Journal of Engineering Research and Reports

Volume 25, Issue 6, Page 40-54, 2023; Article no.JERR.102150
ISSN: 2582-2926

The Use of Machine Learning in Oil Well Petrophysics and Original Oil in Place Estimation: A Systematic Literature Review Approach

Ekemini Johnson [a*], Okure Obot [b], Kingsley Attai [a], Julius Akpabio [b] and Udoinyang Inyang [b]

[a] *Ritman University, Ikot Ekpene, Nigeria.*
[b] *University of Uyo, Uyo, Nigeria.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JERR/2023/v25i6921

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/102150

Review Article

Received: 25/04/2023
Accepted: 29/06/2023
Published: 15/07/2023

ABSTRACT

Machine learning is a form of artificial intelligence that is applicable in all fields of study. It incorporates many algorithms used in carrying out various tasks such as classification, predictions, estimations, comparisons, approximations, optimization and selections. In estimating original oil in place, which affords the explorationist the foresight on the total amount of crude oil that is potentially in reservoir. Machine learning is found to perform reserves estimation with speed and accuracy where insufficient data are available. These among other attributes of Machine Learning motivated a systematic literature review of studies undertaken between 2010 and 2021 and explore the strengths and limitations reported in the studies. In the oil industry, different types of data are gathered from subsurface and surface in order to know the reservoir hydrocarbon potential. Sensorsare known to be able to collect these data in large quantity, analyse and

**Corresponding author: Email: eke5461@gmail.com;*

used to predict the output.3127 articles related to the study were collated from 4 databases and after a series of inclusion and exclusion criteria were conducted on the articles, 104 journal articles met the criteria and were used for the review. Results of the study reveal that between the years under review, 2019 had the greatest number of articles (20 of the 104) pertaining to the topic reviewed. 61% of authors reported inadequate data while 39 % reported under-performance of the algorithm. It was also revealed that machine learning was applied to perform predictions/forecasting in the industry than it was used to solve other problems, while Artificial Neural Network (ANN) was the most used artificial intelligence technique. The study opens another vista of knowledge for researchers to navigate machine learning in the estimation of original oil in place and petrophysics analysis. This emerging technology is smart and makes data evaluation easy and straightforward.

Keywords: Machine learning; petrophysics; oil in place and prediction.

ABBREVIATIONS

AI – Artificial Intelligence
ANN - Artificial Neural Network
BRA - Bayesian Regularization Algorithm
CNN - Convolutional Neural Network
DT - Decision Tree
DNN - Deep Neural Networks
ET - Evolutionary Techniques
EN - Ensemble Techniques
FFBP - Feed Forward Back Propagation
FS - Fuzzy Sets
GA-ANN - Genetic Algorithm-Artificial Neural Network
GRULSTM - Gated Recurrent Units and Long Short-Term Memory
KNN – K Nearest Neighbor
LSSVM Least Square Support Vector Machine
LogitBoost - Logistic Boosting Regression
ML - Machine Learning
MLP- Multi-Layer Perceptron
MSE - Mean Square Error
MELM-PSO -Multiple Extreme Learning Machine used with Particle Swarm Optimization
Multinom - Multinomial Logistic Regression
NN - Neural Networks
NNPC - Nigerian National Petroleum Corporation
OIP - Oil in Place
PSO Particle Swarm Optimizer
RBFs - Radial Basis Functions
RNN - Recurrent Neural Network
RQs – Research Questions
RVM - Relevant Vector Machine
RF – Random Forest
SVM - Support vector Machine
SLR - Systematic Literature Review
XGBoost - Extreme Gradient Boosting

1. INTRODUCTION

"Crude oil is seen to be a combination of volatile liquid hydrocarbons which is composed mainly of hydrogen and carbon but also contains some nitrogen, sulfur, and oxygen. These elements form a large variety of complex molecular structures, of which some cannot be readily identified. Regardless of variations, however, most crude oil ranges from 82% to 87% carbon by weight and 12% to 15% hydrogen by weight" [1]. Crude oil is the major stay of most of the wealthy nation's economy thereby making the oil and gas sector play an important role in the world economy. Many countries endowed with hydrocarbon resources rely on this sector for revenue and growth, and the global economy in its entirety operates on fossil fuels. A good percentage of every major production network starts and ends with burning oil and gas, and depends solely on it to transport commodities from one location to another. Oil and gas have taken centre stage in global economic growth right from the industrial revolution [2].

Crude Oil Exploration and Production involves several activities ranging from seismic survey to the point of actually using the data obtained from seismic survey for ascertaining the volume of oil in place (OIP). The recoverable oil using estimation methods such as Analogy, Volumetric, Material Balance and Decline Curve Analysis with part of data generated during the exploration process being the petrophysical properties of the oil well could as well be ascertained. These known estimation methods are resource consuming and so most exploration firms seek to utilize methods that are cost effective and are able to utilize varying completeness of data to obtain valid results as such, they shifted to machine learning. The volume of data generated during the survey is enormous thereby making processing and handling using the four estimation methods challenging by the oil and gas industries since proper technical analysis of this data generated needs to be carried out to improve performance of oil and gas industries [3].

When the limitations of these estimation methods such as post-production overestimation problem and time consumption were established, researchers shifted to the use of machine learning to improve on the estimation of oil in place by exploring the rich, flexible and computational capabilities provided by machine learning [4]. “Machine learning (ML) is seen to be effective in the area of petroleum exploration and production because it is ideal for addressing those problems where our theoretical knowledge is still incomplete but for which we do have a significant number of observations and other data as it is seen in crude oil exploration processes” [5]. “ML is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment and they are considered the working horse in the new era of big data” [6]. ML has an ever increasing presence and positive effect on a broad variety of research and commercial fields and this has really enhanced productivity in most organizations [7]. “Techniques based on machine learning have been applied successfully in diverse fields ranging from pattern recognition, computer vision, spacecraft engineering, finance, entertainment, and computational biology to biomedical, medical applications and oil and gas. Machine learning has application in different domains such as Computer vision, prediction, semantic analysis, natural language processing and information retrieval” [8]. ML play different but very important roles in different petroleum engineering and geosciences segments which include petroleum exploration, reservoir characterization, oil well drilling, production and well stimulation, which emphasizes the newly emerging field of unconventional reservoirs. According to Tariq et al [9], “the advent of powerful computers, ML algorithms and extensive data generated from different industry tools, a bright future is seen as solutions are developed for complex problems in the oil and gas industry. These were previously beyond the capability of analytical solutions or numerical simulation, since ML is capable of incorporating every detail in the log data and every information connected to the target data”.

