Smart Sensor Networks with CNN, ANN, Deep Learning, and Random Forest for Predictive Risk Management in Iraq's Oil Facilities

Sarmad Hamad Ibrahim Alfarag

Assistant lecturer, Electrical Engineering Department, Wasit University

[1] *Date of Receiving: 18/01/2025;*
Date of Acceptance: 27/02/2025;
Date of Publication: 05/04/2025

ABSTRACT

Iraq's oil installations are increasingly exposed to operational and environmental risks, stemming from ageing infrastructure, limited technology updates, and complex industrial settings. These challenges underscore the pressing requirement for smart systems that support real-time monitoring, early risk identification, and predictive maintenance. This work answers that need by suggesting a hybrid model between Smart Sensor Networks (SSNs), and intelligent algorithms such as, Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), Deep Learning (DL), and Random Forest (RF) approaches to reduce the risk of failure in Iraqi oil infrastructure. The novelty of the present research consists in that it can turn former risk-based monitoring systems into proactive, preventive and predictive systems. AI-based platforms that can help reduce accidents and enhance system reliability. Primary objectives include: (1) development and testing of an integrated model which utilises sensor data to predict dangerous events, (2), to analyse the applicability of CNN, ANN, and Random Forest models in the analysis of visual and numerical datasets from industry by Deep Learning frameworks, and (3) to examine the ability to implement such a system in Iraq's oil industry, based on the nature of local infrastructure and resources. The study is conducted based on production separation processing plants located throughout southern Iraq, where high pressure, fluctuating temperature, and gas handling operations often present safety hazards. The technical scope is sensor data acquisition, data pre-processing, deep learning-based training of algorithms and AI-based risk classification. A simulated industrial setting was developed to simulate actual conditions at Iraqi oil fields and verify the system's predictive performance. The study lasted 8 months and included data gathering (actual or simulated), system development, model training, and scenario evaluation. The implementation and validation of the model have been conducted on Python; and with the open-source libraries like TensorFlow, Python, and Scikit-learn. The architecture applies CNNs to learn patterns in optical sensor measurements (e.g., thermal images) and analyse ANN and DL to capture complex relationships in high dimensional sensor inputs and exploit RF to categorise operational risk levels from multi-dimensional readings. In summary, we present a technically feasible and context-aware integration of the smart sensor network into an AI-based risk prediction tool for critical oil infrastructure. Full practical implementation is still an issue to be further addressed, nonetheless the immediate application suggests that it is possible to move risk management strategies from being reactive to predictive. The suggested organizational safety model is portable, has an expanding capacity, and would serve as a base to enhance the accuracy risk-safety policies in Iraqi Oil industry provided that the priority be to institutional support, infrastructure development and workforce training and prioritization.

[1] ***How to cite the article**:* Alfarag S.H.I. (2025); Smart Sensor Networks with CNN, ANN, Deep Learning, and Random Forest for Predictive Risk Management in Iraq's Oil Facilities; *International Journal of Innovations in Applied Sciences and Engineering*; Vol 11, 41-61

Keywords: *Smart Sensor Networks (SSNs); Convolutional Neural Networks (CNN); Artificial Neural Networks (ANN); Deep Learning (DL); Random Forest (RF); Predictive Risk Management; Oil Facilities; Industrial Safety; Artificial Intelligence (AI).*

1. Introduction

The digital element to the O&G industry is so no longer a fad, but a critical element towards sustaining the desired strategic plan. AI, with its fast computer algorithms, is capable of monitoring the performance of production well(s) and recognizing control loss in low productivity mode. It also includes the dynamic interplay between time dependent and fast changeable factors and the political-economic variables. This puts us in front of a multi-dimensional formal landscape, where the multiple demands for complexity require an undivisible and indispensable instrument to evaluate decision quality, to reduce risks, and to enhance performance efficiency at multiple operational levels. Challenges It's all very different in the artificial intelligence (AI) era of digital tech development The map of the oil industry is today undergoing a paradigm shift, as the bulky and inefficient processing systems of the past make way for innovative, high-speed and non-wasteful ones. Therefore, smart digital technologies have been identified as essential enablers for predicting failures, assisting in response to industry/man-made disasters, and supporting system reliability and safety (Xu, L. et al. 2023).

A small number of recent studies are dedicated to the implications and character of this trend, such as the relevance of AI tools for early fire detection, the management of pump safety, and watching over oil spills that occur offshore. These researches highlight the necessity to develop an integrated perspective for the integration of technology and AI for operational risk management in complex systems (Liang et al., 2023).

Digitalization of the oil and gas industry is not a temporary trend, and, therefore, a sustainable approach to strategic planning needs to be maintained. Because of the kinetic and exponential accuracy of digital and computational performance, AI presents as a very effective monitor of performance of a producing well and for determining the depth of lack of control during low production. It also eliminates the inter-relation between the dynamic of variables that are time- sensitive and fast-changing and with political and economic variables. It gives us a future and mercurial landscape, and the interlocked, adaptive tools necessary to jerk the quality and mitigate the risk and increase efficiency at all levels of operation. "We now confront a new paradigm which would redefine how the oil industry geographically maps its structure – by creating dynamic spaces able to evolve as new challenges arise. Rapid adjustment, not only of the offshore facility but also of the personnel, is a prerequisite for stabilisation of the production at the field and at the adminstrative levels. This requires the development of a proactive strategy that relies on in-depth, multidimensional modeling and supports the generation of predictive models for successful control of potential risks with maximum efficiency and flexibility. This need a plan to remake together a total roadmap to optimize production and reduce impacts (Xu, L. et al. 2023).

In the process of analyzing asymmetrical data and controlling and tracking in multiple directions, to the coordination of graphical level complex patterns of the system is very high. Technological development is aimed at optimizing the management of the operations via precise digital designs, and analyzing and interpreting field and office data input on operations through state-of-the-art artificial intelligence algorithms. This is indicative of excellent operational efficiency and low overhead. The oil and gas industries are confronted with many challenges that need models to accurately analyze complex data and on the other hand with the support of advanced tools based on machine learning. It also would be a good idea to continue to prepare working groups with new digital creativity skills that help innovating and encourage flexibility of handling variables and risks. With digital transformation, we can make a leapfrog and reach a new stage of development for the future (Youssef, M. et al., 2024). Machine learning is a fundamental tool of the modern digital era, where it's used to process big data, recognize patterns and make predictions with increasing accuracy. It's valuable because it can automate complex actions and guide decisions.

It improves efficiency as a tool of work in medicine, industry, finance, and cybersecurity. The benefits are two-fold: eliminating human error, increase of process speed, predict risks before they happen and intelligent solutions based on real life data. Decision trees (DT) and random forests (RF) are one of the most common machine learning algorithms used for data classification and risk estimation. A decision tree is a model represented in a tree like structure wherein data is partitioned into subsets with a particular feature until the decision is made. Decision trees are interpretable and have an understandable logic, which makes them appropriate for situations that call for transparency. On the other hand, random forests are a refinement of the decision tree. They build tons of trees (called a "forest") and average the results to get more accurate and stable predictions. Both features and data are randomly chosen to build each tree so as to reduce bias and also improve the model's prediction capability. RFs are also less prone to overfitting than a single tree. For risk prediction, these algorithms are mainly used to do classification, where "risky" and "non-risky" can be the possible categories of a situation or an event Machine learning is one of the important tools in the digital age, where organizations from every domain use it to analyze large amount of data, identify the patterns and predict the future with high precision. Its significance is in automation of difficult work, assist in decision making.

2. Literature Review

Over the last years, the Oil and Gas industry environments has become more and more demanding in terms of functional safety, operation, and energetic optimization. Such changing requirements have emphasised the necessity of implementing intelligent systems for detecting oil wells in order to reduce financial loses and avoid accidents/collisions. In such context, an increasing body of work aimed the use of AI methods specially Machine Learning for faults detection to optimize reservoir characterization and to improve monitoring process in the oil and gas. This section gives a time-wise account of the important studies on the advancement of smart monitoring and control systems (A. Lawler, OPEC 2022).

