

Research Article

Machine Learning Framework for Real-Time Pipeline Anomaly Detection and Maintenance Needs Forecast Using Random Forest and Prophet Model

Obi Chukwuemeka Nwokonkwo , Nwankwo Uchechukwu Samuel ,
Udoka Felista Eze , Adetokunbo MacGregor John-Otumu* 

Department of Information Technology, Federal University of Technology, Owerri, Nigeria

Abstract

This paper introduces an Intelligent Model for Real-Time Pipeline Monitoring and Maintenance Prediction to enhance infrastructure integrity and operational efficiency in Nigeria's oil and gas sector. Given the country's economic dependence on oil and gas revenue, efficient pipeline transportation is crucial. However, pipelines face challenges such as corrosion, mechanical failures, vandalism, and theft, leading to economic losses and environmental risks. Current monitoring systems are mainly reactive, lacking timely anomaly detection and predictive maintenance capabilities. To tackle these challenges, the study utilized sophisticated machine learning methods by combining the Random Forest classifier for real-time anomaly detection with the Prophet model for predictive maintenance forecasting. Datasets from Kaggle were used. The Random Forest classifier demonstrated robust performance with an accuracy of 93.48%, precision of 93.75%, recall of 96.77%, and an F1-score of 95.24%. The Prophet model provided accurate hourly forecasts of operational parameters, aiding proactive maintenance scheduling. Despite some errors encountered (RMSE: 21.48 and MAE: 18.17), the Mean Absolute Percentage Error (MAPE) of 14.87% indicates relatively minor discrepancies compared to actual values. In conclusion, the Intelligent Model shows significant advancements in pipeline monitoring and maintenance prediction by leveraging machine learning for early anomaly detection and timely maintenance interventions. This proactive approach aims to reduce downtime, prevent environmental damage, and optimize operational efficiency in Nigeria's oil and gas infrastructure. Future research could focus on enhancing system scalability across diverse terrains, employing advanced deep learning techniques such as Transformer Networks and Autoencoders for improved prediction accuracy, and exploring cybersecurity measures like blockchain integration to ensure data integrity and protect critical infrastructure from cyber threats.

Keywords

Pipeline Infrastructure, Machine Learning, Monitoring, Prediction, Random Forest Classifier, Prophet Model

1. Introduction

Monitoring oil and gas pipelines is crucial in Nigeria due to the significant economic and environmental impacts of pipe-

line failures. As Africa's largest crude oil producer, Nigeria relies heavily on oil revenue, which is vital to its economy [1].

*Corresponding author: adetokunbo.johnnotumu@futo.edu.ng (Adetokunbo MacGregor John-Otumu)

Received: 29 June 2024; **Accepted:** 19 July 2024; **Published:** 31 July 2024



Copyright: © The Author(s), 2024. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

However, frequent pipeline vandalism, theft, and accidental leaks pose serious threats to economic stability and environmental sustainability [2]. Nigerian pipelines cross diverse terrains from subsea swamps and rainforests to savannah grasslands exposing them to varying climates, deliberate sabotage, and soil conditions, leading to petroleum leaks that harm host communities and the environment [3].

In the early to mid-1990s, unemployed youths in remote areas vandalized pipelines, siphoning fuel for black market sale, resulting in substantial revenue loss for the government [4]. In February 2007, the Nigerian National Petroleum Corporation (NNPC) estimated a loss of about N10 billion due to pipeline vandalism, including equipment damage and product loss [4]. Illegal oil bunkering causes the Nigerian government an estimated US\$14 billion in annual losses [5]. Within eight weeks, Nigeria lost \$100 million worth of crude oil to theft [6].

Pipeline security in Nigeria's oil and gas industry has faced severe challenges due to vandalism and theft, often by militant groups and oil thieves exploiting the inadequately monitored pipeline network [7]. These activities result in massive financial losses and severe environmental degradation, contaminating soil and water and affecting local communities and wildlife [8]. Traditional pipeline monitoring methods in Nigeria, such as manual patrols and basic surveillance equipment, are outdated and reactive, failing to prevent or quickly respond to breaches [9]. This highlights the need for advanced technologies to safeguard infrastructure. Modern technologies like fiber-optic sensing, drones, and satellite GPS systems have been introduced to enhance pipeline monitoring. For instance, the FALCON Platform uses distributed temperature sensors (DTS) and intelligent distributed acoustic sensors (iDAS) for continuous monitoring and improved detection capabilities [1]. However, Nigeria's challenging terrain and socio-political dynamics often hinder the effective deployment of these technologies [10].

Wireless Sensor Networks (WSNs) have emerged as a promising solution for real-time monitoring of pipeline integrity. WSNs offer a cost-effective and efficient means of detecting leaks and intrusions using low-power, multi-hop communication systems [11]. Despite their potential, WSNs in Nigeria face challenges such as limited bandwidth, inefficiency, and difficulties in real-time monitoring and precise leak localization. The Internet of Things (IoT) provides another innovative approach for monitoring Nigeria's oil pipelines, facilitating remote and automated monitoring with real-time data on pipeline conditions [9]. However, implementing IoT-based solutions requires significant investment in infrastructure and cybersecurity measures to protect critical oil and gas assets from potential cyber-attacks.

Expert systems, developed from AI concepts in the early 1970s, have become crucial innovations in various fields, including the oil and gas industry [12]. They address problems requiring human-like intelligence in fields such as law, biology, engineering, and process control [3]. In the oil and gas

industry, remote real-time monitoring and optimization have become standard operational practices over the past decade [13]. Integrating remote sensing technology or expert systems with software applications for data processing, management, analysis, and understanding helps reduce costs, increase safety, and improve overall performance by quickly identifying and resolving problems [14].

