

Application of Artificial Intelligence to Monitor Leaks from Pumps

Dankwa, O. K.¹, Mensah, J. S.², Amarfio, E. M.³, Amenyah Kove, E. P.⁴

^{1,2,3,4}Petroleum and Natural Gas Engineering Department, University of Mines and Technology (UMaT), Tarkwa

DOI: <https://doi.org/10.51584/IJRIAS.2024.90303>

Received: 07 February 2024; Revised: 18 February 2024; Accepted: 22 February 2024; Published: 27 March 2024

ABSTRACT

During the production of hydrocarbons, offshore platforms frequently leak large volumes of oil and gas. Traditional methods for detecting these leaks are prone to errors, making them inefficient at detecting leaks precisely. These leaks, however, have damaging effects on the environment and humans, posing economic risks to companies as well. This study explores the application of artificial intelligence to monitor pump leaks. The data used in this work was obtained from Kaggle, which contained sensor readings from pumps. Supervised (random forest, support vector machine, and naïve bayes) and unsupervised (isolation forest) machine learning algorithms were employed for leak detection. The results showed that supervised machine learning algorithms were more accurate, with random forest having the greatest F1-score (0.993). Leveraging artificial intelligence for leak monitoring proved effective, offering a promising alternative to traditional methods.

INTRODUCTION

The production of hydrocarbons is key to meeting the energy demands on the oil and gas industry. However, during the production of hydrocarbons, offshore platforms frequently leak significant amounts of oil and gas (Vinnem and Røed, 2015). These oil and gas leaks frequently go unnoticed for a long time because seemingly large amounts of oil and gas are quickly diluted by the surrounding water and atmosphere respectively (Arifin *et al.*, 2018; Mokhatab *et al.*, 2012). Nevertheless, these leaks have damaging effects on the surrounding environment, especially to aquatic organisms. Leaks in pipeline networks are one of the major causes of innumerable losses in pipeline operators and nature. Incidents of pipeline failure can result in serious ecological disasters, human casualties and financial loss (Adegboye *et al.*, 2019). Leakages contribute to substantial financial losses. Over the past three decades, pipeline accidents led to about \$7 billion damages to property and caused the loss of more than 500 lives (Lena, 2012).

Oil and gas leakages are one of the factors that contribute to the industry's carbon footprint. Currently, oil and gas companies are under pressure to significantly decrease their carbon footprint as a result of investors' growing concern over environmental, social, and governance (ESG) issues (Adenubi *et al.*, 2023; Magill, 2021). Traditional methods for monitoring leaks such as visual inspection, pressure testing, ultrasonic testing, dye testing and temperature monitoring are time-consuming and expensive. In addition to that, they are prone to errors making them inefficient in detecting leaks precisely, allowing for gradual release of hydrocarbon into the surrounding ecosystem.

Acoustic methods of leak detection are applicable in varied scenarios but are limited by the inability to be effective in high noise environments and their signal strengths diminish with increased distance from the source of the leak (Meng, 2012). Fiber optic sensing is also another leak detection technique that uses

change in temperature and strain to detect leaks using fiber optic cables. This method is however unable to determine rate or total leakage quantity (Adegboye, 2019). Real-time transient model and computationally based methods are also sometimes used for finding leaks, but they are limited by their poor -quality signals and complexity respectively.

To efficiently monitor oil and gas leaks, the oil and gas industry can implement digital technologies like artificial intelligence and machine learning. According to Magill (2021), technologies like artificial intelligence and machine learning can analyze the past, optimize the present and predict the future. This implies that artificial intelligence can learn and improve over time by analyzing historical data and identify patterns. Artificial intelligence has been utilized in several areas such as application of artificial intelligence to detect leaks in pipelines (Idachaba *et al.*, 2021; Spandonidis *et al.*, 2022; Chernikov *et al.*, 2020). However, no significant research has been done on applying artificial intelligence to monitor leaks from pumps. In this research, a machine learning based approach that utilizes pump sensor data to monitor leaks is explored to monitor leaks from pumps. Supervised and unsupervised machine learning algorithms are used in monitoring leaks.