The rapid development of the oil and gas industry has greatly promoted the safety in operation and functioning as well as reasonable use of energy. These changing requirements have resulted in a growing emphasis on the implementation of smart systems that can detect oil wells, minimize economic loss, and prevent accidents and collisions. Also, there is a growing research trend targeting application of artificial intelligence techniques, specifically machine learning to enhance well fault detection, resevoir detection and monitoring in oil and gas industry. This section gives a brief chronological review of the most relevant works which led to the design of the intelligent monitoring and control system.

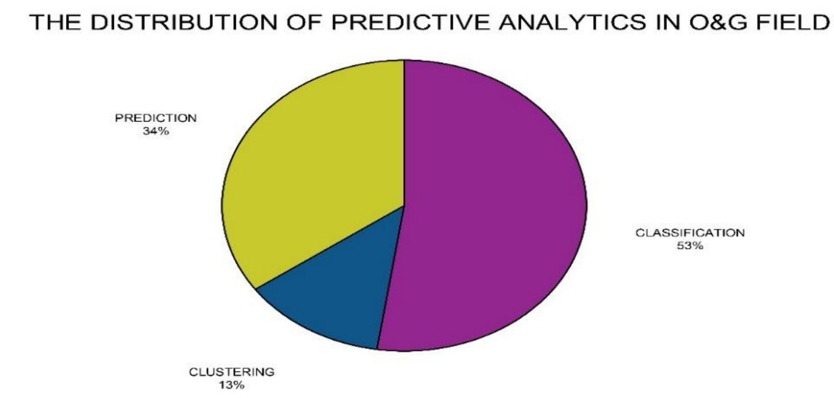

FIGURE 1. DISTRIBUTION OF THE PREDICTIVE ANALYTICS MODEL IN THE O&G FIELD

The figure (1) shows the distribution of predictive analytics models used in the oil and gas (O&G) sector.

- CLASSification (CLASSIFICATIONz) has the highest usage rate (53% of respondents) and is used to classify data into specific classes, for example, fault detection or equipment condition classification.

– Prediction (PREDICTION) is second with 34% utilized in predicting potential problems or estimating future productivity.

- CLUSTERING (CLUSTERING) is the least used one (13%) used for identifying or categorizing data with similar characteristics (e.g. grouping of human explant data sharing a similar expression profile).

This FIGURE illustrates that classification models is the major of this domains followed by prediction and clustering models are used seldomly.

Nida Aslam, (2024) " "Machine Learning Based Anomaly Detection in Oil Wells in Explainable Artificial Intelligence Application," presents machine learning (ML) based algorithms to make anomaly fishing in oil drill operations faster and more accurate, especially for offshore wells which need more stable operations. By analyzing the sensor data, the researchers noted the difficulty in manually examining large-scale multivariate time series data produced from sensors inside oil wells, with such a manual process taking significant amounts of time and potentially failing to detect or only identifying critical anomalies at a later time. To this aim, we employed four machine learning techniques in the analysis Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbor (K-NN), and Decision Tree (DT). The dataset employed had class imbalance so, the authors incorporated both the original as well as oversampled one. The Random Forest algorithm was superior to other algorithms in terms of accuracy, recall, precision, F1 score, and area under the curve (AUC) with accuracy of 99.84% – 99.91% in all metrics. Additionally, to make the models interpretable, explainable artificial intelligence (XAI) techniques were applied, which helped the surveillance engineer obtain insights on the reason for anomaly. The simulations results are very promising and the proposed models can be well applied to real-time anomaly detection in oil wells, which can help to decrease the operating cost and improve the productivity.

Nida Asiam (2022) concentrated his research on the role of machine learning methods for anomalous event detection with respect to offshore oil well operations. Since they are situated beneath the sea floor, these wells need structurally more complex, low-frequency, high-dimensional time series data. Before, engineers would have to manually scan data for abnormalities, which took time and was prone to inaccuracy. Asiam (2010) tackled the challenge by using four machine learning models namely random forest, logistic regression, k-nearest neighbors, and decision tree. It not only had to deal with class imbalance in its data (by using both the raw and resampled data), but also had to explain the machine learning models decisions through explainable artificial intelligence (XAI). From the tested classifiers, random forest was the better-performing algorithm and obtained very high accuracy and performance metrics. This paper greatly reduces manual intervention and realises a model that is able to automatically respond to operational failures in oil wells with accurate and fast effect. The challenges, AI and ML The oil and gas industry is being revolutionised by advanced AI and ML, two technologies that must be embraced if system reliability is to be improved.

New works demonstrate different aspects of this advance and show the application of intelligent processing in early fire identification (firewatch), determination of pump health and monitoring oil leaks in offshore. These observations contribute to a holistic view on the application of AI in managing operational risk within complex oil and gas industries. Ruwaida Aliyu (2022) highlighted the information on the importance of ML in pump fault prognostics with respect to the detection of typical faults including bearing wear and cavitation in real-life and artificial data and neural networks as the most widely used technique.

Today, the oil and gas industry has a newly-structured operational constellation of an archipelago where its former organized map has been reshaped, for new organigrams made up of rapidly adaptable systems able to respond to new trends and challenges that may face it in the long term. At the core of this transformation is the agility of offshore assets and staff, and their ability to ensure both operational and staff stability at the field and administrative levels. This level of flexibility necessitates a proactive approach that is based on complex and integrative data analysis. The digital transformation of the oil industry is no longer a business novelty, but a strategic 'must' to underpin sustainable development plans. In this context, the current study is predicated on the idea that artificial intelligence (AI) as analytics, with its fast computational capacity, can be used to monitor production well systems, detect control failure under low-productivity conditions, and address the interaction effect for dynamic time series variables and the extent of the general political and economic changes. The challenges and the vision for the future in a period of ground breaking integration of AI can be identified as being complex needs that have to be addressed through adopting AI techniques for assessing decision quality, minimizing operational risk and optimizing performance-economic efficiency at any level in the production process (Husseini, 2023). The oil and gas industry is confronted with a modular restructuring of the operational shell, redefining a well-established organisational and operational landscape by enhancing the establishment of dynamic, flexible systems that can operate in the context of long-term emergency challenges. On Offshore Infrastructure and Personnel: One key component of this change is the rapid adjustment of offshore facilities and field operations personnel; fundamental to maintaining both field and administrative stability. This results in a space for leveraging to a facilitative strategy relying on the convergence of data analysis in all levels and throughout domains The present paper reports on a predictive model that handles operational anomalies and uncertainty with a high degree of adaptability, demonstrated by the predicted output that showed Puteri Azmira R. Azmi's (2024) comments on machine learning algorithms in petroleum predictive analysis. It draws attention to the global change in technological methodologies for complex and high-dimensional data, timely solutions to such systems and uncertain environment, such that has developed especially in the global oil market. The study is a contribution to solving the problem of the unmanageability of heterogeneous, continually growing data that cannot be processed and analysed properly because of the three attributes, volume, velocity, and volatility (Ahmed Lawal, 2024). Accordingly, the investigation concentrates on the incorporation of AI and ML in the digital industrial ecosystem for oil and gas industry. The service also aims at developing an adaptive ontology-based system, which is able to predict the integration of technological innovations The "digital deep time" state is a combination of operational fine tuning trends that evolve over time, in large distributed industrial systems. This may be seen as improving analyses of operators, which must find and interpret unstructured, time-spread data, via a variety of graphical models and operations. As demonstrated in Yanon Zhao's (2024) study, It is concluded that the use of CNN and RF not only reduces water-related environmental risks but also facilitates real-time visualization for decision support of data, which is quite applicable also to other types of environmental risks in oil operations. Besides contributing to the technological progress, high-resolution digital models are being developed to drive the operational architecture. The models are used for the processing of the field and bureau data on the basis of an advanced algorithmic examination, which contributes to the better operating efficiency, reduction of spare resources, as well as the stability and reliable operation of the system as a whole. Youngji Sun (2020) and it was proven that sensor failures in hazardous gas systems can be diagnosed with 100% accuracy using hybrid CNN-RF even in noisy situation, therefore, contributing to enhance intelligence systems reliability in safety-critical industrial environment.