The oil and gas industry faces challenges such as operational scale, remoteness, and the need to process vast volumes of data [4]. Pipelines, which transport oil and natural gas from production sites to refineries and markets, face threats from aging equipment, inadequate maintenance, extreme weather, natural disasters, and deliberate sabotage [15]. Expert monitoring systems are essential for detecting and repairing underground pipeline issues, especially in remote locations. These systems differ from conventional programming techniques and consist of knowledge bases and inference procedures to solve problems [14]. They provide quantitative information from research and interpret qualitative data, addressing imprecise data by assigning confidence values to inputs and conclusions. They can explain their reasoning, enhancing user confidence in recommendations [12].

Developing expert systems involves collaboration between programmers and domain experts, with extensive testing to ensure reliability [14]. In Nigeria, the oil and gas industry is a cornerstone of the national economy. The extensive pipeline network is crucial for transporting oil and gas but is vulnerable to threats like corrosion, mechanical failures, environmental factors, and sabotage. Pipeline leaks or failures lead to severe environmental, economic, and safety consequences, exacerbated by Nigeria's challenging terrains. Current systems lack real-time predictive capabilities needed for proactive risk management. To address these issues, this study employed advanced machine learning techniques by integrating the Random Forest classifier for real-time anomaly detection and the Prophet model for predictive maintenance forecasting. The Random Forest classifier was chosen for its robustness in handling large datasets and its capability to model complex interactions between variables, making it highly effective for detecting anomalies in pipeline data. The Prophet model, developed by Facebook, is a powerful tool for time series forecasting, allowing for accurate predictions of pipeline maintenance needs and facilitating proactive management. The anticipated outcome of this research is the development of an intelligent model for real-time pipeline monitoring, anomaly detection, and maintenance prediction, tailored to Nigeria's unique challenges. This model is expected to: (1) Improve the accuracy and timeliness of anomaly detection in pipelines (2) Enhance the predictive capabilities for maintenance, reducing downtime and preventing failures (3) Provide a comprehensive solution that integrates real-time monitoring with predictive analytics, improving the overall safety and efficiency of pipeline operations.

By leveraging on advanced machine learning techniques, this study aims to contribute significantly to the field of

pipeline monitoring, offering practical solutions to the pressing issues faced by Nigeria's oil and gas industry. This study aims to develop an Intelligent Model for Real-Time Pipeline Monitoring, Anomaly Detection and Maintenance Prediction to address these challenges and safeguard Nigeria's vital oil and gas resources.

2. Related Works

Table 1 summarizes studies emphasizing the urgent requirement for advanced monitoring technologies to address pipeline issues, showcasing a variety of technological strategies.

Table 1. Summary of Related Works.

Techniques Used	Outcome
Proteus 8 [16]	Enhanced spillage detection system for intrusions and vandalism.
Distributed optical fiber sensors, random forest model [17]	Achieved an accuracy of 91.30 % in oil and gas pipeline leak detection with temperature and vibration data.
GSM-based monitoring framework [18]	Advocated for nationwide automated pipeline monitoring to curb vandalism and oil theft.
PIR sensors, vibration sensors, sound sensors [19]	Recommended a multi-sensor system for pipeline monitoring to detect and alert authorities to vandalism.
AI-driven system, neural networks [20]	Explored AI-driven intrusion detection for gas pipelines, showing potential for burst detection.
Remote flow valve activation, hardware design [21]	Designed an intrusion monitoring system with effective leakage detection and hardware emphasis.
Proteus software, differential flow equations [22]	Developed an analytical model for detecting single and multiple leaks in oil pipelines.
Wireless Sensor Networks (WSNs) [11]	Reviewed challenges and solutions for WSN-based oil pipeline monitoring in Nigeria.
Modern intrusion detection technologies [7]	Discussed security threats on pipelines and recommended tailored detection technologies.
Internet of Things (IoT) [9]	Proposed IoT solutions for proactive pipeline monitoring to mitigate vandalism's economic impact.
Automated crack detection, vandalism detection, SMS alerts [2]	Proposed an intelligent system for real-time oil spill detection and remote monitoring.

3. Materials and Methods

3.1. Dataset Description

A dataset description provides an essential overview of collected data, detailing field names, definitions, data types, units, and value ranges. It explains the data source, handling of missing values, and includes example records to illustrate structure. Metadata on data collection methods and preprocessing steps is also covered. This comprehensive description

ensures transparency, aiding users in accurate interpretation, effective analysis, and decision-making, thereby enhancing the dataset's utility and reliability.

Table 2 describes the pipeline anomaly dataset from Kaggle.com, with a size of 107KB, containing 1,152 instances and 5 fields. **Table 3** details the pipeline maintenance dataset, also from Kaggle.com, with a larger size of 1.25MB, comprising 4,977 instances and 24 fields. These datasets provide comprehensive information for analyzing pipeline anomalies and maintenance, crucial for developing effective monitoring and predictive maintenance models.

Table 2. Description of pipeline anomaly dataset.

Dataset Filename	Pipeline_anomaly_dataset.CSV
Dataset Size	107Kb
Source	Kaggle.com
Dataset Instances	1152
No of Fields	5

Table 3. Description of pipeline maintenance dataset.

Dataset Filename	Pipeline_maint_dataset.CSV
Dataset Size	1.25Mb
Dataset Instances	4977
Source	Kaggle.com
No of Fields	24

3.2. Model Design and Training Considerations

Developing an intelligent model for real-time pipeline monitoring and maintenance prediction can leverage the strengths of both the Random Forest classifier and the Prophet model. The Random Forest classifier handles complex, non-linear relationships among operational parameters like pressure, flow rates, temperature, and maintenance records, predicting failures and maintenance needs with high accuracy. Its ensemble learning techniques enhance generalization and robustness. Meanwhile, the Prophet model captures time-dependent patterns, including seasonality and trends in

maintenance schedules, and incorporates external factors like holidays.

