World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

ECONOMIC, OPERATIONAL AND ENVIRONMENTAL ASSESSMENT OF MACHINE LEARNING-DRIVEN ASSET INTEGRITY MANAGEMENT: A COMPARATIVE STUDY OF PREDICTIVE AND TRADITIONAL MAINTENANCE STRATEGIES IN A GAS PROCESSING PLANT

Ayodeji Adewale Aribisala [1,*], Igwe Ikechi [1], Fidelis Onuoha Wopara [1] and Bayo Balogun [2]

[1] Department of Petroleum Engineering, Rivers State University, Port Harcourt, Nigeria.
[2] Renaissance Africa Energy Company Limited, Port Harcourt, Nigeria.
* Corresponding Author

ORCID Details
Ayodeji Adewale Aribisala: https://orcid.org/0009-0003-4303-6193
Igwe Ikechi: https://orcid.org/0000-0002-9298-8305
Fidelis Onuoha Wopara: https://orcid.org/0000-0002-4581-6463
Bayo Balogun: https://orcid.org/0009-0005-9419-2828

World Journal of Advanced Engineering Technology and Sciences, 2026, 21(02), 244–254

Publication history: Received on 09 October 2026; revised on 17 November 2026; accepted on 19 November 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.21.2.0504

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Gas processing facilities operating in the Niger Delta region of Nigeria face persistent asset integrity challenges associated with equipment degradation, corrosion, process instability, high humidity, flooding and ageing infrastructure. Conventional maintenance strategies, particularly reactive maintenance and periodic inspection, may not adequately address changing equipment conditions or provide sufficient warning of impending failures. This study evaluates the potential economic, operational and environmental benefits of integrating machine learning (ML) into asset integrity management, using the Adex Gas Processing Plant in the Niger Delta as a case study.

Historical operational and maintenance data were integrated from supervisory control and data acquisition (SCADA) systems, computerised maintenance management records, well performance information and environmental datasets. The analytical framework incorporated Random Forest classification, Isolation Forest anomaly detection, Kaplan–Meier survival analysis and Cox proportional hazards modelling. A maintenance optimisation model was also developed to compare the cost and operational implications of reactive and ML-informed predictive maintenance strategies.

The reconstructed test-set assessment comprised 2,920 observations, including 1,109 confirmed failure events. The reconstructed confusion matrix contained 894 true positives, 215 false negatives, 498 false positives and 1,313 true negatives. These values correspond to an accuracy of 75.58%, precision of 64.22%, recall of 80.61%, specificity of 72.50% and F1-score of 71.49%. The comparative economic scenario estimated that maintenance costs could decrease from US$301,560,000 under the reactive baseline to US$71,950,650 under the ML-informed predictive maintenance scenario, representing a gross estimated saving of US$229,609,350 (76.14%). Estimated maintenance downtime decreased from 54,272 to 18,552 hours, representing a reduction of 65.82%. Using an assumed emission factor of 100

tCO_2e per avoided shutdown, the 894 true-positive detections corresponded to a scenario estimate of 89,400 tCO_2e in potentially avoided emissions.

The findings indicate that integrating machine learning with maintenance planning can provide a quantitative basis for prioritising equipment interventions, reducing exposure to emergency maintenance costs and improving environmental risk management. However, the reported economic and environmental benefits are model-based estimates and require prospective operational validation. The study demonstrates the potential value of an integrated, data-driven asset integrity management framework for gas processing facilities operating under challenging environmental and operational conditions.

Keywords: Machine Learning; Predictive Maintenance; Maintenance Optimization; Economic Assessment; Downtime Reduction; Environmental Sustainability.

1 INTRODUCTION

Asset integrity management (AIM) is essential to the safe, reliable and sustainable operation of oil and gas production and processing facilities. Gas processing plants contain interconnected equipment and systems, including separators, compressors, heat exchangers, pumps, pipelines, pressure vessels, control valves and instrumentation systems. Failure of any critical component can affect production continuity, equipment availability, personnel safety and environmental performance. The consequences may include unplanned shutdowns, emergency maintenance expenditure, hydrocarbon releases and increased emissions. Effective asset management therefore requires a systematic approach to managing asset risks and performance throughout the operational life of facilities [1].

