IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 14, 2021, accepted October 1, 2021, date of publication October 4, 2021, date of current version October 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3117891

A Data-Driven Decision Support Tool for Offshore Oil and Gas Decommissioning

PATTARAMON VUTTIPITTAYAMONGKOL[1], AARON TUNG[2], AND EYAD ELYAN[3]
[1]School of Information Technology, Mae Fah Luang University, Chiang Rai 57100, Thailand
[2]Woodside Energy Ltd., Perth, WA 6000, Australia
[3]School of Computing, Robert Gordon University, Aberdeen AB10 7GE, U.K.

Corresponding author: Pattaramon Vuttipittayamongkol (pattaramon.vut@mfu.ac.th)

This work was supported in part by the Online Visiting Scholars for World Class Research Collaboration Programme under Reinventing University System 2021 funded by the Ministry of Higher Education, Science, Research and Innovation, Thailand.

ABSTRACT A growing number of oil and gas offshore infrastructures across the globe are approaching the end of their operational life. It is a major challenge for the industry to plan and make a decision on the decommissioning as the processes are resource exhaustive. Whether a facility is completely removed, partially removed or left in-situ, each option will affect individual parties differently. Stakeholders' concerns and needs are collected and analyzed to obtain the most compromised decommissioning decision. Engaging with hundreds of stakeholders is extremely complicated, hence time-consuming and costly. This issue can be addressed using a predictive model to provide suggested decommissioning options based on the data of previously approved projects. However, the lack of readily available relevant datasets is the main hindrance of such an approach. In this paper, we introduce a new oil and gas decommissioning dataset extensively covering all types of offshore infrastructures in the UK landscape over a 21-year period. An experimental framework using several learning algorithms on the new dataset for predicting the decommissioning option is presented. Various resampling methods were applied to tackle the imbalanced class distribution of the dataset for improved classification. Promising results were achieved despite the exclusion of some stakeholder-related features used in the traditional approach. This shows signs of a potential solution for the industry to significantly reduce time and cost spent on a decommissioning project, and encourages more efforts put into researching on this timely topic.

INDEX TERMS Classification, decision support tool, machine learning, offshore infrastructure, oil and gas decommissioning, supervised learning.

I. INTRODUCTION

In light of the recent acceleration of energy transition, the upcoming wave of offshore oil and gas decommissioning activities is creating significant anxieties for oil and gas operators and governments. As many fields worldwide are approaching the end of their lifespan, it is estimated that the total oil and gas decommissioning expenditure globally would amount to at least US$400 billion between 2021 and 2050 [42]. In addition to massive costs required to decommission offshore facilities, the decommissioning operations themselves are known to have significant environmental and social impacts. As such, decisions pertaining to oil and gas decommissioning, whether the offshore structure will be fully removed, partially removed or left in-place, tend to attract considerable interests from a large quantum of different local, regional, and global stakeholders.

As currently required by legislative bodies, oil and gas decommissioning activities have to extensively involve stakeholders [61]. The list of common stakeholders reported in the literature is shown in Figure 1 [22], [65]. Hundreds of stakeholders may be involved making the process of gathering and analyzing information costly and time-consuming. In addition, these stakeholders have different interests and preferences, which can pull decommissioning decisions in multiple directions [15]. This complicates the decision-making process and makes the decommissioning project even more lengthy. It has been evidenced that this part

The associate editor coordinating the review of this manuscript and approving it for publication was Xianzhi Wang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

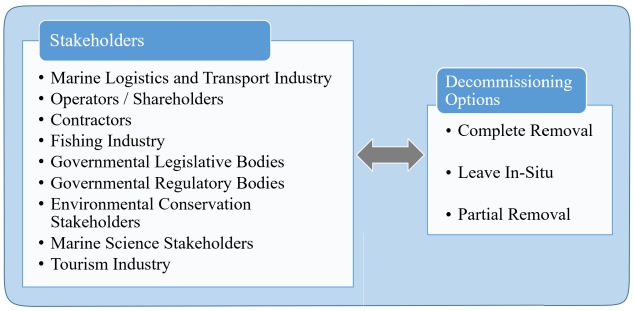


FIGURE 1. Stakeholders involved in oil and gas decommissioning operations.

of a decommissioning project can take years and sometimes up to ten years before an agreement is reached.[1]

Current best practice for balancing multiple stakeholders' views and making decommissioning decisions relies heavily on the use of multi-criteria decision analysis (MCDA) tools [40], [64]. Examples of MCDA tools commonly adopted for oil and gas decommissioning decision-making include the Comparative Assessment (CA), Best Practicable Environmental Option (BPEO), and Net Environmental Benefit Analysis (NEBA). These MCDA tools weighs different decommissioning options against a set of criteria to determine the best decision using a scoring system. However, there is no standard guidance on the score assignment to each parameter in the criteria. The decision output can thus be highly subjective. This has shown to cause public controversy over the decommissioning plan such as in the famous case of the Brent Spar field, which seriously damaged the operator's reputation [37].

As the emergence of technology as well as the growth in oil and gas decommissioning data, machine learning will be a good solution to the aforementioned problems. There has been extensive research on the use of machine learning in the oil and gas industry. In recent years, the number of research works in this area reaches several hundred yearly and is growing exponentially [24]. However, we see very little progress on machine learning-driven approaches in the field of oil and gas decommissioning. This could be mainly attributed to the lack of public datasets. Accessing raw data of decommissioning activities can done through the form of reports, where the Offshore Petroleum Regulator for Environment and Decommissioning (OPRED) database [2] is currently the biggest public source. Even so, extracting the information from reports is a challenging task. Interpreting the data and selecting potential variables to compile a dataset require interdisciplinary knowledge of engineering, management, law, data science, etc.

There had not been any public oil and gas decommissioning dataset readily available for machine learning tasks until 2021 [71]. The authors introduced a new oil and gas decommissioning dataset containing decommissioning activities of pipelines, which is the most common type of offshore infrastructures. A successful application of machine learning on predicting the decommissioning option for pipelines was presented. The overall classification results were promising although the issue of imbalanced class distribution was not addressed leading to low accuracy in the smaller classes. These results are similar to those presented in earlier work of Martin *et al.* [46], which is, to the best of our knowledge, the first published research work in machine learning-related oil and gas decommissioning. However, the limitation of this work is that only one out of several existing types of offshore infrastructures was considered. Similarly, a key weakness of Martin *et al.* [46]'s work is the very limited number of real-world instances used in the experiments. The classification results were based on only 14 oil and gas decommissioning activities.

In this paper, we introduce a new, extensive and up-to-date oil and gas decommissioning dataset and present the use of several machine learning techniques to build predictive models for the decommissioning option. The main contributions are outlined as follows:

- A new oil and gas decommissioning dataset is presented to the research community. The dataset contains 1,846 instances covering all types of offshore oil and gas infrastructures. The data was extracted from the reports of 120 decommissioning programs undertaken by 31 oil and gas companies over a 21-year period. This data is very up-to-date as the last program approval included was granted in June 2021. Exploratory data analysis with in-depth technical discussion of the dataset is provided. This also includes removal of ineffective features and redundant feature identification through correlation analysis.
- We present an experimental framework using machine learning for predicting the oil and gas decommissioning option. Several supervised learning methods have been applied. Results show that this approach is a potential direction for the oil and gas industry in planning and executing decommissioning activities. By following such an approach, some information that usually takes year to collect from hundreds of stakeholders can be dropped. This is the first time it is proved by experimental evidence that some of the key features considered in the traditional approach can be omitted. This crucial finding will facilitate significant reduction of costs and time spent on a decommissioning project.
- Various data resampling techniques for handling class-imbalance have been used to improve the classification. The obtained results will encourage further efforts put into improving the prediction accuracy in

[1] Brent Decommissioning Stakeholder Engagement Report: A Supporting Document to the Brent Field Decommissioning Programmes (https://www.shell.co.uk/sustainability/decommissioning/brent-field-decommissioning.html)

[2] Oil and gas: decommissioning of offshore installations and pipelines (https://www.gov.uk/guidance/oil-and-gas-decommissioning-of-offshore-installations-and-pipelines, Accessed on: Jul. 24, 2021)

order to motivate the adoption of this emerging technology in the industry, not only to save costs and time but also to reduce human biases in determining the decommissioning option. This research findings will also pave the way for future exploration and study on this timely topic.

The remainder of this paper is structured as follows. Section 2 provides recent related applications of machine learning in the oil and gas industry. Section 3 introduces the new oil and gas decommissioning dataset with detailed discussion on the dataset's properties and statistical analysis. In Section 4, an explanation of two sets of experimental setups for predictive decommissioning option is given. Section 5 presents results and discussion in detail. Finally, Section 6 concludes the paper, findings and potential future directions.

II. MACHINE LEARNING IN OIL AND GAS

In recent years, machine learning has attracted considerable attention from the oil and gas industry [7]. The annual number of machine learning research papers in the industry reaches several hundreds and is rising exponentially [24]. Examples of machine learning utilization are petroleum exploration and production forecasting [53], [59], detection and correction of equipment malfunctions [72], maintenance support system [33], reservoir modeling and characterization [21], [31] and drilling performance optimization [30], [63].

In determining locations to develop oil fields takes significant effort to manually process and interpret well log data [60]. Many machine learning-driven approaches have been proposed to address the lengthy and time-consuming issue [10], [32], [53], [60]. However, none had succeeded in a fully automated process without human intervention in interpreting and concluding the results [10], [53], [60]. The significant amount of missing data in well logs is one of the key hindrances in applying machine learning [25]. Nick *et al.* [10] showed the use of boosted trees to estimate missing values in order to improve the lithology classification accuracy of a deep neural network. Geological model matching is another tedious task in field exploration. Roubickova *et al.* [53] proposed a semi-supervised clustering approach to significantly reduce a number of models used in determining locations for developing oil fields. Firstly, regression analysis was used to estimate the amount of oil in place (OIP) in each geological model. This was followed by clustering the models based on OIP. By using representative models from each cluster, they were able to reduce the number of final models to as low as 0.5%. This will help reduce the time spent in model matching; nonetheless, experts are still needed to complete the entire process.

Numerical reservoir simulation is so far the most effective means for oil and gas production forecasting used in the industry [6]. However, it requires accurate prior manual operations and calculations, which is time-consuming. A great deal of machine learning-based methods have been proposed for production forecasting while getting rid of such a limitation. Deep learning techniques have been used for prediction of time-series data. Among several techniques, Long Short-Term Memory Neural Network (LSTM) was often adopted [2], [19], [41], [56], [62]. Similarly, Adaptive Neuro-Fuzzy Inference System (ANFIS) is another efficient algorithm for time-series prediction that was frequently used for such a task [6], [73].

During drilling operations, unexpected hazards and equipment failure can cost severe losses [47]. Attempts to mitigate, early detect, or prevent such events using machine learning have been proposed. A recent survey showed that deep learning, support vector machine and random forest had lately become more popular in the application of hazard prediction [50]. Mamudu *et al.* [43], [44] developed hybrid models based on neural network and Bayesian network algorithms that not only served as a risk monitoring system but also as product optimization. Roy *et al.* [54] utilized ANFIS for predicting fracture toughness to prevent rock failure during drilling. They showed that such an approach provided significantly higher accuracy than the traditional analysis using multiple regression.

Lost circulation, which is loss of drilling fluid into a formation, is one of the most common issues that lead to many other problems in oil and gas productions [34]. Both traditional learning algorithms and neural network-based algorithms were used in prediction of lost circulation [1], [4], [5], [34], [55]. In [55], the authors presented regression analysis on the severity of lost circulation using decision tree and artificial neural network-based models. Since the data size was not sufficiently large, it is not surprising that the decision tree model provided higher accuracy than the other. Similarly, Abbas *et al.* [1] reported superior results of SVM over neural network-based algorithms in predicting lost circulation occurrence. In contrast, when dealing with a large amount of data such as time-series data, Aljubran *et al.* [4] showed that deep learning methods far outperformed traditional ones in lost circulation detection.

In petroleum refining, product quality monitoring is critical for industry's profitability. The concentrations of the top and the bottom streams in the distillation column needs to be well controlled to achieve desired product purity. This is challenging for engineers since the distillation columns are complex and highly unpredictive [51]. Application of machine learning has been proposed to handle such a task; however, not many works have been seen due to limited available data [20], [36], [52]. Fatima *et al.* [20] used ANFIS to estimate the top and bottom compositions in a distillation column. Even with limited samples, the ANFIS model provided good prediction accuracy. Similarly, Ramli *et al.* [51], [52] proposed the use of neural network for composition prediction. Since some variables were not available from the plant, they obtained these missing variables by means of simulation.