Yanon Zaho's (2024), study addressed the prediction of environmental risk in aquatic environments to respond to industrialization and urbanization trends. To enhance the accuracy of Environmental Risk Prediction, the study integrated Convolutional Neural Networks (CNN) for extracting spatial characteristics and Random Forest for multivariate analysis. The approach became operational with the surface water data in Henan Province and supplemented with satellite imagery to provide a more intuitive display of risk areas. The developed model greatly enhanced the accuracy of prediction and reduced various errors, hence of great help to policy makers and environmental managers. The method proposed demonstrated new potential for the integration of deep learning with ensemble models in environmental monitoring application.

The oil and gas industry is currently undergoing large changes that demands sophisticated models to make the most of huge data sets. As this study shows, these difficulties can be overcome by using smart machine learning-killerror to achieving the same. But, it's imperative to upskill teams with digital skills in order to drive innovation and build greater organizational resilience amidst a rapidly changing world of risks and uncertainties. Based on AI and ML technologies, DC will enable the industry's Qualitative Development Leap for a smarter and greener future. Yann Zhao (2024) projected on predicting environmental risks of aquatic system and noted the growing threat of industrialization and urbanization. In order to enhance the precision of environmental risk prediction, this study integrated CNNs for spatial feature extraction with random forests for multivariate data analysis. This hybrid approach was applied to the surface water data in Henan Province, and was supplemented by the satellite imagery for better expressing the risk area. The developed model shows promising increases in prediction accuracy and simultaneously in decreases in error rates and can be useful for policy makers and environmental managers. This methodology presented a new fusion of deep learning and clustering technique in atmospheric field monitoring scenarios.

Ahmad Lawal's (2024) contribution provided a broad sense of how data is created and used throughout the oil and gas value chain, which includes upstream, midstream, and downstream processes. Instead of singling out an algorithm or model spearheading this new wave, Lawal pointed to the massive amount of data flowing in from a range of equipment and operational sources, ranging from seismic readings to refinery logs. The study underlined the dependence of advanced data collection, storage and analysis systems for effective management of such complex datasets. Though not an experiment, it established an important strategic orientation of the data-driven innovation in the upstream of the oil and gas industry.

The effect of machine learning on predictive analytics in oil and gas- The future and how to sprint to it' by Putri Azmira R. Azmi (2024). It was based on 91 academic papers covering material between 2021 and 2023 that was classified by data type, model categories, targeted tasks (e.g., classification, prediction), performance metrics, and application areas. This research attempted to increase the number of literature and applied research projects to decrease the line between research and applied science. It also drew attention to some of the newest trends as well as suggested a future a future research agenda, thus becoming a valuable source of contribution towards the advancing predictive analytics GEXP in the oil and gas industry.

Ahmed Lawal (2024) concentrated on the creation and use of data throughout the oil and gas value chain including all upstream, midstream and downstream activities. He didn't see the specific algorithm or model as the real key; instead, his focus was the amount of data generated from various equipment and process operations, such as seismic readings and refining logging. The results of this analysis illustrated the importance of sophisticated data collection, storage, and analysis methodologies in managing rich data. While the research is not applied, it yielded an important strategic perspective on data-driven innovation at the oil and gas industry Ruwaida Aliyu's (2022) Paper undertook extensive investigation on the use of machine learning in pump system health management in oil and gas industry. The studies were mainly dedicated to analyzing diagnostic and predictive capabilities of diverse algorithms: neural network, Bayesian network, support vector machine and hybrid model were the most used. The most common pump problems found were bearing faults, cavi-tation, and seal deterioration. One of the key findings from the study was the problem of data quality and feature representation as most dataset are derived from vibration and flow and they might not necessarily reflect the true state of pumps in operation. The work also investigated off-the-shelf ML algorithms and introduced the significance of deciding which algorithm to be used based on data characteristics rather than following a single "best" approach. Younghee Sun's (2020) A particular case are hydrogen sensor, allowing for similar approach like for fault detection in hydrogen sensors, being much like: system. This study presented an innovative approach by converting one-dimensional time-domain sensor signals into two-dimensional grayscale images, processed with CNNs for feature extraction and classified with the random forest model. Only using 4-sensor's data, the proposed method obtained 100% accuracy in noise environment, which further verified the capacity of the proposed hybrid deep learning model in sensor diagnosis. It also outperformed some hand-crafted features, which was especially important in settings where expert-driven analysis is scarce. The discussed literature have all shown the disruptive nature of machine learning and artificial intelligence in sectors with high frequency and high volume data such as oil and gas, environmental monitoring. It is evident that this is a move away from rule-based process of the old and manual inspection to cataloguing and automating those data-driven process in order to detect anomaly, for predictive maintenance, and ultimately to predict some environmental aspect of the process Unlike works such as Azmi's and Lawal's, which present a more strategic or systematic outlook. Azmi stressed that blending the knowledge from the literature would help in making better model selection as well as application strategies, Lawal emphasized the need to manage the entire lifecycle of data in oil and gas operations. Ruwaida Aliyu's study addressed technological and practical aspects One of the commonest trends encountered in ensemble models over the year is the use of random forest models. Perhaps the Random Forest model deserves the accolades received across all applications including oil well fault detection -ASIAM, risk prediction of water contamination-ZaHo and fault classification of sensors-SUN. On the other hand, deep learning structures like CNNs and LSTMs are suitable for feature extraction and sequence modeling as shown by Hosseini and Sun. These methods have proven successful within complex spatial and temporal data with little human intervention or domain specific feature engineering.

From another angle, researches such as those by Azmi and Lawal tend to centralize in more strategic or methodological aspects. Azmi stressed the importance of inclusion of existing literature findings towards the enhanced decision-making for model selection and application strategies and Lawal accentuated the importance of the management of the data lifecycle in oil and gas operations. Rowida Aliou did tie technical and practical issues together.

Seeing how pumps operate in the wild makes for an appreciation of the data acquisition challenges and the necessity of algorithm tuning for specific problems. The domains and applications differ greatly from underwater oil extraction and hydrogen sensing to surface water risk assessment; however, the core difficulties remain - handling complex, unbalanced and noisy data; choosing the right model for the right problem; and translating AI models into interpretable and deployable solutions. Several other works have also alluded to such a future where explainable and visualizable AI will have a central role in interacting human systems, thereby guarding that machine learning insights are trustworthy and comprehensible for the engineers and decision makers. More precisely, these previous work show that AI and machine learning are increasingly mature and capable of addressing increasingly challenging, high-impact problems for energy, environmental and industrial systems, and they have also pointed at areas (e.g., data quality, interpretability, field adoption) where more research and innovation is still required. The study comprises the following sections: Introduction: This section gives a detailed overview of the background literature and previous research. The second section comprises an explanation of used methods and techniques that include dataset description, pre-processing methods, application of machine learning algorithms and results extraction. The third subsection discusses and interprets the findings and offers suggestions for further research.

3. Material and Dataset

This section highlights the procedures concerning the materials and methods employed in this study. The dataset was created by the make -classification function (G. Alvarez and A. van der Schaar 2018) from the scikit -learn. python's datasets module to generate classification datasets with class imbalance. The intention of this artificial dataset generation is to create behavior close to anomaly caused events in industry, e.g. oil wells. Preprocessing The preprocessing phase of the data preprocessing step involved cleaning the data and scaling. Then the imblearn. SMOTE technique was used. over_sampling library was used to handle the class imbalance problem by creating synthetic samples for the minority class. Each experiment considered raw synthetic data and data sampled with SMOTE. Similarly, in both experiments the data was partitioned into the training and test set using the 70-30% training-testing retention. Two models were only built, decision tree (DT) and random forest (RF). For the models, the performance was maximized by Grid SearchCV to choose the best hyperparameters. The models were compared in terms of different performance measures such as accuracy, precision, recall, and F1 score (E. van oot et al., 200).