Traditional asset integrity management practices commonly combine scheduled inspections, preventive maintenance, condition monitoring and corrective maintenance. These approaches remain important because they provide structured mechanisms for identifying equipment deterioration and maintaining compliance with engineering and operational requirements. Risk-based inspection provides a systematic approach to prioritising inspection activities according to the probability and consequences of equipment failure [2]. However, fixed inspection intervals and reactive responses may not always reflect the actual condition of equipment or the changing risk profile of operating assets. Where equipment degradation develops between inspection intervals, maintenance personnel may receive insufficient warning to plan interventions before functional failure occurs.

These challenges are particularly relevant to gas processing facilities in the Niger Delta region of Nigeria. The operating environment is characterised by high humidity, seasonal rainfall, flooding, saltwater intrusion in some locations and other environmental conditions that may contribute to corrosion and infrastructure deterioration. Research on climate-related risks in the Niger Delta highlights the vulnerability of the region to extreme weather events and associated environmental pressures [3]. In addition, the complexity of gas processing operations requires continuous monitoring of process conditions, equipment performance and maintenance history. The availability of operational data provides an opportunity to improve conventional asset integrity practices through advanced analytical techniques.

Machine learning offers the capability to identify patterns in historical operational and maintenance data, classify potential failure conditions, detect abnormal operating behaviour and estimate equipment failure risk. Previous research has demonstrated the application of machine learning techniques in predictive maintenance, particularly in equipment fault identification and maintenance decision support [4]. When combined with engineering knowledge and maintenance planning, these capabilities can support a transition from predominantly time-based or reactive interventions towards condition-informed and risk-based maintenance decisions.

Unplanned equipment failures generate significant costs beyond direct repairs, including production losses, emergency logistics, and environmental liabilities. While predictive maintenance mitigates these expenses through earlier risk-based interventions, it also introduces implementation, infrastructure, and false-alarm costs. Consequently, its economic value must be evaluated against a defined maintenance baseline rather than assumed from model accuracy alone. Furthermore, equipment failures and process upsets drive hydrocarbon releases and emergency flaring, a critical environmental concern highlighted in global gas flaring assessments [5]. Earlier detection can reduce these emissions, provided benefits are quantified using transparent assumptions.

However, few case-based studies connect predictive performance with maintenance costs, downtime, and environmental outcomes in gas processing facilities within developing oil and gas regions. This study addresses this gap by evaluating an integrated machine learning (ML)-based asset integrity management framework against a traditional reactive maintenance baseline at the Adex Gas Processing Plant in Nigeria's Niger Delta. Specifically, this research develops ML models for failure prediction, anomaly detection, and survival analysis, while assessing their potential reductions in maintenance expenditure, operational downtime, and environmental emissions.

2 MATERIALS AND METHODS

2.1 Study area and operational context

The study was conducted using historical operational and maintenance data from the Adex Gas Processing Plant in the Niger Delta region of Nigeria. The facility processes associated and other natural gas streams through interconnected processing and utility systems. Its operations depend on the availability and integrity of process equipment, instrumentation, rotating machinery, pipelines and supporting infrastructure.

The environmental and operational conditions of the Niger Delta provide an important context for asset integrity management. High humidity, seasonal rainfall, flooding and possible saltwater exposure may increase environmental stresses on exposed equipment and infrastructure. These conditions were considered through environmental risk indicators incorporated into the analytical framework. The study used existing facility operating information and equipment operating-envelope records as contextual references for interpreting process and equipment data. The study did not independently recalculate the thermodynamic phase envelopes or equipment design capacities.

