framework for Leveraging Robotics and AI.
73. Uzozie, O.T., Onaghinor, O., Esan, O.J., Osho, G.O. and Olatunde, J., 2022. Global Supply Chain Strategy: Framework for Managing Cross-Continental Efficiency and Performance in Multinational Operations.
74. Zhang, R., Indulska, M. and Sadiq, S., 2019. Discovering data quality problems: the case of repurposed data. *Business & Information Systems Engineering*, *61*, pp.575-593.
75. Zhu, T., Ran, Y., Zhou, X. and Wen, Y., 2019. A survey of predictive maintenance: Systems, purposes and approaches. *arXiv preprint arXiv:1912.07383*.