Volume 2, Issue XI, November 2023,**No. 43, pp. 470-482**

Submitted 29/9/2023; Final peer review 16/11/2023

Online Publication 29/11/2023

Available Online at http://www.ijortacs.com

A MODEL TO DETECT OIL THEFT IN NIGERIAN PIPELINES USING ARTIFICIAL INTELLIGENCE

[1]Chiedu Raphael S., [2]Ugwu Edith A., [3]Asogwa Tochukwu C.,[4]Oliokwe Bibian N.
[1,2,3,4] Department of Computer Science
[1,2,3]Enugu State University of Science and Technology, [4]Enugu State Polytechnic, Iwollo, Enugu State, Nigeria
[1]raphaelstephen111@gmail.com; [2]edith.ugwu@esut.edu.ng, [3]Tochukwu.asogwa@esut.edu.ng, [4]ibianemehelu@gmail.com

Abstract

This study delves into the critical issue of pipeline oil theft in Nigeria, a nation heavily reliant on its oil industry. By harnessing the power of artificial intelligence (AI), mainly through the utilization of machine learning algorithms and neural networks, the research aims to counteract the economic and environmental impact of oil theft. The study conducts an extensive analysis of various machine learning models, focusing on accuracy, actual positive rate, false negative rate, and Receiver Operating Characteristics (ROC). Leveraging feature scaling, permutation, and advanced techniques like PCA, the study develops an Artificial Neural Network (ANN)-based model to detect pipeline oil theft. The proposed system showcases remarkable promise in seeing pipeline oil theft incidents with an accuracy rate of 90.75%. The integration of crucial attributes like time, pressure, flow rate, and temperature, enriches the model's precision, even if it leads to a slightly moderated ROC value. The study contributes to the broader knowledge by emphasizing the potential of A.I securenatural resources, reduce losses, and fosterthe sustainability of Nigeria’s oil industry.

Keywords: Pipeline, oil theft, Nigeria, neural network, feature extraction, accuracy

1. INTRODUCTION

Nigeria is a major oil producer in Nigeria. This valuable resource can bring in significant income, but unfortunately, oil theft has harmed the nation's economy through lost revenue. Oil is the world's most important energy source, making up approximately 33% of global energy consumption (Hanga & Kovalchuk, 2019).Nigeria is an economy heavily reliant on the oil industry itsprimary source of income and sustenance. Nigeria was once the fifth largest producer of oil in the world, but due to technical difficulties that have caused its production output to decline and a decrease in the global oil demand, it has now dropped to thirteenth place (Chika & Ndidi, 2022).Understanding the issues in addressing oil theft in Nigeria, artificial intelligence, according to (Hanga & Kovalchuk, 2019), has gained significant traction in various fields including the oil sector, providing effective ways to handle such complex issues). Oil theft severely threatens the Nigerian economy, resulting in substantial revenue losses for the government and oil companies. Furthermore, oil theft has significant environmental implications, leading to the pollution of water bodies, the destruction of ecosystems, and long-term ecological damage. Lastly, oil theft has the issues of insecurity and the perpetuation of illegal practices.The issue of oil theft from pipelines breeds the events resulting in oil leakage and spillage in environments which leads to significant economic losses, environmental degradation, and perpetuation of illicit practices. Addressing this problem requires integrating artificial intelligence such as the artificial neural network (ANN) and advanced technologies to enhance surveillance, data analysis, and response mechanisms, thereby protecting Nigeria's oil resources, mitigating the consequences, and fostering a secure and sustainable energy industry. Furthermore, Artificial neural networks according to (Puri, etal., 2016), are a computational models inspired by networks of biological neurons, wherein the neurons compute output values from inputs.This study looks into how artificial intelligence can be incorporated into the existing infrastructure of the oil industry and other sectors to help reduce oil theft (bunkering) in Nigeria, reduce financial losses, and strengthen the economy.