Our study attempts to examine the application of machine learning in oil well petrophysical properties analysis and original oil in place estimation. The specific objectives of the study are to: i. review the machine learning techniques used in the oil well petrophysical properties analysis and original oil in place estimation. ii. Determine the most used machine learning tool in petrophysical properties analysis and original oil in place estimation between 2011 and 2021. iii. Explore the strength and limitation reported in the studies.

The remaining part of the paper is arranged as follows: Section II presents recent and dated literature onOil Well petrophysical-properties and original oil in place estimation. Section III and IV present the Research Method and Results with discussion respectively.

2. LITERATURE REVIEW

This section presents related literature and from the literature gathered, the authors searched and examined the studies performed between the years 2011 and 2021 in digital libraries to develop the SLR.

Several researchers have carried out development of petrophysical properties prediction and characterization models and their works have helped in forecasting, optimization of production, characterization of reservoir properties and oil in place estimation. These models are machine learning models that include regression models, artificial neural networks (ANNs), radial basis functions (RBFs) etc.

"The petrophysics of any oil field include reservoir fluid properties and reservoir rock properties which could affect oil recovery and volume of oil production" [10]. The properties include porosity, permeability, fluid saturation, mobility among others. Evaluating these properties successfully is necessary for determining the hydrocarbon potential of a reservoir system and also helps to predict the behavior of complex reservoir situations. Rock porosity and fluid saturations are most times seen as the principal factors involved in determining the amount of oil and gas originally in place while permeability is seen as a measure of the ease with which fluid flows through the pore spaces of rock. Oil is recovered (extracted) using several methods, mostly depending on geology. Conventional oil is extracted from underground reservoirs using traditional drilling and pumping methods. In recent times, Researchers have really applied Artificial Intelligence to seek to reduce all resources used in exploration and drilling of oil. Okon et al [11] aimed to "predict reservoir petrophysical properties of porosity, permeability and water saturation using Artificilal Neural network Network. The developed Artificial Neural Network (ANN model is a feed-forward back-propagation (FFBP) network with 12 neurons in its hidden layer with the Levenberg–Marquardt algorithm as the best learning algorithm, better than Bayesian regularization and Scaled conjugate gradient".

Getting to know the estimated volume of oil in place is necessary so as to help when actually computing the recoverable volume of oil in a particular reservoir. According to Sowiżdżał et al [12], methods for oil in place estimation includes Volumetric, Material balance, Production history and Analogy methods. All of these methods use the petrophysical properties of the well to estimate the oil in place. Prediction of reservoir petrophysical properties from well-logs data has evolved from the use of experts' knowledge and statistics to the use of artificial intelligence (AI) models. In 2015, Ahmadi and Pournik [13] built a predictive model of chemical flooding for enhanced oil recovery purposes. To achieve this goal, a new support vector machine method which was developed by Suykens and Vandewalle [14] was employed. In the work, high precise chemical flooding data banks reported in previous works were employed to test and validate the proposed vector machine model. According to the mean square error (MSE), correlation coefficient and average absolute relative deviation, the suggested Least Square Support Vector Machine (LSSVM) model has acceptable reliability; integrity and robustness. Thus, the proposed intelligent based model can be considered as an alternative model to monitor the efficiency of chemical flooding in oil reservoir when the required experimental data are not available or accessible. The result obtained showed that the most important parameters affecting the recovery factor are surfactant concentration and surfactant slug size.

It is found in literature that some of the models for petrophysical properties and oil in place estimation made use of a single tool; some are either a hybrid or ensemble of tools while some are a comparison of the results of different tools.

2.1 Studies with Single Tools

27 out of 104 of the literature gathered used a single tool for their work and they were mostly ANN based. Some of the works are tabulated as Table 1.

2.2 Studies with Ensemble or Hybrid Tools

In the studies under review 33 out of 104 used an ensemble or hybrid tool for their work. Some of the works are tabulated in Table 2.

2.3 Studies that Compared Results of Different Tools

44 out of 104 actually set out to use tools that are either single or ensemble for a task and compare the results thereafter. Some of the works are tabulated as Table 3.

Table 1. Tabulated ANN based literature

Reference	Machine learning methodology	Description	Problem solved
Fegh et al [15]	ANN	Used artificial neural network in predicting Permeability and construction of 3D geological model from data obtained from Iranian gas reservoir with the objective to construct a comprehensive and quantitative characterization of a carbonate gas reservoir in marine gas field to help in decision making process.	Prediction and Characterization
Mahmoud et al [16]	ANN	Estimated oil recovery factor for Water drive sandy reservoirs using artificial neural network. ANN model was trained using data collected from 130 water drive sandstone reservoirs. The developed ANN-based equation outperformed the available equations in terms of all the measures of error evaluation considered in this study.	Estimation
Zerrouki and Baddari [17]	ANN	Estimated natural fracture porosity from well log data using artificial neural network in HassiMessaoud oil field, Algeria. In this paper, fracture porosity using four conventional log data (deep resistivity, density, neutron porosity and gamma ray) from well #1 and well #2 in HassiMessaoudoil field was estimated. The structure of the ANN was trained using the back-propagation algorithm,	Estimation
Zolotukhin and Gayubov [18]	Random Forest	Showed how machine learning can be used in reservoir permeability prediction and modelling of fluid flow in porous media. The study describes the methodology for determining the permeability of a porous medium using machine learning. The study evaluates the permeability of porous medium samples using artificial neural networks with a very high correlation of predicted values with the available data in the sample using a very limited number of available experimental data	Prediction