2.2 Data sources and preparation

Historical data covering ten years were obtained from multiple facility information sources. These included SCADA process measurements, corrective and preventive maintenance records, well performance information, equipment records, chemical treatment information and relevant environmental data. The maintenance dataset contained 3,120 corrective maintenance records classified into ten equipment categories. The SCADA dataset included 30 primary process measurement tags sampled at six-hour intervals for the real-time monitoring and anomaly detection layer. The integrated modelling dataset also contained engineered variables derived from historical sensor measurements and maintenance information.

Data preparation involved timestamp standardisation, duplicate removal, equipment-status encoding, sensor validity checks and temporal alignment of the different data sources. Maintenance work orders were structured to establish equipment failure events and their relationship with operational measurements. Features were developed from the historical data to represent equipment condition and operating behaviour. These included rolling statistical measurements, rates of change, lag variables, historical failure frequency, time since preventive maintenance, calendar indicators and environmental risk indicators. Historical failure counts over 30-day and 90-day periods were included to represent recurring equipment problems.

The feature engineering process expanded the original variables into a larger analytical feature set. The final modelling configuration reported 167 input variables for the overall failure prediction model. The engineered variables were intended to improve the representation of equipment condition, temporal behaviour and maintenance history.

2.3 Machine learning model development

The predictive framework incorporated supervised classification, unsupervised anomaly detection and survival analysis.

2.3.1 Supervised classification

Random Forest (RF) was developed for equipment failure prediction. Random Forest was selected because of its ensemble structure and its ability to capture nonlinear relationships between process measurements, maintenance history and failure outcomes [6]. Class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) on the training data [7]. A time-based 80:20 training and testing arrangement was adopted, with the training period ending before the test period. This arrangement was intended to provide a more realistic assessment of model performance on subsequent operational observations.

The held-out test dataset contained 2,920 observations. The models were evaluated using accuracy, precision, recall and F1-score. Particular attention was given to recall because undetected equipment failures may have significant operational and safety implications. Mechanism-specific Random Forest classifiers were also developed for ten degradation or failure categories.

2.3.2 Anomaly detection

Isolation Forest was used as an unsupervised method for identifying unusual patterns in process sensor measurements. The algorithm identifies observations that differ from the learned normal operating behaviour through an isolation-based approach [8]. The anomaly detection layer was designed to complement supervised failure prediction by identifying unusual sensor patterns that may require further engineering investigation. An anomaly score was used to

support monitoring and alert classification. An identified anomaly was not automatically treated as a confirmed equipment failure. Instead, it represented an indication of abnormal behaviour requiring assessment against process conditions, equipment history and operational knowledge.

2.3.3 Survival analysis

Kaplan–Meier survival analysis and Cox proportional hazards modelling were applied to estimate time-to-failure behaviour for selected equipment categories. The Kaplan–Meier method provided non-parametric survival estimates [9], while the Cox model examined the relationship between selected sensor covariates and equipment failure risk [10].

The survival models were intended to complement the classification framework by providing additional information about the timing of potential failure events. This information can assist maintenance planners in prioritising interventions where equipment condition and failure history indicate increased risk.

2.4 Maintenance optimisation and operational assessment

The maintenance planning component translated predicted equipment risk into intervention priorities. A mathematical maintenance optimisation model was developed to minimise expected maintenance-related costs over a 30-day planning horizon, subject to maintenance resource and scheduling constraints. The model considered maintenance decisions across ten equipment categories. The decision framework incorporated predicted failure risk, planned maintenance expenditure, emergency repair implications and available maintenance capacity.

A risk threshold was used to determine when equipment should be prioritised for intervention. The optimisation model also examined the sensitivity of maintenance cost to changes in the risk threshold. The reported optimal threshold of $\theta = 0.10$ represented the minimum-cost solution within the tested model configuration. This value should be interpreted as a model-specific result rather than a universal operating threshold for gas processing facilities. Operational impact was evaluated through estimated failure detection, maintenance downtime and the potential reduction in unplanned shutdown events. The comparison used a reactive maintenance baseline and an ML-informed predictive maintenance scenario.