METHODS

The objective of this study was to develop an Artificial Intelligence (AI) model that can detect leaks from pump in real-time. A combination of pump sensor data and machine learning algorithms were used to train an AI model to detect leaks.

Dataset

The dataset used in this project was a pump sensor data obtained from Kaggle (Anon, 2023). The data consisted of a total of 220,320 sensor readings. The sections of the dataset were; timestamp (time sensor readings were taken), sensors (52 sensors), machine status (normal, recovering and broken). The following were some limitations in using the dataset for the purpose of leak monitoring in the oil and gas industry:

1. The pump was a centrifugal water pump.
2. The dataset was not very specific to leaks (focused on pump health).

Correlating the dataset to leaks

Due to inaccessibility to data specific to oil pumps leaks, some correlations were made to fine-tune the available data to be leak specific. The following were the parameters considered for correlation:

- Pump type. The dataset available was a centrifugal water pump. Interestingly, the oil and gas industry also use centrifugal pump. Table 1 shows the difference between a centrifugal water pump and a centrifugal oil pump. The differences between water centrifugal pump and oil centrifugal pumps were analyzed.

Table 1 Difference between centrifugal water pump and centrifugal oil pump (Anon., 2020)

Centrifugal Water Pump	Centrifugal Oil Pump
Impellers are designed for low viscosity	Impellers are designed for high viscosity
Materials are designed to withstand corrosive properties of water	Materials are designed to withstand high temperature and pressure

From Table 1, the differences between centrifugal water pump and centrifugal oil pump are not very significant in terms of leaks. The reason being that, in the oil industry, oil pumps come in the different

specifications due to the properties of fluids such as viscosity, temperature, pressure and many more. Hence, the impact of using centrifugal water pump for monitoring leaks instead of centrifugal oil pump may not be very significant. However, in terms of performance, these differences may be significant (Siddique *et al.*, 2017). The differences in terms of leak monitoring can be properly addressed by fine-tuning the trained model with sensor data of leaks from an oil pump.

- Correlation between pump health and pump leakage. Leaks are early signs of pump failure (Anon., 2024). Sensitivity analysis was performed to correlate the dataset to be leak specific. Random forest was first used to extract important features of the dataset. Secondly, the type of sensors used to collect data for leak detection such as vibration, temperature, pressure, and flowrate were used to streamline the remaining features. As such, machine status of broken and recovering were mapped to leaks. Machine status of normal was correlated to no leak.

Model Development

Rather than depending on a single algorithm to monitor leaks, supervised and unsupervised machine learning algorithms were used to monitor leaks.

Data Preprocessing

After loading the dataset into the Integrated Development Environment (IDE) which is jupyter notebook, the dataset was preprocessed by filling the null values in the dataset using linear interpolation method. Equation 1 represents the linear interpolation equation used. In addition to that, the dataset was also normalized within a range of 0 and 1 to ensure consistency across different sensor readings. Equation 2 represents the min-max normalization equation used.

$$y = y_1 + \frac{(x-x_1)(y_2-y_1)}{x_2-x_1} \quad (1)$$

where x_1 and y_1 are the first coordinates, x_2 and y_2 are the second coordinates, x is the point to perform the interpolation, and y is the interpolated value.

$$X_N = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Where X_{min} and X_{max} are minimum and maximum values in the dataset, X is a value from the dataset, and X_N is normalized value.

Supervised learning algorithms

The next phase of the work after preprocessing the data was implementing the machine learning algorithms. The supervised learning algorithms used in this project were random forest, support vector machine, and naïve bayes. The supervised learning algorithms were trained using the sensor data and labeled examples. The performance of the supervised learning algorithms were evaluated using leave-one-out cross-validation technique. The dataset was split into ten equally disparate sets. Nine of the subsets were used to train the models, and the last subset was used to assess how well the models' classified data. To ensure that all of the ten equally disparate data had been used as test data, this procedure was repeated for nine more times. Hence, the classification models were trained and tested in ten trails. The performance was calculated using the average outcomes from the training and testing process.