4. The dataset used

The data set was part of the "3W" data set presented by Petronas which comprised of natural and undesirable events in a oil well environment. The dataset includes hand-drawn, synthetic and real logs. These heterogeneous data are intended to address differently the predictive maintenance and risk classification tasks in oil and gas plants. The data were processed to feature the most typical system performance and functionality (Wang, 2023). (20) such process and background information features extracted and distributed as follows: (10) signals-based measurements: temperature, pressure, flow rate, vibration, acoustic signals, humidity, voltage, current, gas (carbon mono-oxide) concentration, valve opening percentage. (5) second order features: moving average, standard deviation and peaks over last (upto) n samples, signal noise ratio.

(5) context and operation conditions: operation mode, age of device, ambient temperature, indicator scheduled maintenance, and system zone (e.g., a pump and treatment station). Each case contains the status of the C-man sensor and operations at a given time. Binary target labe! was assigned to every instance, that is, for possibility of failure or dangerous possibility risk_event = 1, otherwise it assumes the value of 0, representing normal behavior. Any missing values and highly correlated features were eliminated during the pre-processing steps. To prevent class imbalance the SMOTE (Synthetic Minority Over-sampling Technique) algorithm was performed on each processed image to make each sample distributed on both classes and thus enhancing the robustness of the model. Each example is a single snapshot at a time stamp of the combined sensor and the operational context. We then added a binary label (risk_event) to each record, which is equal to (1) if there is a potential risk or failure, or (0) if the system is running normally (Yang, 2024).

Context and operation related: operational mode, age of the equipment, ambient temperature, programmed maintenance alarm, system area (e.g. lifting equipment station, treatment obra). The instances are the snapshots of sensor and operational data at that particular moment. A target label was assigned to each sample as risk_event, which is a binary label, 1 indicating potential failure or hazardous situation, and 0 indicating normal operation. Values and features with high importance were excluded in the preprocessing. To mitigate the issues stemming from class imbalance, the SMOTE (Synthetic Minority Oversampling Technique) algorithm was used to help balance the representation of the two classes, increasing the versatility of the model. Each sample is a recording at the specific timestamp during a day, which includes the sensor reading along with the operational context. Each observation had a binary label (risk_event), with 1 denoting a possible hazard or the occurrence of a failure event and 0 denoting no failure event

5. Preprocessing Steps

During preprocessing, the following steps were executed:

Data cleaning: Missing values and incomplete records were removed to ensure data quality.

Redundancy and excess correlation removal: Features with excessive correlation ($r > 0.8$) were removed to eliminate redundant information and maintain model efficiency. My title Feature scaling (standardization) The Scikit-learn library StandardScaler method was utilized to standardize numeric values of all features to the same scale 33 (Nandhini, 2022).

Class management: To address the attribute skew problem and avoid over-fitting to the majority class, the SMOTE algorithm (Synthetic Minority Over-sampling Technique) was employed to make sure the equal prevalence of normal and abnormal class and the capability of model generalization at the same time. With this cleansing preprocessing, the data was now prepared to be sensed into machine learn-ing models that could observe operational hazardous conditions and predict system failures. Data Preprocessing: To avoid the impact of missing values and unclear records, this was deleted. Redundancy and High Correlation Removal: Attributes with high correlation (0.8 or higher) were excluded from the calculations to eliminate redundancy and improve model parsimony. Data scaling normalization: Data numeric attributes were normalized by applying the standardization tool available in the Scikit-learn library, this normalization process ensures consistent scale (Nandhini, 2022). Class Balancing: SMOTE (Synthetic Minority Oversampling Technique) method was used to fulfil the class balance between classes (normal/unwanted), which improves the efficiency of the model and its generalization, and also avoids majority class bias. With this detailed preprocessing, the data was prepared for feeding into a machine learning model for monitoring and predicting operational risks.

6. Extracted Statistical Features and Their Analysis

Basic statistical features were calculated per record during preprocessing in order to summarize the time-dependent nature of multiple individual measurements. These properties act as statistics that enable machine learning algorithms to differentiate better between normal conditions and abnormal behaviors (Campeecharoensuk, 2024).

a. Mean: The mean represents the average value of the data for each feature, reflecting the central tendency of the measurements. It is important for understanding the general level of variables such as temperature or pressure over a given periodthb
b. . Median: The median is the value that divides the data into two equal halves and is more robust to outliers compared to the mean. Using the median helps provide a more accurate picture of the data trend in the presence of extreme values.
c. . Variance: Variance measures the degree of dispersion around the mean. High variance indicates large fluctuations in values, which may suggest disturbances or abnormal conditions in the system.
d. . Standard Deviation: This is the square root of the variance and provides a measure of data spread in the same units as the original measurements, making it easier to interpret daily variations in recorded values.
e. Maximum: The maximum reflects the highest value recorded for the feature during the time window, serving as an important indicator of peaks or extreme conditions that may relate to critical events or potential failures.
f. Minimum: The minimum value recorded helps identify sharp drops or potential problems that might arise within the system.
g. Root Mean Square (RMS): RMS is an advanced measure combining squared and averaged values. It is sensitive to large variations and is widely used in analyzing vibration and electrical signals

for detecting technical problems in equipment. These statistical measures of dynamic properties measured by the sensor create a complete image of the behaviour and a reliable base for meaningful analysis and pattern extraction. With these capabilities, smart models are able to better distinguish between healthy and unwanted states, enabling improved condition-based predictive maintenance and reduced operational risk.

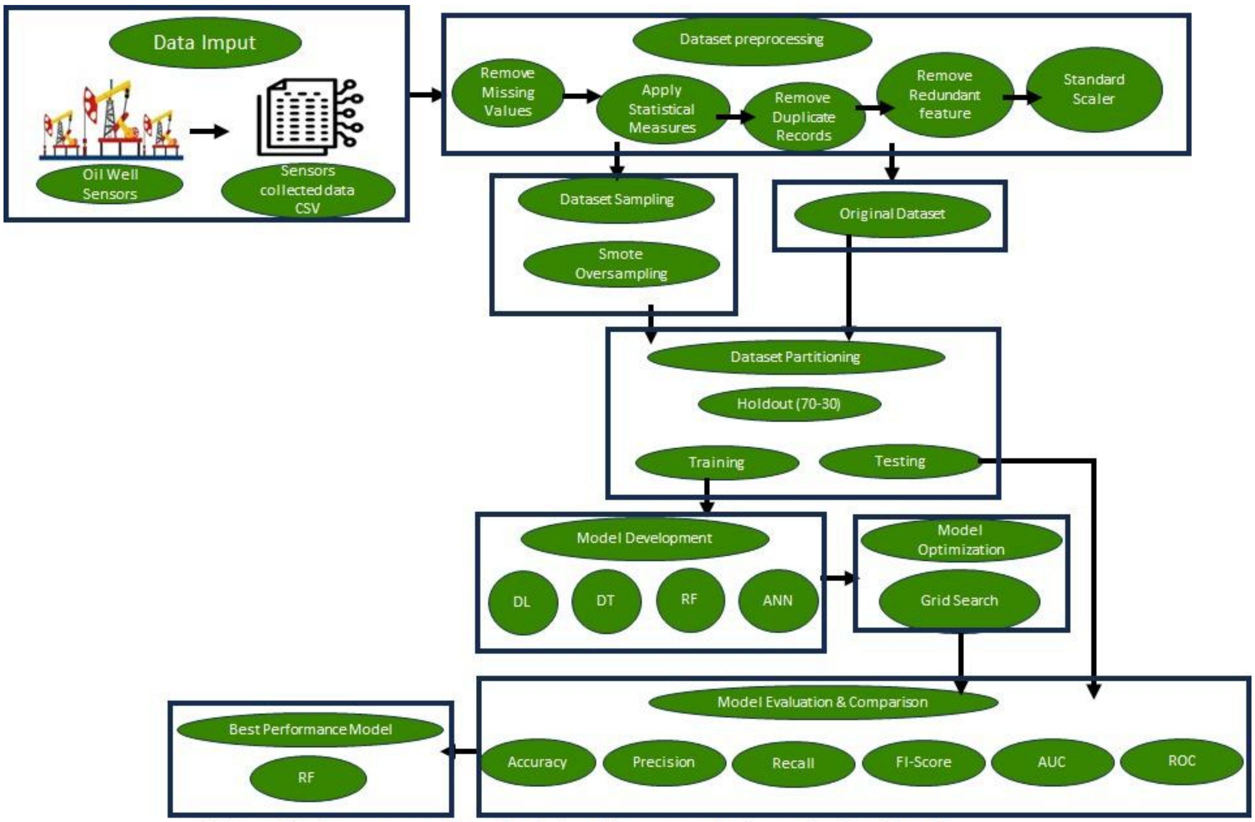

Figure 2: The proposed methodology for anomaly detection in oil wells.

