



## Intelligent Directional Survey Data Analysis to Improve Directional Data Acquisition

Mohammed Imran Basheer Ahmed<sup>1</sup>, Rim Ali Zaghdoud<sup>1</sup>, Mohammed Salih Ahmed<sup>1</sup>, Mousa Alrabea<sup>1</sup>,  
Abdullatif Alsuwaidi<sup>1</sup>, Nawaf Alzaid<sup>1</sup>, Ahmed Alyousef<sup>1</sup>, Mohammad Aftab Alam Khan<sup>1</sup>,  
Atta-ur Rahman<sup>2\*</sup>, Sghaier Chabani<sup>3</sup>, Gomathi Krishnasamy<sup>4</sup>, Abdullah Alturkey<sup>1</sup>

<sup>1</sup> Department of Computer Engineering (CE), College of Computer Science and Information Technology (CCSIT), Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia

<sup>2</sup> Department of Computer Science (CS), College of Computer Science and Information Technology (CCSIT), Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia

<sup>3</sup> Department of Networks and Communications (NC), College of Computer Science and Information Technology (CCSIT), Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia

<sup>4</sup> Department of Computer Information System (CIS), College of Computer Science and Information Technology (CCSIT), Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia

Corresponding Author Email: [aaurahman@iau.edu.sa](mailto:aaurahman@iau.edu.sa)

<https://doi.org/10.18280/mmep.100214>

### ABSTRACT

**Received:** 10 June 2022

**Accepted:** 20 November 2022

#### Keywords:

*machine learning, deep learning, HSE, ANN, DNN, directional surveys data*

Many countries rely on oil and gas production as it is an essential part of the global economy. As a result, various challenges may thrive from the process of extracting oil and gas from the ground that may affect the operational aspects of the construction process. So, it is important to maintain the production and Health, Safety & Environment (HSE). The project aims to automate the process of the Directional Survey Data (DSD) in a way that can be cost-effective for the operational process and more stable for future use. DSD relates to the process of horizontal directional drilling (HDD) and raw data obtained from the surveys using survey stations on the way to bore hole like azimuth and inclination etc. In this work, we propose a fully automatic Directional Survey Data Analysis system based on the recognition patterns. The dataset comprised of 34069 real-time instances has been used. Two machine learning algorithms and four deep learning algorithms were investigated in this regard. For the deep learning approach RNN, LSTM, BI-LSTM, and Extreme Learning Machine (ELM) were used, and for the machine learning approach SVM and Naïve Bayes have been investigated. Selection of these candidate approaches was based on their promising nature in the related fields of study in terms of accuracy and precision. The experimental result demonstrated that Naïve Bays got 100% accuracy, ANN, LSTM and GRU managed to get 100% accuracy, BI-LSTM had a slightly lower accuracy achieving 98.7%, Simple RNN was lower than BI-LSTM achieving 82% accuracy, SVM got 81.1% accuracy, while ELM had the lowest performance receiving 55.3% accuracy. Overall, the scheme outperforms state-of-the-art techniques in the literature.

## 1. INTRODUCTION

The oil and gas are one of important industry in the world today, the prices directly affect the health of global the economy, as the oil and natural gas are considered one of the necessary resources that provide over half of the world's energy [1]. Since there is a high demand in oil and gas throughout the world, there are many risks oil and gas companies face like price risks, supply, and demand risks, and most importantly the operational cost risks. The process of drilling tends to be a difficult process due to the uncertainty of the worldwide production prices being beyond the control of any company in this industry [2]. This study focuses on automating the process of extracting oil and gas using Artificial intelligence techniques since the directional survey has been around since the start of the drilling in the oil and gas industry, this field is continuously developing to make the process of surveying, locating underground stations easier and cost-efficient as the tools have been constantly developed

throughout the years. However, by using artificial intelligence methods and techniques such as machine learning can reduce the costs even more and achieve very high and consistent results. Machine learning helps us in solving real-world problems by training a set of data, training the data continuously allows us to automatically improve the results. In today's world, many problems can be solved by using Machine learning techniques and algorithms. Such problems machine learning can solve have had a tremendous contribution to individuals. There are many machine learning techniques to be used. In Machine learning, the problems can be either classified into a regression problem or a classification problem. A regression problem means that the output must be a continuous number such as salary or weight. A classification problem means that the output is either "Yes" or "No", "0" or "1". By creating models such as regression or classification we can predict valuable information that will reduce costs and provide a reliable model for future use in the industry. There are many algorithms to be used in Machine learning,