Table 2. Ensemble–based models

Reference	Machine learning methodology	Description	Problem solved
Anifowose et al [19]	ANN-bagging and RF	Predicted petroleum reservoir properties from downholesensor data using an ensemble model of neural networks. The study presented an ensemble model of ANN that combines the diverse performances of seven "weak" learning algorithms to evolve an ensemble solution in the prediction of porosity and permeability of petroleum reservoirs. When compared to the individual ANN, ANN-bagging and Random Forest, the proposed model performed bestby having the highest R-Square consistently for all the datasets.	Prediction
Tian et al [20]	GA–ANN	Predicted Permeability of porous media using a combination of computational fluid dynamics and hybrid	Prediction

Reference	Machine learning methodology	Description	Problem solved
		machine learning methods. The pore structure parameters were extracted as input parameters and the permeability was calculated as the output parameter. For the ML modeling, a hybrid ML method was proposed using a combination of artificial neural network (ANN) and genetic algorithm (GA). The ANN was employed to learn the nonlinear relationships and GA was used to tune ANN architecture for the best performance. The prediction results show that the GA–ANN was robust in predicting permeability based on pore structure parameters.	
Abad et al [21]	Multiple Extreme Learning Machine used with Particle Swarm Optimization (MELM-PSO)	Predicted condensate viscosity in the near wellbore regions of gas condensate reservoirs using a hybrid machine learning algorithms. The analysis done indicates that the Multiple Extreme Learning Machine used with Particle Swarm Optimization (MELM-PSO) model provides the highest prediction accuracy	Prediction
Jian et al [22]	deep neural networks (DNN)	Integrated deep neural networks and ensemble learning machines for missing well logs estimation. Experiential results showed that the proposed method can really estimate missing logs more accurately than traditional ones, and the performance is promising.	Estimation
Oloso et al [23]	Ensemble SVM and standalone SVM	Characterized oil pressure-volume-temperature (PVT) using Ensemble systems. The study develops ensemble support vector regression and ensemble regression tree models to predict two important crude oil PVT properties: bubble point pressure and oil formation volume factor at bubble point.	Characterizatior /Classification,

Table 3. Single or ensembled based models

Reference	Machine learning methodology	Description	Problem solved
Al-Mudhafar W. [24]	Multinomial Logistic Regression (Multinom), Logistic Boosting Regression (LogitBoost), and Extreme Gradient Boosting (XGBoost)	used several tools for prediction of the characteristics of clastic reservoirs. The Multinomial Logistic Regression (Multinom), Logistic Boosting Regression (LogitBoost), and Extreme Gradient Boosting (XGBoost) were all comparatively adopted for lithofacies classification. After that, the resulting most accurate discrete facies distribution by LogitBoost were included along with the well logging interpretations into the multivariate permeability modeling through advanced machine learning approaches to model and predict the corrected core permeability given well logging interpretations for all wells in the reservoir. More specifically, the Multivariate Adaptive Regression Splines (MARS) and Smooth Generalized Additive Models (SGAM) were comparatively adopted to model and predict the core permeability at all wells.	Prediction

Reference	Machine learning methodology	Description	Problem solved
Otchere et al [25]	Artificial Neural Network (ANN), Support Vector Machine (SVM) and Relevant Vector Machine (RVM)	Did a comparative analysis of ANN and SVM models. The review focuses on the use Artificial Neural Network (ANN), Support Vector Machine (SVM) and Relevant Vector Machine (RVM) in petroleum industry. The Support Vector Machine (SVM) and Relevant Vector Machine (RVM) outperformed the ANN. This makes them preferable than the ANN when there are limited data sets.	Comparison and Optimization
Rahmanifard et al [26]	Bayesian regularization algorithm (BRA)	Did a comparative study of the application of supervised machine learning techniques for multivariate modelling of gas component viscosity. The study explore gas component viscosity prediction versus molecular weight, critical properties, acentric factor, normal boiling point, dipole moment, and temperature using 38 supervised machine learning algorithms. The algorithms are tested by using 4673 data sets for 1602 organic and inorganic gas components collected from the literature. The study does a comparison of the outputs of the best predictive model with the viscosity models provided in the literature.	Comparison and Optimization
Haagsma et al [27]	Random forest(RF) and Decision Tree (DT)	Carried out Secondary porosity prediction in complex carbonate reefs using 3D CT scan image analysis and machine learning. The objective of this study was to quantify secondary porosity from CT scans of complex carbonate reservoirs and develop a predictive model using machine learning to predict secondary porosity on readily available wireline log data. Data was obtained from six well locations over the Brown Niagaran and A-1 Carbonate intervals	Prediction
Hsu and Zhang [28]	Convolutional neural network based classifier (CNN), k-nearest neighbor (KNN), random forest (RF), support vector machines (SVM), recurrent neural network (RNN), gated recurrent units and long short-term memory (GRULSTM).	Proposed a petroleum engineering data text classification using convolutional neural network based classifier. Researchers conducted experiments with real world petroleum engineering data text,labelled into 6 categories.the research was implemented using CNN-based model, and also the study was implemented with other machine learning approaches which includes k-nearest neighbor, random forest, support vector machines, recurrent neural network, gated recurrent units and long short-term memory.	Classification
El-Amin and Subasi [29]	SVM, k-NN, ANN and RF.	Developed a generalized scaling-law for oil recovery using machine learning techniques. The research aimed at using a scaling law to predict oil recovery from a laboratory sample and developed machine learning techniques to predict the dimensionless scaled time to mimic the actual time of recovery in terms of physical primary parameters of porosity, permeability and viscosity	Prediction

3. MATERIALS AND METHODS

In this section, a summary of the steps followed in the literature review, the goals, research questions, inclusion and exclusion criteria and the framework of the study are discussed.

The first subsection (A) gives a summary of the steps followed in reviewing literature. The second subsection (B) discusses the goals of the study and the research questions. The third subsection (C) explains the article selection strategy and the fourth subsection (D) presents the final repository of the papers used in the study.

3.1 Overview

This systematic literature Review (SLR) follows the guidelines proposed by Kitchenham and Charters [30].

The research was undertaken based on the steps proposed by Kitchenham and Charters [30] to survey the existing knowledge about the use of machine learning techniques in petrophysical properties and oil in place estimation. The SLR process applied is shown in Fig. 1.