2.5 Economic assessment

The economic assessment used actual maintenance cost records from the facility to derive representative cost parameters. Of the available maintenance work orders, 379 contained non-zero actual cost values and were used to establish the cost distribution for the economic model.

Five cost parameters were derived using percentile-based estimates:

Unplanned shutdown cost: 90th percentile.
Emergency repair cost: 75th percentile.
Preventive maintenance cost: median.
False-alarm investigation cost: 25th percentile.
The resulting cost parameters are presented in Table 1.

Table 1: Facility-derived economic cost parameters

Cost parameter	Estimated value
Unplanned shutdown cost	US$134,100
Emergency repair cost	US$27,700
Preventive maintenance cost	US$7,370
False-alarm investigation cost	US$2,110
Environmental damage cost proxy	US$24,592,700
Downtime cost per hour	US$577

The reactive maintenance scenario assumed that confirmed failures in the test dataset resulted in unplanned shutdowns and emergency repairs. The predictive maintenance scenario incorporated model detection outcomes, planned interventions and false-alarm investigation costs. The analysis used a one-time implementation capital expenditure (CAPEX) of US$650,000 and an annual operating expenditure (OPEX) of US$112,000 for monitoring and model

retraining. These investment costs were considered separately from the comparative maintenance scenario totals and should be included in a complete net-present-value and return-on-investment assessment.

Calculation of Maintenance Cost Savings

The gross maintenance saving was calculated as the difference between the estimated cost of reactive maintenance and the estimated cost of the ML-informed predictive maintenance scenario:

$$S_g = C_r - C_p \quad (1)$$

where:

S_g = Gross estimated maintenance saving (US$);
C_r = Estimated reactive maintenance cost (US$);
C_p= Estimated predictive maintenance scenario cost (US$).

The percentage reduction in maintenance cost was calculated as:

$$R_c = \left(\frac{(C_r - C_p)}{C_r}\right) * 100 \quad (2)$$

where R_c represents the percentage reduction in maintenance cost.

To account for the initial implementation investment and first-year operating expenditure, the indicative first-year adjusted saving was calculated as:

$$S_n = S_g - CAPEX - OPEX \quad (3)$$

where:

S_n = Indicative first-year adjusted saving (US$);

$CAPEX$ = One-time implementation capital expenditure (US$);

$OPEX$ = First-year operating expenditure (US$).

2.6 Environmental assessment

Environmental impact was assessed by estimating the potential emissions avoided through earlier identification of failures and prevention of unplanned shutdown events. The environmental assessment used 894 true-positive detections from the 1,109 confirmed failure events. An assumed emission factor of 100 tCO_2e per avoided unplanned shutdown was applied to estimate potential avoided emissions.

$$Potential\ avoided\ emissions = 894 * 100$$

$$= 894000\ tC0_2e$$

This represents a scenario estimate rather than a direct measurement of emissions avoided at the facility. It assumes that each true-positive detection enabled an intervention that prevented an otherwise expected unplanned shutdown. Actual emissions reductions would require validation using plant-specific flare, venting and process-event data.

The environmental damage proxy was US$24,592,700, based on the maximum observed work-order cost. This figure is used only as a proxy for potential environmental liability; it is not a measured environmental damage cost or a verified financial benefit.

3 RESULTS

3.1 Predictive model performance

The overall Random Forest classifier was assessed using a reconstructed held-out test dataset of 2,920 observations. The confusion matrix was reconciled using 1,109 confirmed failure events, 894 true-positive detections and 498 false-positive detections. The resulting false-negative and true-negative counts were 215 and 1,313, respectively.

Table 2: Reconstructed overall Random Forest classification performance

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.756	0.642	0.806	0.715

The reconstructed results indicate that the classifier identified 80.61% of the confirmed failure events. Its precision of 64.22% indicates that some predicted failure events were not confirmed as actual failures. The resulting F1-score of 71.49% reflects the balance between precision and recall. Mechanism-specific Random Forest models achieved F1-scores of up to 0.90, indicating the potential value of equipment- or mechanism-specific classification.