Unsupervised learning algorithm

The unsupervised learning algorithm used in this project was isolation forest. The algorithm was trained on sensor data without labeled examples. The intuition was to find unusual patterns in the dataset and identify

them as a leak. The parameters of the isolation forest used were; 100 number of trees with a contamination of 0.1 and maximum features of 1.0. The performance metrics used to evaluate the isolation forest algorithm were precision, recall, and f1-score. Precision was used to measure the performance of the models by quantifying the number of positive class prediction that actually belong to the positive class. The precision metric focuses on minimizing the false positives. Equation 3 represents the equation for precision. Recall on the other hand, was used to measure the performance of the models by quantifying the number of positive class predictions made out of all the positive examples in the dataset. Recall metric focuses on minimizing the false negatives. Equation 4 represents the equation for recall. The f1-score measured the performance of the model by combining both the precision and recall properties into a single measure. Equation 5 represents the f1-score equation. The performance metrics ranges from 0 to 1, with 1 having the highest accuracy and 0 having the least accuracy.

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})} \quad (3)$$

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})} \quad (4)$$

$$\text{F1-score} = \frac{(2 \times \text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (5)$$

RESULTS

Dataset

Fig 1 shows the results after sensitivity analysis was conducted on the 52 sensors. From Fig 1, out of the 52 sensors, only 19 sensors had impact on the label. After factoring in sensor types for monitoring leaks, the sensors were further streamlined to 10. Fig 2 shows a pie chart of streamlined sensors and their importance in leak detection. From Fig 2, the motor coupling vibration sensor had the most important feature, with pump casing vibration sensor having the least important. Fig 3 shows a count plot of the streamlined dataset. From Fig 3, the percentage of normal and leaks were approximately 94% and 6% respectively. From Fig 3, the dataset is said to be imbalance. Imbalanced dataset may lead to overfitting the model. The imbalanced dataset was handled by using a leave-one-out cross-validation technique to ensure an accurate model evaluation.