2. Literature review

Martins, etal., (2023)presented the application of machine learning to the oil problems on pipelines. The studypresents five machine learning algorithms that detect oil theft on pipelines: logistic regression, random forest, XGboost, cat boost, and multilayer perceptron (MLP). In the first one, logistic regression has the accuracy, precision, recall, f1 score, and AUC-ROC (area under the receiver operating characteristic curve) of; 85%, 78%, 90%, 83%, 0.92; XG boost has accuracy-89%, precision-82%, recall-91%, f1 score-86%, AUC-ROC-0.95%; cat boost has accuracy-87%, precision-80%, recall-89%, f1score-84%, AUC-ROC-0.93; MLP has an accuracy of 91%, precision-85%, recall-93%, f1 score-89%, AUC-ROC-0.96;the random forest has the accuracy-88%, precision-81%, recall-90%, f1 score-85%, AUC-ROC-0.94;the results showed theMLP having the highest accuracy, precision, recall, and recall, f1 score, and AUC-ROC, making it the best algorithm for the detections. Despite the success of these algorithms, there is still roomfor improvementsin the f1-score.Al Jameel,etal., (2022), studied an anomaly detection model for oil and gas pipelines using machine learning. Five machine learning algorithms were used to develop detection models for pipeline leaks namely: random forest, support vector machine, k-nearest neighbor, gradient boosting, and decision tree. In the models for developing the detection of pipelines leaks, the first on our list;the random forest was tested and itshowed an accuracy percentageof 91.81% while the recall, precision, ROC-AUC, and f1-score were at 0.92; the next model was tested which is the SVM, the model presented an accuracy percentage of 97.43% while the recall, precision, ROC-AUC, and f1-score were at 0.97; followed by the next model KNN showing the accuracy of 89.37% while the recall, precision, ROC-AUC, and f1-score were at 0.89; following the KNN was the gradient boost presenting the accuracy of 90.25% while the recall, precision, ROC-AUC, and f1-score were at 0.90; and the final model the decision tree has been tested, showed an accuracy percentage of 84.97% while the recall, precision, ROC-AUC, and f1-score were at 0.85. The results presented that the support vector machine (SVM) learning algorithms, with an accuracy of 97.43%, conquered the other algorithms in detecting pipeline leakage. Thus, it proved its efficiency as an accurate model for detecting leakage in oil pipelines.Despitethe success, there is a need for improvements on the f1-score and ROC-AUC.Li, etal., (2022), researched the oil pipeline leakage detection method based on a novel(SSA-CNN)sparrow search algorithm and convolutional neural network.Presentingthat the (SSA-CNN) method is proposed for oil pipeline leakage detection;Furthermore, the proposed SSA-CNN method converts the input data from time series to two-dimensionalmatrixes, and the classification conditions of different convolution kernel sizes and different pooling sizes are compared., while the SSA algorithms areused to optimize the parameters of CNN. The simulation results show that using two-dimensional data as input with the traditional machine learning method can enhance the neural network extraction of eigenvalues. Also, the proposed SSA-CNN method was able to accurately classify 148 sample points in 150 test sets with an accuracy rate of 98.67%, which is not only higher than traditional machine learning methods but also further improves the classification capability of CNN, while the SSA-CNN method can use the parameters already learned in the test set to ensure real-time detection.Sircar, etal., (2021), On the application of machine learning and artificial intelligence in the oil and gas industry.Revealed the existence of an intelligent system that can eliminate the risk factor and cost of maintenance; Furthermore, the development and progress in using the technologies have become smart-making the judgementprocedures easy and straightforward like Artificial Neural Networks (ANN) and Feed-Forward (FF). ANN model, with synaptic weights, is associated with several inputs and outputs i.e., Summing the product of inputs and their corresponding weights to pass through a transfer function to get the outputs of the layer. The model is an effective machine-learning method to solve complicated problems in the oil and gas industries. It helps predict pipeline conditions, thereby allowing the user to predict pipe failure rate and mechanical real ability. On the other hand(FF) according to (Aeshna and Thonhauser, 2015), transfers the information forwardincluding hidden neurons to the model. Sun, etal., (2016), introduced the intelligent crude oil anti-system for detecting oil leakage based on IOT under different scenarios.Presenting the methodology levels in implementing this technology: Sensor Level, Transmission Level, And Application Level. AttheSensor Level, kinds of sensors, pan-tilt and fixed cameras, and remote terminal units (RTU) are installed to capture situ signals, images and videos. Radiofrequency identification (RFID) tags automatically identify anything attached to acting as an electronic barcode (Gubbi et al., 2013), and the FRID system composed of readers can help monitor these objects in real-time. AttheTransmission Level, sensor data such as signals, texts, images, and videos will be transmitted to the server and application system via a wired, wireless, or satellite communication network. While, attheApplication Level, the sole aim is in the construction of two parts: one is data storage and analysis, and the other is a system application. Supervisory control and data acquisition (SCADA) areexcellentremote monitoring and control system that operates coded signals by integrating computer technology, cybernation technology, and communication channels.