3.2 Anomaly detection and monitoring

The Isolation Forest monitoring layer identified abnormal operating patterns in the SCADA dataset. The analysis reported 514 plant-wide anomalies, corresponding to 31.2% of the observations used in the anomaly assessment. For the selected high-frequency degradation mechanisms, anomaly intensity was reported at 74.0%. These results demonstrate the potential of unsupervised learning to complement labelled failure prediction. The anomaly detection layer provides an additional monitoring mechanism for identifying unusual sensor behaviour that may not be represented in the supervised training data.

However, the detected anomalies should not be interpreted as confirmed equipment failures. Their operational significance depends on engineering review and comparison with equipment condition, process limits and maintenance records.

3.3 Time-to-failure assessment

The survival analysis identified differences in the reported failure intervals of selected equipment categories. The median time-to-failure was estimated at approximately three days for Instrumentation and Control equipment and four days for Rotating Equipment and Pumps.

The findings indicate that equipment categories may have different failure-risk profiles and maintenance urgency. The survival analysis therefore provides a complementary basis for prioritising interventions beyond binary failure classification. The combination of survival analysis and supervised classification provides additional information for maintenance planning. While classification estimates the likelihood of a failure condition, survival analysis provides information about the expected timing of failure events.

3.4 Operational impact: failure detection and downtime

The confusion matrix used for the operational assessment was reconstructed to ensure consistency with the 2,920 test observations and 1,109 confirmed failure events.

Table 3: Failure detection and operational impact

Indicator	Result
Total test observations	2,920
Confirmed failure events	1,109
True-positive detections	894
False-positive detections	498
False-negative events	215
True-negative observations	1,313

Failure detection rate	80.61%
Accuracy	75.58%
Precision	64.22%
Specificity	72.50%
F1-score	71.49%
Reactive maintenance downtime	54,272 hours
ML-informed scenario downtime	18,552 hours
Estimated downtime reduction	35,720 hours
Percentage downtime reduction	65.82%

The 894 true-positive detections represented 80.61% of the 1,109 confirmed failure events. The remaining 215 events were classified as false negatives, representing 19.39% of the confirmed failures. These undetected events reinforce the need to retain conventional inspection, condition monitoring and engineering safeguards alongside machine learning predictions.

Estimated downtime decreased from 54,272 hours under the reactive baseline to 18,552 hours under the predictive maintenance scenario. This represented a reduction of 35,720 hours, equivalent to 65.82%. The estimated reduction depends on the scenario assumption that earlier detection enables planned intervention and prevents some unplanned shutdowns.

3.5 Economic assessment

The economic assessment compared the estimated maintenance expenditure under the conventional reactive maintenance strategy with that of the machine learning (ML)-informed predictive maintenance scenario at the Adex Gas Processing Plant. The analysis focused on the potential financial benefits of early failure detection and planned maintenance interventions, taking into consideration maintenance expenditure, false-alarm investigation costs, implementation capital expenditure and annual operating expenditure.

The comparative assessment indicated that the estimated total maintenance cost under the reactive maintenance baseline was US$301,560,000, compared with US$71,950,650 under the ML-informed predictive maintenance scenario. This represents a substantial reduction in the estimated maintenance expenditure and suggests that identifying equipment failure risks before functional failure could reduce dependence on emergency repairs and unplanned maintenance interventions.

Table 4: Comparative economic assessment of reactive and ML-informed predictive maintenance

Economic indicator	**Reactive maintenance baseline**	**ML-informed predictive maintenance scenario**
Total maintenance cost	US$301,560,000	US$71,950,650
False-alarm investigation cost	—	US$1,050,780
Gross estimated maintenance saving	—	US$229,609,350
Percentage cost reduction	—	76.14%
One-time implementation CAPEX	—	US$650,000
Annual operating expenditure (OPEX)	—	US$112,000

The gross maintenance saving was calculated as the difference between the estimated reactive maintenance cost and the ML-informed predictive maintenance scenario cost using equation (1) above.