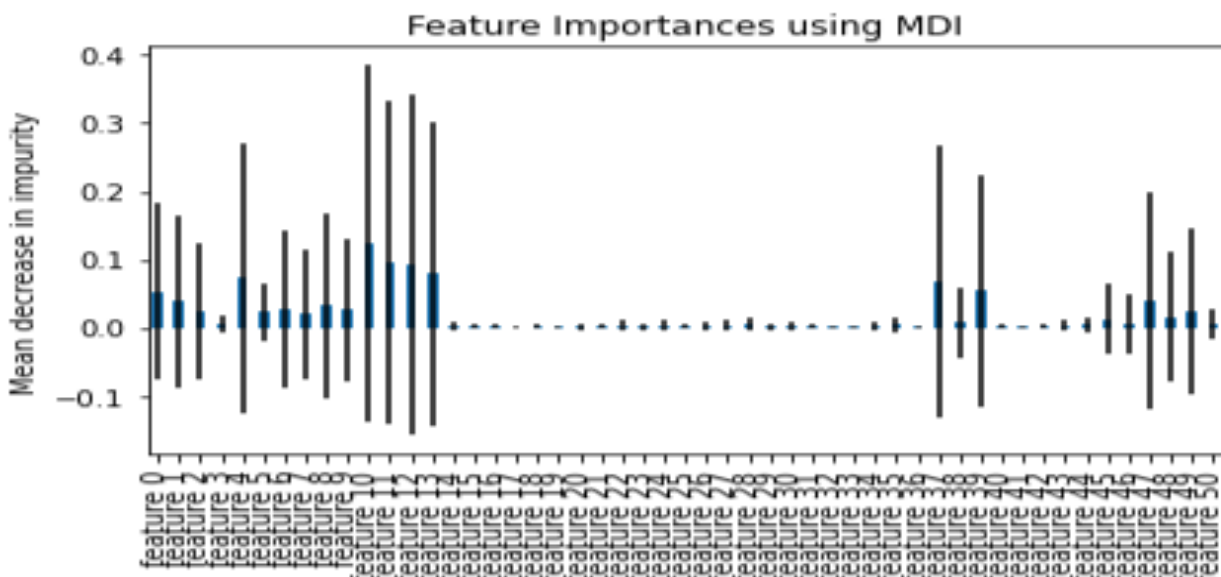


Fig 1 Sensitivity analysis on features

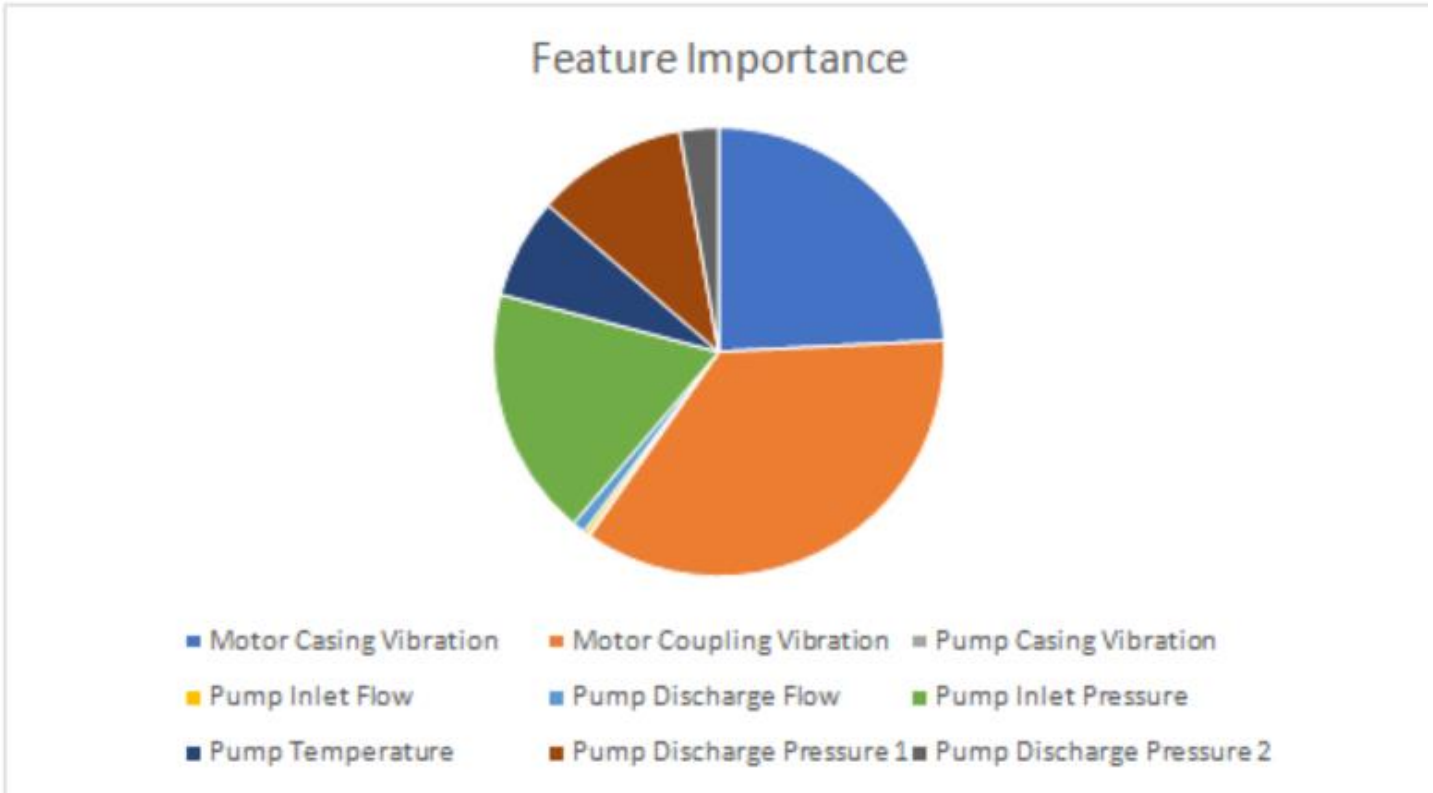


Fig 2 Pie chart of streamlined sensors and their importance in leak detection

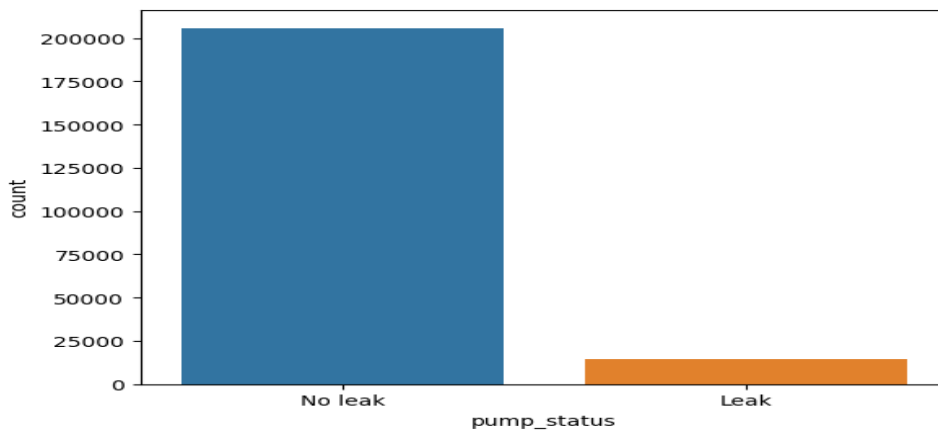


Fig 3 Count plot of streamlined dataset

Model Development

Supervised and unsupervised machine learning algorithm were used in detecting leaks.

Supervised machine learning algorithm

Table 2 shows the results of the supervised machine learning algorithms after performing leave-one-out cross-validation technique. From Table 2, naïve bayes had a precision score of 0.998, which was the highest precision score among the two other algorithms. This simply implies that the naïve bayes algorithm returned more relevant results as compared to irrelevant results. On the contrary, random forest had the highest recall score, scoring 0.994, which simply implies that the algorithm returned most of the relevant results. The f1-score metrics, being a balance of precision and recall, was used to score the best model as it's an improved version of precision and recall. Random forest had f1-score of 0.993, recording the highest amongst the

other two algorithms. This implies that, random forest had the best accuracy followed by naïve bayes, then finally support vector machine. It is therefore proposed that, random forest be used to monitor leaks from pumps as it provides the best performance relative to naïve bayes and support vector machine.

Table 2 Results of LOOCV

Performance metric	Random Forest	Support Vector Machine	Naïve Bayes
Precision	0.993	0.996	0.998