3. METHODOLOGY

While addressing the issues of oil theft in Nigeria using artificial intelligence, certain procedures were strongly applied in the study which are as follows; firstly, an intensive literature review was conducted listing out the different models used by the writer in identifying the necessary problems and the solutions in the development of the algorithms. In the review, the study aims to find out the gap in the reviews mentioned above and their respective solutionsconcerning the issues of oil theft. Furthermore, the solutions and gap can be filled by ushering the AI models thereby giving it the necessary dataset to train and implement. i.e., knowing the difference between a stable flowing pipeline and a not stable flowing pipeline. Finally, recommendations were made for the further detection of pipeline oil theft.

3.1 Data Collection

The data used was collected from the Nigerian National Petroleum Corporation (NNPC). The datamodel the behavior of oil flow in the pipeline during theft such as leakages, bunker, and explosionsfrom April 2018 to July 2023. The data attributes are reported in Table 1;

Table 1: Data Attributes of Dataset

ATTRIBUTES	Data Description	Data Format
Time (hh: mm: ss)	The period at which measurements were recorded.	Time and Date
Fluid Pressure (PSI)	The exerted pressure by the fluid in the pipeline.	Integers
Flow Rate (m^3/h)	The flow of fluid through the pipeline.	Integers
Temperature (C^o)	The degree (hotness or coldness) of the fluid in the pipeline.	Integers

3.2 Data Processing

The data processing section explains the approaches taken to process the collected data.The data collected needs to be processed because of its enormous quantities; thereby giving rise toData Imputation. There are several approaches used in data imputation but one will be mentioned which is the Exception-Maximization (EM) Imputation; the EM technique according to (Chukwu, et. al.,), is a maximum likelihood-based approach that works with the relationship between the unknown parameters of the data model and the missing data.(Howell, 2008), pointedout in his study that the EM corrects problems by estimating variances and covariances that incorporate the residual variance.

3.3Data Augmentation

Due to the largeamount of data collected, to enhance the diversity and variability of the dataset even if the dataset is substantial, there is a need for data augmentation;A study by (Shorten and Khoshgoftaar, 2019) demonstrated the effectiveness of data augmentation techniques in improving the performance ofthe models on various datasets. On the other hand, augmentation of datais a technique used to artificially increase the size of a dataset by creating new samples via various techniques. These techniques are Image Augmentation, Text Augmentation, Audio Augmentation, Data Augmentation in Time Series, and Feature Scaling and Permutation. Feature Scaling and Permutation are essential data augmentation techniques that contribute to better model performance and generalization when working with tabular data related to oil exploration or production processes. They can be used in combination with other data augmentation techniques to enhance the dataset's diversity and reduce the risk of overfitting, leading to more accurate and robust models for oil-related tasks.

3.4Feature Extraction

Feature extraction refers to transforming raw data into numerical features that can be processed while preserving the original dataset, which can be accomplished manually or automatically. The technique aims to reduce the dimensionality of the data while retaining the most essential information; some commonly used feature extraction techniques are Principal Component Analysis (PCA)(Jolliffe, 2002), Linear Discriminant Analysis (LDA) (Kumar, 2021), Independent Component Analysis (ICA) (Tharwat, 2018), Autoencoders (Jeremy

Jordan, 2018), Feature Scaling (Bael dung, 2022), Bag of Words (BoW/ Term Frequency-Inverse Document Frequency (TF-IDF) (Wisam, et.al., 2019), Mel-Frequency Cepstral Coefficients (MFCC) (Logan, 2000), Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2001),and Wavelet Transform (Pooja and Vishnu, 2011).PCA is one of the best featureextractiontechniques for oil exploration. PCA according to (Jolliffe, 2002), is a statistical technique used to transform high-dimensional data into lower-dimensional space while preserving the most important patterns and relationships in the data. It also helps to reduce high-dimensional exploration data while maintaining important patterns and relationships. Applying PCA, scientists gained insights into the structures, formation, and potential hydrocarbon reservoirs, making better decision-making during exploration and reservoir characterization. The effectiveness of PCA depends on the specific dataset and exploration objectives.

3.5Artificial Neural Network (ANN)

This section presents the machine learning algorithm used for the training of the processed and extracted data. The algorithm adopted is the Artificial Neural Network. This ANN is made up of some fundamental building blocks or rather components that are interconnectedwhich are: neurons (nodes), layers, weight matrix, activation function, bias, optimization algorithms, loss/cost function, parameters and hyperparameters.Combining these components, ANN can learn data, make predictions, natural language processing, and decision-making. (Ravish,2023). Figure 1, presents the diagrammatic representation of the ANN model.