$$S_g = C_r - C_p$$

S_g= 301,560,000-71,950,650

S_g=US$229,609,350

The percentage reduction in maintenance cost was calculated using equation (2):

$$R_c = \left(\frac{(C_r - C_p)}{C_r}\right) * 100$$

Therefore:

R_c={{301,560,000-71,950,650}/{301,560,000}}*100

$$R_c = 76.14\ \%$$

The result indicates that the ML-informed predictive maintenance scenario produced an estimated 76.14% reduction in maintenance expenditure relative to the reactive maintenance baseline. This potential reduction reflects the assumed economic advantage of identifying equipment failure risks early enough to support planned maintenance interventions rather than relying predominantly on corrective actions after equipment failure.

The cost associated with false-positive predictions was also considered in the economic assessment. Based on 498 false-positive detections and an estimated investigation cost of US$2,110 per investigation, the total false-alarm investigation expenditure was calculated as:

$$C_{FA} = N_{FP} * C_i \quad (4)$$

where:

C_{FA} = Total false-alarm investigation cost (US$);

N_{FP} = Number of false-positive detections;

C_i = Cost per false-alarm investigation (US$).

Thus:

$$C_{FA} = 498 * 2110$$

$$C_{FA} = US\$1{,}050{,}780$$

This expenditure represents an important consideration in evaluating the practical economics of predictive maintenance. Although false-positive alerts may result in additional inspection and investigation activities, their associated costs should be assessed against the potential consequences of undetected equipment failures, emergency repairs and unplanned shutdowns. The investigation cost must, however, be accounted for only once in the overall predictive maintenance cost calculation.

In addition to the comparative maintenance expenditure, the proposed ML-informed framework requires an initial implementation investment and continuing operational expenditure. The estimated implementation capital expenditure (CAPEX) was US$650,000, while the annual operating expenditure (OPEX) was US$112,000. These costs were incorporated into an indicative first-year financial assessment to provide an estimate of the potential benefit after accounting for the initial investment and first-year operating requirements.

The indicative first-year adjusted saving was calculated using equation (3) as:

$$S_n = S_g - CAPEX - OPEX$$

$$S_n = 229{,}609{,}350 - 650{,}000 - 112{,}000$$

$$S_n = \text{US\$228,847,350}$$

The resulting US$228,847,350 represents the indicative first-year adjusted saving under the stated cost assumptions. It should not be interpreted as a realised financial benefit or a complete net-present-value (NPV) result. A

comprehensive investment appraisal would require consideration of the project evaluation period, discount rate, future operating expenditure, system integration costs, model maintenance costs and the timing of expected cash flows.

Overall, the economic assessment indicates that the ML-informed predictive maintenance scenario has the potential to reduce maintenance expenditure substantially compared with the reactive maintenance baseline. The estimated gross saving of US$229,609,350 and the 76.14% reduction in maintenance cost demonstrate the potential financial value of integrating predictive analytics into asset integrity management. Nevertheless, these findings are based on a retrospective, modelled comparison. The actual financial benefits would depend on the operational effectiveness of the framework, the implementation of recommended maintenance interventions, equipment availability, spare-parts procurement and the ability to sustain model performance during routine plant operations.

3.6 Maintenance optimisation results

The mathematical maintenance optimisation model was applied over a 30-day planning horizon across 10 equipment categories. The model evaluated maintenance intervention priorities under different risk thresholds and resource constraints. The reported optimal risk threshold was $\theta = 0.10$, at which the model estimated a minimum total cost of US$2,653,200 within the tested optimisation configuration. This result demonstrates the sensitivity of the maintenance plan to the selected intervention threshold.

The optimisation model provides an additional decision layer beyond failure prediction. Whereas the classification models estimate failure risk, the optimisation model seeks to allocate maintenance interventions in a manner that considers risk, cost and resource availability. The US$2,653,200 optimisation result represents a separate 30-day planning scenario. It should not be directly combined with the US$71,950,650 comparative scenario total because the two figures arise from different analytical configurations and cost-assessment bases.