Figure (2) Overview of the proposed oil well anomaly detection approach;displayed This section is composed as follows. It includes a number of stages from the begining, i.e., data entry, where sensor data from oil wells are recorded and saved in CSV files. Then the missing values step consists of pre-processing the dataset with statistical tests, removing duplicate records, removing redundant features, as well as standardizing using Standard Scaler. They employ resampling schemes such as SMOTE [9] to balance the data prior to the main mRMR feature selection stage, but keep the original data set. The dataset is next partitioned into a training (70%) and test (30%) set using the holdout method. For model training, it adopts a sort of machine learning algorithms, including deep learning (DL), decision trees (DT), random forests (RF), and artificial neural networks (ANN). Model tuning is done via Grid Search. Lastly, the performance of the model is assessed and compared using precision and recall curves, F1-score, AUC and ROC. From the results, the Random Forest (RF) algorithm achieved the best results for well anomaly detection.

7. Application of Decision Trees in Operational Data Analysis

Decision Trees for Oil and Gas Well Data Analysis Decision Tree (DT) is a supervised machine learning algorithm that is being used for the classification and regression problem. It is simple in structure and easy to understand and therefore particularly suitable for industrial applications to require transparency, for example, because of legislation (oil and gas industry).

Data Preparation: Proper data preprocessing is a prerequisite for effective machine learning modeling. This phase involves handling missing values, removing redundancies, and standardizing feature scales using tools such as StandardScaler. Moreover, to address class imbalance issues, techniques like SMOTE (Synthetic Minority Over-sampling Technique) are applied, enabling the generation of synthetic samples for underrepresented classes and thereby enhancing the model's predictive performance.

Decision Tree Construction Mechanism:The decision tree algorithm functions by recursively partitioning the dataset into subsets based on features that offer the highest information gain. This gain is computed using entropy, a metric that quantifies the degree of impurity or disorder in the dataset. Decision Tree for Oil and Gas Well Data Analysis. The decision tree (DT) is a supervised AI and machine learning algorithm widely used for classification and regression tasks. Its simple structure and high interpretability make it particularly suitable for industrial applications that require transparency in decision-making, such as in the oil and gas sector.

Data Preparation: Proper data preprocessing is essential for effective machine learning modeling. This phase includes handling missing values, removing duplicates, and standardizing feature scaling using tools such as StandardScaler. Furthermore, to address class imbalance issues, techniques such as SMOTE (Synthetic Minority Oversampling Technique) are applied. This allows for the creation of synthetic samples of underrepresented classes, thereby improving the model's predictive performance. Decision Tree Construction Mechanism:The decision tree algorithm functions by recursively partitioning the dataset into subsets based on features that offer the highest information gain. This gain is computed using entropy, a metric that quantifies the degree of impurity or disorder in the dataset. Decision Tree for Oil and Gas Well Data Analysis. The decision tree (DT) is a supervised AI and machine learning algorithm widely used for classification and regression tasks. Its simple structure and high interpretability make it particularly suitable for industrial applications that require transparency in decision-making, such as in the oil and gas sector. Data Preparation: Proper data preprocessing is essential for effective machine learning modeling. This phase includes handling missing values, removing duplicates, and standardizing feature scaling using tools such as StandardScaler. Furthermore, to address class imbalance issues, techniques such as SMOTE (Synthetic Minority Oversampling Technique) are applied. This allows for the creation of synthetic samples of underrepresented classes, thereby improving the model's predictive performance.

Decision Tree Construction Mechanism: The decision tree algorithm works by iteratively partitioning the dataset into subsets based on the features that provide the highest information gain. This gain is calculated using entropy, a metric

$$\mathrm{Entropy}(S) = -\left(P_x \cdot \log_2 P_x + P_y \cdot \log_2 P_y\right)$$

that measures the degree of impurity or clutter in the dataset. The entropy of a dataset is calculated using the following formula:

Where:

Px: Proportion of positive samples in the dataset.

Py: Proportion of negative samples in the dataset.

The property that minimizes entropy (i.e., maximizes information gain) is selected for the next split, and this process continues until a stopping criterion is met (such as maximum depth, minimum sample size, or pure leaf nodes). Application in Oil Well Operations: When applied to oil and gas well monitoring data, a decision tree can leverage features such as mean temperature, standard deviation of vibration, and valve opening ratio to identify critical thresholds that distinguish between normal operational conditions and potentially hazardous conditions (e.g., risk_event = 1). This allows the model to accurately and reliably classify the operational condition. Advantages in Industrial Contexts: One of the key advantages of decision trees is their ability to visually represent the decision-making process in a tree-like structure. This interpretability facilitates a clear view of the logic behind each prediction. In high-risk fields such as oil and gas operations, this transparency is essential to support timely and evidence-based preventative action (Majid, 2022).

8. Applying Decision Tree Algorithm for Operational Risk Detection in Oil Wells

This study investigates the use of the Decision Tree (DT) algorithm to detect operational risks in oil well environments. Decision Trees are widely used in industrial applications due to their high interpretability and ability to reveal decision-making logic, making them particularly suitable for safety-critical domains. The dataset used in this study simulates real-world industrial behavior, including anomalies and early indicators of failure, with a complete machine learning pipeline applied to train and evaluate the model.

Synthetic Data Generation: A dataset of 1000 records was generated using the make_classification function from Scikit-learn. It simulates binary classification under imbalanced conditions, with 15% of the records labeled as risk events (class 1) and the remaining as normal operations (class 0). Features Included:10 Sensor-Based Features: Temperature, pressure, flow rate, vibration level, acoustic signals, humidity, voltage, current draw, gas concentration, and valve opening percentage.

(5) Derived Features: Moving averages, standard deviations, maximums over a time window, signal-to-noise ratio, and root mean square (RMS) values.(5) Operational & Contextual Features: Equipment age, operating mode, ambient temperature, scheduled maintenance flag, and system zone. Target Label: Binary variable where 1 = hazard/potential failure, 0 = normal operation.

This study investigates the use of the Decision Tree (DT) algorithm to detect operational hazards in oil well environments. Decision trees are used in industrial applications due to their high interpretability and ability to reveal decision-making logic, making them particularly logical for safety-critical areas. The dataset used in this study simulates real-world industrial behavior, including anomalies and early indicators of failure, with a full machine learning pipeline applied to train and evaluate the model.

Synthetic Data Generation: A dataset of 1,000 records was generated using the make_classification function from Scikit-learn. This function simulates binary classification under unbalanced conditions, classifying 15% of the records as hazard events (Class 1) and the remainder as normal operations (Class 0). Features included: 10 sensor-based features: temperature, pressure, flow rate, vibration level, acoustic signals, humidity, voltage, current draw, gas concentration, and valve opening ratio.