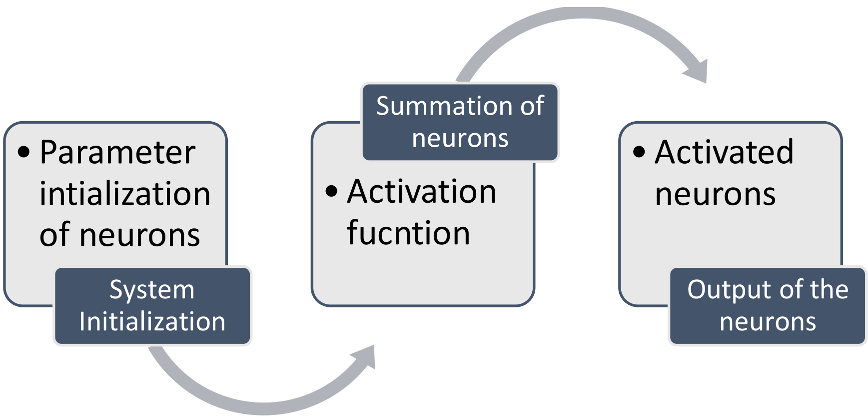

Figure 1: The Artificial Neural Network Model.

Figure 1 presents the model of the ANN which was utilized for the development of the pipeline oil theft detection system. The ANN was trained with the data of oil flow collected from busted oil pipelines in Nigeria. This data was imported into the neural network toolbox and the train using a back-propagation optimization algorithm. This algorithm adjusts the hyper-parameters of the neurons such as weight, learning rate, momentum and bias functions while monitoring the gradient loss. As the dataset was imported to the neural network model, it was split into training, test and validation sets before proceeding with training, while the test and validation sets were applied to test and validate the results of the trained model. During this training process, the evaluation process takes place at various epochs, iteratively until the desired model output is determined based on the convergence of neurons. Algorithm 1 was used to model the neural network training process until the oil theft detection model was generated.

Algorithm 1: Oil theft detection generation model

1. Start
2. Initialization of neurons
3. Import data on oil flow
4. Data attribute identification
5. Data splitting 70:15:15
6. Configuration of neurons
7. Training of neurons % adjustment of hyper-parameters
8. While loss function is tolerable
9. End training; generate a model for oil theft detection
10. Else
11. Back-propagation to step 7
12. End

3.6 The Development of Pipeline Oil Theft Detection Model

In the development of the model, the data collected was imported into the ANN and trained using the back-propagation (BP) optimization algorithm.During the period of training, the data was sectioned into three: training, validation, and test data giving the ratio of 75:15:15. TheBP algorithm was responsible for the adjustment of the weight and bias of the neurons while the error was monitored. The algorithm's training stopped when the error was close to zero, and the model for the detection of pipeline oil theft was generated. The flow chart of the model generated was presented in Figure 2;