Optimizing the Random Forest involves tuning tree depth, number of trees, and feature selection, while Prophet adjustments include seasonality settings and relevant events. Integrating these models provides accurate, proactive monitoring and maintenance predictions, optimizing pipeline reliability and reducing downtime. Table 4 describes in workflow for random forest model, while Table 5 describes the workflow for Prophet Model. Continuous updates and validation against real-time data ensure the model's adaptability and reliability in dynamic conditions.

Table 4. Workflow for Random Forest Model.

Step	Description
Model Initialization	Initialize the Random Forest classifier with specified parameters.
Feature Preparation	Handle missing values using Simple Imputer for numeric features and encode categorical features using Label Encoder within a Pipeline.
Data Splitting	Split the dataset into training and testing sets using train_test_split, with a test size of 20% and a random state of 42.
Model Training	Train the Random Forest classifier (clf) on the training data (X_train, y_train).
Prediction	Predict the target variable (y_pred) and obtain predicted probabilities (y_proba) on the test set (X_test).
Evaluation Metrics Calculation	Calculate and print evaluation metrics: Accuracy, Precision, Recall, F1-Score, and ROC-AUC using functions from sklearn.metrics.

Table 5. Workflow for Prophet Model.

Step	Description
Data Loading	Load the dataset from 'forecast_maint_data.csv'.
Data Preprocessing	Convert the 'date' column to datetime format (pd.to_datetime). Rename columns to 'ds' for timestamp and 'y' for measurements.
Train-Test Split	Split the data into training and testing sets (80% train, 20% test).
Model Initialization	Initialize the Prophet model (model = Prophet()).
Model Training	Train the Prophet model on the training data (model.fit(train)).
Prediction	Generate future date ranges for prediction (future = model.make_future_dataframe(periods=len(test))). Make predictions using the trained model (forecast = model.predict(future)).
Evaluation Metrics Calculation	Calculate Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (mae, mse, rmse, r2).
Visualization - Forecast Plot	Plot the forecasted values (model.plot(forecast)).
Visualization - Components Plot	Plot the components of the forecast (model.plot_components(forecast)).
Visualization - True vs Predicted Plot	Plot true vs predicted values (plt.plot(test['ds'], y_true, label='True'), plt.plot(test['ds'], y_pred, label='Predicted')).

3.3. System Architecture

System architecture is a structured framework detailing software and hardware components, their interactions, and properties. For the Intelligent Model for Real-Time Pipeline Monitoring and Maintenance Prediction, it integrates components to enhance safety and reliability by gathering sensor data, detecting anomalies, forecasting maintenance, and ensuring timely interventions, thus maintaining efficient pipeline operations.

Figure 1 illustrates the framework for an intelligent model aimed at real-time pipeline monitoring, anomaly detection and maintenance prediction. This framework incorporates several components to ensure efficient anomaly detection and accurate maintenance forecasting, enhancing pipeline safety and reliability. Data is initially collected from sensors along the pipeline, capturing parameters like flow rate, temperature, pressure, and vibrations, forming the dataset for model training. Data pre-processing involves cleaning, removing duplicates, and handling missing values to ensure consistency. The dataset is split into training and test sets in an 80/20 ratio for model training and evaluation.

Two machine learning models are trained: the Random Forest (RF) classifier for anomaly detection and the Prophet model for maintenance prediction. The RF classifier learns from labeled historical data to detect anomalies, while the Prophet model forecasts future failures and maintenance needs based on trends and seasonal patterns. The models' performance is evaluated using metrics such as accuracy, recall, precision, and F1-score.

High metric values indicate effective models.

In the prediction phase, the models analyze real-time sensor data. The RF classifier monitors for anomalies, and the Prophet model forecasts maintenance needs. An integrated system triggers alerts and updates stakeholders, ensuring timely interventions. An SMS alert system provides real-time notifications, while a SCADA server manages data communication and integration, ensuring effective pipeline management. This framework enhances pipeline safety by addressing immediate and future issues proactively.

The methodology for developing an intelligent model for real-time pipeline monitoring, anomaly detection, and maintenance prediction is focused on enhancing pipeline safety and reliability through advanced machine learning techniques. This approach addresses the complexities of pipeline operations and the limitations of traditional monitoring methods.

The Random Forest (RF) classifier was selected for anomaly detection due to its ability to manage high-dimensional data and perform well with complex and noisy datasets. Its ensemble nature aggregates decisions from multiple decision trees, increasing accuracy and reducing overfitting risks. The Prophet model was chosen for maintenance prediction because it effectively handles time series data with strong seasonal and trend components. Pipelines exhibit seasonal variations and long-term trends, making Prophet ideal for forecasting future maintenance needs. Its ability to incorporate holidays and significant events enhances the accuracy and context-awareness of its predictions.

Compared to deep learning methods, the RF classifier and Prophet model offer unique advantages. They are more in-

interpretable, providing insights into feature importance, which is crucial for gaining trust from stakeholders and regulatory bodies. They are also more computationally efficient, requiring fewer resources and smaller labeled datasets, making them suitable for real-time applications. Additionally, the Prophet model's design to handle missing data and outliers ensures robust performance in challenging conditions.

To optimize the performance of the RF classifier and Prophet model, several enhancements were implemented. Feature engineering created derived features that capture the

dynamics of pipeline operations, such as moving averages and rate of change for sensor parameters. Hyperparameters of the RF classifier was optimized using grid search and cross-validation to ensure effective anomaly detection. The Prophet model was enhanced with domain-specific knowledge, including known maintenance schedules and significant events, to improve forecast accuracy. Extensive data pre-processing was conducted to clean the sensor data, remove duplicates, and handle missing values, ensuring consistency and reliability in the training and evaluation datasets.