4 DISCUSSION

4.1 Economic implications of machine learning-driven asset integrity management

The principal economic finding of this study is the estimated reduction in maintenance expenditure from US$301.56 million under the reactive baseline to US$71.95 million under the ML-informed predictive scenario. The corresponding gross saving of US$229.61 million represents a 76.14% reduction within the modelled comparison. This result illustrates the potential economic value of identifying equipment failure risks before functional failure occurs. Under reactive maintenance, equipment interventions are initiated after a failure has developed, potentially requiring emergency logistics, expedited procurement, overtime and unplanned production shutdowns. In contrast, predictive maintenance can provide an opportunity to plan interventions, coordinate maintenance resources and reduce the likelihood of emergency responses.

The economic advantage is particularly relevant where emergency repair and shutdown costs are substantially greater than planned maintenance expenditure. The facility-derived cost parameters indicate that unplanned shutdown and emergency repair events represent significant cost exposures. Consequently, even where predictive models generate false-positive alerts, the cost of investigation may be lower than the potential cost of an undetected failure.

However, the economic result should be interpreted within the assumptions of the scenario model. The estimated saving is not equivalent to a realised cash saving because the predictive maintenance framework was assessed retrospectively rather than through a controlled, prospective operational deployment. Actual savings would depend on the accuracy of future predictions, maintenance execution, spare-parts availability, production constraints and the extent to which detected risks can be mitigated.

Furthermore, implementation and annual operating expenditure should be incorporated into a complete financial appraisal. The reported US$650,000 implementation cost and US$112,000 annual operating cost are relevant to determining the net present value, payback period and return on investment of the proposed system.

4.2 Operational implications and equipment reliability

The test-set results indicate a failure detection rate of 80.61%, with 894 true-positive detections out of 1,109 confirmed failure events. The remaining 215 false-negative events demonstrate that the predictive framework cannot independently identify every failure. Conventional inspection, condition monitoring and engineering safeguards therefore remain necessary.

The overall classification results produced an accuracy of 75.58%, precision of 64.22%, recall of 80.61% and F1-score of 71.49%. These metrics provide a consistent account of the stated test-set counts. They should be distinguished from

the mechanism-specific model results, which achieved F1-scores of up to 0.90. Machine learning should support, rather than replace, engineering judgement. Predictions should be assessed alongside inspection findings, process operating limits, equipment criticality, maintenance history and established safety procedures.

4.3 Environmental sustainability implications

Integrating machine learning (ML)-informed predictive maintenance into asset integrity management supports environmental risk reduction by enabling early failure detection and preventing unplanned plant shutdowns, which typically increase flaring, venting, and energy consumption. For the Adex Gas Processing Plant, the environmental assessment evaluated 894 true-positive failure detections from the reconstructed test set, representing 80.61% of the 1,109 confirmed failure events. Using an assumed emission factor of 100 tonnes of carbon dioxide equivalent tCO_2e per avoided unplanned shutdown, the potential avoided greenhouse gas emissions

$$E_a = N_{TP} * EF \tag{5}$$

where:

E_a = Estimated potential avoided greenhouse gas emissions (tCO_2e);

N_{TP} = Number of true-positive failure detections;

EF = Assumed emission factor per avoided shutdown (tCO_2e).

Substituting the values:

$$E_a = 894 * 100$$

$$E_a = 89400 tCO_{2e}$$

The resulting estimate of 89,400 tCO_2e represents the potential emissions that could be avoided if each of the 894 correctly detected failure events enabled timely maintenance intervention and prevented an otherwise expected unplanned shutdown. This provides an indication of the possible environmental benefit of predictive maintenance when integrated with effective operational decision-making.

In addition, an indicative environmental damage proxy of US$24,592,700 was established from the maximum observed maintenance work-order cost in the historical dataset to represent potential environmental liability. Because both figures are scenario-based estimates rather than direct measurements, future studies should validate these outcomes using plant-specific flare records, direct methane and venting measurements, and verified ecological remediation costs.