Recall	0.994	0.925	0.942
F1-score	0.993	0.945	0.961

Unsupervised machine learning algorithms

Table 3 shows the accuracy of the unsupervised machine learning algorithm. Precision had the highest score amongst the two other performance metrics. Based on the f1-score, the isolation forest algorithm was said to be accurate. However, comparing the supervised machine learning algorithms, to the unsupervised machine learning algorithms (isolation forest), the lowest accuracy of the supervised machine learning algorithm, support vector machine, had a better f1-score than the isolation forest algorithm. Therefore, the supervised machine learning algorithms had better results than the unsupervised machine learning algorithms.

Table 3 Scores of unsupervised machine learning algorithms

Performance metric	Score
Precision	0.999
Recall	0.880
F1-score	0.936

CONCLUSION

In this study, supervised and unsupervised machine learning algorithms were used to monitor pumps and detect leaks. The supervised machine learning algorithms had higher accuracy than the unsupervised machine learning algorithm. Combining the supervised and unsupervised machine learning algorithms can improve the accuracy of leak detection and prediction. The results of this study showed the effectiveness of this approach and its potential to improve the reliability and efficiency of pump monitoring systems thereby making this approach cost-effective. Lastly, pumps come in different specifications. To utilize this algorithm in real-time, the model needs to be fine-tuned with data from existing pump monitoring systems.