From well exploration to petroleum refining, the literature shows that machine learning is capable of diminish human effort in many processes. Despite, one missing important

piece is the application of machine learning in decommissioning of offshore infrastructures [46]. The complicated and time-consuming nature of planning and decision making for decommissioning [11] makes machine learning a potential candidate for addressing such an issue. To the best of our knowledge, there has been only two publications recently in the topic of machine learning-driven approaches for oil and gas decommissioning [46], [71]. This limitation was due to the lack of oil and gas decommissioning datasets readily available for researchers. Martin *et al.* [46] showed that it was feasible to predict decommissioning option using machine learning techniques. However, their experimental results were based on bootstrapping of 14 real-world samples, which was a very small number of data and hence prone to cause errors and overfitting [67]. Another key drawback of their approach is the inclusion of CA assessment scores. In so doing, they did not get rid of time and resources required to gather and analyze the information from several hundreds of stakeholders. Moreover, assessment scoring is known to be subjective since there is no standard prescriptive guidance to follow [71]. These weaknesses were addressed in the work of Vuttipittayamongkol *et al.* [71]. The authors introduced the first publicly available oil and gas decommissioning dataset, where promising results on predictive decommissioning option were presented. As opposed to the earlier work of Martin *et al.* [46], CA scores were not taken into account and hundreds of real samples were used in the experiments. However, the limitation of this work is that out of many types of oil and gas infrastructures [15], only pipeline was considered.

III. DATASET

The new oil and gas decommissioning dataset is composed of 1,846 instances, each of which represents the decommissioning activity of an offshore infrastructure. Table III shows part of the dataset in the CSV file. The full dataset is made available online (See GitHub[3]). The class label is the final decommissioning option: Full Removal, Partial Removal, Leave In-Situ. Selection of features that potentially influence decommissioning decision-making was based on an expert's review of decommissioning guidelines of various industries worldwide. All types of offshore infrastructures, which comprises a total of 17 types, are included. It has to be noted that we group these structure types into two categories regarding the difference in features for classification purposes. Detailed analysis and discussion on the dataset is given below.

A. DATA COLLECTION AND EXTRACTION

We extracted 1,846 decommissioning activities from 120 decommissioning program reports. The reports are open to the public in the OPRED database,[2] the sole source of oil and gas decommissioning reports in the United Kingdom landscape. The 120 decommissioning programs were under-

[3]https://github.com/fonkafon/OilGas/blob/main/Dataset_Full_Updated_18June2021.csv

TABLE 1. Example of the new oil and gas decommissioning dataset.

Company Name	Structure	Type	Technical	Safety	Environmental	Societal	Cost	Weight	Size	Dia.	Length	Metal	Plastic	Concrete	Residues	Position	Decision
BP Exploration (Alpha) Limited	PL148	Pipeline	Full Removal	Partial Removal	Full Removal	Full Removal	Leave In-Situ			6	24.968	Yes	Yes	No	Chemicals	Trenched and Burried	Partial Removal
Shell U.K. Limited	PL80	Pipeline	Full Removal	Partial Removal	Full Removal	Full Removal	Leave In-Situ			20	3.88	Yes	No	No	Hydrocarbon	Trenched and Burried	Leave In-Situ
Marathon Oil UK LLC	PLU4217	Cable	Full Removal	Full Removal	Full Removal	Full Removal	Leave In-Situ			5.3	14.5	Yes	Yes	No	None	Trenched and Burried	Leave In-Situ
Repsol Sinopec Resources UK	PLU2652	Cable	Full Removal	Full Removal	Full Removal	Full Removal	Leave In-Situ			6	10.95	Yes	Yes	No	None	Trenched and Burried	Leave In-Situ
Spirit Energy North Sea Limited	PL1099	Pipeline	Full Removal	Partial Removal	Full Removal	Full Removal	Leave In-Situ	328.78	1080	4	15.1	Yes	Yes	No	None	Trenched and Burried	Partial Removal
Ithaca Energy (UK) Limited	Athena Manifold	Manifold	Full Removal	Full Removal	Full Removal	Full Removal	Leave In-Situ	200				Yes	No	No	None	Seabed Laid	Full Removal
Tullow Oil SK Limited	Orwell Manifold	Manifold	Full Removal	Full Removal	Full Removal	Full Removal	Leave In-Situ	25				Yes	No	No	None	Seabed Laid	Full Removal
Perenco UK Limited	Gawain Manifold	Manifold	Full Removal	Full Removal	Full Removal	Full Removal	Leave In-Situ	1.8	3.978			Yes	No	No	None	Seabed Laid	Full Removal
Nexen Petroleum UK Limited	Ettrick Valve Skids 1	Subsea	Partial Removal	Partial Removal	Full Removal	Full Removal	Leave In-Situ	6760				Yes	Yes	No	None	Seabed Laid	Full Removal
Spirit Energy North Sea Limited	DP4 Platform Topside	Topside	Full Removal	Full Removal	Full Removal	Full Removal	Leave In-Situ	6500				Yes	No	No	None	Surface	Full Removal
Neptune E&P UKCS Limited	Minke Rock Cover	Rock Cover	Partial Removal	Partial Removal	Partial Removal	Full Removal	Leave In-Situ					No	No	Yes	None	Seabed Laid	Leave In-Situ

Description	Pipeline Number (as per PWA)	Diameter (inches)	Approx. Length (km)	Description of Component Parts	Product Conveyed	From – To End Points	Burial Status	Pipeline Status	Current Content
Chemical and Control Umbilical Jumper	PLU1496	75mm	0.027	Electro / Hydraulic / Chemical Control Umbilical	Methanol, Mineral Based Hydraulic Fluid	From Kingfisher Manifold UTA to Well HP2	Surface laid and covered in concrete mattresses	Non-Operational	Sea water, trace methanol, water based hydraulic fluid, mineral based hydraulic fluid

FIGURE 2. Example of part a decommissioning program report from which we extracted the data.

taken by 31 different oil and gas companies and approved by OPRED during 2000-2021. Each program report contains four documents: 1) decommissioning proposal, 2) comparative assessment report, 3) environmental statement and 4) stakeholder engagement report. Figure 2 shows part of an information table in a decommissioning program report from which we extracted data. In column 2 of the example, PLU means umbilical. Electro in column 5 refers to electrical parts, which implies that the structure was made of metal. Information in column 6 and 10 indicates that chemical residues were present. It is also worth noting that the reports are in different formats. Thus, extracting data from hundreds of these documents requires not only a great deal of effort but also multidisciplinary knowledge of an expert in engineering, management, law, data science, etc. This is one of the main reasons that oil and gas decommissioning datasets are scarcely available to the research community.

The rationale that we selected the oil and gas UK landscape for the study is as follows. Firstly, the source is publicly available unlike other landscapes such as Thailand, where data is only accessible to oil and gas operators in a Production Sharing Agreement. Secondly, its public database of decommissioning project reports is the largest source in the world.

Table 2 contains the description of 17 features and the class label. These features were selected based on extensive literature review [11], [15], [18], [64], [66] and the analysis of the report documents. It is found that the type of oil and gas offshore infrastructure influences the decommissioning option. This is because each type of structures also presents its own safety, technical, environmental, social and economic challenges [15], [23], [64]. Similarly, other technical specifications, namely, weight, size, diameter, length, materials (metal, plastic, concrete), residues and position of the structure also impact the determined outcome of the CA process.

Numerical values of Weight, Size, Diameter and Length needed to be converted into the same units to allow appropriate comparison. In general, the larger and heavier the offshore infrastructure, the more difficult it is to be fully removed [18]. The material that makes up the infrastructure can have an effect on the decommissioning option because of potential environmental impacts. Plastic materials, for example, are preferably removed in full as they can degrade and release harmful chemicals into the marine environment [58].

With regard to the residues, they are determined by the function of the infrastructure. Storage tanks, for example, are used for the purpose of storing hydrocarbons prior to being transported to the shore. As such, despite cleaning and flushing efforts, storage tanks are expected to contain hydrocarbon residues. Similarly, umbilicals are used to transport chemicals, e.g. methanol, and hence are expected to contain some chemical residues. Residues would have an impact on the decommissioning option as removing the infrastructure would eliminate the risk of these residues leaking into the marine environment, in the case the integrity of the infrastructure fails [17].

The position of the structure can either be surface (above the waterline), seabed laid (on the seabed) or trenched and buried. Buried pipelines, for example, are more difficult to be fully removed as compared to surface laid pipelines. Extensive dredging of the seabed is required to expose the pipelines so that they can be accessed by cranes for removal. Such an activity, which may cause leakage in the pipelines, is an environmental concern that affects the decommissioning decision [13].

As can be seen in Table 2, the qualitative analysis values of the five aspects in the CA, namely, technical, safety, environmental, societal and cost, are also included in the dataset. Following the OPRED guidance,[2] decommissioning program reports of oil and gas fields in the UK landscape must include the comparative assessment to incorporate stakeholders' opinions. As discussed earlier, these variables are prone to be subjective and require tedious efforts. In later sections, extensive analyses will be carried out to determine the redundancy of the five aspects with other features and the plausibility of excluding them in the classification task.

Finally, the class label is the decommissioning decision approved and adopted for the activity. While there are different decommissioning sub-methodologies, the eventual decommissioning decision can largely be classify into three main categories: full removal, partial removal and leave in-situ [11]. Single-lift, piece-small, and multiple-lift methodologies, for example, are all sub-categories of full removal. The reason for not considering multiple sub-categories of decommissioning options is because firstly, they can be largely influenced by external factors such as the availability of tools and vessels, rather than actual features of the infrastructure itself. Secondly, many of these sub-methodologies are only assessed after the comparative assessment process through further front end engineering and design, and negotiations with the supply chain.

TABLE 2. Feature description.

Feature	Description
Company Name	Oil and gas company undertaken the decommissioning activity
Structure	Name of the offshore structure
Type	Type of the offshore structure
Technical	Qualitative analysis value of the technical aspect
Safety	Qualitative analysis value of the safety aspect
Environmental	Qualitative analysis value of the environmental aspect
Societal	Qualitative analysis value of the societal aspect
Cost	Qualitative analysis value of the cost aspect
Weight	Weight of the structure (ton)
Size	Volume of the structure (m^3)
Diameter	Diameter of the structure (inch)
Length	Length of the structure (km)
Metal	The structure was made of metal: 0 = no, 1 = yes
Plastic	The structure was made of plastic: 0 = no, 1 = yes
Concrete	The structure was made of concrete: 0 = no, 1 = yes
Residues	Type of toxic fluids in the structure: hydrocarbons, Chemicals or None
Position	Position of the structure: Surface, Seabed Laid or Trenched and Buried
Decision	The determined outcome of the CA process: Full Removal, Partial Removal, Leave In-Situ

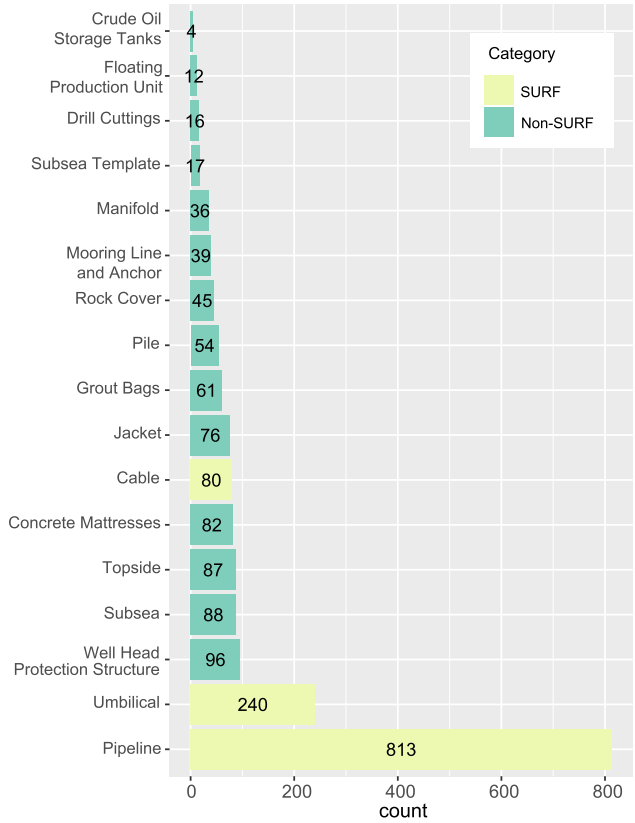


FIGURE 3. Data distribution by type and category.

B. EXPLORATORY DATA ANALYSIS

In this section, we will explore in detail the characteristics of each feature in the dataset. Since different types of structures have different forms, some features will be different and hence in later sections classification will have to be carried out separately. We subset the dataset into 2 categories that are 1) Subsea Umbilicals, Risers and Flowlines (SURF) and 2) Non-SURF. SURF will have Diameter and Length due to their cylindrical shapes whereas other types of structures, which are at location, have Weight and Size. Detailed discussion of the categorical and discrete features will be provided followed by discussion of the continuous features.