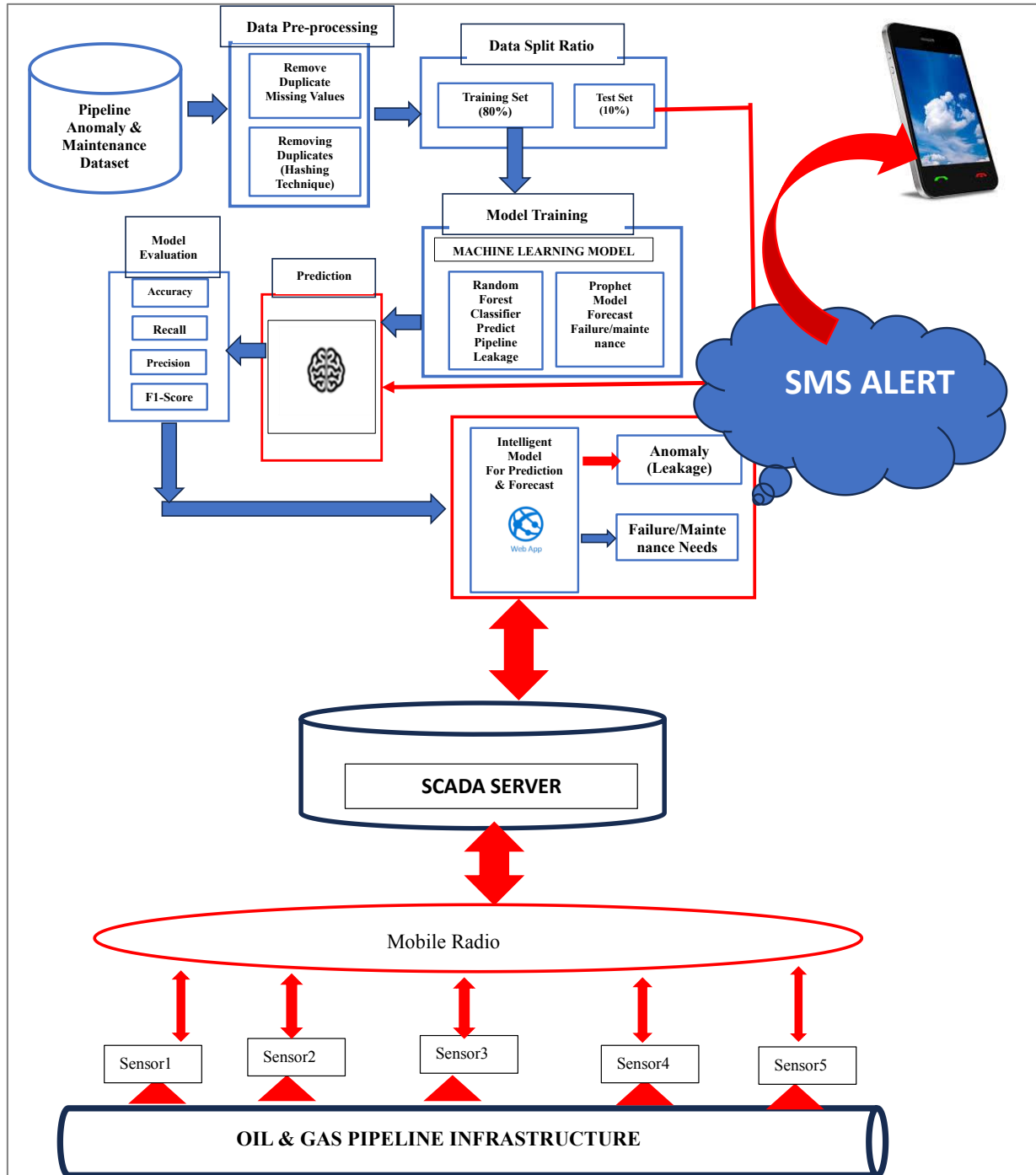


Figure 1. Proposed System Framework.

3.4. Notational Expression for Pipeline Anomaly Detection and Failure/Maintenance Forecast

This section addresses the notational expression that represents the workflow and interactions within the intelligent model for real-time pipeline anomaly detection and maintenance prediction.

1. Data Collection:

Let D represent the dataset collected from pipeline sensors.

$D = \{(t_i, F_i, T_i, P_i, V_i, L_i)\}$ where t_i is the timestamp, F_i is the flow rate, T_i is the temperature, P_i is the pressure, V_i is the vibration, and L_i indicates if an anomaly (leak) is present.

2. Data Pre-processing:

Clean and pre-process dataset D .

Remove duplicates: $D_{clean} = \text{remove_duplicates}(D)$

Handle missing values: $D_{processed} = \text{impute_missing}(D_{clean})$

3. Dataset Splitting:

Split $D_{processed}$ into training and test sets.

$D_{train} = 0.8 \times D_{processed}$

$D_{test} = 0.2 \times D_{processed}$

4. Model Training:

Train Random Forest (RF) classifier RF for anomaly detection.

Train Prophet model PM for maintenance prediction.

$RF \leftarrow \text{train}(D_{train})$

$PM \leftarrow \text{train}(D_{train})$

5. Performance Evaluation:

Evaluate models using metrics such as accuracy (A), recall (R), precision (P), and F1-score (F1).

$A_{RF}, R_{RF}, P_{RF}, F1_{RF} = \text{evaluate}(RF, D_{test})$

$A_{PM}, R_{PM}, P_{PM}, F1_{PM} = \text{evaluate}(PM, D_{test})$

6. Prediction Phase:

Apply models to real-time data D_{real}

Anomaly detection: $A_{score} = RF(D_{real})$

Maintenance forecast: $M_{forecast} = PM(D_{real})$

7. Alert System:

If $A_{score} > \text{threshold}$, trigger anomaly alert.

If $M_{forecast}$ predicts maintenance need, notify stakeholders.