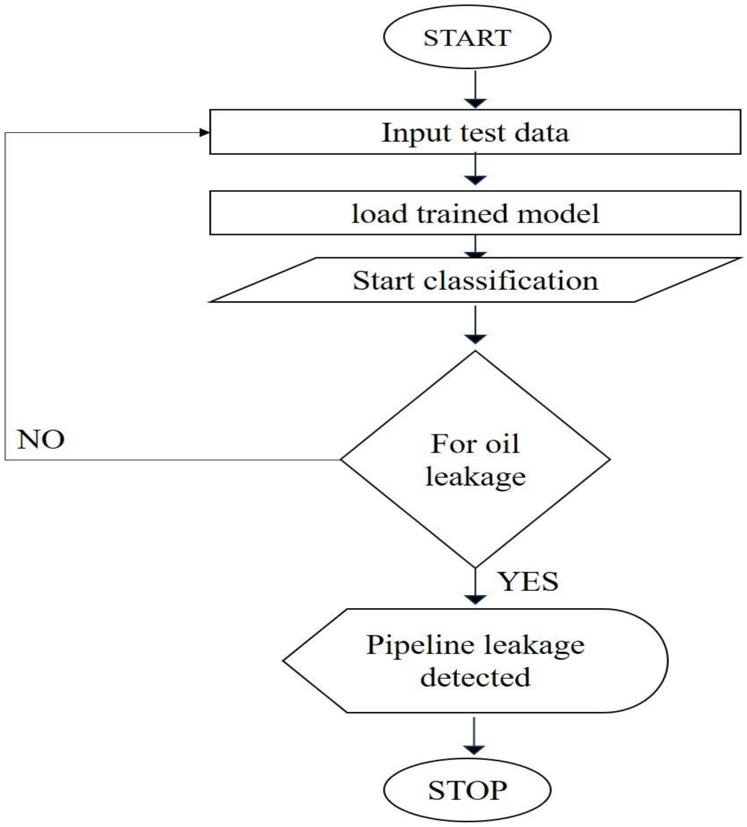

Figure 2: The diagrammatic representation of the model development

4. RESULTS AND DISCUSSIONS

This section provides an in-depth analysis of the outcomes from the neural network training, which was undertaken to effectively classify instances of pipeline oil theft. Central to this evaluation is the utilization of a confusion matrix, an efficient tool that offers a comprehensive breakdown of the classification model's prediction even the actual labels within the oil theft data sets. By outlining the occurrences of the true positives rate, true negatives rate, false positives rate, and false negatives rate, this matrix serves as an invaluable instrument for measuring the model's performance, accuracy, and efficacy. The demonstration of these results was reported in the confusion matrix of Figure 3;

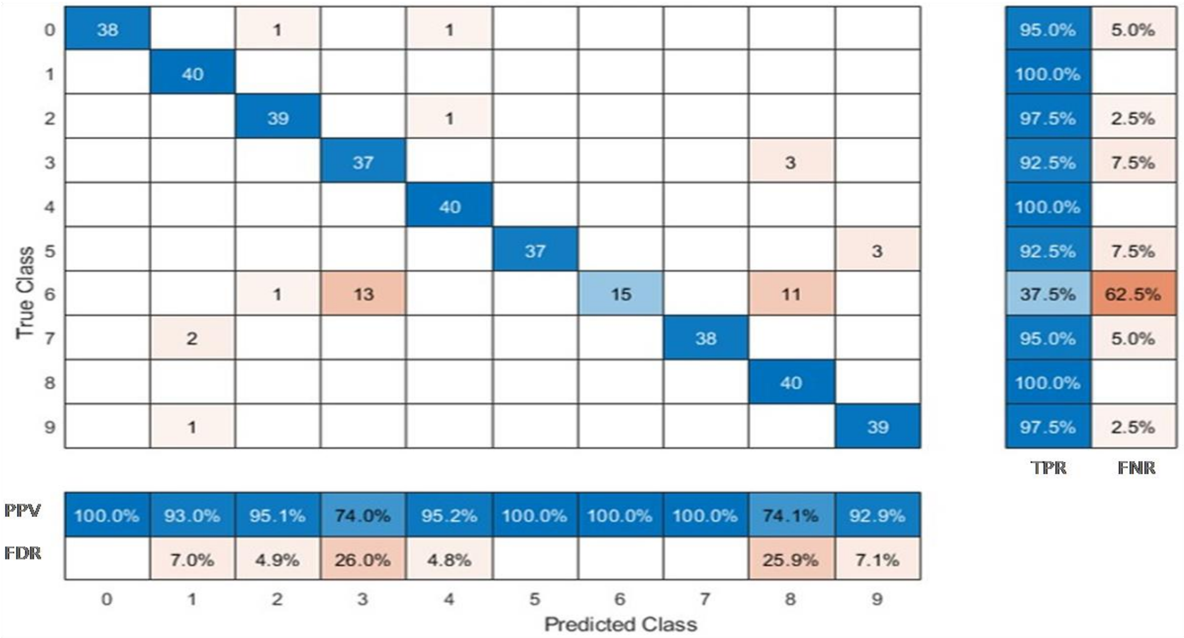

Figure3: Shows the validation confusion matrix of the ANN

In the illustrative Figure 3, the matrix comes to life, highlighting the intricate interplay between the identified features of pipeline oil leakage and those associated with non-pipeline oil leakage. Digging into specifics, following the model's comprehensive training regimen, it emerged with a commendable accuracy of 90.75%.Within the true positive ratequadrant, the model effectively captured and correctly labeled 90.75% of instances involving pipeline oil leakage, showcasing its ability to accurately detect actual cases of leakage. However, it is important to note that in the false negative realm, the model encountered a challenge, inaccurately identifying 9.25% of instances as non-leakage when they corresponded to pipeline oil leakage. Similarly, the PPV reported an average of 83.14% and FDR of 7.5%. Overall these results implied that the model is good and was able to correctly detect and classify theft in the pipeline. Similarly, the Area under curve analysis was presented in Figure 4;