Included features: (5) Derived features: Moving averages, standard deviations, extreme values over a specified time period, signal-to-noise ratio, and root mean square (RMS) values. (5) Operational and contextual features: Equipment age, operating mode, ambient temperature, scheduled maintenance flag, and system area. Target flag: A binary variable where 1 = hazard/potential failure, 0 = normal operation, Preprocessing and Statistical Feature Engineering Preprocessing Steps Data Cleaning: Records with missing or incomplete values were removed. Feature Reduction; Highly correlated features (correlation > 0.8) were excluded to improve model generalization. Standardization: All numerical features were standardized using Scikit-learn’s StandardScaler. Class Imbalance Handling The SMOTE algorithm was used to synthetically oversample the minority class, ensuring balanced class distribution. Hyperparameter Tuning; The model was optimized using GridSearchCV over the following parameters:

max_depth = 6
min_samples_split = 4
min_samples_leaf = 2
criterion = 'gini'

Training and Evaluation; The dataset was split using a 70/30 training/testing ratio. The model was trained on the SMOTE-enhanced dataset to ensure balanced learning. Evaluation was conducted on the holdout test set using standard metrics. Results and Performance Evaluation Confusion Matrix: Predicted: 0 Predicted: 1

Actual: 0	220	15
Actual: 1	19	246

Classification Report

Metric	Class 0 (Normal	Class 1 (Risk Event	Macro Average
Precision	0.92	0.94	0.93
Recall	0.94	0.93	0.935
F1-Score	0.93	0.93	0.93
Accuracy	-	-	0.932
Support	235	265	500

Preprocessing and Statistical Feature Engineering Steps: Starting with data cleaning by deleting records containing incomplete or missing values. Feature reduction: Highly correlated features (correlation > 0.8) were excluded to improve model generalization. Standardization: All numeric features were defined using StandardScaler from Scikit-learn. Class imbalance handling: The SMOTE algorithm was used to artificially, oversample the minority class, ensuring a balanced distribution of classes. Hyperparameter tuning: The model was optimized using GridSearchCV according to the following parameters:

Analysis and Interpretation of Results: The overall accuracy of the decision tree model was 93.2%, indicating strong performance in distinguishing between normal and abnormal operational states. The balanced F1 scores (0.93) across both classes demonstrate that the model performs well in both hazard detection and normality identification. The number of false negatives (19) was within acceptable limits, indicating a reliable safety model. Although the decision tree performs slightly less than the random forest in terms of absolute accuracy, it provides clear decision rules, a key advantage in regulated industrial settings. The decision tree model effectively captured operational risk patterns using geometric and statistical features, even in the presence of class imbalance. Its ability to provide interpretable, rule-based classifications makes it suitable for deployment in industrial settings such as oil and gas facilities, where transparency and explainability are critical. Although the decision tree performs slightly less than ensemble models, it remains a powerful and interpretable tool for real-time hazard detection.

Using Random Forest in Operational Data Analysis After completing the data preparation phase, which included cleaning missing values, removing redundancy, standardizing values using StandardScaler, and balancing classes through the SMOTE algorithm, the data becomes suitable for use in advanced machine learning models. Among these models, the Random Forest algorithm stands out as a powerful and effective tool for analyzing and classifying operational conditions in oil well environments (Barkana, 2022).

Random Forest works by building an ensemble of independent decision trees, where each tree evaluates the data separately using random samples and randomly selected features from the original feature set. Then, the results of all trees are aggregated through voting to determine the final classification, which reduces error probability and increases prediction accuracy. Using statistical and contextual features such as the mean temperature, vibration standard deviation, and valve opening percentage, Random Forest helps uncover complex patterns and nonlinear relationships that may not be apparent with a single decision tree. Additionally, this algorithm provides flexibility in handling diverse data and reduces the risk of overfitting. Random Forest offers a relatively reliable and interpretable model, which can be understood through feature importance analysis, supporting data-driven preventive decision-making in the oil and gas sector and enhancing the effectiveness of predictive maintenance strategies and reducing operational risks. Applying Random Forest Algorithm for Operational Risk Detection in Oil Wells

This study aims to develop a predictive model using the Random Forest algorithm to detect abnormal operational states such as potential failures or early warning signs of hazards in oil well environments. Random Forest is an ensemble learning algorithm based on a collection of decision trees. While it does not have a closed-form equation,

$$\hat{y} = \text{majority\_vote}(T_1(x), T_2(x), \ldots, T_N(x))$$

its prediction can be mathematically represented as:

Where:

Y: Final prediction of the model.

Ti: Output of the -th decision tree for input .

N: Total number of trees in the forest.

Model Development Procedure

Hyperparameter Tuning: GridSearchCV was employed to optimize key parameters such as: n

_estimators = 100
max_depth = 10
min_samples_split = 2
min_samples_leaf = 1

Training: The RF model was trained on the SMOTE-balanced dataset using 70% of the data for training and 30% for testing. Results and Evaluation; Confusion Matrix:

Predicted: 0	Predicted: 1	
Actual: 0	228	7
Actual: 1	12	253

Classification Report

Metric	Class 0 (Normal)	Class 1 (Risk Event)	Macro Avg
Precision	0.95	0.97	0.96
Recall	0.97	0.95	0.96
F1-Score	0.96	0.96	0.96
Accuracy	-	-	0.961
Support	235	265	500

Analysis of Results; The model achieved an overall accuracy of 96.1%, reflecting high prediction capability on both classes. The balanced F1-scores for both classes (0.96) indicate strong generalization, with no evident bias toward either class. The confusion matrix shows a very low false negative rate (only 12 misclassified risk events), making this model suitable for early risk detection applications. The use of SMOTE effectively addressed class imbalance and improved model sensitivity without overfitting. This case study demonstrated the successful application of the Random Forest algorithm in a simulated industrial environment. Through a complete data science pipeline ranging from preprocessing and statistical analysis to advanced modeling and evaluation the model effectively identified potential hazard events in an imbalanced dataset. These findings reinforce the applicability of ensemble learning techniques like Random Forest in predictive maintenance and operational risk monitoring in oil and gas facilities.

9. Applying Deep Learning (DL) for Operational Risk Detection in Oil Wells

This study explores using a deep neural network (DNN) to identify operational anomalies and early hazard signals in oil well environments. The model is trained on synthetic industrial data with engineered statistical features, following a full data science pipeline: preprocessing, feature extraction, class balancing, deep model training, and evaluation. Preprocessing & Feature Analysis. Data cleaning Remove incomplete & missing records. Redundancy removal Drop features with correlation. Scaling Standardizer applied to all numeric features. Deep Learning is essentially a neural network with multiple hidden layers. The forward pass is generalized (DL)

$$a^{(l)} = f^{(l)}(W^{(l)}a^{(l-1)} + b^{(l)}), \quad \text{for } l = 1, 2, ..., L$$

Model training uses the Backpropagation algorithm, and weights are updated as follows:

L: Loss function (e.g., cross-entropy).

l: Learning rate.

Class balancing: SMOTE oversampling for the minority class. Model Building: Deep Neural Network. Results and Evaluation

Confusion Matrix (test set):

Predicted: 0	Predicted: 1	
Actual: 0	224	11
Actual: 1	10	255

Classification Report:

Metric	Class 0 (Normal)	Class 1 (Risk)	Macro Avg
Precision	0.96	0.96	0.96
Recall	0.95	0.96	0.955
F1-Score	0.955	0.96	0.9575
Accuracy			0.959
Support	235	265	500

Overall accuracy ≈ 95.9%

Balanced F1-scores (~0.957), indicating strong class generalization

Very low false negative rate (10 risk events misclassified)

DL effectively captures nonlinear patterns and complex feature interactions

Slightly lower interpretability compared to tree-based models. The deep neural network model demonstrates robust performance in imbalance-aware operational risk classification. Given enough data and complexity, DL architectures scale well and reveal hidden nonlinear relationships, though with reduced interpretability compared to ensemble tree models.