5 CONCLUSION

This study evaluated the economic, operational and environmental implications of machine learning-driven asset integrity management using historical data from the Adex Gas Processing Plant in the Niger Delta region of Nigeria. The integrated framework combined supervised machine learning, unsupervised anomaly detection, survival analysis, maintenance optimisation and a decision-support approach to support equipment risk assessment and maintenance planning.

The reconstructed held-out test-set assessment comprised 2,920 observations, including 1,109 confirmed failure events. The reconstructed confusion matrix contained 894 true positives, 215 false negatives, 498 false positives and 1,313 true negatives. These values produced an accuracy of 75.58%, precision of 64.22%, recall of 80.61%, specificity of 72.50% and F1-score of 71.49%. The results indicate that the framework identified a substantial proportion of confirmed failure events, although the false-negative and false-positive classifications demonstrate the continuing importance of conventional inspection and engineering review.

The comparative economic scenario estimated a reduction in maintenance costs from US$301,560,000 under the reactive baseline to US$71,950,650 under the ML-informed predictive maintenance scenario. This represents a gross estimated saving of US$229,609,350, equivalent to a 76.14% reduction. Estimated downtime decreased from 54,272 to 18,552 hours, representing a reduction of 65.82%. These results indicate the potential operational and economic value of using predictive information to prioritise maintenance interventions. The estimated benefits remain scenario-based and should not be interpreted as realised savings.

The environmental assessment estimated potential avoided emissions of 89,400 tCO_2e, based on 894 true-positive detections and an assumed emission factor of 100 tCO_2e per avoided shutdown. This estimate requires validation using plant-specific emissions and shutdown records. Similarly, the US$24,592,700 environmental damage proxy represents the maximum observed work-order cost and should not be interpreted as a verified valuation of environmental damage.

Overall, the study demonstrates the potential of integrating machine learning with conventional asset integrity practices to support equipment risk assessment, maintenance planning and performance evaluation. The findings support further development and prospective validation of ML-informed maintenance planning in gas processing facilities operating under challenging environmental conditions. Future work should include validation against the original model prediction records, prospective plant deployment, independent testing across additional facilities, plant-specific emissions measurement and a comprehensive lifecycle financial assessment.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors sincerely express their gratitude to Rivers State University Port Harcourt for providing the academic environment, institutional support, and resources that made this work possible. We also extend our heartfelt appreciation to Renaissance Africa Energy Company Limited for the invaluable opportunity, encouragement, and support provided during the development and preparation of this manuscript. Their commitment to fostering research, innovation, and scholarly excellence has significantly contributed to the successful completion of this study.

Disclosure of conflict of interest

The authors declare that they have no financial, personal, professional, or institutional conflicts of interest that could have influenced the research or publication of this manuscript.

REFERENCES

[1] International Organization for Standardization, *Asset Management—Vocabulary, Overview and Principles*, ISO 55000:2024. Geneva, Switzerland: ISO, 2024.

[2] American Petroleum Institute, *Risk-Based Inspection*, API Recommended Practice 580, 3rd ed. Washington, DC, USA: API, 2016.

[3] I. Hassan, R. M. Kalin, J. A. Aladejana, and C. J. White, "Potential impacts of climate change on extreme weather events in the Niger Delta part of Nigeria," *Hydrology*, vol. 7, no. 1, Art. no. 19, 2020, doi: 10.3390/hydrology7010019.

[4] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. T. V. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, Art. no. 106024, 2019, doi: 10.1016/j.cie.2019.106024.

[5] World Bank, *2025 Global Gas Flaring Tracker Report*. Global Flaring and Methane Reduction Partnership, 2025.

[6] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[8] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proceedings of the Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422, doi: 10.1109/ICDM.2008.17.

[9] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, 1958, doi: 10.1080/01621459.1958.10501452.

[10] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–220, 1972, doi: 10.1111/j.2517-6161.1972.tb00899.x.