Following Kitchenham and Charters [30], the research problem, objective, and questions were defined. Thereafter, the search string was defined, followed by inclusion and exclusion criteria. Based on the search string defined and inclusion and exclusion criteria, a general search in all relevant databases (ACM, Science Direct, Google Scholar, IEEE) was done. From the results of examination, the duplicate articles were eliminated, thereby giving result to a list of selected papers which were read by title, abstract and keywords to finally have a list with the selected final articles which were thoroughly read and analyzed.

3.2 Goal and Research Questions

The goal of this research is to identify the extent to which machine learning has impacted the results obtained in Petrophysical Properties Characterization and Original Oil in place Estimation. The research also determines the cause behind the use and adoption of machine learning techniques from region to region. This research is a systematic literature review that follows the sequence stated by Kitchenham and Charters [30].

The scope of the study is to identify, analyze and synthesize works that were published in the last 10 years (from 2011 to 2021) in the area of machine learning as applied to petrophysical properties and original oil in place estimation [31-33].

To achieve the research goal, the following four (4) Research Questions are formulated:

- **RQ1:**Why is machine learning relevant in Oil Well Petrophysical Properties and Original Oil in Place Estimation?
  - **RQ1.1:** What machine learning algorithms are adopted in the studies under review?
  - **RQ1.2:** Which regions of the world used Machine Learning?
  - **RQ1.3:** To what extent has the use of Machine Learning translated into practical implementation beyond the theoretical design?
- **RQ2:** How many studies have been conducted on this topic from 2011 to 2021?
- **RQ3:** What were the problems investigated and presented in the literature of oil well petrophysical properties and Original Oil in place estimation?
- **RQ4:** What are the key limitations and strengths reported in the studies related to petrophysicalproperties and Oil in Place Estimation?

3.3 Article Selection

This Section briefly explains the following aspects of article selection:

- Source selection search keywords.
- Application of inclusion/exclusion criteria.

3.3.1 Source selection and search keywords

The following digital libraries were used for the search results: 1. Google Scholar 2. ACM 3. Science Direct 4. IEEE Xplore.

Search query was formulated with logical operators used to link the search keys as shown in Table 4.

The number of articles downloaded from each of the digital library before the inclusion and exclusion criteria were applied for final selection of papers used for the systematic literature review is shown in Table 5.

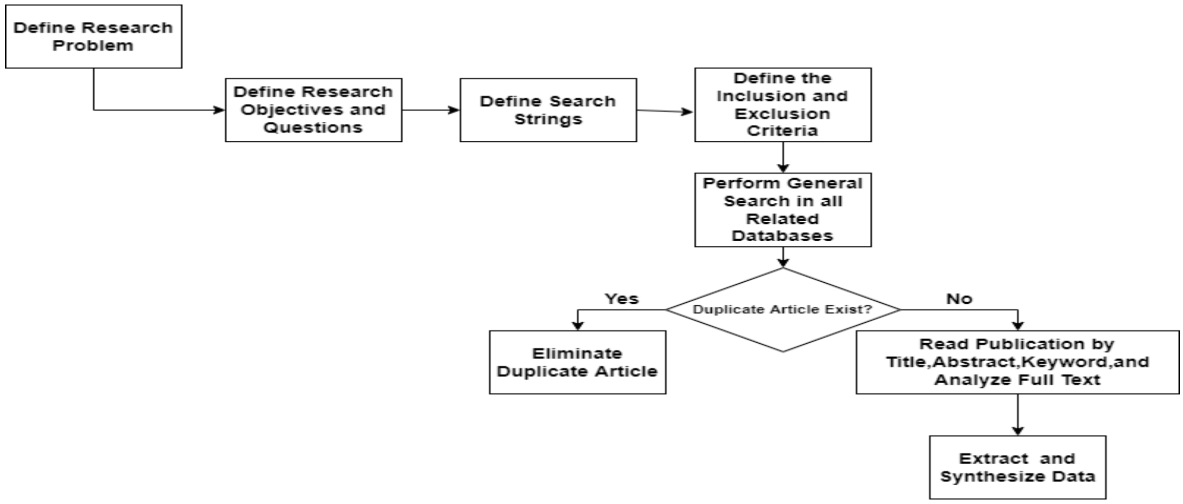

Fig. 1. Framework of SLR

Table 4. Search keywords

"Oil "OR "Petroleum" AND "Petrophysical properties" AND "Machine learning" OR "Intelligent System"
"Oil in Place" OR "Petroleum Volume Estimation" OR "Petrophysical Properties characterization" OR "Machine Learning" OR "Intelligent System"
"Machine Learning" AND "Estimation" OR "Prediction" OR "Computation" AND " Original Oil in Place" OR "Oil in Place" AND "Petrophysical Properties "

Table 5. Number of articles based on each database

Database	# of Articles
Google Scholar	1830
ACM	1130
Science Direct	157
IEEE	10
Total	3127

Table 5 show the available articles using the mentioned keywords.

3.3.2 Application of inclusion/exclusion criteria

In this work, the following inclusion criteria is considered:

a. Period of publication from 1st January 2011 to 31st December 2021.
b. Publications published in English language.
c. Publications that were peer-reviewed.
d. Publications that focus on Oil in Place Estimation and application of machine learning to Petrophysical Properties.
e. Publication that presented the keywords which belong to the string determined in this SLR.

The following are the exclusion criteria:

a. Conference/ Poster abstract.
b. Duplicate instances of the same study.
c. Focus of study does not answer RQs.
d. Focus is not Machine Learning for oil in place estimation and petrophysical properties characterizations.
e. Not written in English.

3.4 Final Pool of Articles and the Online Repository

After search and follow-up analysis, so that unrelated papers are excluded, a pool of 104 studies was selected. 3023 papers were excluded based on our exclusion criteria, where 444 were conference abstract, 404 were duplicates, 2001 studies did not answer RQs, 119 of the papers focus were not on oil in place estimation and petrophysical properties characterization. 15 were not written in English and 40 were not publicly available to be downloaded. The analysis of the final pool of articles is captured in Fig. 2.