depending on how complicated the problem is, some algorithms might not be optimal and will lead to poor results. Deep learning is descended from Machine learning, and it is mainly used to solve very complicated problems that Machine learning cannot solve. Algorithms used in Deep learning are much more powerful. We have conducted several comparisons between related works under specific criteria for the methods that were used, the accuracy rate, and the dataset used in the related work. However, no previous work has addressed this approach, which is to construct a machine learning model named Directional survey correction model (DSCM) using the directional data Survey analysis grounded data to improve the directional data acquisition and reporting system. The DSCM is focusing on automating human-related jobs like Rig Operator and Equipment Operator. Therefore, implementing the machine model in the directional survey will be as accurate as it can be, and ensures the safety of the operation while running the directional survey, to avoid false positive and false negative rates. The proposed work has utilized the recurrent neural network algorithms known as the Long Short-Term Memory (LSTM) and BI-LSTM. Furthermore, simple Recurrent Neural Networks (RNN), the support vector machine algorithm (SVM), and Naïve Bays are used in the experiment to construct various models and select the model that that proves to be accurate based on best possible result based on the dataset. The results were compared based on the testing accuracy rate. The highest is Naïve Bayes at 100% outperforming the deep learning algorithms like BI-LSTM, and Simple RNN while the ELM has the lowest performance of 55.3% accuracy rate.

The rest of the paper is organized as follows: section 2 highlights the related work and summarizes the related studies by finding the research gap. Section 3 provides the description of the proposed model. Section 4 provides the empirical results and discussion while section 5 concludes the paper.

## 2. RELATED WORK

In the oil and gas industries, some problems might cost a lot of money and effort to solve. With Artificial intelligence and machine learning techniques, many problems can be solved by using various algorithms such as Artificial neural networks (ANN). The literature review provides a lot of problems discussed in the field of oil and gas. It also provides a perspective on what type of algorithms are used and which works better. Working on the literature review is a critical factor in solving any problem. The developed item UGI42.03 provided some solutions for some problems that are divided into two parts. The first part used is mainly working with operations, and the second part is about improving the algorithms. Binder et al. [3] used an operational algorithm and implemented some equations. After the development, the researchers have gotten their results for the geophysical surveys they achieved to create a tool that is resistant to shock and high temperatures, as well as the issues of surveying for wellbores that are vertical, were solved by using micromechanical mechanics. In conclusion, by working on the study, they were focused on improving new methods for some types of surveys.

Seibi et al. [4] discusses the complexity of drilling wellbores that has been increasing over time, and it includes some of the techniques for the existing technologies of drilling, measurements, and data transmission. In this article, there will

be a comparison of the advantages and disadvantages of each technique. MWD A method used to provide wellbore path and other measurements. Engineers are looking for new methods and techniques for surveying and drilling; directional drilling and surveying were mentioned in this article, including their techniques and transmission of the data. Researchers used many techniques like MWD System and bending moment measurement, which are used mainly for surveying, Gravity MWD System, which is used for directional drilling, curvature deduction, and dogleg survey.

Akinpelu et al. [5] points out a mapping of the pollution in the environment using a system of artificial intelligence, including different algorithms and techniques. For the model, they used algorithms such as support vector regression (SVR), which can be enhanced by applying (MS), which is searching manually, and genetic algorithm (GA) which is considered a searching algorithm. To predict, they supported the following model by using correlation coefficient (CC), root means error, mean absolute percentage, and error. As shown, GA and SVR-Gauss had better performance than MS and SVR, GA and SVR Poly by the accuracy of 63% and 536%, also MS and SVR performed better than GA and SVR Poly by 288%. While ranking the algorithms, it will be as the following, GA-SVR-Gauss > MS-SVR > GA-SVR-Poly. In conclusion, this work used the algorithm SVR as a hybrid for the parameters GA and MS. After Using the Correlation Coefficient, and RMSE, MAE, and MAPD, it appeared that GA-SVR-Gaussian is better than other algorithms as listed.

Xu [6] focused mainly on learning algorithms for machine learning, such as the Extreme Learning Machine (ELM), which is used for hidden layers. In the paper, they compared two algorithms which are Extreme Learning Machine (ELM) and Least Squares Support Vector Machine (LSSVM). For a regression model, Neural Networks (NN) had many risks for the work; on the other hand, Support Vector Machine was the right algorithm to be used in these conditions. Neural Network (NN) takes a longer time to create a regression model, and nowadays, they have brought a new algorithm that is called Extreme Learning Machine (ELM). ELM is much faster in learning and easier than CNN, as well as solving different problems of engineering. As discussed, this paper points out the impact of algorithms on the localization of a plate structure, and models view the location of the issue. For a security system there must be some types of tools as detection tools to prevent unwanted outcomes.