$\text{SMS_alert}(A_{score}, M_{forecast})$

8. Integration and Communication:

SCADA server integrates real-time data, model outputs, and communication.

$\text{SCADA} = \text{integrate}(D_{real}, A_{score}, M_{forecast})$

9. Pipeline Monitoring:

Continuous data flow from sensors to SCADA server.

$S = \{S1, S2, S3, S4, S5\}$

$\text{Monitor}(S) \rightarrow \text{SCADA} \rightarrow RF, PM \rightarrow \text{Alert}$

3.5. Deployment Specifications

Tables 6, 7, and 8 provides comprehensive details on the deployment specifications for pipeline monitoring systems. Table 6 covers the necessary computer specifications. Table 7 outlines key sensor specifications, including types, measurement ranges, accuracy, and environmental conditions for Distributed Fiber Optic Sensors (DFOS). Table 8 describes sensor configuration aspects, such as system components, installation procedures, data acquisition and processing, and maintenance requirements.

Table 6. Computer Specification.

Component	Recommendation
CPU	Intel Core i7
GPU	NVIDIA RTX 3080
RAM	16GB DDR4
Storage	1TB SSD
Operating System	Windows 10 Pro
GPU Driver	NVIDIA CUDA Toolkit 11.0
Deep Learning Framework	TensorFlow, Keras
Python Distribution	Python 3.9.7
IDE	Jupyter Notebook, PyCharm, Anaconda 3
Web Interface Development	JavaScript, HTML
Local Hosting Platform	Flask

Table 7. Key Sensors Specifications.

Specification	Details
Sensor Type	Distributed Fiber Optic Sensors (DFOS) Example: DTS, DAS, DSS Temperature: -200 °C to +850 °C
Measurement Range	Pressure: Up to 700 bar (10,000 psi) Strain: $\pm 10,000 \mu\epsilon$ (micro-strains) DTS: 0.1 to 1 meter
Spatial Resolution	DAS: 1 to 10 meters DSS: 0.1 to 1 meter Temperature: ± 0.1 °C to ± 1 °C
Measurement Accuracy	Pressure: $\pm 0.1\%$ of Full Scale Strain: $\pm 1\%$ of measurement
Response Time	Real-time data acquisition with latency typically less than 1 second
Sensor Length	Up to 50 kilometers per sensing unit, extendable beyond 100 kilometers Temperature Range: -40 °C to +85 °C for the sensing unit
Operating Environment	Humidity: 0% to 95% non-condensing Environmental Protection: IP67 or higher for field-deployable units

Table 8. Sensor Configuration.

Configuration Aspect	Details
System Components	Interrogator Unit: Processes and analyzes light pulses Fiber Optic Cable: Single-mode or multi-mode, ruggedized for harsh environments Connectors and Splices: Ensure minimal signal loss Power Supply: 12V DC or 24V DC, with AC options Deployment: Fiber optic cables deployed alongside or within pipelines
Installation	Protection: Armored or in protective conduits to prevent damage Calibration: Initial and periodic recalibration for accuracy Software: Advanced solutions for data interpretation, visualization, alarms
Data Acquisition and Processing	Integration: Compatible with SCADA and other monitoring systems Machine Learning: Analyzes patterns to predict leaks or structural issues Durability: High durability with minimal maintenance
Maintenance and Reliability	Diagnostics: Continuous system diagnostics Redundancy: Redundant systems for critical applications

3.6. Model Evaluation

Evaluating the model in the context of pipeline monitoring is vital for gauging the performance and reliability of the machine learning models employed in the Intelligent Model for Real-Time Pipeline Anomaly Detection and Maintenance Forecast. Key metrics in this evaluation include accuracy, precision, recall, F1-score, Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Each of these metrics offers valuable insights into the model's effectiveness.

Accuracy measures the proportion of correct predictions made by the model out of all predictions. In the context of pipeline monitoring, high accuracy indicates that the model reliably detects both anomalies (such as leaks) and normal operating conditions. The accuracy is represented in equation (1):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision evaluates the proportion of true positive predictions among all positive predictions. For pipeline anomaly detection, high precision means that when the model predicts an anomaly, it is likely correct, minimizing false alarms and ensuring maintenance resources are efficiently used. The precision is represented in equation (2):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall (or sensitivity) measures the proportion of actual positives correctly identified by the model. In this context, high recall ensures that most actual anomalies are detected, which is critical for preventing pipeline failures and ensuring safety. The recall or sensitivity is represented in equation (3):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

F1-Score is the harmonic mean of precision and recall, providing a balanced measure that considers both false posi-

tives and false negatives. A high F1-score indicates that the model maintains a good balance between precision and recall, making it robust for real-time monitoring and prediction. The F1-score is represented in equation (4):

$$\text{F1-score} = 2 * \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (4)$$

The MAPE measures forecast accuracy as a percentage. It reflects how much the forecast deviates from actual values, on average. It is calculated by averaging the absolute percentage errors for each prediction period. The MAPE is represented in equation (5):

$$\frac{1}{n} \sum_{t=1}^n \frac{|Pt - At|}{At} \quad (5)$$

The MAE gauges a forecasting model's accuracy by averaging the absolute differences between predicted and actual values. It disregards the direction of errors (positive or negative) and focuses on their magnitude. A lower MAE indicates a model with predictions generally closer to reality. The MAE is represented in equation (6):

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |Pt - At| \quad (6)$$

The RMSE measures prediction error. It calculates the average squared difference between forecasts and actual values, then takes the square root. This metric helps compare models for the same data, but depends on the measurement scale. The RMSE is represented in equation (7):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Pt - At)^2} \quad (7)$$

4. Results

This section details the results derived from the experimental study conducted.