1) TYPE

As shown in Figure 3, which presents the distribution of data by type, there are 17 types of offshore structures. Pipeline, umbilical and cable are in the categories of SURF whereas the remaining types belong to Non-SURF. There are a total of 1,133 instances in SURF and the remaining 713 are in Non-SURF. Not surprisingly, pipeline, which is the most common types of offshore infrastructures [61], is the majority in the dataset.

2) CA ASPECTS

Figure 4 shows the distribution of each of the fives aspects in CA in the SURF category. Interestingly, all 1,133 instances have the same values in each of Technical, Environmental, Societal and Cost. This can be justified as follows. Based on the sizes and weights of SURF structures, technically, all could be fully removed. However, fully removing a SURF structure requires cutting, dredging, and exposure of personnel to harsh offshore environment for a long period of time hence compromising safety. As such, it is better to remove the structure partially. Environmentally, fully removing all SURF would be better to eliminate the risks of residues leaking into the marine environment, in the case that the SURF element degrades over time. It will also revert the seabed to a pristine condition prior to oil and gas exploration. Societal-wise, removing SURF structures would ensure the safety of fishermen and other users of the sea. In terms of cost, the operation will be the cheapest if the structure is just left in-place. Therefore, in the classification task, it is clear that the four features with single values, namely, Technical, Environmental, Societal and Cost should be dropped as they will not have any contribution.

For the Non-SURF category, as can be seen in Figure 5, there are different values in each aspect. Even though there are some values with small frequency, they should not be ignored when performing classification. Technically, most infrastructures in the Non-SURF category could be fully removed. With further exploration, we found that the remaining structures that should not be fully removed are extremely large and heavy jackets. These jackets were designed to be installed; hence, no considerations was given to its removal.

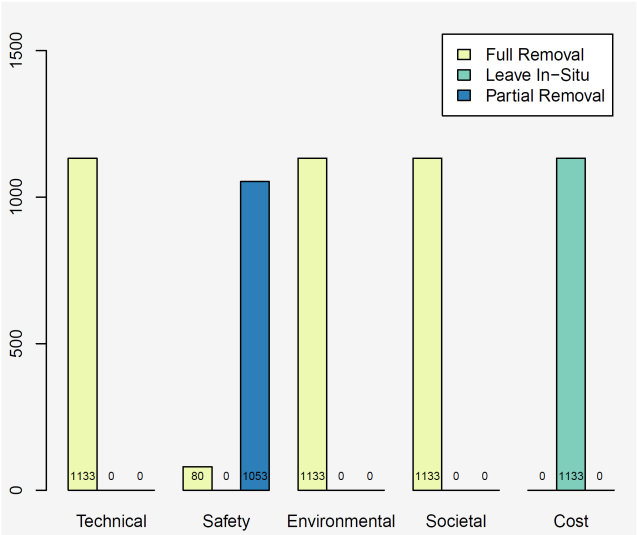


FIGURE 4. Distributions of the five aspects in CA in SURF.

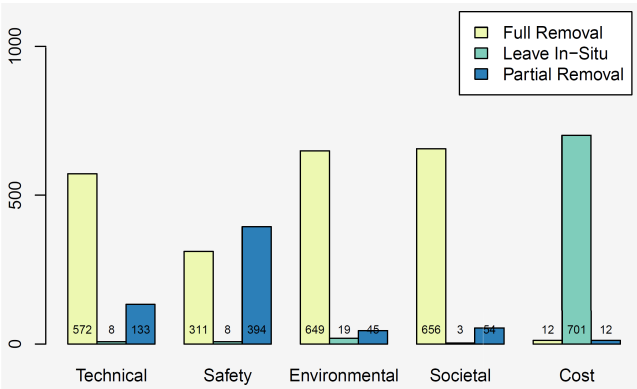


FIGURE 5. Distributions of the five aspects in CA in Non-SURF.

Based on current technology, it is technically not feasible to remove them fully.

As for Safety, the 311 samples of full removal are the smaller subsea equipment and floating production units. All these equipment can be easily removed without significantly compromising the safety of the personnel conducting the removal work. The same eight cases of leave in-situ for Technical and Safety are significantly heavy concrete gravity-based jackets, which were not even feasible to be partially removed. Similarly, the 394 partial removal are large structures. Full removal would pose risks to the personnel conducting the removal.

In the environmental aspect, full removal is generally preferred because it eliminates the risk of residues leaking. The 19 Leave In-Situ cases consists of concrete gravity-based jackets and drill cuttings. Removing concrete gravity-based jackets requires a great deal of effort and power, which could result in significant carbon emissions and disturbance to the marine environment whereas disturbing drill cuttings will release toxic materials into the marine environment. As such, it is environmentally best to leave them in-situ. The 45 partial removal items are drilling piles. Because they were hammered deep into the seabed, removing them causes seabed disturbance. As such, in the environmental aspect, it would be best to partially cut the drilling piles and remove them.

The societal aspect is mainly driven by the impact to commercial fisheries. Full removal is generally the most preferred when possible. The partial removal cases are piles, which could only be partially removed. The three leave in-situ cases are large concrete-based jackets, which are preferred to be left in-situ so that the legs can be seen above the waterline by other users of the sea. In such cases, visual aids are usually installed on the concrete-based jackets to further enhance visibility.

In terms of cost, for the majority of the oil and gas facilities, it is cheaper to not remove them; thus, leaving in-situ is dominating here. However, the 12 floating production units are better to be fully removed as there would not be much economical benefits of leaving them in-situ. For the other 12 partial removal cases, these are the moorings and anchor chains linked to the 12 floating production units. Removing them fully would cost more than partially removing, but there is not much difference in terms of cost between partial removal and leaving in-situ. As such, partial removal was preferred.

3) MATERIALS: METAL, PLASTIC AND CONCRETE

Three common materials of offshore oil and gas structures are metal, plastic and concrete. A structure may be made with a combination of two or more types of materials. Different types of materials affect the decommissioning decision differently. For example, a metal structure may have a very high weight, or a structure containing plastic can degrade over time [58].

All SURF items are generally metal pipes (made of steel, aluminum or other composites of high grading). As such, almost all of the SURF items contain metal as evident in Figure 6. There are few exceptions where structures are solely

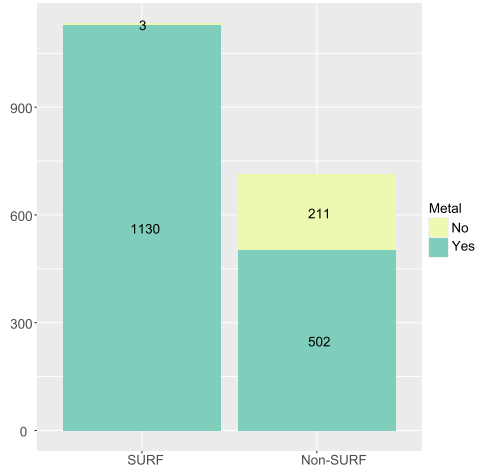


FIGURE 6. Distributions of the metal and non-metal structures.

made of plastic. For Non-SURF items, the majority contain metal as the main build-up material. Steel jackets, for example, as its name suggests are made from high-grade steel materials. Subsea components, anchors and moorings, for example, are also made of metal so that they can withstand higher pressure without deforming as compared to plastic materials. However, there are Non-SURF structures that do not contain metal. These are concrete-based items such as concrete gravity-based jackets, concrete mattress and grout bags.

As seen in Figure 7, the majority of SURF structures contain plastic as part of their coating to prevent direct exposure of the metal component to the marine environment. These plastic coatings prevent corrosion and erosion of the metal component to maintain the integrity of the infrastructure throughout its operational lifetime. Another usage of plastic is its flexibility to allow some degree of movements from both transportation of fluids under high pressure flow rates and/or to prevent SURF from buckling under high external metocean forces. However, there are some SURF items that do not contain plastic. These are generally older SURF structures coated with layers of concrete, which were used early in the 1970s and 1980s where plastic manufacturing was not yet popularized.

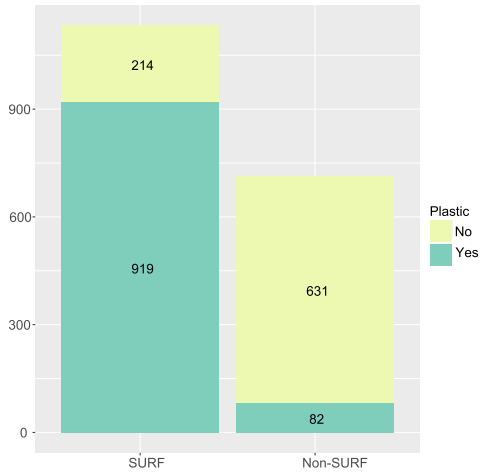


FIGURE 7. Distributions of the plastic and non-plastic structures.

Most Non-SURF items do not contain plastic. Non-SURF structures such as manifolds, jackets, and topsides utilize metal-based protective coatings as a corrosion/erosion protection method. Metal-based protective coatings are used rather than plastic to prevent movements of these items. Jackets and topsides movements should be restricted as much as possible in order to ensure the safety of the workers on the platforms.

From Figure 8, it can be seen that the majority of SURF contain no concrete whereas there are a few exceptions, which is discussed above. Non-SURF materials have a higher concrete-to-non-concrete ratio as compared to SURF because there are elements such as concrete gravity-based structures,

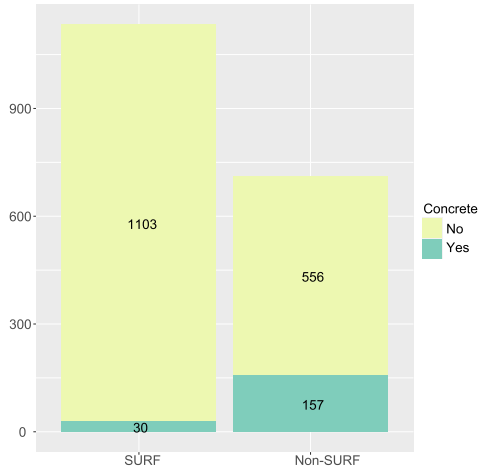


FIGURE 8. Distributions of the concrete and non-concrete structures.

grout bags and concrete mattresses that are solely made of concrete. The majority of the non-SURF items, however, are made of steels because they are much cheaper, quicker and technically easier to design, transport and install.

4) RESIDUES

Hydrocarbons and chemicals are common residues left in offshore structures after cleaning and flushing. Typically, hydrocarbons, which contain radioactive materials, are considered more toxic than chemicals. The presence and type of residues influence the decommissioning decision since they can harm the environment.

As mentioned earlier, SURF are used to transport hydrocarbons and chemicals between the wells and the surface facilities for processing or the shore. It is highly likely that residues of the transport materials would remain in the SURF structures as reflected by Figure 9. Exceptions are cables, which are used for transporting electrical signals to and from

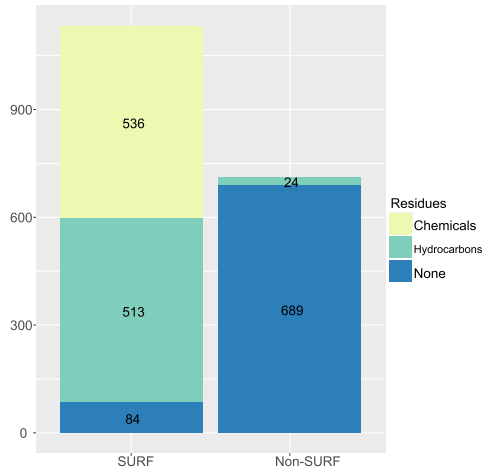


FIGURE 9. Distributions of residue types.

the topside controls for the purpose of controlling the flow rate of hydrocarbons and or chemicals in the other SURF elements. Thus, they do not have any residues.

Figure 9 shows clearly that the majority of non-SURF do not contain residues because they have little to no contact with hydrocarbons or toxic chemicals. Mattresses and grout bags, for example, are just stabilizing features holding SURF items in place. Jackets and piles are support structures that hold up topsides.

5) POSITION

The position of the structure is also important in determining the decommissioning option as it directly impacts the difficulty and safety in the removal. As shown in Figure 10, all SURF are either seabed laid, or trenched and buried. This is due to the fact a SURF item is an infrastructure connecting a surface oil and gas facility to another located on the seabed. The variation in burial status for SURF items largely depends on the metocean conditions in the region, e.g. wind speed, cyclone occurrences and pathways and wave conditions.