Ahmad et al. [7], many algorithms were used, such as Support Vector Machine (SVM), to solve many issues like false alarms by decreasing it and the rate of detection by increasing it. Some techniques weren't efficient for large datasets, and for that, some of the classification techniques must be used like Support Vector Machine (SVM), Random Forest (RF), and Extreme Learning Machine (ELM), as they applied these techniques while working on this issue. After training and testing, the full data sample had 65,535 records. Support Vector Machine (SVM) had a 98.7% accuracy, Random Forest (RF) had a 97.7% accuracy, and Extreme Learning Machine (ELM) had 99.5% accuracy. These accuracies are received using full records of the Dataset, and it points out that ELM got the best result.

Škrjanc and Vulić [8] provides information about directional wells and the anti-collision because collisions can be a big issue while planning for a position for a wellbore as well as surveying and drilling. It was in Alaska on the north slope, which is a huge space between seas. A comparison was

made for error models, which are Operator Wellbore Survey Group (OWSG) and Industry Steering Committee for wellbore surveying accuracy (ISCWSA). The goal of this study was to improve anti-collision methods and guides for drilling to avoid collisions, which means drilling in a safe direction due to the problems occurring while drilling in a wrong direction or place. The separation factor (SF) indicates collision avoidance as its equation is by dividing the center-to-center distance by minimum separation. As a result, it appears that the Operator Wellbore Survey Group (OWSG) has better performance, and it should be used as a default model.

Mahajan et al. [9] provides information about drilling operations for fluids that are mud. There are many objectives to achieve while drilling mud, mainly avoiding blowouts and kicks and by keeping the wellbore stable. Researchers have used the machine learning (ML) approach to achieve some of their goals. Different Machine learning methods were discussed and compared for either supervised or unsupervised methods to predict the issues that could appear while drilling. Machine learning algorithms used are artificial neural networks (ANN), random forest (RF), which is used for a regression approach, XGBoost (XGB), support vector machines (SVM), and K-nearest neighbors (KNN), and lastly, Markov decision process (MDP). As a result, they applied these algorithms in many approaches, and some of the mentioned techniques performed the best and achieved the greatest result, such as ANN and RF.

Gul [10], investigates a random forest algorithm was applied that observed the productivity of field-wide, including the data injection as a given value to estimate the timelapse oil profiles within wellbore locations. The progress is estimated using calculated field data by getting the complex structure, heterogenous, and heavy faulted offshore reservoir. This algorithm was obtained to apply the inverse modeling standard, which uses the productivity of the field, and it is an injection to estimate the time-lapse saturation. Wireline logging has limitations not only in the productivity and it is injection but includes interruption, limitation of the operation tools, difficulties in deviated wells, and pulling the tubing pump. The limitations and risks could include well intervention and how it can face an increase in the cost besides risks of its operations that could bring limitations for the frequency of data acquisition. Data acquisition could cause some limitations while applying the model since the given data is not dependable when it comes to visibility, and it is a sufficient perspective. The given input features in this model are wide-field injection and productivity, and the produced output will represent how deeply the time-lapse of oil saturation inside the location of the well by using the reservoir matching history technique. The outcome of the model shows the efficiency of applying the random forest algorithm for getting matched results between productivity and injection with the oil saturation.

Wang et al. [11], there is a concentration on the progress of novel-driven data and their models. By using deep learning with its analysis techniques of the data to get productivity progresses using good horizontal pairs and their temperature from vertical wells observation to customize the barriers. Data were calculated and gathered from many different fields which are working on the oil, such as geometrics, proportions, and spatial distributions of barriers for the building of SAGD models. The neural network of deep learning that was applied is known as CNN to catch the complexity of nonlinear relations between inputs and their outputs parameters. Because

of how the thickness of bitumen oil sands, conventional recovery methods like water flooding do not work in all situations. There were two horizontal wells drilled parallelly in the ground of the reservoir, the upper well considered as the injector and the lower well considered as the producer. Usually, it is difficult to locate the barriers and customize their sizes and dimensions. The CNN model was implemented to solve this issue in two ways. Getting a bunch of geological models by the static data and applying the changed operations of models each time by the dynamic data, which in addition will be helpful to decrease the mismatch of productivity between models and their data. To analyze the inputs and outputs of the driving of the data in the models, there were several parameterization phases applied. One of the main purposes behind doing this experiment is to verify if the deep learning techniques could be applied within the dynamic integration of the data for reservoir customization.