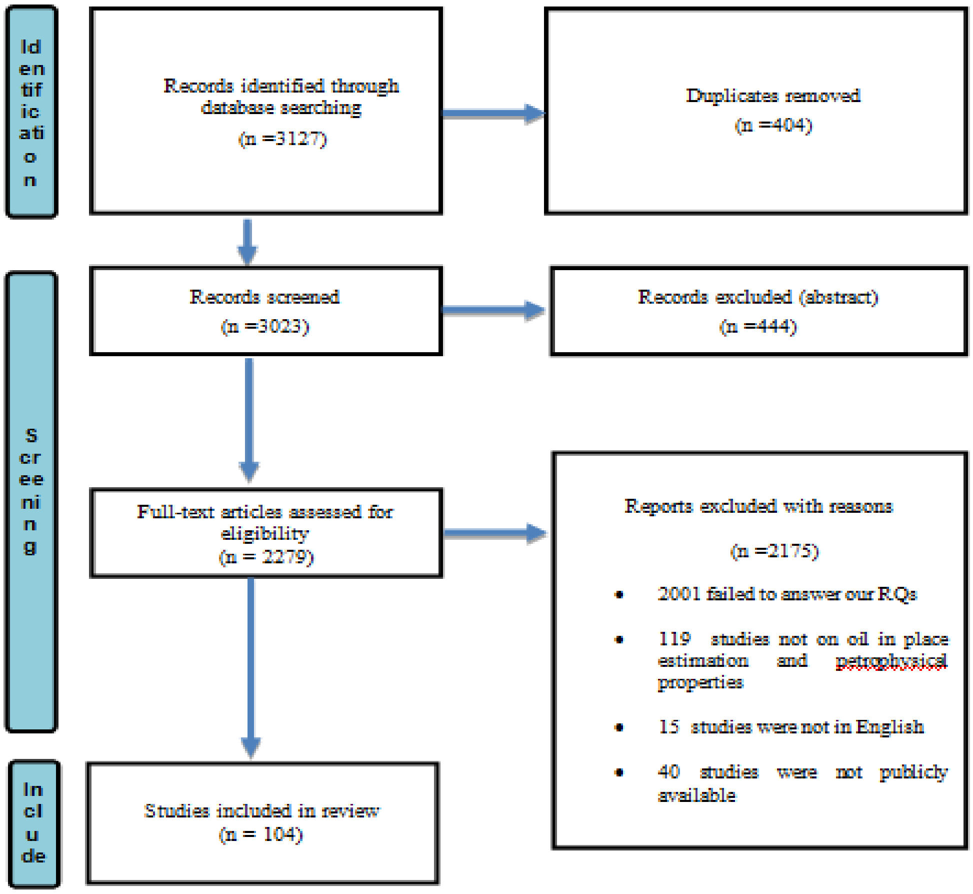

Fig. 2. Prisma flow diagram of study selection

4. RESULTS AND DISCUSSION

Here the results are presented according to the research questions.

RQ1: Why is machine learning relevant in Oil Well Petrophysical Properties and Original Oil in Place Estimation?

RQ1 is answered using the sub-questions RQ1.1 to RQ1.3.

RQ1.1: What machine learning algorithms are adopted in this study?

The machine learning algorithm used in the research is categorized into Support Vector Machines (SVM), Evolutionary Techniques (ET), Neural Networks (NN), Ensemble Techniques (EN) and Fuzzy Sets (FS). The Support Vector Machines (SVM) category comprises the support vector machines, support vector regression, least square support vector machines and optimized support vector machines. The Evolutionary Techniques comprises the particle swarm optimization, genetic algorithm, differential evolution, Ant colony optimization, The Runner-Root Algorithm, Artificial Bee Colony Algorithm and Bees Algorithm, the Neural Networks category comprise the Artificial Neural Networks, Multi-Layer Perceptron, Echo State Networks, Radial Basis Neural Network, Probabilistic Neural Network, Functional Networks, Recurrent Neural Network, Deep Neural Network and Modified Feed Forward Neural Network. The Ensemble Techniques comprise Gradient Boost, Random Forest, Adaptive Neuro-fuzzy Inference System, Adaptive boosting, Hybrid Self-Adaptive Artificial Neural Network, Multiple Extreme Learning Machine, Ensemble Model of Artificial Neural Network, Artificial Neural Network Bagging, Logistic Boosting Regression, Extreme Gradient Boosting, Stacked Generalization Ensemble Model of Support Vector Model and Sum Ensemble Model based on Conventional Bagging Method. The Fuzzy set comprise Fuzzy Cognitive Map, Fuzzy Logic, Type-2 Fuzzy Logic, Optimized Fuzzy Logic, Clustering Algorithm based on Fuzzy Sets, The classification and the frequency of each class is shown in Fig. 3.

RQ1.2: Which regions of the world use Machine Learning?

The usage of the machine learning by these countries is captured in Fig. 4 with countries on the X axis while Frequency is on the Y axis. According to Fig. 4, Iran used machine learning most, followed by China, US and Canada. Libya, Oman, Kuwait, Korea, Portugal in that order and Yemen used machine learningleast.

RQ1.3: To what extent has the use of Machine Learning translated into practical implementation beyond the theoretical design?

Studies that data were gathered, preprocessed, trained, tested and validated is 96 which represents 92.31% of the studies and the number of studies where there was no form of training, testing and validation which was seen as theoretical framework is 8 which represents 7.69% of the total studies. The number of practical implementation and theoretical framewok is depicted graphically in Fig. 5 with the blue portion representing the practical implementation and orange portion representing the theoretical frameworks.

RQ2: How many studies have been conducted on this topic from 2011 to 2021?

A total of 104 studies have been conducted within this period. The number of studies on yearly basis is depicted in Fig. 6.