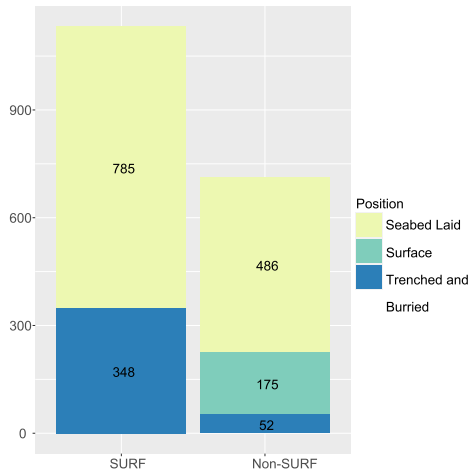


FIGURE 10. Distributions of structure's positions.

The dominance of seabed-laid Non-SURF elements is attributed to mattresses and grout bags being used to stabilize SURF structures on the seabed. Surface facilities, which are visible above the waterline, include floating production units, topsides and jacket. Trenched and buried structures are mostly piles, which are hammered deep into the seabed to act as a secure foundation holding up the surface facilities such as topsides and jackets.

6) DECISION

The distributions of decommissioning decisions, which are the class labels of the dataset, are provided in Figure 11. It is worth noting that the classes in both SURF and Non-SURF categories are not equally distributed. This will be explored in the experiment section.

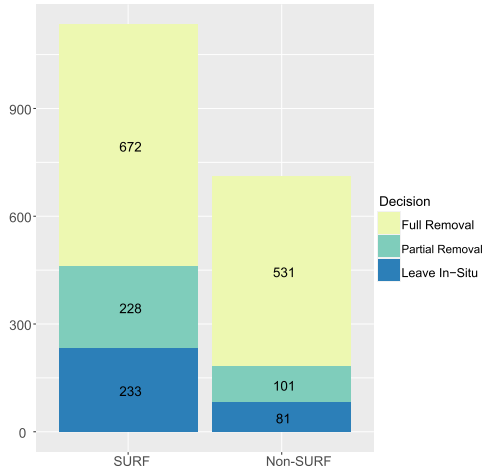


FIGURE 11. Distribution of decommissioning decisions.

7) WEIGHT AND SIZE

As discussed earlier, Weight and Size will be considered in Non-SURF elements only. Figure 12 and Figure 13 present the density plots of weights and sizes, respectively. The graphs suggest that the average weight and size are on the lower end and there are scattered quantities towards the upper end. For clearer visibility of the majority of the values, an inset is given in the figure. Table 3, which provides statistics of the continuous features, clearly informs that there are huge gaps among values in Weight and Size. The largest value is extremely high and the smallest one is extremely low, and the highest value is also significantly far from the mean and the median. However, it has been confirmed by an oil and gas decommissioning expert that these extreme values are valid and are not outliers. They are primarily topsides and concrete gravity-based jackets, which are significantly larger in weight and size compared to

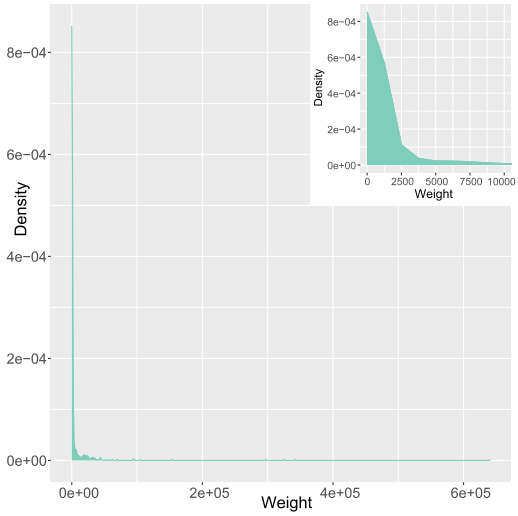


FIGURE 12. Probability distribution of weights.

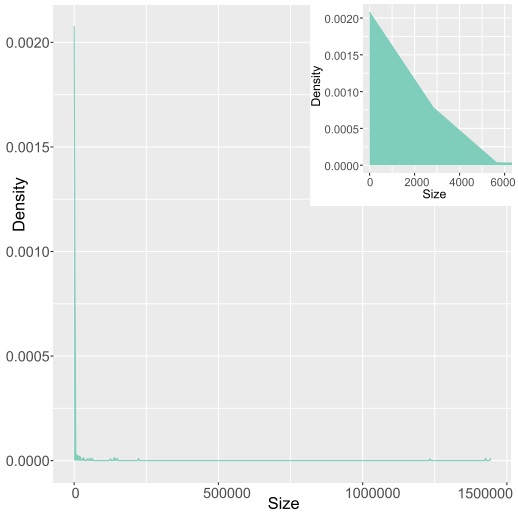


FIGURE 13. Probability distribution of sizes.

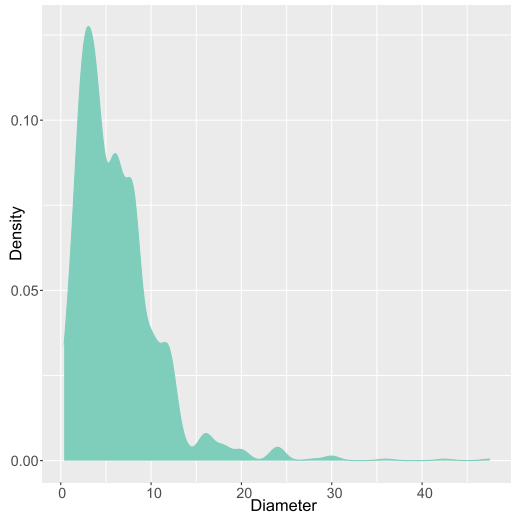


FIGURE 14. Probability distribution of diameters.

TABLE 3. Statistics of the continuous features.

	Weight (ton)	Size (m^3)	Diameter (inch)	Length (km)
Minimum	0.2	0.1	0.375	0.001
Median	196.4	170.2	5.433	0.150
Mean	5,851.8	22,308.9	6.211	17.869
Maximum	640,717	1,443,256	47.5	7,074

other Non-SURF structures. Thus, when training a predictive model, these extreme values should be included but handled carefully. Moreover, the skewed distributions of Weight and Size shown in the insets in Figure 12 and Figure 13 suggest that normalization will be needed in the preprocessing step.

8) DIAMETER AND LENGTH

Diameter and Length are considered for SURF, which are pipe-like structures. The probability distribution of Diameter and Length are presented in Figure 14 and Figure 15, respectively. Although diameters are not normally distributed, fortunately, there is no extreme values in diameters. As can be seen in Figure 14, there are fewer larger structures. These are found to be the main production pipelines, where hydrocabons from smaller in-field pipelines flow into and get transported to the shore. Table 3 shows that the average diameter of SURF structures is 5-6 inches. Values in Length are highly diverse and there exist extreme cases as shown in Figure 15. Very long structures are the main production pipelines whereas very short ones are umbilicals, which are typically congregated less than 2 kilometers.

IV. DATA PREPROCESSING

This section discusses in detail the preprocessing steps we carried out to prepare the data for building predictive models. The steps include handling missing values, redundant features removal and data normalization.

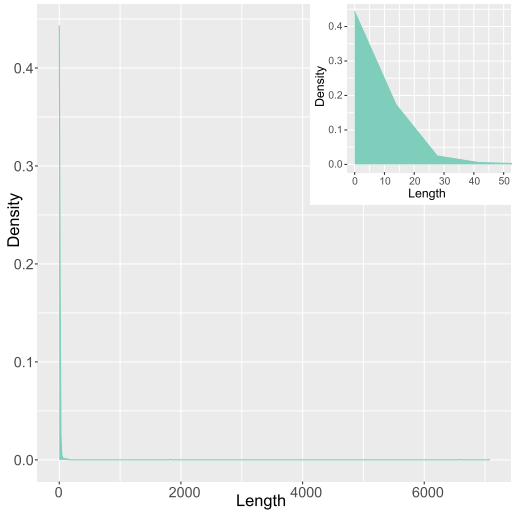


FIGURE 15. Probability distribution of lengths.

A. MISSING VALUES

For SURF, 24 and 2 missing values are found in Diameter and Length, respectively. There are a total of 24 instances with missing values, which accounts for 2.12% of the dataset (1,133 instances). For the ease of convenience and since the remaining instances would be sufficient for the classification purposes, we decided to remove those 24 instances from the dataset. This resulted in 1,109 remaining instances in SURF.

As for Non-SURF structures, there are 19 missing values in Weight and 475 missing values in Size. The missing values in Size is large compared to a total of 713 instances in the dataset. Fortunately, we learned that the Pearson's correlation coefficient of Weight and Size is 0.944 suggesting a high linear correlation between the two features. These are considered redundant features, and one of them should be removed to avoid poor performance of learning algorithms [39]. Thus, we dropped the Size feature from the

dataset and then removed instances with missing values in Weight. As a result, there are 649 remaining instances in the Non-SURF dataset.

B. REDUNDANT FEATURES REMOVAL

We performed feature reduction by eliminating any redundant features from the dataset. This was expected to reduce computational time and improve the learning accuracy [14]. Following a common approach [28], [49], correlations among features were examined to discover redundant features. Correlation analyses were carried out on the remaining instances after missing values and the Size feature were removed from the dataset.

Since there are both numerical and categorical features in the dataset, we performed three different sets of correlation tests: 1) numerical - numerical, 2) numerical - categorical and 3) categorical - categorical. For the correlation between two numerical features, Pearson's correlation coefficient was used. The correlation between numerical and categorical features was tested using Intraclass Correlation Coefficient (ICC). Finally, the Chi-square test was performed to obtain the relationship degree between two categorical features. Note that ordinal features such as Metal, Plastic and Concrete (presence or absence) have been one-hot encoded and were treated as numerical. Nominal features with only two different values, such as Safety in SURF and Cost and Residues in Non-SURF, were also transformed into numerical values of 0 and 1 using one-hot encoding.

In this study, a pair of features that has a correlation coefficient above 0.8 or below −0.8 will be considered redundant [45]. Such a threshold was selected to ensure a sufficiently high correlation while preventing excessive elimination of features.

1) SURF

As can be seen in Figure 16, the Pearson's correlation coefficients among all numerical features in SURF are low. Hence, there is no concerning linear relationship among these variables. In Figure 17, Residues and Safety have a high ICC of 0.97. This high degree of relationship coincides with our discussion about the two features in Section III. To pursue our objective of reducing costs and time in planning a decommissioning project and considering that the residue information can be readily obtainable, it was clear that Safety should be the choice for elimination. In Figure 18, there is no correlation coefficient above the elimination threshold among the categorical features. Thus, only Safety was further removed from the SURF dataset.

2) NON-SURF

Figure 19 shows that there is no concerning linear relationship among the numerical features in Non-SURF. In Figure 20 and Figure 21, it can be seen that Type has high correlations with many other features, namely, Metal, Concrete, Technical, Safety and Position. Thus, removing Type would get rid of the redundancies. Moreover, in Figure 20, Weight appears highly

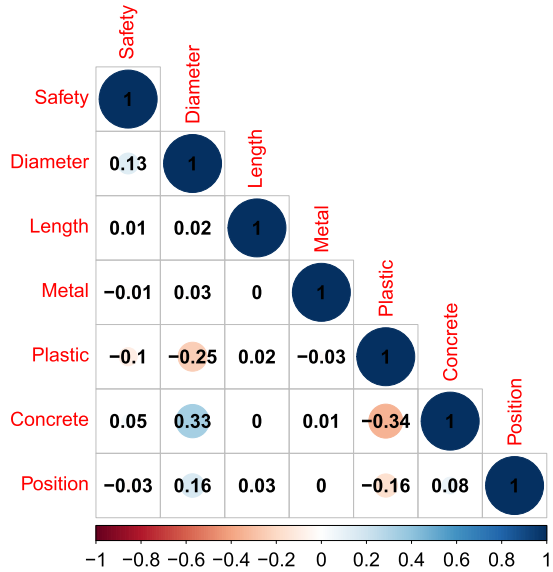


FIGURE 16. Correlations among numerical features in SURF.

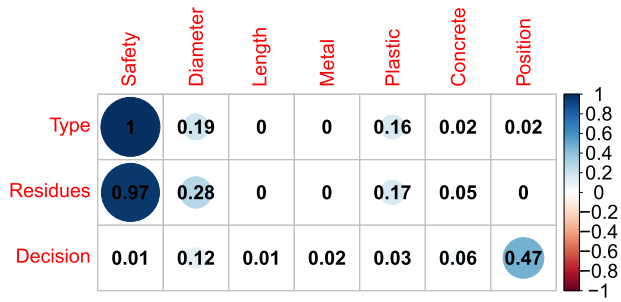


FIGURE 17. Correlations between numerical and categorical features in SURF.