**Table 1.** Summary of related work

Ref.	Method/Algorithm	Accuracy	Dataset
[3]	Operational algorithm MWD System, bending moment measurement,	-	-
[4]	Gravity MWD System, curvature deduction, dogleg survey	-	-
[5]	support vector regression (SVR), MS Genetic Algorithm (GA)	-	-
[6]	Extreme Learning Machine (ELM), Least Squares Support Vector Machine (LSSVM), ANN, CNN	-	-
[7]	SVM, Random Forest (RF), ELM	98.7%, 97.7%, 99.5%	65535 records
[8]	Tangential, Average angle Balanced tangential, Radius of curvature, Minimum curvature, Helical arc	-	-
[9]	OWSG, ISCWSA, Separation factor (SF)	-	-
[10]	ANN, RF, XGBoost, SVM, KNN, MDP	-	-
[11]	Random forest, Functional networks, LSTM, SVM, Decision trees. Linear regression, KNN, Gradient boosting, Ada boost, multi-layer perceptron	90%.	PF-1C: 23680 samples PF-14: 18648 samples PF-12: 5835 samples PF-11: 14644 samples
[12]	BP neural network, Genetic algorithm with BP NN (GA-BP model)	GA-BP is higher than BP.	-

Liang et al. [12], a Genetic algorithm (GA) and Back Propagation neural network (BPNN) are applied to analyze the relations functions between predicted outputs and the number of customized parameters in the drilling related data. The model decreased error occurred by the usage of one feature parameter, and higher the speed of BP using genetic algorithm, additionally avoiding dealing with local extreme value. By that, the overflow drilling diagnosis accomplished and decreased the percentage of wrong and missed judgments of drilling

kicks. Drilling kicks are considered as big damage to the drilling speed, and it is produced quality. Also, it could cause fluid and leakage of the drilling progress, which results can be considered as a large amount of loss in the cost. These risks of drilling can be prohibited and decreased in their possibility of occurring if it is discovered at the beginning of their occurrence. We will compare the literature reviews by first comparing the key elements in every paper that we have reviewed such as the method, the accuracy rate, and the number of samples in the dataset, and discuss the findings of the studies alongside the techniques that performed well in the studies and managed to achieve the highest accuracy. This comparison will be between related works under specific criteria for the methods that were used, the accuracy rate, and the dataset used in the related work. A summary of the related work is listed in Table 1.

From the comprehensive literature review, we have noticed that problems that were solved using ANN have shown satisfiable and reliable results. However, the Dataset plays a huge role. Some datasets may perform poorly using ANN.

For the undergoing study, we believe that trying several techniques such as ANN, Recurrent neural networks (RNN), Deep neural networks (DNN), and Long Short-term memory (LSTM) will help us to get the best and most reliable model that is possible by using these techniques. Some of the papers consisted of surveying and drilling [3]; it was noted that to maintain a high accuracy during the operations, many techniques were used to achieve higher accuracy and to maximize the efficiency of drilling [4]. Methods used such as MWD System, Gravity MWD System, curvature deduction, dogleg survey [8] and the tangential method and their improved versions. Some of these methods were used for surveying, and others were used for drilling [5]. The algorithm Support Vector Regression provided a good result for predicting discrete values, as well as being enhanced by other algorithms like (MS) that are manual searching, or genetic algorithm (GA), that is used as a searching algorithm [6]. After comparing different types of algorithms, it appeared that Extreme Learning Algorithm (ELM) has better performance than algorithms like CNN because ELM is faster [7]. ELM had the best accuracy result in some of the reviewed papers, which proves how useful the algorithm [9]. Issues like collision occurrence should be solved by methods such as OWSG that had better performance and result, and ISCWSA that are considered anti-collision methods [10]. ANN was one of the best algorithms used during searching and studying these papers because it had some of the best accuracies and results [11]. The results of the model show the efficiency of applying the random forest algorithm for getting matched results between productivity and injection with the oil saturation [12]. Nonlinear relations contain a lot of complex computations. Depending only on the general algorithms, the model fit of the data will not be accomplished as it is supposed to be. Therefore, neural networks will be helpful to get the diagnosis faults and their recognition patterns.

### 3. DESCRIPTION OF THE PROPOSED TECHNIQUES

The algorithms that will be used in this Proposed work are RNN, LSTM, BI-LSTM, Extreme Learning Machine (ELM), SVM and Naïve Bayes. Based on the literature reviews, these techniques performed well in the studies and managed to achieve the highest accuracy. In the following sections the proposed techniques will be explained technically.

### 3.1 Feed-forward neural network

Feed-forward neural network (FFNN) is one of the main techniques that are used in Deep learning. Feedforward networks are mainly used for supervised learning tasks, they are known to be easy to build. FF networks do not have any memory structure which makes it a drawback when dealing with sequence data. FFNN introduces a new concept called layers. There are mainly three layers in FFNN, input, output, and hidden layers [13, 14]. Figure 1 highlights the layered structure of FFNN. The algorithms that will be used in this Proposed work are RNN, LSTM, BI-LSTM, Extreme Learning Machine (ELM), SVM and Naïve Bayes. Based on the literature reviews, these techniques performed well in the studies and managed to achieve the highest accuracy. In the following sections the proposed techniques will be explained technically.