Table 9. Summary of Prophet Model Forecast Results for Maintenance Needs.

ds	yhat	yhat_lower	yhat_upper
1/1/2024 0:00	101.477612	95.025488	108.17543
1/1/2024 1:00	101.293297	95.228491	107.600144
1/1/2024 2:00	101.108981	94.439863	106.662997
1/1/2024 3:00	100.924665	94.312921	107.452197
1/1/2024 4:00	100.740349	94.462473	106.603424
1/1/2024 5:00	100.556033	93.556064	106.768901
1/1/2024 6:00	100.371718	93.775845	106.39061
1/1/2024 7:00	100.187402	93.587873	106.646444

ds	yhat	yhat_lower	yhat_upper
1/1/2024 8:00	100.003086	93.836199	106.221952
1/1/2024 9:00	99.81877	93.806844	106.230431
1/1/2024 10:00	99.634455	93.308373	105.852675
1/1/2024 11:00	99.450139	93.158037	105.734058
1/1/2024 12:00	99.265823	92.735753	106.344609
1/1/2024 13:00	99.081507	92.59336	105.499479
1/1/2024 14:00	98.897191	92.219593	105.351828
1/1/2024 15:00	98.712876	92.338281	105.149699
1/1/2024 16:00	98.52856	92.170087	104.945244
1/1/2024 17:00	98.344244	91.78075	104.979772
1/1/2024 18:00	98.159928	91.465187	104.733737
1/1/2024 19:00	97.975613	91.22553	104.455174
1/1/2024 20:00	97.791297	91.578203	103.7677
1/1/2024 21:00	97.606981	91.139953	104.23382
1/1/2024 22:00	97.422665	90.960262	103.640478
1/1/2024 23:00	97.238349	91.006873	103.397636
1/2/2024 0:00	97.054034	90.790693	103.321047
1/2/2024 1:00	96.869718	90.206936	102.498589
1/2/2024 2:00	96.685402	90.177687	102.880591
1/2/2024 3:00	96.501086	90.137298	102.559735
1/2/2024 4:00	96.31677	89.788627	102.909365
1/2/2024 5:00	96.132455	89.724287	102.595168
1/2/2024 6:00	95.948139	89.128862	102.368184
1/2/2024 7:00	95.763823	89.762832	102.208097
1/2/2024 8:00	95.579507	89.574003	102.02685
1/2/2024 9:00	95.395192	89.189907	101.30719

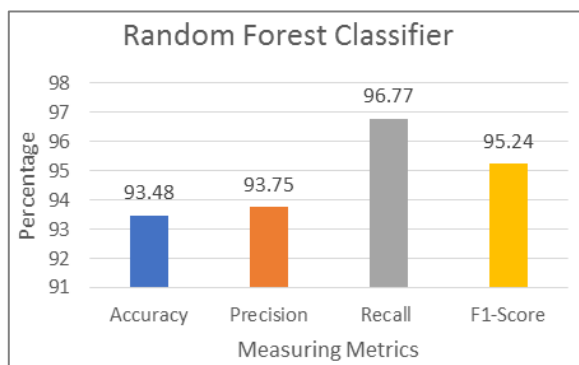


Figure 2. Random Forest Classifier Result for Pipeline Anomaly (Leakage) Detection.

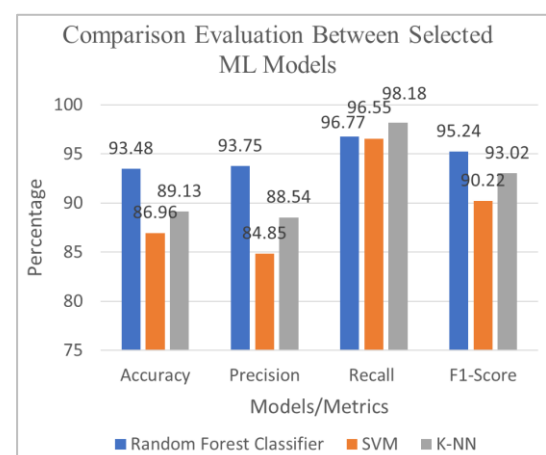


Figure 3. Representation of ML Models Comparison.

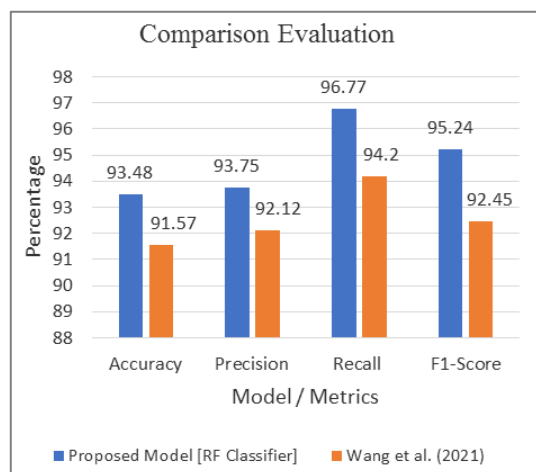


Figure 4. Graphical Representation of Comparison Evaluation with Previous Study.

Table 10. Model Evaluation of Prophet Model.

RMSE	21.48
MAE	18.17
MAPE	14.87%

5. Discussion

Figure 2 illustrates the performance metrics of the Random Forest Classifier for detecting pipeline anomalies, specifically leakages, using a test dataset of 230 instances. The classifier achieved an overall accuracy of 93.48%, significantly higher than the 91.30% accuracy reported in previous research [17]. The precision rate of 93.75% indicates the classifier's high accuracy in identifying true positives while predicting pipeline anomalies, thereby reducing false positives. The recall rate of 96.77% demonstrates the classifier's effectiveness in detecting actual anomalies, ensuring most true anomalies are identified. The F1-score of 95.24%, which balances precision and recall, highlights the classifier's robustness in both identifying and reporting pipeline issues accurately.