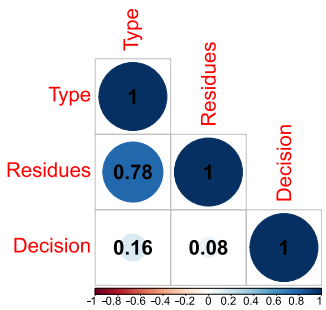


FIGURE 18. Correlations among categorical features in SURF.

correlated with Environmental and Societal. Following our objective to minimize the use of comparative assessment analysis and considering that the weight information can be readily available, Environmental and Societal were the better choice for feature reduction. Thus, Type, Environmental and Societal will be excluded for the classification of the Non-SURF dataset.

C. NORMALIZATION

Since varying scales of numerical features can cause biases during learning of an algorithm, we applied normalization

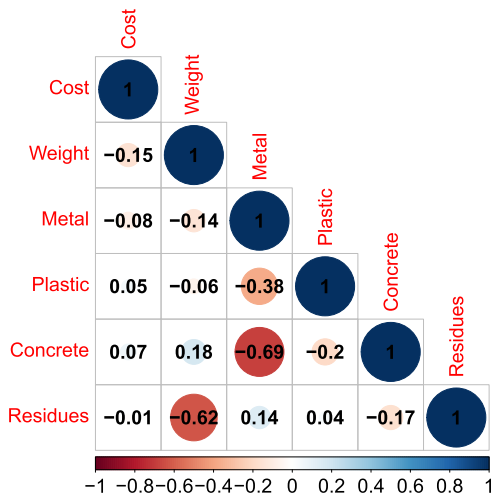


FIGURE 19. Correlations among numerical features in non-SURF.

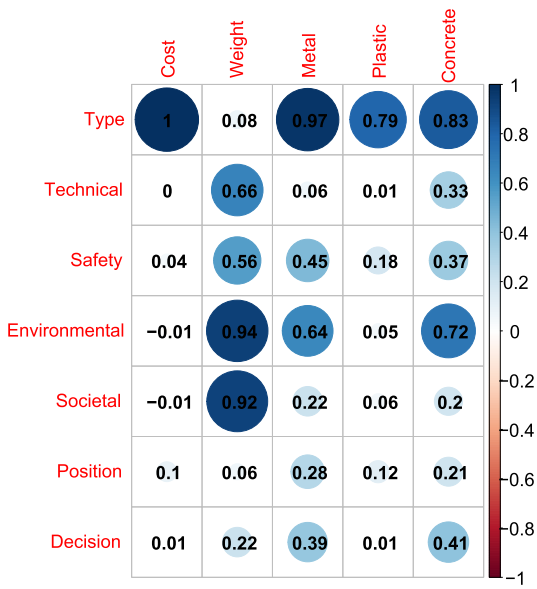


FIGURE 20. Correlations between numerical and categorical features in non-SURF.

to suppress this flaw. Min-Max scaling was used for this purpose. Equation 1 expresses the formula of the Min-Max scaling, where x_{max} and x_{min} are the maximum and minimum values of the feature, respectively. This will scale the feature values to be in the range of 0 and 1.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

V. EXPERIMENTAL SETUP

To validate the applicability of the new dataset for predictive oil and gas decommissioning using machine learning approaches, two sets of experiments were carried out. In Experiment I, several standard machine learning algorithms were used for classification. Firstly, classification results before and after feature reduction were statistically

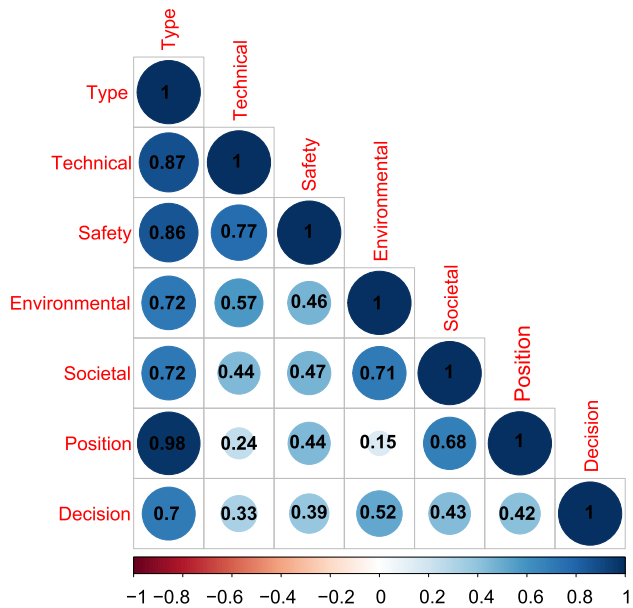


FIGURE 21. Correlations among categorical features in non-SURF.

compared to validate the removal of redundant features. Secondly, results were compared among the classification models to find out the best outcomes on the datasets. Experiment II involved improving the classification results using data resampling methods to tackle the class-imbalance problem. Details of the setups of the experiments including data partitioning, the lists of learning algorithms and resampling methods used along with their parameter settings, and evaluation metrics are provided below.

A. DATA PARTITIONING

For all experiments, the same training and testing sets were used. The dataset was partitioned into 80:20 of training and testing sets. In the training phase, 10-fold cross-validation was employed for the purpose of model selection based on accuracy. Thus, in each round of model building, 72% and 8% of the dataset were used for training and validation, respectively. Lastly, the testing set, which was unseen data, was used for model evaluation.

B. EXPERIMENT I SETUP

In this experiment, selected standard learning algorithms were Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Naive Bayes (NB) and Neural Network (NN). The objective of the experiment is two-fold. First, to present results that will ascertain the validity of the approach to the oil and gas industry since some key features used in the traditional approach such as comparative assessment scores were excluded in the predictive model building. Paired T-test was used to evaluate significance of the differences in the results with and without redundant feature removal. Second, to determine

the best classification result on the presented dataset using commonly-used supervised learning methods.

The default parameter settings of the learning algorithms in the caret package [35] in *R* were used. Some parameters were automatically tuned and selected during the cross-validation. For RF, the number of features determined at each split, $mtry = 2, 3, 5, 7, 9$. The number of trees, $mtree = 500$. For DT, C4.5 decision tree, which is an improved extension of the Iterative Dichotomiser (ID3) algorithm, was chosen. The confidence threshold in the range of [0.01, 0.5] with a step of 0.1225 was examined. The minimum instances per leaf (M) was set to 1, 2,..., 5. The radial bias function kernel was used for SVM with cost $C = 0.25, 0.5, 1, 2, 4$ and $\gamma = \frac{1}{f}$, where f is the number of features in the dataset. In NB, Laplace correction, $fL = 0$ and bandwidth adjustment $= 1$, and models with and without a kernel were compared. For NN, the number of hidden units, $size = 1, 3, \ldots, 9$ and the weight decay, $decay = 0$ and 10^{-d}, where $d = 1, 2, 3, 4$.

C. EXPERIMENT II SETUP

In experiment II, eight resampling methods were applied on the training data in order to address the class-imbalance problem and improve classification results. The methods used were random oversampling (ROS), the condensed nearest neighbor decision rule (CNN) [29], oversampling using Gaussian Noise (GN) [38], WEighted Relevance-based Combination Strategy (WERCS) [9], Synthetic Minority Oversampling Technique (SMOTE) [16], Borderline-SMOTE (BLSMOTE) [26], Density-Based Synthetic Minority Oversammpling Technique (DBSMOTE) [12] and the combination of SMOTE and GN methods (SMOGN) [8].

It should be also noted that there are limited choices of available resampling methods for multi-class datasets [68], [70]. The rationales behind the selection of these methods are their notability, applicability to multi-class datasets, and suitability to the problem. In many classification tasks of imbalanced datasets, the minority class is the most important class where the cost of misclassification can be unacceptably high as compared to that of the majority classes [69]. However, in predictive oil and gas decommissioning, the classes are equally important; hence, the goal is to achieve high accuracy for all classes. These chosen resampling methods can serve such a purpose making them suitable for our problem. We followed the parameter settings for all methods as presented in their original works. For further details of these methods, readers are referred to the references provided.

D. EVALUATION METRICS

To evaluate the classification results, common evaluation metrics for multi-class problems were adopted. In Experiment I, results were compared using average measures. The measures include two different average measures of recall – the geometric mean of recall (G-mean) and the arithmetic mean of recall (mean accuracy), the overall accuracy, the mean precision, the mean F1-score and the overall area under the receiver operating characteristics curve (AUC). In addition to these average measures, detailed results of each class are also presented in Experiment II.

Equation 2 expresses the formula of the recall of class i, where TP_i is the true positive of class i and n_i is the number of test instances in class i.

$$recall_i = \frac{TP_i}{n_i} \tag{2}$$

The three different average accuracies, namely, G-mean, mean accuracy and overall accuracy, were used for extensive evaluation and comparison. The measures are expressed in Equation 3, 4 and 5, respectively, where N_c is the number of classes. The overall accuracy provides a good picture of the total portion of correctly classified cases; however, it can be highly influenced by the majority class [70]. G-mean always gives values less than or equal to mean accuracy [70]. This is because the geometric mean is more affected by lower values, but the mean accuracy weighs all values equally. Thus, G-mean will be useful when detecting significantly low recalls among all classes, especially when there is an occurrence of zero recall. In other situations, the mean accuracy may be preferable as it will not have a bias towards lower values providing more accurate average of the class accuracies.

$$G\text{-}mean = (\prod_{i=1}^{N_c} recall_i)^{\frac{1}{N_c}} \tag{3}$$

$$mean\ accuracy = \frac{\sum_{i=1}^{N_c} recall_i}{N_c} \tag{4}$$

$$overall\ accuracy = \frac{\sum_{i=1}^{N_c} TP_i}{\sum_{i=1}^{N_c} n_i} \tag{5}$$

The precision of class i ($precision_i$) is calculated as in Equation 6. The mean precision of all classes, which we will refer to as *precision* for the ease of convenience, follows Equation 7. Similarly, the formula for the mean F1-score, *F1-score*, is given in Equation 9, where $F1\text{-}score_i$ is F1-score of class i (Equation 8).

$$precision_i = \frac{TP_i}{TP_i + FP_i} \tag{6}$$

$$precision = \frac{\sum_{i=1}^{N_c} recall_i}{N_c} \tag{7}$$

$$F1\text{-}score_i = 2 * \frac{precision_i * recall_i}{precision_i + recall_i} \tag{8}$$

$$F1\text{-}score = \frac{\sum_{i=1}^{N_c} F1\text{-}score_i}{N_c} \tag{9}$$

For AUC, we adopted the calculation of multiclass AUC defined by Hand and Till [27], which is a widely recognized method for multi-class problems. The formula is given in Equation 10, where i and j are two different classes and $AUC_{i,j}$ is the AUC of the class i and class j pair.

$$AUC = \frac{2}{N_c * (N_c - 1)} * \sum_{i,j} AUC_{i,j} \tag{10}$$

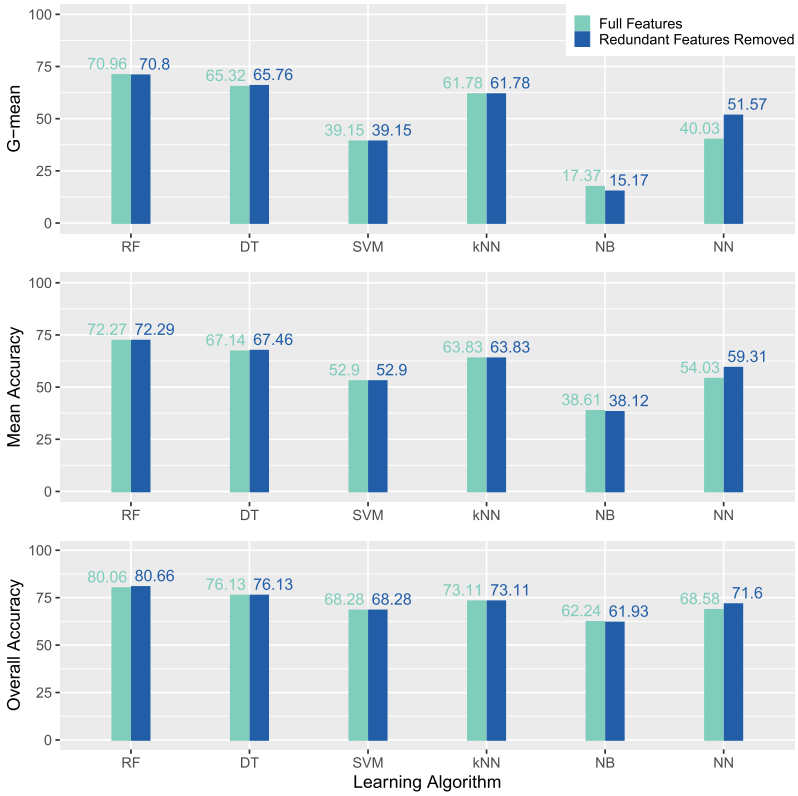


FIGURE 22. Average accuracies of different learning algorithms on SURF.