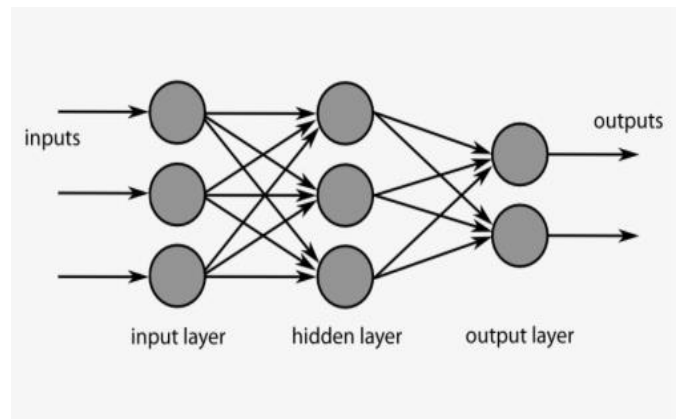


Figure 1. Feedforward neural networks layers

Input layer is the first layer in the feedforward neural network which is used to read the data and feed it into the network. Its sole purpose is to get the data into the system. Output layer is the last layer of the feedforward neural network which is used to output the predictions. The number of neurons in the output layer is determined based on the nature of the problem whether it was a classification problem or regression problem. For the classification problem, the number of neurons is equal to the number of classes. for the regression problem: the aim is to predict a single value and set that value into a single neuron in the output layer. It can take advantage of an activation function depending on the nature of the problem whether it was a linear activation for regression, sigmoid for binary and SoftMax for multi-classification problems. Hidden layer: consists of multiple layers that are created to ensure the approximation of the nonlinear functions. It introduces the bias terms, as it constantly adds calculations that will affect each neuron in the next layer. Bias is there to shift the activation function. It can play a crucial role in the convergence of the network. The Hidden layer takes advantage of the activation functions like Sigmoid, Tanh, and Rectified Linear Unit (ReLU). The number of each neuron can increase or decrease, so it is much more flexible than the input and output layers when it comes to the number of neurons being modified [15, 16]. The number of neurons in the hidden layer varies from application to application and is usually determined by the hit and trial method. The same approach is used to find the optimum number of the hidden layer neuron in the proposed technique.

### 3.2 Recurrent neural networks

Recurrent neural network (RNN) is a class of artificial neural networks. It is derived from feed forward neural networks which exhibit temporal dynamic activity. It can deliver predicted outcomes like the way the human brain functions. RNN is like the traditional neural networks, however it has a memory state that is added to the neurons. That helps the neuron to get optimized faster and hence the overall convergence and/or learning rate improves significantly. It is also known as neural network with memory. It keeps on recurring the previous state of memory to get gradually better. The following Figures 2-3 demonstrate the key difference between the Recurrent Neural Network (RNN) and FFNN. Figure 2 provides the RNN architecture.

RNN can produce outputs, copy that output and loop back to itself several times. Where in feed-forward neural networks (FFNN) the data moves only in one direction, so it produces an output, but it is not able to loop back to the neural network. There are different types of RNN, and these types are One to One, One to Many, Many to One and Many to Many. Recurrent Neural Networks handle short-term dependencies when the resources are limited because it is not computationally intensive. However, simple RNN suffers from various technical problems of Vanishing/Exploding Gradients in the Deep Neural Networks [17, 18].

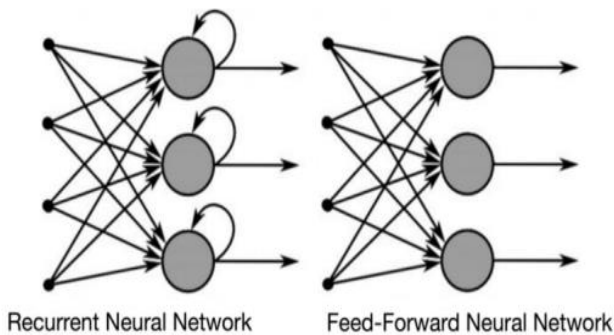


Figure 2. RNN architecture

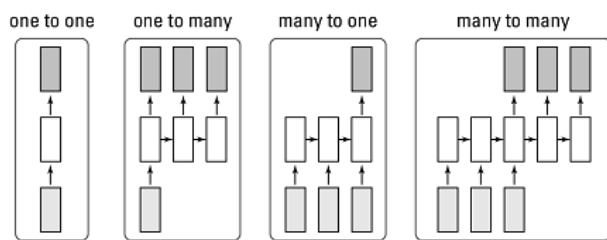


Figure 3. Types of RNN

### 3.3 Long short-term memory network

Long Short-Term-Memory (LSTM) is another type of Recurrent neural networks. An improved version that achieves higher accuracy and upstanding performance, mainly focused on sequence data problems. Unlike RNN the structure of LSTM is slightly different. LSTM consists of three gates, input, output, and forget gate. These gates regulate the flow of information in and out of the LSTM unit. The forget gate is responsible for either throwing unnecessary information keeping them for the next process by applying the sigmoid function, if the result is near 0 it will be overlooked and