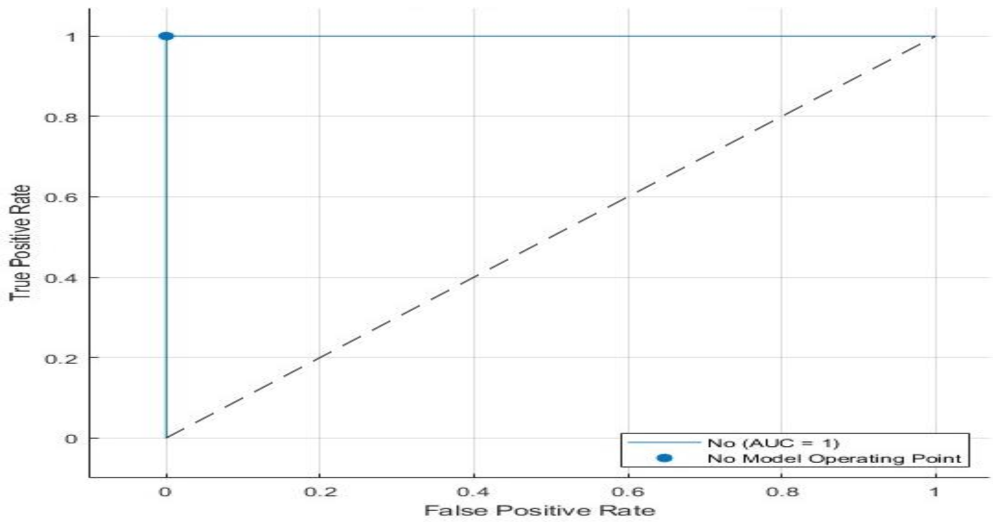

Figure 4: Shows the validation ROC curve of the ANN

In Figure 4, we encounter the Receiver Operating Characteristics (ROC) curve, a graphical representation that provides insightful details about the performance of binary classification algorithms. This curve serves as a visual tool to illustrate the delicate balance between two crucial factors: thetrue positive rate (also known as 1 sensitivity) and the false positive rate (referred to as 1 specificity). As we observe the depicted curve, it becomes evident that the true positive rate is notably high. This implies that the algorithms demonstrate a remarkable capability to correctly identify positive cases. This is further emphasized by the area under the curve (AUC), a numerical measure of the ROC curve's overall performance. In this instance, the AUC equals the maximum value of 1, which indicates a near-perfect discrimination between the two classes. Furthermore, it distinguishes between pipeline oil leakage and non-oil pipeline leakage instances, bolstering its credibility and utility for such classification tasks.

4.1 Validation of ANN pipeline leakage detection model using K fold approach

Taking into careful consideration key evaluation metrics such as accuracy, true positive rate, false negative rate, and Receiver Operating Characteristics (ROC), this section comprehensively introduces the cross-validation approach employed for the rigorous validation of the ANN-based pipeline leakage detection model.Table 2 presents the validation results.

Table 2: Validation results

FOLD	ACCURACY (%)	TPR (%)	FNR (%)	ROC
1	90.75	90.75	9.25	1
2	95.56	96.58	3.56	0.93
3	96.75	96.46	3.44	0.92
4	90.8	90.78	3.26	0.94

5	94.77	95.02	3.87	0.86
6	95.97	96.06	3.44	0.92
7	93.65	93.67	3.46	0.95
8	92.85	92.8	2.9	0.81
9	96.82	96.54	3.22	0.92
10	92.75	92.25	2.88	0.8
Average	**94.67**	**94.91**	**3.928**	**0.905**

Table 2 presents the results of the system validation considering the accuracy, TRP, FNR and ROC. From the result it was deducted that comparatively, the average accuracy reported 94.67%, TRP was 94.91%, FNR was 3.928% and ROC scored 0.905 respectively. This implied that during the repetitive training, testing and validation process, respectively. What this means is that the model was able to detect theft in the pipelines automatically and with a high success rate when collectively considering the results. Similarly, the result was comparatively evaluated considering other state-of-the-art algorithms and the results were presented in Table 3;