Figure 3 compares the performance of three machine learning models selected; Random Forest Classifier, SVM, and K-NN for detecting pipeline leakage anomalies. The Random Forest Classifier outperformed the other models, achieving an accuracy of 93.48%, precision of 93.75%, recall of 96.77%, and an F1-score of 95.24%. The SVM model, while competitive, achieved an accuracy of 86.96%, precision of 84.85%, recall of 96.55%, and an F1-score of 90.22%. The lower precision of the SVM suggests it may produce more false positives compared to the Random Forest. K-NN demonstrated a strong recall of 98.18%, indicating its ability to identify actual positive instances, but its overall accuracy and precision were lower than those of the Random Forest. In essence, the Random Forest Classifier emerged as the most

balanced and effective model for developing an intelligent system for real-time pipeline monitoring and maintenance prediction.

The high performance of the Random Forest Classifier is consistent with findings from related studies [16, 18, 21, 22]. The comparison in Figure 4 with previous literature further highlights the superiority of the proposed Random Forest model, which outperforms the approach by Salihu *et al.* in all evaluated metrics.

Tables 9 and 10 provide insights into the performance of the Prophet model for predicting pipeline maintenance needs. The model's ability to forecast hourly predictions with relatively narrow uncertainty intervals enhances real-time monitoring and proactive maintenance scheduling. The RMSE of 21.48 and MAE of 18.17 indicate a reasonable error level, while the MAPE of 14.87% shows that the predictions are quite close to actual values. These results suggest that the Prophet model can be effectively integrated into a real-time system to predict maintenance needs, potentially preventing costly failures and minimizing downtime.

While this study demonstrates the effectiveness of the Random Forest and Prophet models for pipeline monitoring and maintenance prediction, several avenues for future research can be explored:

1. **Hybrid Models:** Future research could focus on developing hybrid models that combine the strengths of multiple machine learning techniques to improve predictive accuracy and robustness.
2. **Deep Learning Approaches:** Investigating the application of deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), could further enhance anomaly detection and predictive maintenance capabilities.
3. **Real-Time Adaptation:** Developing adaptive models that continuously learn from real-time data and adjust their predictions accordingly can improve the system's responsiveness to changing conditions.
4. **Comprehensive Sensor Networks:** Expanding the sensor network to include additional types of sensors (e.g., acoustic, electromagnetic) and enhancing data integration techniques could provide a more comprehensive monitoring solution.
5. **Cybersecurity Measures:** As IoT and advanced monitoring technologies are increasingly used, ensuring robust cybersecurity measures to protect the integrity and confidentiality of pipeline data is essential.
6. **Field Trials and Validation:** Conducting extensive field trials and validation studies in various operational environments can provide valuable insights into the models' real-world performance and help fine-tune their parameters.

6. Conclusions

This study emphasizes the creation and implementation of

an intelligent model for real-time pipeline monitoring, anomaly detection, and maintenance prediction, with a focus on the Nigerian oil and gas sector. Utilizing advanced machine learning techniques, specifically the Random Forest classifier and Prophet model, this research addresses critical challenges in pipeline monitoring and maintenance, ensuring enhanced safety, reliability, and operational efficiency.

The Random Forest classifier has shown outstanding performance in detecting pipeline anomalies, demonstrating its efficacy in accurately identifying true anomalies while minimizing false positives and ensuring most actual anomalies are detected. Its balanced performance further attests to the model's robustness and reliability in real-time anomaly detection, making it a superior choice compared to other models like SVM and K-NN.

Complementing the anomaly detection capabilities of the Random Forest classifier, the Prophet model effectively forecasts pipeline maintenance needs. The model's ability to provide timely predictions with narrow uncertainty intervals facilitates proactive maintenance scheduling. Performance metrics indicate a reasonable level of accuracy in the predictions, demonstrating the model's potential to prevent costly failures and minimize downtime through timely maintenance interventions.

Integrating these models into a comprehensive framework for real-time pipeline monitoring and maintenance prediction marks a significant advancement in the field. By leveraging real-time data from various sensors, the system enhances the ability to detect anomalies and predict maintenance needs, ensuring timely responses to potential issues. This proactive approach not only mitigates the risk of pipeline failures but also enhances overall operational efficiency, contributing to the economic stability and environmental sustainability of Nigeria's vital oil and gas industry.

The study recognizes the importance of exploring future research directions to further refine and enhance the proposed models. Potential areas of investigation include developing hybrid models that combine multiple machine learning techniques, applying deep learning approaches like CNNs and RNNs, and creating adaptive models capable of learning from real-time data. Additionally, expanding the sensor network to include diverse types of sensors and ensuring robust cybersecurity measures are crucial for comprehensive and secure pipeline monitoring.

Abbreviations

BMI	Body Mass Index
PM	Pipeline Monitoring
OGI	Oil and Gas Industry
PV	Pipeline Vandalism
EI	Economic Impact
ES	Environmental Sustainability
RL	Revenue Loss
IOB	Illegal Oil Bunkering

SC	Security Challenges
AT	Advanced Technologies
FOS	Fiber-Optic Sensing
WSN	Wireless Sensor Networks
IOT	Internet of Things
ES	Expert Systems
RST	Remote Sensing Technology
RTM	Real-Time Monitoring
PC	Predictive Capabilities
PRM	Proactive Risk Management

Acknowledgments

The authors express their gratitude for the support and assistance provided by the entire team at the Production Monitoring Centre (PMC) of Shell Petroleum Development Company (SPDC) in Port Harcourt, Nigeria.

Author Contributions

Obi Chukwuemeka Nwokonkwo: Conceptualization, Resources, Writing – review & editing.

Nwankwo Uchechukwu Samuel: Conceptualization, Data curation, Investigation, Methodology, Visualization, Methodology, Writing – original draft,

Udoka Felista Eze: Formal Analysis, Investigation, Supervision, Writing – review & editing.