deleted, on the other hand if the result is approaching toward 1 it will be sufficient and saved. Input gate is responsible for updating the cell state. The output gate decides what the hidden state should be. The LSTM model is held by a memory cell called the cell state that can maintain its state over time. This transfers the related information into the next states of the neural network. In LSTM, two main activation functions are used, Tanh and Sigmoid. The downside of using LSTM is the exploding gradient problem, where sometimes gradient values may go to infinity. This is prevented by adding a little constant or boosting factor. Another downside is that applying LSTM is computationally expensive [19, 20]. Figure 4 shows the LSTM architecture.

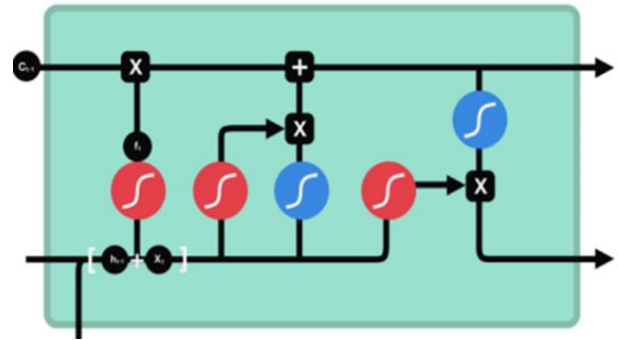


Figure 4. LSTM architecture

That can be overcome in the bi-directional LSTM. The Unidirectional LSTM addresses the problem of being time consuming and computationally expensive, due to it preserving the learning progress of the past only because the only input that is seen is from the past. The Bidirectional long-short term memory (BI-LSTM) overcomes this issue through combining two LSTM layers that preserve information from both past and future context by enabling the neural networks to have the sequence information that appear in both directions backwards (future to past), or forward (past to future). BI-LSTM provides good results that can be used for text classification and speech recognition models [21, 22]. Both architectures are given in Figure 5. As depicted in the figure, in Bi-LSTM one LSTM unrolling from left to right (say LSTM1/upper hidden layer) on the input (say X). While other LSTM form right to left (say LSTM2/lower hidden layer) and same input dataset stream is fed to both.

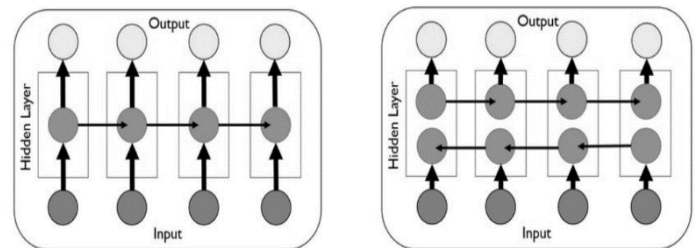


Figure 5. LSTM & BI-LSTM architectures

### 3.4 Extreme learning machines

The Extreme Learning Machine (ELM) is a learning algorithm for the single hidden layer feedforward neural network. Compared to the traditional methods ELM has a better performance in terms of accuracy, efficiency, and



stability, it is faster and harvests a high performance that has the capability to overcome the over-fitting problems and the slow training speed. ELM is based on the empirical risk minimization theory, and it needs a single iteration for its learning process. Because of its robustness and fast learning rate, ELM is applied in various real-time learning for classification, clustering, and regression tasks [23-25]. The architecture of ELM is depicted in Figure 6. The complete mathematical derivation of ELM can be found in similar work [23, 24] where ELM were applied to two different problems namely software define network and predicting diabetes type II, respectively.