VI. RESULTS AND DISCUSSION

This section provides detailed results and discussion in the two sets of experiments.

A. EXPERIMENT I

In this experiment, results are presented and discussed according to the two aforementioned objectives. Firstly, classification results before and after removing redundant features are thoroughly examined and validated using a statistical tool. This is followed by a comparison among the results achieved using different standard learning algorithms.

1) FEATURE REDUCTION VALIDITY

In Section IV, we determined correlations among the features to minimize redundancy in the dataset. Results suggest that some features related to the comparative assessment should be excluded in classification. Since these features are key aspects in the traditional approach of resolving the final oil and gas decommissioning option, we validated the removal by comparing the classification results carefully using paired T-tests. Results are presented in Figure 22 - Figure 25 and Table 4, where the first three rows show the accuracy of each class.

In the SURF dataset, only Safety was found to be redundant with some other features and hence removed. As can be seen in Figure 22 and Figure 23, the classification results on the SURF subset using full features and reduced features were comparable on most algorithms in all measures. The p-values of paired T-tests in Table 4 also confirm this finding. That is, all p-values on SURF are greater than 0.05 suggesting that at the significance level of 0.05 there is no strong evidence to support that the results on full and reduced features are statistically different.

TABLE 4. P-values of paired T-tests.

Measures	SURF	Others
Full Removal	0.792613	0.363217
Leave In-Situ	0.414747	0.136624
Partial Removal	0.4352	0.418716
G-mean Accuracy	0.464019	0.268421
Mean Accuracy	0.381502	0.193
Overall Accuracy	0.327298	0.152847
Precision	0.94434	0.356875
F1-Score	0.408029	0.211028
AUC	0.610028	0.33686

Figure 22 and Figure 23 show that results with SVM and kNN remained unchanged whereas RF and DT were hardly affected by the removal. This can be attributed to the fact that only one feature was removed from SURF. Moreover, most of these algorithms have some advantageous properties to deal

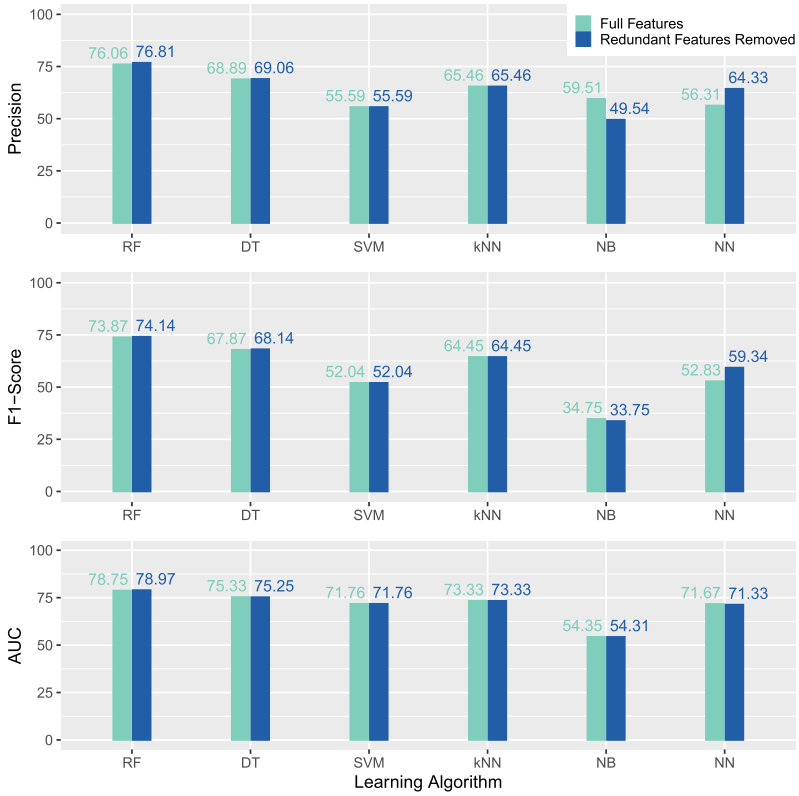


FIGURE 23. Precision, F1-score and AUC of different learning algorithms on SURF.

with redundant features. For example, the C4.5 decision tree algorithm can prune redundant trees whereas RF adopts bootstrapping and random feature selection. SVM is independent of the feature space dimensionality and uses regularization to avoid over-fitting [48]. Fortunately, in our experiments, the regularization parameter C was tuned and properly set during cross validation. This helped reduce the issue of redundant features in the dataset. Similarly, the results with NB were slightly impacted by the feature removal. This is evidenced by small changes in most measures. The reduction in G-mean of NB, which was relatively more noticeable, was due to the bigger influence of the smaller class accuracy as discussed in Section IV. Similarly, the reduction in precision was greatly impacted by the domination of FP of the bigger class even though the accuracy of the class of interest (TP) contributed to a higher class accuracy rate (recall). Interestingly, results with NN were clearly improved in most measures. This could be attributed to performance improvement of NN once the redundancy was eliminated. Automatic feature selection is known to be one of the main advantages of NN. However, its low performance on SURF with the full features could be due to the use of small-sized training samples, which reduced its ability in feature selection.

In the Non-SURF dataset, Type, Environmental and Societal were found to have high correlations with some other features and hence excluded from the classification process. Results for Non-SURF are shown in Figure 24 and Figure 25. Similar to the results on SURF, most of the algorithms had quite stable performance regardless of the feature removal. These algorithms were RF, DT, kNN and NN, which provided unchanged or slightly changed results in all measures. In contrast, classification results using SVM and NB clearly decreased. This is not surprising as classification results can be dependent on both learning algorithms and the dataset. Since on some datasets, more features were preferred for SVM to produce the best separating hyperplane. It is evidenced in the report of Salimi *et al.* [57], where reducing features resulted in significant decreases in classification accuracy on most datasets. In the same manner, it was demonstrated in [3] that removing redundant features sometimes greatly hurt the performance of NB.

In conclusion, we have shown that it is practical to reduce features in determining the oil and gas decommissioning option. This could be achieved using machine learning algorithms that are robust to the changes such as RF, DT, kNN and NN. It is worth noting that the removed features are the key features used in the traditional approach for decommissioning decision-making. Specifically, they are factors in the CA process, which involve gathering requirements and opinions from several hundreds of stakeholders. This process usually takes years to complete. Thus, the findings in this experiment would be useful in convincing the oil and gas

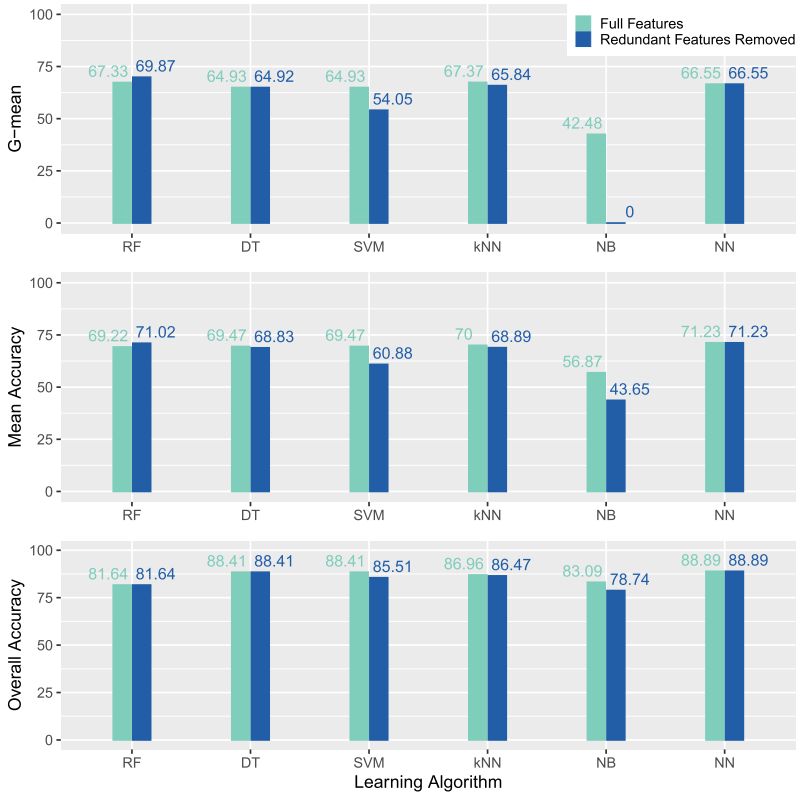


FIGURE 24. Average accuracies of different learning algorithms on Non-SURF.

industry and its stakeholders that some traditional practices could be eradicated to save time and costs significantly.

2) LEARNING ALGORITHM PERFORMANCE COMPARISON

In this part, classification results achieved on the datasets with reduced features using standard supervised learning methods are compared. For SURF, Figure 22 and Figure 23 show clearly that RF provided the best results in all measures. It gave G-mean of 70.8%, the mean accuracy of 72.29%, the overall accuracy of 80.66%, precision of 76.81%, F1-score of 74.14% and AUC of 78.97%. The next best overall results were achieved using DT, kNN and NN, respectively whereas SVM and NB had the lowest accuracy.

For Non-SURF, as can be seen in Figure 24, RF and NN provided competitive accuracy and were among the best algorithms. RF gave the highest G-mean of 69.87% whereas NN had the highest mean accuracy of 71.23% and the highest overall accuracy of 88.89%. Figure 25 shows that NN also achieved the highest precision of 93.36% and the highest F1-score of 78.05% while its AUC of 71.4% was competitive with that of RF and kNN. Thus, it can be said that the best overall results were of NN. The results with DT and kNN were also among the top, however with lower G-mean and mean accuracy than RF and NN suggesting that the accuracy of a smaller class, i.e. Partial Removal or Leave In-Situ, was relatively low. Lastly, results with SVM and NB were the lowest.

B. EXPERIMENT II

The objective of Experiment II was to improve the classification results using different resampling techniques to address the imbalanced class distribution of the decommissioning dataset. In Experiment I, it was shown that promising classification results on the decommissioning dataset can be obtained using standard learning algorithms. RF provided the highest accuracy on SURF and was among the algorithms that gave the best classification results on Non-SURF. Moreover, it was shown to be robust to the feature reduction. For this reason, we selected RF as the baseline for the purpose of demonstration in this experiment.

Table 5 and Table 6 present detailed results of RF with several resampling methods. *None* denotes the baseline, which is RF with no data resampling applied. Recall, Precision, F1-score of each class along with their means and AUC are provided. The **bold** results indicate results that were improved from the baseline.

For SURF, as can be seeen in Table 5, all resamping methods but WERCS improved the accuracy in predicting the minority class(es), i.e. Leave In-Situ and Partial Removal. ROS, DBSMOTE and SMOGN helped increase the accuracy in both minority classes leading to higher precision of the majority class (Full Removal) and improvement in the overall AUC. ROS achieved such improvements while maintaining competitive mean recall, mean precision and mean F1-score with the baseline. This suggests that ROS

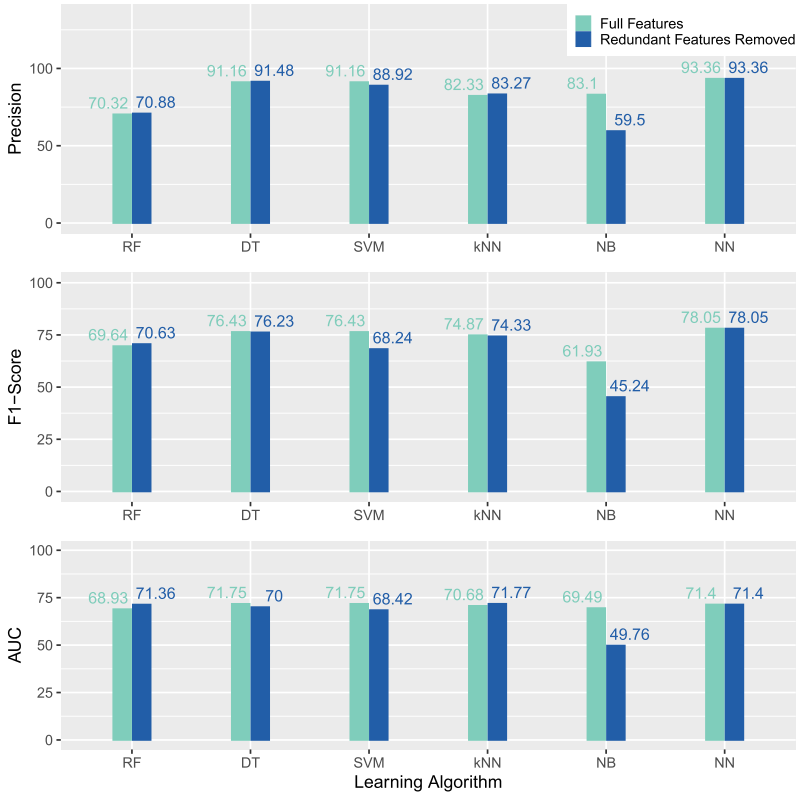


FIGURE 25. Precision, F1-score and AUC of different learning algorithms on Non-SURF.