REFERENCES

1. Adegboye, M. A., Fung, W. K. and Karnik, A., (2019), “Recent advances in pipeline monitoring and oil leakage detection technologies: Principles and approaches”, *Sensors*, 19(11), p. 2548.
2. Adenubi, S., Appah, D., Okafor, E., and Aimikhe, V. (2023), “A Review of Leak Detection Systems for Natural Gas Pipelines and Facilities”, *Journal of Energy Technologies and Policy*, Vol. 13, No.2, 19 pp.
3. Anon. (2024), “What is Pump Failure”, <https://www.globalpumps.com.au/blog/what-is-pump-failure>. Accessed: February 14, 2024.
4. Anon. (2023), “Pump-sensor-data”, *pump_sensor_data (kaggle.com)*. Accessed: August 30, 2023.
5. Anon. (2020), “What are the Differences Between Water and Oil Pumps?”, <https://www.otomasyonavm.com/en/what-are-the-differences-between-water-pumps-and-oil-pumps>. Accessed: February 14, 2024.

6. Arifin, B., Li, Z. Shah, S. L., Meyer, G. A., and Colin, A. (2018), “A novel data-driven leak detection and localization algorithm using the Kantorovich distance”, *Computers and Chemical Engineering*, Vol 108, pp. 300–313.
7. Chernikov, A. D., Eremin, N. A., Stolyarov, V. E., Sboev, A. G., Semenova-Chashchina, O. K. and Fitsner, L. K. (2020), “Application of artificial intelligence methods for identifying and predicting complications in the construction of oil and gas wells: problems and solutions”, *Georesources*, 22(3), pp.87-96.
8. Idachaba, F. and Rabiei, M. (2021), “Current technologies and the applications of data analytics for crude oil leak detection in surface pipelines”, *Journal of Pipeline Science and Engineering*, 1(4), pp.436-451.
9. Lena, V.G. Pipelines Explained: How Safe Are American’s 2.5 Million Miles of Pipelines. Available online: <https://www.propublica.org/article/pipelines-explained-how-safe-are-americas-2.5-million-miles-of-pipelines>. Accessed: February 17, 2024.
10. Magill, J. (2021), “Oil Industry Turns to AI to Help Confront Daunting Challenge”, <https://www.forbes.com/sites/jimmagill/2021/03/26/oil-industry-turns-to-ai-to-help-confront-daunting-challenges/>. Accessed: June 19, 2022.
11. Meng, L.; Yuxing, L.; Wuchang, W.; Juntao, F. (2012), “Experimental study on leak detection and location for gas pipeline based on acoustic method”, *Journal of Loss Prevention in the Process Industries*, 25, pp. 90–102.
12. Mokhatab, S., Poe, W. A., and Mak, J. Y. (2012), “Raw Gas Transmission”, In *Handbook of Natural Gas Transmission and Processing*, 2nd, Gulf Professional Publishing: Waltham, MA, USA, pp. 103–176.
13. Siddique, M.H., Bellary, S.A.I., Samad, A., Kim, J.H. and Choi, Y.S., 2017. Experimental and numerical investigation of the performance of a centrifugal pump when pumping water and light crude oil. *Arabian Journal for Science and Engineering*, 42, pp.4605-4615.
14. Spandonidis, C., Theodoropoulos, P., Giannopoulos, F., Galiatsatos, N. and Petsa, A. (2022), “Evaluation of deep learning approaches for oil & gas pipeline leak detection using wireless sensor networks”, *Engineering Applications of Artificial Intelligence*, 113, p.104890.
15. Vinnem, J.E. and Røed, W., (2015), “Root causes of hydrocarbon leaks on offshore petroleum installations”, *Journal of Loss Prevention in the Process Industries*, Vol. 36, pp. 54-62.