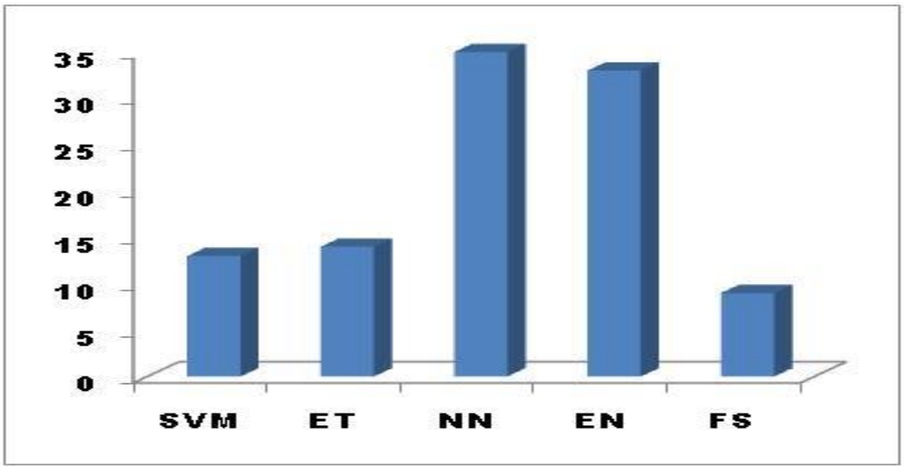

Fig. 3. Bar chart of the classes of algorithm and their corresponding frequencies

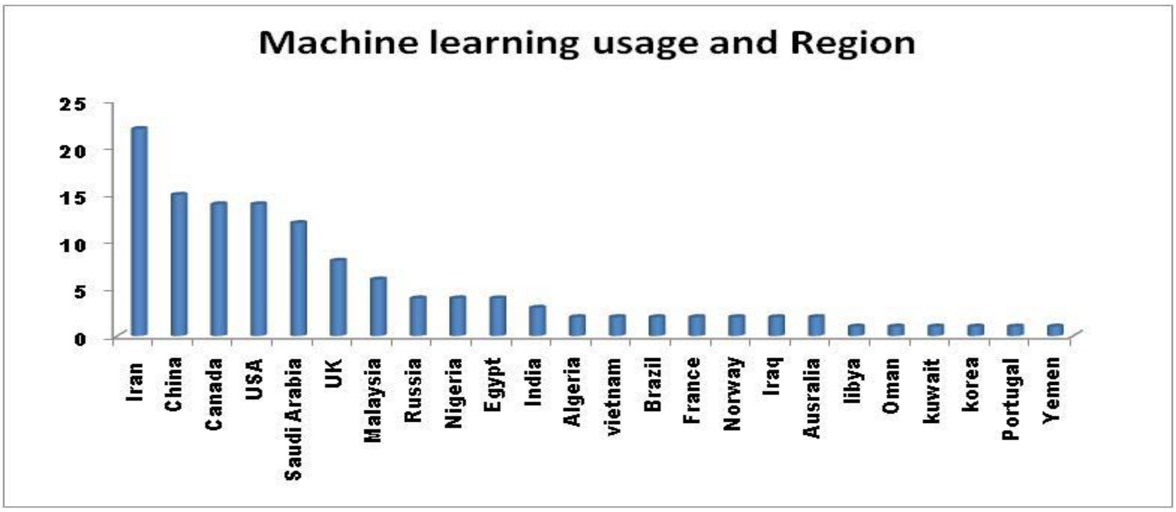

Fig. 4. Machine learning tool based on regions usage

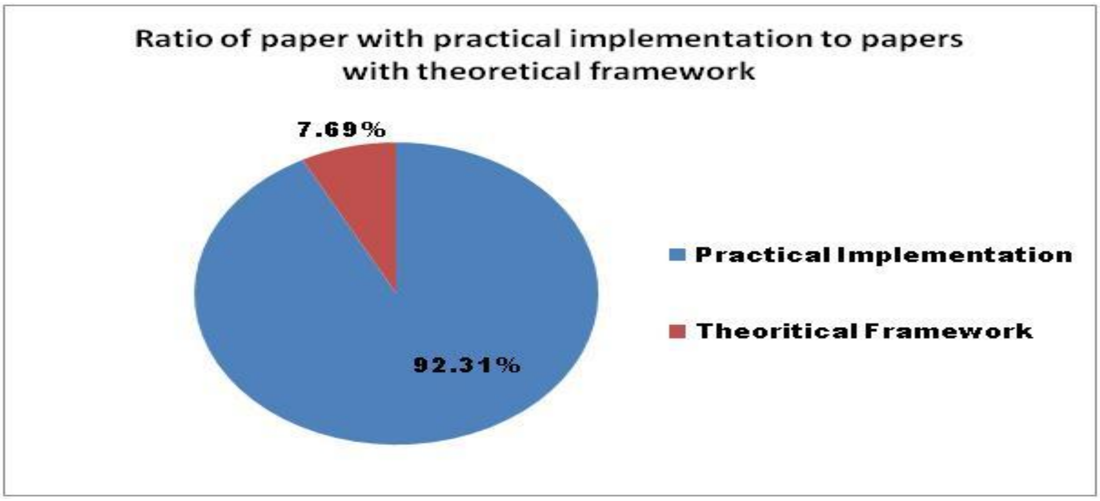

Fig. 5. A Pie chart showing the ratio of papers that had a practical implementation to the paper that had theoretical framework