Adetokunbo MacGregor John-Otumu: Data curation, Investigation, Methodology, Resources, Visualization, Writing – review & editing.

Funding

This work is not supported by any external funding.

Data Availability Statement

The data supporting the outcome of this research work has been reported in this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Fagbami, D., Echem, C., Okoli, A., Mondanos, M., Bain, A., Carbonneau, P., and Amarquaye M. A Practical Application of Pipeline Surveillance and Intrusion Monitoring System in the Niger Delta: The Umugini Case Study. In Proceedings at the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, July 2017.
<https://doi.org/10.2118/189056-MS>

- [2] Igbajar, A. and Barikpoa, A. N. Designing an Intelligent Microcontroller based Pipeline Monitoring System with Alarm, Sensor, International Journal of Emerging Technologies in Engineering Research (IJETER), 2015, 3 (2): 22-27.
- [3] Agbaeze, K. N. Petroleum Pipeline Leakages in PPMC' Report for Chief Officers Mandatory Course 026, Lagos, 2002.
- [4] Yiu, C. S., Grant, K., & Edgar, D. Factors affecting the adoption of Internet Banking in Hong Kong—implications for the banking sector. International journal of information management, 2007, 27 (5), 336-351.
- [5] Ikokwu, T. Oil Spill Management in Nigeria: Challenges of Pipeline Vandalism in the Niger Delta Region of Nigeria, 2007, 16.
- [6] Ojiaku et al. Pipeline Vandalisation Detection Alert with SMS, Int. Journal of Engineering Research and Applications, 2014, 4 (9). pp. 21.
- [7] Achilike C. M. N. Securing Nigeria's Crude Oil and Gas Pipelines –Change in Current Approach and Focus on the Future, Scientific Research Journal (SCIRJ), 2017, 5 (1): 1-9.
- [8] Oseni, W. Y., Akangbe, O. A., & Abhulimen, K. Mathematical modelling and simulation of leak detection system in crude oil pipeline. Heliyon, 2023, 9 (4), e15412.
<https://doi.org/10.1016/j.heliyon.2023.e15412>
- [9] Ayeni, T. P., and Ayogu, B. A. Intelligent Pipeline Monitoring System Based on Internet of Things, Scientific Research Journal (SCIRJ), 2020, 8 (8): 44-49.
- [10] Idachaba, F., Wokoma, E., Okuns, G., Brown, C., and Ian W. Fiber Optic Based Pipeline Oil and Gas Leak and Intruder Detection System with Security Intervention Plan. In Proceeding at the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, August 2013.
<https://doi.org/10.2118/167537-MS>
- [11] Ezeja, O. M. and Ahaneku, M. A. Review on the Use of Wireless Sensor Network Systems for Oil Pipeline Surveillance, 2020 LGT-ECE-UNN International Conference: Technological Innovation for Holistic Sustainable Development (TECHISD2020), 2020, 137-145.
- [12] Muggleton, M. J. and Brennan, Linford, P. W. Axisymmetric wave propagation in fluid-filled pipes: wavenumber measurements in in vacuo and buried pipes, Journal of Sound and Vibration, 2004, 270(1), 171-190.
[https://doi.org/10.1016/S0022-460X\(03\)00489-9](https://doi.org/10.1016/S0022-460X(03)00489-9)
- [13] Adebayo S. HDD Technology takes pipeline across 1.7km Escravos River, 'Fortune Business, 2010.
<http://www.nigerianbestforum.com/general>
- [14] Ajakaiye, B.A Combating Oil spill in Nigeria: Primary role and responsibility of the National Oil Spill Detection and Response Agency (NOSDRA) Consultative workshop August 4 -6, 2008, Calabar, Nigeria, 2008.
- [15] Okorodudu, O. F., Okorodudu, P. O., & Irikefe, E. K. A Model of Petroleum Pipeline Spillage Detection System for use in the Niger Delta Region of Nigeria. International Journal of Research - Granthaalayah, 2016, 4 (12), 1-16.
<https://doi.org/10.5281/zenodo.221441>
- [16] Wang F., Zhen Liu, Z., Zhou X., Li S., Yuan X., Zhang Y., Shao L., & Zhang X, (INVITED) Oil and Gas Pipeline Leakage Recognition Based on Distributed Vibration and Temperature Information Fusion, Results in Optics, 2021, 5 (1): 1-9.
<https://doi.org/10.1016/j.rio.2021>
- [17] Olaiya, O. O., Oduntan, O. E., and Ehiagwina, F. O. A Short Overview of Pipeline Monitoring Technologies for Vandalism Prevention with a Proposed Framework for a GSM Based System, Journal of Engineering & Research Tech, 2020, 13 (5): 175-188.
- [18] Salihu, O. A., Agbo, I. O., Saidu M., and Onwuka, E. N. Multi-Sensor Approach for Monitoring Pipelines, International Journal of Engineering and Manufacturing, 2017, 6, 59-72.
<https://doi.org/10.5815/ijem.2017.06.06>
- [19] Umbre, S., and Gaikwad S. Practicality of Gas Pipeline Inspection System, International Journal of Research Publication and Reviews, 2024, 5 (3): 125-131.
- [20] Okpo, N. C., Itaketo, U. T., and Udofia, K. Analytical Model of Leak Detection and Location for Application in Oil Pipeline Intrusion Detection System, Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2023, 10 (1): 16029-16038.
- [21] Okpo, N. C., Udofia, K. M. and Friday, S. A. Design of Oil Pipeline Intrusion Monitoring System with GSM Module-Based Remote Flow Valve Activation Mechanism, Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2023, 10 (1): 15999-16007.
- [22] McAllister, W. Pipeline Rules of Thumb Handbook: Quick and Accurate Solutions to your everyday Pipeline Problems, 6th Edition, 2005.