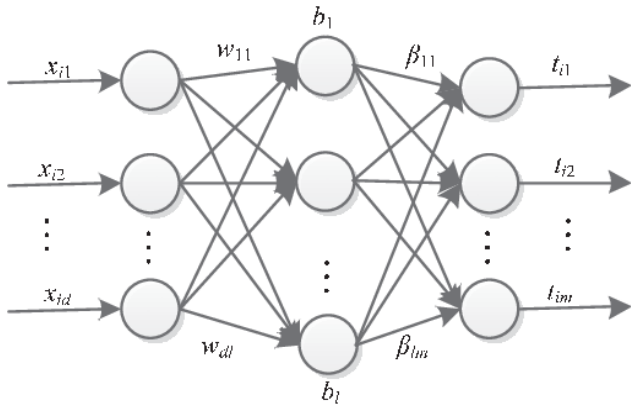


Figure 6. ELM architecture

### 3.5 Support vector machine

Support Vector Machine (SVM), also known as Support Vector Classification, is a linear and supervised machine learning algorithm. The algorithm is mostly used for classification problems, but it can apply to regression and classification problems. It can perform the classification test by creating a hyperplane that separates the data into different classes, for all the points of one category to appear on one side, and the points of the other category appear on the other side of the hyperplane. The main objective of SVM is to achieve the highest separation between two categories by maximizing the margin of the hyperplane. The data points that appear to be closer to the hyperplane are called support vectors. One of the greatest advantages of the algorithm is that it functions well when the number of dimensions appears to be greater than the number of samples [26-28]. However, the main drawback of SVM is that it functions poorly while dealing with a large data set, as it requires more training time to provide the best performance. The function for optimal hyperplane is given in Eq. (1).

$$hx = wTx + b \quad (1)$$

### 3.6 Naïve bayes

The Naïve Bayes classifier is a probabilistic machine learning model that is used for classification tasks, it is based on the Bayes theorem. The algorithms are used in many fields like spam filtering, and recommendation systems. the pros of the Naïve Bayes algorithms that it provides a high accuracy to show results due to its efficiency [29, 30]. However, the main setback of it is that it requires the predictors to be independent, and that can restrict the performance of the classifier. The formula for using this algorithm is as shown in Eq. (2).

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)} \quad (2)$$

where,  $P(c)$ : Probability of hypothesis;  $P(x)$ : Probability of data,  $P(c|x)$ : Probability of hypothesis to this data, and  $P(x|c)$ : Probability of data which proves this hypothesis.

## 4. EXPERIMENT AND RESULTS

This section describes the dataset and how it was cleaned. It also includes a detailed explanation of the experimental setup and how the models are constructed. Lastly, the optimization process and the performance measures that were applied in the proposed work.

### 4.1 Dataset description

The dataset comprises 34068 rows allocated for 157 wellbores and 13 columns which represent the wellbore number, and it is main measurements which are MD, Inclination, Azimuth, and the validity of the station based on those three measurements. The rest columns are representing the eight rules which clarify why the station is not considered a valid station, with one column for each rule. All the algorithms considered in the study have been trained with the same set of instances after due preprocessing was conducted on the dataset. Such as normalization, eliminating redundant instances and missing values and handling outliers.

### 4.2 Experimental environment

The experiment was implemented using Python, machine learning algorithms were done using Jupyter notebook, and deep learning algorithms were done using Google Colab due to the simplicity of libraries in installation on Google Colab.

### 4.3 Performance measures

The proposed model is evaluated using the confusion matrix, precision, and recall, accuracy, and F1-score measures. The confusion matrix will contain a summary of the model predictions, to identify the mistakes for each class in the model and calculate the correct and incorrect predictions [31, 32]. In the Precision, true positive values are required to determine the precision. In the recall, false positive values are required to determine the recall. The accuracy rate is another type of performance measure, the classification accuracy can be computed by dividing the number of true positives and false positives by the total number of samples (in the formula N indicates the total number of samples). the F1-score can be defined as the harmonic mean between precision and recall, it indicates how stable and how accurate is the classifier [33-40]. The measures are given in Eq. (3) to Eq. (6).