TABLE 5. Classification results on SURF with different resampling methods.

		None	ROS	CNN	GN	WERCS	SMOTE	BLSMOTE	DBSMOTE	SMOGN
Recall	Full Removal	93.88	89.29	5.61	86.73	89.29	82.14	88.78	86.22	81.63
	Leave In-Situ	60.29	**64.71**	22.06	**67.65**	55.88	**63.24**	58.82	**63.24**	**66.18**
	Partial Removal	62.69	**67.16**	**85.07**	62.69	62.69	62.69	**64.18**	**68.66**	**68.66**
Mean Recall		72.29	**73.72**	37.58	**72.36**	69.28	69.35	70.59	**72.71**	72.16
Precision	Full Removal	84.79	**87.94**	44	**87.63**	**87.06**	**90.96**	**87.44**	**89.42**	**90.91**
	Leave In-Situ	73.21	70.97	18.52	66.67	66.67	53.75	66.67	65.15	60
	Partial Removal	72.41	64.29	25.33	61.76	57.53	56.76	59.72	60.53	57.5
Mean Precision		76.81	74.4	29.28	72.02	70.42	67.16	71.28	71.7	69.47
F1-Score	Full Removal	89.1	88.61	9.95	87.18	88.16	86.33	88.1	87.79	86.02
	Leave In-Situ	66.13	**67.69**	20.13	**67.15**	60.8	58.11	62.5	64.18	62.94
	Partial Removal	67.2	65.69	39.04	62.22	60	59.57	61.87	64.34	62.59
Mean F1-Score		74.14	74	23.04	72.18	69.65	68	70.82	72.1	70.51
AUC		78.97	**79.66**	56.36	78.03	77.41	**79.01**	78	**80.01**	**79.17**

was an effective resampling method for the SURF dataset, which provided improvements in the minority classes with a desirable trade-off with the majority class' accuracy. GN, SMOTE and BLSMOTE resulted in higher accuracy in one of the minority classes. However, the overall improvements and tradeoffs were not as good as those of ROS and DBSMOTE. Lastly, CNN was the only method that failed to improve the results appropriately. It led to severe decreases in all average measures.

For Non-SURF, Table 6 shows clearly that the result of applying GN was outstanding. The method led to improvements in all measures of all classes. DBSMOTE also contributed to the increases in all average measures. Similarly, WERCS and SMOTE provided competitive average results

TABLE 6. Classification results on Non-SURF with different resampling methods.

		None	ROS	CNN	GN	WERCS	SMOTE	BLSMOTE	DBSMOTE	SMOGN
Recall	Full Removal	87.97	86.08	75.95	**88.61**	**92.57**	83.54	85.44	82.28	82.28
	Leave In-Situ	68.42	68.42	36.84	**73.68**	**70**	**73.68**	52.63	68.42	68.42
	Partial Removal	56.67	56.67	**66.67**	**63.33**	48.72	**70**	**60**	**70**	**73.33**
Mean Recall		71.02	70.39	59.82	**75.21**	70.43	**75.74**	66.02	**73.57**	**74.68**
Precision	Full Removal	91.45	91.28	**95.24**	**93.33**	86.71	**94.29**	**93.1**	**92.86**	**96.3**
	Leave In-Situ	76.47	72.22	36.84	**77.78**	73.68	66.67	71.43	**81.25**	43.33
	Partial Removal	44.74	42.5	32.26	**48.72**	**63.33**	**45.65**	37.5	41.18	**52.38**
Mean Precision		70.89	68.67	54.78	**73.28**	**74.57**	68.87	67.34	**71.76**	64
F1-Score	Full Removal	89.68	88.6	84.51	**90.91**	89.54	88.59	89.11	87.25	88.74
	Leave In-Situ	72.22	70.27	36.84	**75.68**	71.79	70	60.61	**74.29**	53.06
	Partial Removal	50	48.57	43.48	**55.07**	**55.07**	**55.26**	46.15	**51.85**	**61.11**
Mean F1-Score		70.63	69.15	54.94	**73.89**	**72.13**	**71.28**	65.29	**71.13**	67.64
AUC		71.36	70.66	70.85	**75.12**	**74.6**	**76.39**	71.02	**73.91**	**80.1**

with the baseline. SMOGN improved the mean recall and AUC but did not give a good trade-off among the accuracy of classes as the mean precision and F1-score decreased. ROS, CNN and BLSMOTE did not help improve any average measures.

In this experiment, we have shown a potential approach in improving the classification on the decommissioning dataset that is rebalancing the class distribution. The selected resampling methods use different techniques to achieve a balanced distribution of the classes and led to different results. Nonetheless, it was shown that most of these methods helped improve the classification. Our experimental findings suggest that ROS and GN resulted in the highest improvements in the classification of SURF and Non-SURF, respectively.

VII. CONCLUSION AND FUTURE WORK

In this paper, we introduced a new oil and gas decommissioning dataset and presented an experimental framework using machine learning for predictive decommissioning options. The new dataset comprises extensive information of decommissioning activities during 2000 - 2021. All types of offshore oil and gas infrastructures were considered. Data exploratory including correlation analyses to remove potential redundant features were carried out. This was followed by classification of the dataset using several standard machine learning algorithms. Promising accuracies were achieved even with some key features used in the traditional approach removed. Collecting and analyzing information of these features usually takes many years and a great deal of resources to complete. Thus, these results provide a good level of confidence to the oil and gas industry and its stakeholders in incorporating machine-learning approaches in the decommissioning planning. This would eradicate some unnecessarily lengthy processes and reduce time and costs in the activity significantly. Moreover, we presented the use of data resampling techniques to tackle the imbalanced class distribution of the dataset and enhance the classification results. Various rebalancing methods that are capable of handling multi-class problems were employed, and most showed favorable result improvements. The introduction of the new dataset fills the lack of datasets readily available to the research community, which is one of the main causes of extremely limited machine learning-based studies in this topic. The encouraging results on the predictive decommissioning option sheds light on further efforts on this timely and challenging problem.

Potential future direction of this work includes considering decommissioning activities in other oil and gas landscapes. Different criteria may be used for decision making and hence for predictive decommissioning. Accessing industries' reports is challenging since not all are publicly available. However, more studies and successful outcomes of machine learning-based approaches in oil and gas decommissioning will eventually allow the government and related bodies around the world to see the importance of publicizing all useful information. Another interesting direction would be to further improve the classification results from the baseline given in this work. More recent and complicated resampling techniques such as genetic algorithm-based and deep-learning based techniques may be explored.

REFERENCES

[1] A. K. Abbas, N. A. Al-Haideri, and A. A. Bashikh, "Implementing artificial neural networks and support vector machines to predict lost circulation," *Egyptian J. Petroleum*, vol. 28, no. 4, pp. 339–347, Dec. 2019.

[2] F. Abdullayeva and Y. Imamverdiyev, "Development of oil production forecasting method based on deep learning," *Statist., Optim. Inf. Comput.*, vol. 7, no. 4, pp. 826–839, Dec. 2019.

[3] J. Abellán and J. G. Castellano, "Improving the naive Bayes classifier via a quick variable selection method using maximum of entropy," *Entropy*, vol. 19, no. 6, p. 247, Jun. 2017.

[4] M. Aljubran, J. Ramasamy, M. Bassam, and A. Magana-Mora, "Deep learning and time-series analysis for the early detection of lost circulation incidents during drilling operations," *IEEE Access*, vol. 9, pp. 76833–76846, 2021.

[5] H. H. Alkinani, A. T. T. Al-Hameedi, and S. Dunn-Norman, "Data–driven decision–making for lost circulation treatments: A machine learning approach," *Energy AI*, vol. 2, Nov. 2020, Art. no. 100031.

[6] A. M. AlRassas, M. A. Al-qaness, A. A. Ewees, S. Ren, M. A. Elaziz, R. Damaševičius, and T. Krilavičius, "Optimized ANFIS model using Aquila optimizer for oil production forecasting," *Processes*, vol. 9, no. 7, p. 1194, Jul. 2021.

[7] L. AlSuwaidan, "The role of data management in the industrial Internet of Things," *Concurrency Comput., Pract. Exper.*, vol. 2020, p. e6031, Sep. 2020.

[8] P. Branco, L. Torgo, and R. P. Ribeiro, "SMOGN: A pre-processing approach for imbalanced regression," in *Proc. 1st Int. Workshop Learn. Imbalanced Domains, Theory Appl.*, 2017, pp. 36–50.

[9] P. Branco, L. Torgo, and R. P. Ribeiro, "Pre-processing approaches for imbalanced distributions in regression," *Neurocomputing*, vol. 343, pp. 76–99, May 2019.

[10] N. Brown, A. Roubíková, I. Lampaki, L. MacGregor, M. Ellis, and P. V. de Newton, "Machine learning on crays to optimize petrophysical workflows in oil and gas exploration," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 20, p. e5655, Oct. 2020.

[11] A. S. Bull and M. S. Love, "Worldwide oil and gas platform decommissioning: A review of practices and reefing options," *Ocean Coastal Manage.*, vol. 168, pp. 274–306, Feb. 2019.

[12] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "DBSMOTE: Density-based synthetic minority over-sampling TEchnique," *Int. J. Speech Technol.*, vol. 36, no. 3, pp. 664–684, Apr. 2012.

[13] D. Burdon, S. Barnard, S. J. Boyes, and M. Elliott, "Oil and gas infrastructure decommissioning in marine protected areas: System complexity, analysis and challenges," *Mar. Pollut. Bull.*, vol. 135, pp. 739–758, Oct. 2018.

[14] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018.

[15] J. Chandler, D. White, E. J. Techera, S. Gourvenec, and S. Draper, "Engineering and legal considerations for decommissioning of offshore oil and gas infrastructure in Australia," *Ocean Eng.*, vol. 131, pp. 338–347, Feb. 2017.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[17] T. Cresswell, S. Brown, H. Wong, and S. Apte, "Assessing the impacts of scale residues from offshore oil and gas decommissioning on marine organisms," *APPEA J.*, vol. 61, no. 2, pp. 379–383, 2021.

[18] M. Day and A. Gusmitta, "Decommissioning of offshore oil and gas installations," in *Environmental Technology in the Oil Industry*. Hampshire, U.K.: Oxoid, 2016, pp. 257–283.

[19] D. Fan, H. Sun, J. Yao, K. Zhang, X. Yan, and Z. Sun, "Well production forecasting based on ARIMA-LSTM model considering manual operations," *Energy*, vol. 220, Apr. 2021, Art. no. 119708.

[20] S. A. Fatima, N. Ramli, S. A. A. Taqvi, and H. Zabiri, "Prediction of industrial debutanizer column compositions using data-driven ANFIS and ANN-based approaches," *Neural Comput. Appl.*, vol. 2021, pp. 1–13, Jan. 2021.

[21] D. Ferraretti, G. Gamberoni, and E. Lamma, "Unsupervised and supervised learning in cascade for petroleum geology," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9504–9514, Aug. 2012.

[22] S. Genter, "Stakeholder engagement in the decommissioning process," in *SPE Symposium: Decommissioning Abandonment*. London, U.K.: Society of Petroleum Engineers, 2019.

[23] R. Gibson and M. Christou, "Metocean criteria for the fatigue analysis of subsea pipelines," in *Proc. Int. Conf. Offshore Mech. Arctic Eng.*, vol. 58851. New York, NY, USA: American Society of Mechanical Engineers, 2019, Art. no. V07BT06A010.

[24] Y. Hajizadeh, "Machine learning in oil and gas; A SWOT analysis approach," *J. Petroleum Sci. Eng.*, vol. 176, pp. 661–663, May 2019.

[25] B. Hall, "Facies classification using machine learning," *Lead. Edge*, vol. 35, no. 10, pp. 906–909, Oct. 2016.

[26] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-smote: A new over-sampling method in imbalanced data sets learning," in *Proc. Int. Conf. Intell. Comput.* Berlin, Germany: Springer, 2005, pp. 878–887.

[27] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Mach. Learn.*, vol. 45, no. 2, pp. 171–186, 2001.

[28] A. U. Haq, D. Zhang, H. Peng, and S. U. Rahman, "Combining multiple feature-ranking techniques and clustering of variables for feature selection," *IEEE Access*, vol. 7, pp. 151482–151492, 2019.

[29] P. Hart, "The condensed nearest neighbor rule (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 3, pp. 515–516, May 1968.

[30] C. Hegde, M. Pyrcz, H. Millwater, H. Daigle, and K. Gray, "Fully coupled end-to-end drilling optimization model using machine learning," *J. Petroleum Sci. Eng.*, vol. 186, Mar. 2020, Art. no. 106681.