Table 3: Comparative Analysis

S/N	AUTHORS	TECHNIQUES	ACCURACY (%)	AUC-ROC (%)
1.	Martins, et.al., (2023)	Multilayer perception	91.0	0.94
		Logistic regression	85.0	0.92
		XG Boost	89.0	0.93
		Cat Boost	87.0	0.93
		Random Forest	88.00	0.94
2.	Al Jameel, et. al., (2022)	Support vector machine	97.43	0.97
		Random Forest	91.81	0.92
		Support vector machine	97.43	0.97
		k-nearest Neighbor	89.37	0.89
		Gradient Boost	90.25	0.90
		Decision Tree	84.97	0.85
3.	Li et. al., (2022)	Sparrow search algorithms and convolutional neural network (SSA-CNN)	98.67	0.00
4.	New System	New Artificial Neural Network	94.67	0.905

Table 3 constitutes a comparative analysis between the existing reviewed models and the novel system, taking into account essential attributes such as time, pressure, flow rate, and temperature, in terms of their accuracy. A meticulous inspection of the figure unequivocally reveals that the ANN-based model achieved a commendable accuracy rate of 90.75%. While it's apparent that this specific accuracy percentage might not claim the highest position within the spectrum of models, a more comprehensive evaluation underscores the multifaceted nature of success. Indeed, accuracy alone does not paint the complete picture. The ANN's accuracy of 94.67%, coupled with its meticulous consideration of crucial parameters like time, pressure, flow rate, and temperature, underscores its robustness in pipeline oil leakage detection. This comprehensive approach enables the model to account for nuanced variations and potential fluctuations in these critical attributes, enhancing its applicability in real-world scenarios.Furthermore, while another model might boast a marginally higher accuracy, it's imperative to consider the trade-offs. The ANN, with its proficiency and adaptability, presents a formidable advantage. Its accuracy, albeit marginally lower, is offset by its potential for further fine-tuning and optimization, which can potentially bridge the gap. This forward-looking perspective underscores the ANN's capacity for continuous improvement, hinting at a trajectory that promises not only higher precision but also an innate ability to evolve and adapt to changing conditions.In conclusion, a discerning evaluation, encompassing the attributes of time, pressure, flow rate, and temperature, illuminates the robust nature of the ANN-based model; It prompts us to appreciate the convergence of accuracy, adaptability, and potential for refinement, ultimately positioning the ANN as a formidable and compelling solution in the realm of pipeline security and oil leakage detection.In addition, a comparative analysis between the previously reviewed models and the innovative new system, focusing on their Receiver Operating Characteristic (ROC) performance was presented. A careful examination of the figure no doubt showcases that the model under consideration has achieved not-so-high ROC value compared to its counterparts. This discrepancy in ROC values could be attributed to the model's comprehensive consideration of critical attributes, such as time, pressure, flow rate, and temperature. By incorporating these variables into its decision-making process, the model demonstrates a heightened level of caution and precision, which can manifest as a slightly lower ROC value. This cautious approach underscores the model's commitment to minimizing false positives, even if it potentially leads to a relatively moderate ROC outcome.In essence, while the ROC value might not soar to unparalleled heights, the model's calibrated focus on accuracy and the intricate attributes at play underscores its judiciousness in safeguarding against false alarms. This nuanced approach manifests the model's unique strength, assuring a robust and dependable performance in real-world pipeline security and oil leakage detection scenarios.

4.2 Contribution to Knowledge

This study significantly advances our comprehension of leveraging machine learning algorithms to address the complex challenge of pipeline oil theft. Through this endeavor, the proposed system stands to offer considerable benefits to affiliated companies, particularly in mitigating the substantial income losses incurred due to oil theft and illegal bunkering. By strategically harnessing the capabilities of machine learning, this research promises to usher in an era of enhanced detection and prevention strategies, thereby bolstering the economic resilience of these organizations. The knowledge generated herein not only sheds light on the intersection of technology and security but also underscores the potential of innovative solutions in curbing illicit activities and safeguarding critical energy resources.

5. CONCLUSION

In conclusion, this study has delved into the critical issue of pipeline oil theft in Nigeria, a challenge that has significant economic, environmental, and security implications. By harnessing the power of machine learning, mainly through the implementation of Artificial Neural Networks (ANNs), this research has made notable strides in addressing this complex problem. The study's extensive literature review has unveiled the potential of AI and ANN in detecting and preventing oil theft, offering innovative solutions for an industry heavily reliant on oil revenue. The developed ANN-based pipeline leakage detection model, trained on a comprehensive dataset encompassing attributes like time, pressure, flow rate, and temperature, showcases promising results. While the accuracy and ROC values might not stand as the highest among various models, the model's cautious approach and meticulous consideration of crucial attributes underscore its robustness and adaptability. The contribution of this study lies in bridging the gap between technology and oil sector security, presenting a pathway towards curbing illicit practices, safeguarding valuable resources, and fostering a secure and sustainable energy industry for Nigeria's future.