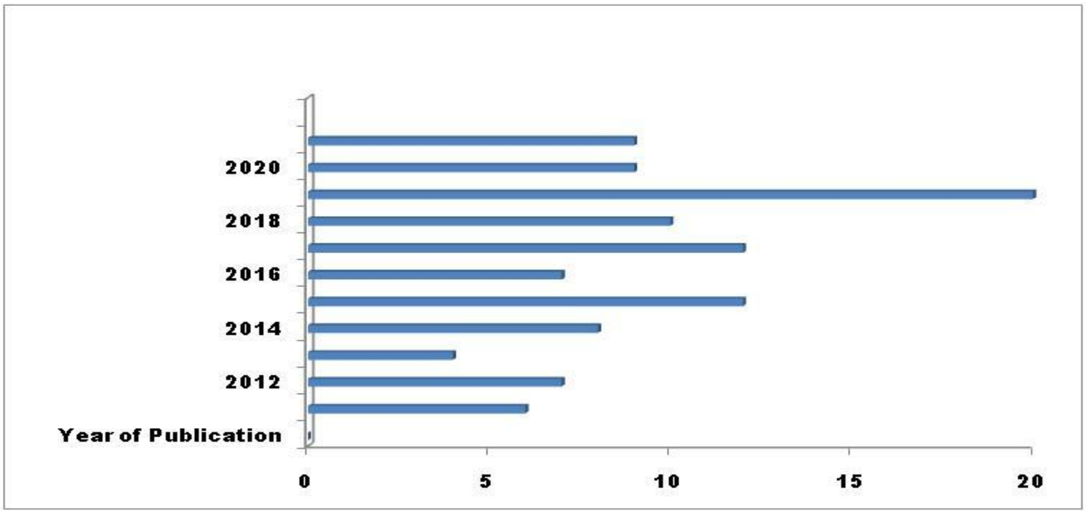

Fig. 6. Yearly breakdown of paper publication

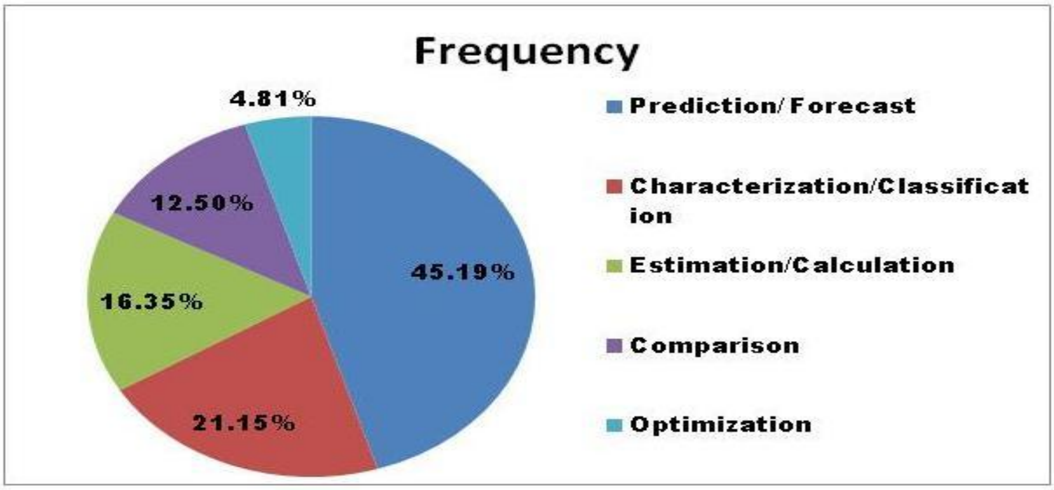

Fig. 7. Bar chart of classes of problem

RQ3: What were the problems investigated and presented in the literature of oil well petrophysical properties and Original Oil in place estimation?

The review shows that problems investigated were classified into: Prediction/ Forecast, Characterization/ Classification, Estimation/ Calculation, Comparison and Optimization. The classes of the problems and their percentages are shown in Fig. 7 with the problem of Prediction/ Forcast having the highest percentage and the problem of Optimization having the least percentage. Fig. 7 is a Pie Chart showing the classes of problems and their corresponding portions.

RQ4: What are the key strengths and limitations reported in the studies related to petrophysical properties and Oil in Place Estimation?

5. CONCLUSIONS

This paper carried out a systematic literature review on papers written in the area of Oil Well Petrophysics and Original Oil in Place Estimation from 1st January 2011 to 31st December 2021. Different techniques are employed at different times in estimating oil in place for optimal outputs and recently machine learning has gained currency. Machine learning has been applied in this endeavor without ascertaining the extent and its impact on oil well petrophysics. This study was therefore aimed at educating researchers of oil well pterophysics on the extent at which machine learning has been applied with the view of escalating the impact of machine learning. The different methodologies of machine learning have been explored, the regions in terms of countries and continents, the years of publication and the frequency of the type of problems solved are among the areas reported in the study. Iran, China and United States of America in that order were the most published countries, 2019 had the highest number of publications (20) out of the 104 publications used in the study. Neural networks (NN) was the most used techniques followed by Ensemble technique (EN) while prediction/forecasting was the most problem solved. The results obtained in the study are capable of prompting other researchers into exploring more capabilities of machine learning that could be useful in oil well petrophysics analysis and estimation of oil in place. In all, our contributions have been in the area of highlighting the development of machine learning in terms of publications and contributions from different regions of the world in the area of oil in place estimation. The dominant methodologies and research focus were also identified. These will assist scholars and researchers to identify the area of machine learning to navigate in the direction of solving petrophysics analysis problems. Our major limitation was the inability to search using the criteria of the synonyms of petrophysics, oil in place and machine learning. Few papers that used such synonyms would have been excluded from the study. In addition to this, our research resources could only support four (4) databases of Google Scholar, IEEE, ACM and Direct Science leading to the few (104) datasets (journal articles). We recommend that in future research, a broaden search be made to include such criteria and more databases be consulted.

6. STRENGTH OF THE STUDIES UNDER REVIEW

As reported in the studies under review, the review show that 12.5% of studies was capable of using a little data set on support vector machines to achieve great results while 87.5% of studies had strength in terms of performance. According to Omary and Mtenzi (2010), in an experiment conducted using different sample sizes from the University of California Irvine (UCI) machine learning repository, the sample sizes are categorized in relation to the number of attributes and number of instances available in the dataset. The dataset is categorized into small (little), medium and large datasets. With dataset that is less than 1473 instances classified as small (little), greater than 1473 but less than 2536 as medium and dataset that is equal to or greater than 4177 as large dataset.

7. LIMITATIONS REPORTED IN THE STUDIES

Limitations reported in the studies under review were data and performance limitations.