[31] T. D. Jobe, A. Research Center, E. Vital-Brazil, and M. Khait, "Geological feature prediction using image-based machine learning," *SPWLA J. Formation Eval. Reservoir Description*, vol. 59, no. 6, pp. 750–760, 2018.

[32] R. Kanfar, O. Shaikh, M. Yousefzadeh, and T. Mukerji, "Real-time well log prediction from drilling data using deep learning," in *Proc. Int. Petroleum Technol. Conf.*, Jan. 2020, pp. 1–12.

[33] L. Kirschbaum, D. Roman, G. Singh, J. Bruns, V. Robu, and D. Flynn, "AI-driven maintenance support for downhole tools and electronics operated in dynamic drilling environments," *IEEE Access*, vol. 8, pp. 78683–78701, 2020.

[34] S. Krishna, S. Ridha, P. Vasant, S. U. Ilyas, and A. Sophian, "Conventional and intelligent models for detection and prediction of fluid loss events during drilling operations: A comprehensive review," *J. Petroleum Sci. Eng.*, vol. 195, Dec. 2020, Art. no. 107818.

[35] M. Kuhn, "Building predictive models in R using the caret package," *J. Statist. Softw.*, vol. 28, no. 5, pp. 1–26, 2008. [Online]. Available: https://www.jstatsoft.org/v028/i05

[36] D. Kundu, T. Khanolkar, T. Shah, and S. Bangad, "Application of machine learning technique to predict crude distillation column inlet temperature/furnace coil outlet temperature in order to maximize distillate yield and to minimize fuel firing in furnaces," *Int. J. Comput. Appl.*, vol. 975, p. 8887, Mar. 2021.

[37] A. T. Lawrence, "The drivers of stakeholder engagement: Reflections on the case of royal dutch/shell," in *Unfolding Stakeholder Thinking*. Evanston, IL, USA: Routledge, 2017, pp. 185–199.

[38] S. S. Lee, "Noisy replication in skewed binary classification," *Comput. Statist. Data Anal.*, vol. 34, no. 2, pp. 165–191, Aug. 2000.

[39] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in E-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020.

[40] Y. Li and Z. Hu, "A review of multi-attributes decision-making models for offshore oil and gas facilities decommissioning," *J. Ocean Eng. Sci.*, May 2021, doi: 10.1016/j.joes.2021.05.002.

[41] W. Liu, W. D. Liu, and J. Gu, "Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network," *J. Petroleum Sci. Eng.*, vol. 189, Jun. 2020, Art. no. 107013.

[42] W. Mackenzie. (Jul. 2021). *Global Decommissioning Anxiety Rises*. Accessed: Jul. 22, 2021. [Online]. Available: https://www.woodmac.com/reports/upstream-oil-and-gas-global-decommissioning-anxiety-rises-507552

[43] A. Mamudu, F. Khan, S. Zendehboudi, and S. Adedigba, "Dynamic risk assessment of reservoir production using data-driven probabilistic approach," *J. Petroleum Sci. Eng.*, vol. 184, Jan. 2020, Art. no. 106486.

[44] A. Mamudu, F. Khan, S. Zendehboudi, and S. Adedigba, "A hybrid intelligent model for reservoir production and associated dynamic risks," *J. Natural Gas Sci. Eng.*, vol. 83, Nov. 2020, Art. no. 103512.

[45] M. Mannil, K. Kato, R. Manka, J. von Spiczak, B. Peters, V. L. Cammann, C. Kaiser, S. Osswald, T. H. Nguyen, J. D. Horowitz, H. A. Katus, F. Ruschitzka, J. R. Ghadri, H. Alkadhi, and C. Templin, "Prognostic value of texture analysis from cardiac magnetic resonance imaging in patients with Takotsubo syndrome: A machine learning based proof-of-principle approach," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 20537.

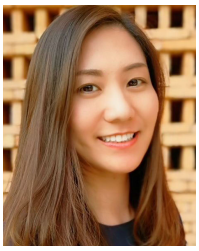
[46] I. D. Martins, L. Bahiense, C. E. D. Infante, and E. F. Arruda, "Dimensionality reduction for multi-criteria problems: An application to the decommissioning of oil and gas installations," *Expert Syst. Appl.*, vol. 148, Jun. 2020, Art. no. 113236.

[47] C. F. Mason, "Policy brief—Regulating offshore oil and gas exploration: Insights from the deepwater horizon experience in the Gulf of Mexico," *Rev. Environ. Econ. Policy*, vol. 13, no. 1, pp. 149–154, Jan. 2019.

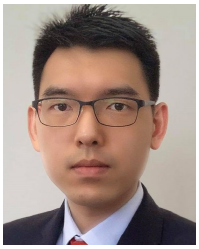
[48] A. Miller, *Subset Selection Regression*. Boca Raton, FL, USA: CRC Press, 2002.

[49] I. M. Nasir, M. A. Khan, M. Yasmin, J. H. Shah, M. Gabryel, R. Scherer, and R. Damaševičius, "Pearson correlation-based feature selection for document classification using balanced training," *Sensors*, vol. 20, no. 23, p. 6793, Nov. 2020.

[50] A. U. Osarogiagbon, F. Khan, R. Venkatesan, and P. Gillard, "Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations," *Process Saf. Environ. Protection*, vol. 147, pp. 367–384, Mar. 2021.
[51] N. M. Ramli, M. A. Hussain, B. M. Jan, and B. Abdullah, "Composition prediction of a debutanizer column using equation based artificial neural network model," *Neurocomputing*, vol. 131, pp. 59–76, May 2014.
[52] N. M. Ramli, M. A. Hussain, and B. M. Jan, "Multivariable control of a debutanizer column using equation based artificial neural network model inverse control strategies," *Neurocomputing*, vol. 194, pp. 135–150, Jun. 2016.
[53] A. Roubickova, N. Brown, and O. Brown, "Using machine learning to reduce ensembles of geological models for oil and gas exploration," in *Proc. IEEE/ACM 5th Int. Workshop Data Anal. Reduction Big Sci. Data*, Dec. 2019, pp. 42–49.
[54] D. Guha Roy, T. N. Singh, and J. Kodikara, "Predicting mode-I fracture toughness of rocks using soft computing and multiple regression," *Measurement*, vol. 126, pp. 231–241, Oct. 2018.
[55] M. Sabah, M. Talebkeikhah, F. Agin, F. Talebkeikhah, and E. Hasheminasab, "Application of decision tree, artificial neural networks, and adaptive neuro-fuzzy inference system on predicting lost circulation: A case study from Marun oil field," *J. Petroleum Sci. Eng.*, vol. 177, pp. 236–249, Jun. 2019.
[56] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, Jan. 2019.
[57] A. Salimi, M. Ziaii, A. Amiri, M. Hosseinjani Zadeh, S. Karimpouli, and M. Moradkhani, "Using a feature subset selection method and support vector machine to address curse of dimensionality and redundancy in hyperion hyperspectral data classification," *Egyptian J. Remote Sens. Space Sci.*, vol. 21, no. 1, pp. 27–36, Apr. 2018.
[58] M.-L. Schläppy, L. M. Robinson, V. Camilieri-Asch, and K. Miller, "Trash or treasure? Considerations for future ecological research to inform oil and gas decommissioning," *Frontiers Mar. Sci.*, vol. 8, pp. 1–8, Jun. 2021.
[59] F. Silva, S. Fernandes, J. Casac ao, C. Libório, J. Almeida, S. Cersósimo, C. R. Mendes, R. Brand ao, and R. Cerqueira, "Machine-learning in oil and gas exploration: A new approach to geological risk assessment," in *Proc. 81st EAGE Conf. Exhib.*, 2019, pp. 1–5.
[60] H. Singh, Y. Seol, and E. M. Myshakin, "Automated well-log processing and lithology classification by identifying optimal features through unsupervised and supervised machine-learning algorithms," *SPE J.*, vol. 25, no. 5, pp. 2778–2800, Oct. 2020.
[61] B. Sommer, A. M. Fowler, P. I. Macreadie, D. A. Palandro, A. C. Aziz, and D. J. Booth, "Decommissioning of offshore oil and gas structures—Environmental opportunities and challenges," *Sci. Total Environ.*, vol. 658, pp. 973–981, Mar. 2019.
[62] X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, L. Jiang, and Z. Cheng, "Time-series well performance prediction based on long short-term memory (LSTM) neural network model," *J. Petroleum Sci. Eng.*, vol. 186, Mar. 2020, Art. no. 106682.
[63] S. E. Z. Lashari, A. Takbiri-Borujeni, E. Fathi, T. Sun, R. Rahmani, and M. Khazaeli, "Drilling performance monitoring and optimization: A data-driven approach," *J. Petroleum Explor. Prod. Technol.*, vol. 9, no. 4, pp. 2747–2756, Dec. 2019.
[64] A. Tung, "Co-creation of knowledge—An effective mechanism for managing decommissioning stakeholders," *APPEA J.*, vol. 61, no. 1, pp. 48–57, 2021.
[65] A. Tung, "A comparison of stakeholder engagement strategies for offshore decommissioning projects in the united kingdom and Australian landscape," in *Proc. Offshore Technol. Conf.*, May 2020.
[66] A. W. J. Tung, "An exploration of stakeholder impacts on the decommissioning of offshore oil and gas facilities–the design, development, and analysis of stakeholder oriented critical paths for United Kingdom and Australia," Ph.D. dissertation, School Manage., Curtin Univ., Bentley, WA, Australia, 2021.
[67] G. Varoquaux, "Cross-validation failure: Small sample sizes lead to large error bars," *NeuroImage*, vol. 180, pp. 68–77, Oct. 2018.
[68] P. Vuttipittayamongkol and E. Elyan, "Improved overlap-based undersampling for imbalanced dataset classification with application to epilepsy and Parkinson's disease," *Int. J. Neural Syst.*, vol. 30, no. 8, Aug. 2020, Art. no. 2050043.
[69] P. Vuttipittayamongkol and E. Elyan, "Neighbourhood-based undersampling approach for handling imbalanced and overlapped data," *Inf. Sci.*, vol. 509, pp. 47–70, Jan. 2020.
[70] P. Vuttipittayamongkol, E. Elyan, and A. Petrovski, "On the class overlap problem in imbalanced data classification," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106631.
[71] P. Vuttipittayamongkol, A. Tung, and E. Elyan, "Towards machine learning-driven practices for oil and gas decommissioning—Introduction of a new offshore pipeline dataset," in *Proc. 9th Int. Conf. Comput. Commun. Manage.*, Singapore, 2021, doi: 10.1145/3479162.3479179.
[72] T. R. Wanasinghe, L. Wroblewski, B. K. Petersen, R. G. Gosine, L. A. James, O. De Silva, G. K. I. Mann, and P. J. Warrian, "Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges," *IEEE Access*, vol. 8, pp. 104175–104197, 2020.
[73] M. Wei, B. Bai, A. H. Sung, Q. Liu, J. Wang, and M. E. Cat her, "Predicting injection profiles using ANFIS," *Inf. Sci.*, vol. 177, no. 20, pp. 4445–4461, 2007.



PATTARAMON VUTTIPITTAYAMONGKOL was born in Lampang, Thailand. She received the B.Sc. degree in electrical engineering from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 2011, the M.Sc. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2013, and the Ph.D. degree from the School of Computing, Robert Gordon University, Aberdeen, U.K., in 2020. Since 2013, she has been a Faculty Member at the School of Information Technology, Mae Fah Luang University, Thailand. Her research interests include data analytics and machine learning with focus on supervised learning on class-imbalanced datasets.



AARON TUNG was born in Johor Bahru, Malaysia, in 1993. He received the B.Eng. degree in petroleum engineering from Curtin University, Australia, in 2016, the Ph.D. degree in law from the University of Aberdeen, U.K., in 2021, and the Ph.D. degree in management from Curtin University, in 2021. He is currently working as a Project Engineer at Woodside Energy. His research interests include several disciplines revolving around the topic of oil and gas decommissioning, including engineering, project management, stakeholder engagement, regulatory law, and data science. He was a finalist of the 2021 Decom North Sea Rising Star in Decommissioning Award.

EYAD ELYAN received the bachelor's degree in computer science from Al-Quds University, Palestine, in 1999, and the M.Sc. degree in software engineering and the Ph.D. degree in 3D facial modeling and recognition from Bradford University, in 2004 and 2008, respectively. He is currently a Professor in machine learning and computer vision with the School of Computing, Robert Gordon University. He is leading the machine learning and vision applications research theme at the School. His primary research interests include machine learning, deep learning, and applied computer vision. He is a fellow of British Higher Education Academy. He also serves as the Scotland Data Laboratory Innovation Centre Ambassador.