6. REFERENCES

Alaa Tharwat, (2018). “Independent Component Analysis”. Faculty of computer science and engineering Frankfurt University of Applied Sciences. https://www.emerald.com.

Ashen, R., Thonhauser, G., (2015). Application of artificial neural networks in geoscience and petroleum industry. Artificial intelligent approaches in petroleumgeosciences.

Bael Dung, (2022). “Feature Scaling”. Bael dung. https://www.bealdung.com.

Beth Logan, (2000). “Mel-Frequency Cepstral Coefficients”. Institute of Electrical and Electronics Engineers.https://www.researchgate.com.

Chika C. and A. F. Ndidi, (2022) “Illegal oil bunkering sabotage and vandalism in Niger Delta area of Nigeria,” Acad. J. Current research.

Chukwu, A. U., Ezichi, O. N, and Dike, A. O. (2015), “On Comparison of Some Imputation Techniques in Multivariate Data Analysis” International Institute for Science and Technology and Education. https://www.iiste.org.

Dalal, T. and Cordelia S., (2001). “Histogram of Oriented Gradient”. Bot Square, https://www.vision.standford.edu.

Gubbi, J., Buyya, R., Marusic, S., et al. (2012). Internet of Things(IoT): A vision, architectural elements, and future directions. Future generationcomputer systems, 29(7), 1645–1660.

Howell, D. (2008), “The analysis of Missing data Handbook of Social ScienceMethodology”, International Institute for Science and Technology and Education, London.https//www.iiste.org.com.

Jeremy J., (2018). “Introduction to Autoencoders”.Data science. https://www.jeremyjordan.me.com.

Jinfeng S., Zhiyue Z., and Xiaoli S. (2016), "The intelligent crude oil anti-theft system based on IOT under different scenarios" 20th International Conference on Knowledge-based and Intelligent Information and Engineering Systems, New York, United Kingdom.

Jolliffe, I.T. (2002). “Principal Component Analysis (2nd ed.). Springer”. https://www.link.springer.com.

K. M. Hanga and Y. Kovalchuk, (2019) “Machine learning and multi-agent systems in oil and gas industry applications: a survey,” computer. Sci. Rev.

Pooja Saxena, and Vishnu Narayan S. (2011). "The Wavelet Transform an Introduction". Government Polytechnic College Ujjain India. https://www.researchgate.net.

Puri, M., Solanki, A., Padawer, T., Tipparaju, S. M., Moreno, W. A., & Pathak, Y. (2016). "Introduction to Artificial Neural Network (ANN) as a Predictive Tool for Drug Design,Discovery, Delivery, and Disposition: Basic Concepts and Modeling.https://www.sciencedirect.com.

Qi Li, Yaru Shi, Rui qi Lin, Wenxiu Qiao, And Wei Ba, (2022) "A novel oil pipeline leakage detection method based on the sparrow search algorithm and CNN, measurement". (https://www.sciencedirect.com).

Rachel Martins Ventriglia, Leila Figueredo Dantas, Bianca Brandão, Silvio Hamacher, Marcos Vinicius Bellé Rocha, André Silveira David, And Fredrico Chalita Ribeiro (2023), "Applying machine learning to fuel theft problem on pipelines" journal of pipelines systems engineering and practice.https://www.ascelibrary.org.

Ravish, R. (2023). "Components of An Artificial Neural Network". https://www.enjoyalgorithms.com.

Shorten, C., and Khoshgoftaar, T. M. (2019). "A survey on Image Data Augmentation for Deep Learning. Journal of Big Data". 6(1), 1-48. https://www.doi.org.com.

Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, h. (2021). Application of machine learning and artificial intelligence in the oil and gas industry. Petroleum Research,https://doi.org.com.

Sumayh S. AL Jameel, Dorieh M. Alomari, Shatha Alismail, Fatimah Khawaher, Aljawharah A. Alkhudhair, Fatimah Aljubran and Razan M. Alzannan (2022),"An Anomaly Detection Model for Oil and Gas Pipelines Using Machine Learning". https://www.doi.org.

Sunil Kumar Dash, (2021). "A Brief Introduction to Linear Discriminant Analysis" Analytics Vidhya, https://www.analyticsvidhya.com.

Wisam A. Q., Musa M. A., Bilal I. Ahmed, (2019). "An Overview of Bag of Words". International Engineering Conference.https://www.researchgate.com.