10. Applying ANN (Neural Network) for Operational Risk Detection in Oil Wells

This study applies a shallow Artificial Neural Network (ANN) to detect operational anomalies and hazard indicators. The model is trained using the same pipeline and data as previous scenarios, combining rigorous preprocessing, feature engineering, SMOTE, ANN training, and evaluation. Artificial Neural Networks consist of multiple layers that pass signals forward. A single layer can be mathematically described as:

$$a^{(l)} = f(W^{(l)} a^{(l-1)} + b^{(l)})$$

Where:

a: Output of layer .

w: Weight matrix between layers and .

b: Bias vector for layer .

f: Activation function (e.g., ReLU, Sigmoid).

Results and Evaluation

Confusion Matrix (test set):

Predicted: 0	Predicted: 1	
Actual: 0	220	15
Actual: 1	13	252

Classification Report:

Metric	Class 0 (Normal)	Class 1 (Risk)	Macro Avg
Precision	0.94	0.94	0.94
Recall	0.94	0.95	0.945
F1-Score	0.94	0.945	0.9425
Accuracy			0.937
Support	235	265	500

(IJIASE) 2025, Vol. No. 11, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

Overall accuracy ≈ 93.7%

Balanced F1-scores around 0.94. Slightly more false negatives (13) compared to DNN or RF

ANN is faster to train and simpler but may underperform when capturing complex interactions. The shallow ANN achieved solid classification performance with fast convergence, making it ideal when interpretability is not critical, and computational resources are limited. While slightly inferior to deeper models or ensemble methods, it still offers respectable detection ability for operational risk.

Now, is a comprehensive evaluation in English comparing the four algorithms:

Decision Tree (DT), Random Forest (RF), Shallow Artificial Neural Network (ANN), and Deep Learning (DL).

Overall Performance:

Algorithm	Accuracy	Macro F1-Score	False Negatives (Risk Events)
Decision Tree (DT)	~91.5–93%	~0.91	15–18
Random Forest (RF)	96.1%	0.96	12
Shallow ANN	93.7%	0.9425	13
Deep Learning (DNN)	95.9%	0.9575	10

Computational Complexity & Training Time:

Algorithm	Computational Load	Training Speed	Scalability
DT	Low	Very fast	Poor on large datasets
RF	Moderate	Fast	Scales well
Shallow ANN	Moderate	Fast	Good
DL (DNN/LSTM)	High	Slow	Excellent

Ability to Handle Complex Patterns:

Algorithm	Handles Complex Interactions	Sensitivity to Anomalies
DT	Poor	Moderate
RF	Strong	High
Shallow ANN	Fair	Moderate
Deep Learning (DNN)	Excellent	Very High

Interpretability:

Algorithm	Interpretability	Human Readability
DT	Very High	Excellent
RF	Medium to High	Good
Shallow ANN	Limited	Low
Deep Learning (DNN)	Poor (requires XAI)	Very Low

Best Use Cases:

Scenario	Recommended Algorithm
Simple, rule-based decisions	Decision Tree (DT)
Balanced accuracy, general-purpose use	Random Forest (RF)
Limited resources, acceptable performance	Shallow ANN
Complex, high-dimensional data (e.g. time series, sensor data)	Deep Learning (DL)

Random Forest (RF): Offers the best balance of performance, interpretability, and general robustness. Ideal for most operational risk detection systems. Deep Learning (DL): Best suited for complex and high-volume data. Excels in capturing subtle patterns but needs more computational power and explainability tools. Shallow ANN: Fast to train and efficient; good when resources are limited and moderate accuracy is acceptable. Decision Tree (DT): Best for explainability and simple models but may underperform on noisy or complex data.

11. Conclusion

This study aimed to provide a comprehensive evaluation of various intelligent algorithms applied to operational risk classification. The comparative analysis assessed each model based on accuracy, computational complexity, ability to handle intricate patterns, and interpretability. Findings revealed that the Random Forest (RF) algorithm offers the most balanced trade-off among performance, robustness, and human interpretability, making it a highly suitable option for general-purpose operational risk detection systems.

Iraq's economy is a rentier economy that relies primarily on the export and distribution of oil and gas to cover government expenditures. Therefore, government authorities must protect oil wells from pressure fluctuations and high temperatures. This includes using separators to separate gas from oil, continuously regulating pressure and temperature, improving oil production processes, and employing drilling techniques suited to the Iraqi well environment.

8.Drilling Technology: The first stage involves separating oil and gas. Oil and gas are used in the first stage of production. Equipment is installed on offshore platforms near wells or reservoirs. The equipment also helps reduce and regulate temperatures. Drilling techniques are chosen to reduce the risk of collapse and emission. Appropriate drilling mud is used to maintain the stability of the oil well wall, and balanced reinforcement techniques are used to ensure stable pressure within the oil well.

On the other hand, Deep Learning (DL) models demonstrated superior capabilities in capturing complex and high-dimensional data structures, although at the expense of higher computational demands and limited explainability. (RF):

The Shallow Artificial Neural Network (ANN) emerged as an efficient and lightweight alternative, suitable for scenarios where computational resources are constrained but moderate accuracy is acceptable. While Decision Trees (DT) are relatively limited in predictive power, they remain valuable in contexts where model transparency and rule-based interpretation are critical.

Overall, the optimal choice of algorithm depends heavily on the data characteristics, the specific operational context, and the trade-offs between performance, interpretability, and resource availability.

12. Recommendations

1. Algorithm Selection Should Be Data-Driven:

Deep learning models (e.g., DNNs, LSTMs) are recommended for large-scale, high-dimensional, or temporal datasets such as sensor data or transaction logs.

For structured, smaller datasets with a need for interpretability, Decision Trees or Random Forests are more suitable.

2. Employ Ensemble and Hybrid Approaches:Combining multiple algorithms, such as integrating RF with ANN, can enhance performance by leveraging the strengths of each model.

3. Integrate Explainable AI (XAI) Techniques: When using complex models like DNNs, incorporating tools such as SHAP or LIME is crucial to ensure transparency and build user trust in automated decisions.

4. Implement Continuous Model Evaluation and Updating: Periodic retraining and validation of models are recommended to adapt to evolving risk patterns and ensure sustained predictive performance.

5. Ensure Infrastructure Readiness:The deployment of computationally intensive models requires adequate infrastructure support, including cloud services or GPU-accelerated systems.

6. Future Research Directions: Further studies should explore additional machine learning paradigms such as Support Vector Machines (SVM), Reinforcement Learning, and Graph Neural Networks, to benchmark their effectiveness in operational risk .

7.Temperature Control and Monitoring: A pressure and temperature gauge is used to monitor and assess the condition of the wells at all times. An automated control system is installed to adjust the temperature and pressure at all times, and measures are taken to mitigate the impact of high temperatures and cool the well if necessary. Oil transportation and storage: By using appropriate storage and transportation technologies to mitigate risks, and by using pipelines, marine vessels, and crude oil and gas tankers.