As reported in studies under review, a total of 63 studies had data inadequacy, 41 had performance limitation.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Benjamin S. Crude oil petroleum product. In: Encyclopedia Britannica Online; 2023. Available:https://www.britannica.com/science/crude-oil
2. Graham E, Ovadia JS. Oil exploration and production in Sub-Saharan Africa, 1990-present: Trends and developments. The Extractive Industries and Society; 2019. DOI: 10.1016/j.exis.2019.02.001
3. Sircar A, Yadav K, Rayavarapu K, Bist N, Oza H. Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Research. 2021;6(4):379-491.
4. Ivan M, Denis O, Dmitry K, Evgeny B, Aram K, Dmitry A. Machine learning for recovery factor estimation of an oil reservoir: A tool for de-risking at a hydrocarbon asset evaluation. Journal of Petroleum Science and Engineering. 2021;184:106513.
5. Lary DJ, Alavi AH, Gandomi AH, Walker AL. Machine learning in geosciences and remote sensing. Geoscience Frontiers. 2016;7(1):3-10.
6. Muhamedyev R. Machine learning methods: An overview. Computer Modeling & New Technologies. 2015;19(6):14-29.
7. McCoy JT, Auret L. Machine learning applications in minerals processing: A review. Minerals Engineering. 2019;132:95-109.
8. Shinde PP, Shah S. A review of machine learning and deep learning applications. In 2018 Fourth international conference on computing communication control and automation (ICCUBEA) IEEE. 2018:1-6.
9. Tariq Z, Aljawad MS, Hasan A, Murtaza M, Mohammed E, El-Husseiny A, Alarifi SA, Mahmoud M, Abdulraheem A. A systematic review of data science and machine learning applications to the oil and gas industry. Journal of Petroleum Exploration and Production Technology. 2021;11(12):4339-4374.
10. Amigun JO, Odole OA. Petrophysical properties evaluation for reservoir characterisation of seyi oil field (Niger-Delta). International Journal of Innovation and Applied Studies. 2013;3(3):756-773.
11. Okon AN, Adewole SE, Uguma EM. Artificial neural network model for reservoir petrophysical properties: Porosity, permeability and water saturation prediction. Journal of Modeling Earth Systems and Environment. 2020;7:2373-2390. Available:https://doi.org/10.1007/s40808-020-01012-4
12. Sowiżdżał K, Słoczyński T, Kaczmarczyk W. Estimation of the oil-in-place resources in the liquid-rich shale formations exploiting geochemical and petrophysical data in a 3D high-resolution geological model domain: Baltic Basin case study. Geofluids; 2020.
13. Ahmadi MA, Pournik M. A predictive model of chemical flooding for enhanced oil recovery purposes: Application of least square support vector machine. Petroleum. 2016;2(2):177-182.
14. Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters, 1999;9(3):293-300.
15. Fegh A, Riahi MA, Norouzi GH. Permeability prediction and construction of 3D geological model: application of neural networks and stochastic approaches in an Iranian gas reservoir. Neural Computing and Applications. 2013;23(6):1763-1770.
16. Mahmoud AA, Elkatatny S, Chen W, Abdulraheem A. Estimation of oil recovery factor for water drive sandy reservoirs through applications of artificial intelligence. Energies. 2019;12(19):3671.
17. Zerrouki AA, Aifa T, Baddari K. Prediction of natural fracture porosity from Well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoudoil field, Algeria. Journal of Petroleum Science and Engineering. 2014;115:78-89.
18. Zolotukhin AB, Gayubov AT. Machine learning in reservoir permeability prediction and modelling of fluid flow in porous media. In IOP Conference Series: Materials Science and Engineering. IOP Publishing. 2019;700(1):012023.
19. Anifowose F, Labadin J, Abdulraheem A. Predicting petroleum reservoir properties from downhole sensor data using an ensemble model of neural networks. In Proceedings of Workshop on Machine Learning for Sensory Data Analysis. 2013:27-34.
20. Tian J, Qi C, Sun Y, Yaseen ZM, Pham BT. Permeability prediction of porous media using a combination of computational fluid dynamics and hybrid machine learning methods. Engineering with Computers. 2021;37(4):3455-3471.

21. Abad ARB, Mousavi S, Mohamadian N, Wood DA, Ghorbani H, Davoodi S, et al. Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs. Journal of Natural Gas Science and Engineering. 2021;95:104210.
22. Jian H, Chenghui L, Zhimin C, Haiwei M. (). Integration of deep neural networks and ensemble learning machines for missing well logs estimation. Flow Measurement and Instrumentation. 2020;73:101748.
23. Oloso MA, Hassan MG, Bader-El-Den M, Buick JM. Hybrid functional networks for PVT characterization. In 2017 Intelligent Systems Conference (IntelliSys). IEEE. 2017:936-941.
24. Al-Mudhafar WA. Integrating machine learning and data analytics for geostatistical characterization of clastic reservoirs. Journal of Petroleum Science and Engineering; 2020. Available:http://www.elsevier.com/locate/petrol
25. Otchere DA, Ganat TOA, Gholami R, Ridha S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. Journal of Petroleum Science and Engineering. 2021;200: 108182.
26. Rahmanifard H, Maroufi P, Alimohamadi H, Plaksina T, Gates I. The application of supervised machine learning techniques for multivariate modelling of gas component viscosity: A comparative study. Fuel. 2021;285:119146.
27. Haagsma A, Scharenberg M, Keister L, Schuetter J, Gupta N. Secondary porosity prediction in complex carbonate reefs using 3D CT scan image analysis and machine learning. Journal of Petroleum Science and Engineering. 2021;207:109087.
28. Hsu T, Zhang Y. Petroleum engineering data text classification using convolutional neural network based classifier. In Proceedings of the 2018 International Conference on Machine Learning Technologies. 2018:63-68.
29. El-Amin MF, Subasi A. Developing a generalized scaling-law for oil recovery using machine learning techniques. Procedia Computer Science. 2019;163: 237-247.
30. Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering; 2007.
31. Omary Z, Mtenzi F. Machine learning approach to identifying the dataset threshold for the performance estimators in supervised learning. International Journal for Infonomics (IJI). 2010;3(3):314-325.
32. Okon AN, Adewole SE, Uguma EM. Artificial neural network model for reservoir petrophysical properties: Porosity, permeability and water saturation prediction. Journal of Modeling Earth Systems and Environment. 2020;7:2373-2390. Available:https://doi.org/10.1007/s40808-020-01012-4
33. Omary Z, Mtenzi F. Machine learning approach to identifying the dataset threshold for performance estimators in supervised learning. International Journal for Infonomics (IJI). 2010;3(3).

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/102150