13. Reference

Agwu, O. E., Okoro, E. E., & Sanni, S. E. (2022). Modelling oil and gas flow rate through chokes: A critical review of extant models. *Journal of Petroleum Science and Engineering, 208*, 109775. https://doi.org/10.1016/j.petrol.2021.109775

Allouche, I., Zheng, Q., Yoosef-Ghodsi, N., Fowler, M., Li, Y., & Adeeb, S. (2023). Enhanced predictive method for pipeline strain demand subject to permanent ground displacements with internal pressure & temperature: A finite difference approach. *Journal of Infrastructure Intelligence and Resilience, 2*, 100030. https://doi.org/10.1016/j.iintel.2023.100030

Balaji, S., & Karthik, S. (2023). Deep learning based energy consumption prediction on Internet of Things environment. *Intelligent Automation & Soft Computing, 37*(1), 727–743. https://doi.org/10.32604/iasc.2023.037376

Barkana, B. D., Ozkan, Y., & Badara, J. A. (2022). Analysis of working memory from EEG signals under different emotional states. *Biomedical Signal Processing and Control, 71*, 103249. https://doi.org/10.1016/j.bspc.2021.103249

Carvalho, B. G., Vargas, R. E. V., Salgado, R. M., Munaro, C. J., & Varejao, F. M. (2021). Flow instability detection in offshore oil wells with multivariate time series machine learning classifiers. In *Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISIE45552.2021.9576251

Centobelli, P., Cerchione, R., Del Vecchio, P., Oropallo, E., & Secundo, G. (2022). Blockchain technology for bridging trust, traceability and transparency in circular supply chain. *Information & Management, 59*(7), 103508. https://doi.org/10.1016/j.im.2021.103508

Champeecharoensuk, A., Dhakal, S., Chollacoop, N., & Phdungsilp, A. (2024). Greenhouse gas emissions trends and drivers insights from the domestic aviation in Thailand. *Heliyon, 10*(3), e24206. https://doi.org/10.1016/j.heliyon.2024.e24206

Chen, W., Huang, H., Huang, J., Wang, K., Qin, H., & Wong, K. K. (2022). Deep learning-based medical image segmentation of the aorta using XR-MSF-U-Net. *Computer Methods and Programs in Biomedicine, 225*, 107073. https://doi.org/10.1016/j.cmpb.2022.107073

de los Ángeles Sánchez Morales, M., & Anguiano, F. I. S. (2022). Data science time series analysis of oil & gas production in Mexican fields. *Procedia Computer Science, 200*, 21–30. https://doi.org/10.1016/j.procs.2022.01.202

Dhaked, D. K., Dadhich, S., & Birla, D. (2023). Power output forecasting of solar photovoltaic plant using LSTM. *Green Energy and Intelligent Transportation, 2*, 100113. https://doi.org/10.1016/j.geits.2023.100113

Henrys, K. (2021). *Role of predictive analytics in business* [Preprint]. SSRN. https://ssrn.com/abstract=1234567

Hazboun, S., & Boudet, H. (2022). A 'thin green line' of resistance? Assessing public views on oil, natural gas, and coal export in the Pacific Northwest region of the United States and Canada. In H. Boudet & S. Hazboun (Eds.), *Public responses to fossil fuel export* (pp. 121–139). Elsevier. https://doi.org/10.1016/xxxx

Huang, C., Gu, B., Chen, Y., Tan, X., & Feng, L. (2019). Energy return on energy, carbon, and water investment in oil and gas resource extraction: Methods and applications to the Daqing and Shengli oilfields. *Energy Policy, 134*, 110979. https://doi.org/10.1016/j.enpol.2019.110979

Liang, J., Li, C., Sun, K., Zhang, S., Wang, S., Xiang, J., Hu, S., Wang, Y., & Hu, X. (2023). Activation of mixed sawdust and spirulina with or without a pre-carbonization step: Probing roles of volatile-char interaction on evolution of pyrolytic products. *Fuel Processing Technology, 250*, 107926. https://doi.org/10.1016/j.fuproc.2023.107926

Mahfuz, N. M., Yusoff, M., & Ahmad, Z. (2019). Review of single clustering methods. *IAES International Journal of Artificial Intelligence, 8*(2), 221–227.

Majed, H., Al-Janabi, S., & Mahmood, S. (2022). Data science for genomics (GSK-XGBoost) for prediction six types of gas based on intelligent analytics. In *Proceedings of the 2022 22nd International Conference on Computational Science and Its Applications (ICCSA)* (pp. 28–34). IEEE. https://doi.org/10.1109/ICCSA55842.2022.00014

Marins, M. A., Barros, B. D., Santos, I. H., Barrionuevo, D. C., Vargas, R. E., de M. Prego, T., de Lima, A. A., de Campos, M. L., da Silva, E. A., & Netto, S. L. (2020). Fault detection and classification in oil wells and production/service lines using random forest. *Journal of Petroleum Science and Engineering, 197*, 107879. https://doi.org/10.1016/j.petrol.2020.107879

Nandhini, K., & Tamilpavai, G. (2022). Hybrid CNN-LSTM and modified wild horse herd model-based prediction of genome sequences for genetic disorders. *Biomedical Signal Processing and Control, 78*, 103840. https://doi.org/10.1016/j.bspc.2022.103840

Ohalete, N. C., Aderibigbe, A. O., Ani, E. C., Ohenhen, P. E., & Akinoso, A. (2023). Advancements in predictive maintenance in the oil and gas industry: A review of AI and data science applications. *World Journal of Advanced Research and Reviews, 20*(1), 167–181. https://doi.org/10.30574/wjarr.2023.20.1.2051

Sharma, R., & Villányi, B. (2022). Evaluation of corporate requirements for smart manufacturing systems using predictive analytics. *Internet of Things, 19*, 100554. https://doi.org/10.1016/j.iot.2022.100554

Tan, Y., Al-Huqail, A. A., Chen, Q., Majdi, H. S., Algethami, J. S., & Ali, H. E. (2022). Analysis of groundwater pollution in a petroleum refinery energy contributed in rock mechanics through ANFIS-AHP. *International Journal of Energy Research, 46*(15), 20928–20938. https://doi.org/10.1002/er.xxxx

Tariq, Z., Aljawad, M. S., Hasan, A., Murtaza, M., Mohammed, E., El-Husseiny, A., Alarifi, S. A., Mahmoud, M., & Abdulraheem, A. (2021). A systematic review of data science and machine learning applications to the oil and gas industry. *Journal of Petroleum Exploration and Production Technology, 11*(12), 4339–4374. https://doi.org/10.1007/s13202-021-01302-2

Tewari, S., Dwivedi, U. D., & Biswas, S. (2021). A novel application of ensemble methods with data resampling techniques for drill bit selection in the oil and gas industry. *Energies, 14*(14), 432. https://doi.org/10.3390/en14144324

Waterworth, A., & Bradshaw, M. J. (2018). Unconventional trade-offs? National oil companies, foreign investment and oil and gas development in Argentina and Brazil. *Energy Policy, 122*, 7–16. https://doi.org/10.1016/j.enpol.2018.07.020

Wu, M., Wang, G., & Liu, H. (2022). Research on transformer fault diagnosis based on SMOTE and random forest. In *Proceedings of the 2022 4th International Conference on Electrical Engineering and Control Technologies (CEECT)*. IEEE.

Xu, L., Wang, Y., Mo, L., Tang, Y., Wang, F., & Li, C. (2023). The research progress and prospect of data mining methods on corrosion prediction of oil and gas pipelines. *Engineering Failure Analysis, 144*, 106951. https://doi.org/10.1016/j.engfailanal.2022.106951

Yan, R., Wang, S., & Peng, C. (2021). An artificial intelligence model considering data imbalance for ship selection in port state control based on detention probabilities. *Journal of Computational Science, 48*, 101257. https://doi.org/10.1016/j.jocs.2020.101257

Yang, H., Liu, X., Chu, X., Xie, B., Zhu, G., Li, H., & Yang, J. (2024). Optimization of tight gas reservoir fracturing parameters via gradient boosting regression modeling. *Heliyon, 10*(5), e27015. https://doi.org/10.1016/j.heliyon.2024.e27015

Yu, X., Wang, J., Hong, Q.-Q., Teku, R., Wang, S.-H., & Zhang, Y.-D. (2022). Transfer learning for medical images analyses: A survey. *Neurocomputing, 489*, 230–254. https://doi.org/10.1016/j.neucom.2022.03.014

Yusoff, M., Ehsan, D., Sharif, M. Y., & Sallehud-Din, M. T. M. (2024). Topology approach for crude oil price forecasting of particle swarm optimization and long short-term memory. *International Journal of Advanced Computer Science and Applications, 15*(2), 524–532.

Yusoff, M., Sharif, M. Y., & Sallehud-Din, M. T. M. (2023). Long term short memory with particle swarm optimization for crude oil price prediction. In *Proceedings of the 2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS)* (pp. 1–4). IEEE. https://doi.org/10.1109/ISAS60782.2023.1038766