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

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

$$Accuracy = \frac{\#Correct\ Predictions}{N} \quad (5)$$

$$F1 = 2 * \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (6)$$

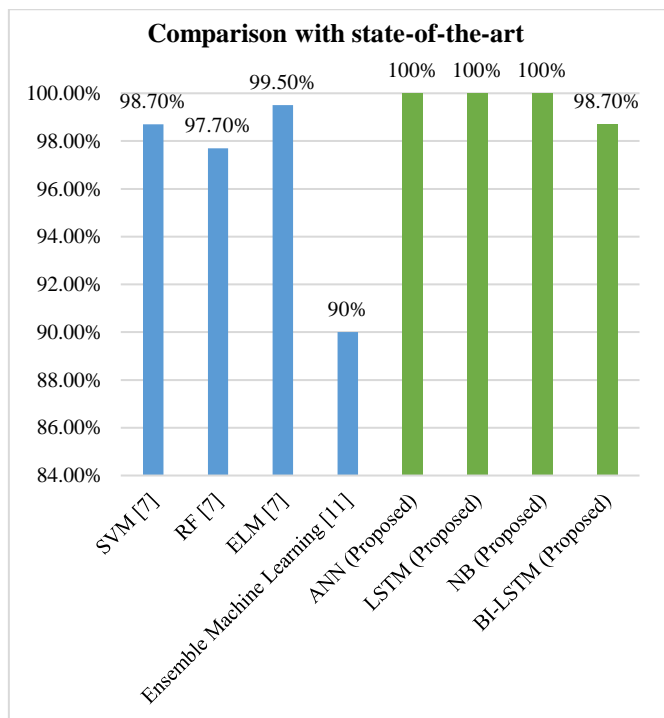
#### 4.4 Results and discussion

Table 2 indicates the result of the seven considered algorithms, most of the algorithms managed to score a high accuracy rate above 80%. The parameters for the network's structures are already explained in the previous sections. The deep learning algorithms ANN, LSTM, and GRU managed to get a 100% on all the evaluation metrics including the accuracy rate, Precision, Recall and F1 score. BI-LSTM had a slightly lower accuracy rate achieving 98.7%, BI-Simple RNN was lower than BI-LSTM achieving 82% accuracy rate, and ELM had performed poorly making it the lowest score across all the algorithms reaching an accuracy rate of 55.3%. The machine learning algorithms SVM and Naïve Bayes provided better results, SVM managing to get 100% in recall, 81.1% in accuracy rate, and 98.6% in F1 score. Naïve Bayes has provided exceptional results managing to get 100% in Accuracy, Precision, Recall, and F1 Score.

**Table 2.** Summary of the result

Algorithms	Accuracy	Precision	Recall	F1 Score
ANN	100%	100%	100%	100%
Simple RNN	82%	82%	99.5%	89.9%
LSTM	100%	100%	100%	100%
GRU	100%	100%	100%	100%
BI-LSTM	98.7%	84.5%	100%	91.6%
ELM	55.3%	-	-	-
SVM	81.1%	81.1%	100%	98.6%
Naïve Bayes	100%	100%	100%	100%

Figure 7 presents a comparison of the proposed scheme with the state-of-the-art schemes in the literature namely [7] and [11]. Apparently the proposed scheme outperforms in terms of accuracy, precision, recall and F1-score, respectively.



**Figure 7.** Comparison with the state-of-the-art schemes

#### 5. CONCLUSIONS AND FUTURE WORK

This paper concentrates on directional survey analysis to improve directional data acquisition and reporting systems using machine learning techniques. A dataset contains 34069 instances. Several deep learning and machine learning classifiers were evaluated to the proposed features including ANN, Simple RNN, LSTM, BI-LSTM, GRU, ELM, SVM, and Naive Bayes. We managed to predict whether the oil and gas station is valued or invalid based on the classification machine learning model. For evaluation, the scheme has been compared with similar approaches in the literature and results are promising. For future work, it is recommended to implement the machine learning model to suggest multi-station analysis corrections to the coordinates based on a regression machine learning model. The survey data quality can be assessed and corrected to improve survey database quality by leveraging aspects such as consistency, integrity, accuracy, and timeless elements. Meanwhile, optimizing well construction life cycle modeling such as wellbore placement, ROI, geological and petrophysics simulations. Further hybrid intelligent systems and deep learning models and frameworks can be investigated on the same dataset to further fine tune the results [41-50].

#### REFERENCES

- [1] What are the top five facts everyone should know about oil exploration? <https://www.forbes.com/sites/quora/2013/04/03/what-are-the-top-five-facts-everyone-should-know-about-oil-exploration/?sh=328418f3d501>, accessed on May 1, 2022.
- [2] 5 of the biggest risks faced by oil and gas companies. <https://www.investopedia.com/articles/fundamental-analysis/12/5-biggest-risks-faced-by-gas-and-oil-companies.asp>, accessed on May 01, 2022.
- [3] Binder, Y.I., Paderina, T.V., Rozentsvein, V.G. (2010). High-rate precision directional surveys in small-diameter wellbores: Results from practical implementation. *Gyroscopy and Navigation*, 1(1): 66-72. <https://doi.org/10.1134/S2075108710010104>
- [4] Seibi, A., Karrech, A., Boukadi, F., Pervez, T. (2009). Wellbore path estimation using measurement while drilling techniques: a comparative study and suggestions for improvements. *Energy Sources, Part A*, 31(14): 1205-1216. <https://doi.org/10.1080/15567030802087502>
- [5] Akinpelu, A.A., Ali, M.E., Owolabi, T.O., Johan, M.R., Saidur, R., Olatunji, S.O., Chowdbury, Z. (2020). A support vector regression model for the prediction of total polyaromatic hydrocarbons in soil: An artificial intelligent system for mapping environmental pollution. *Neural Computing and Applications*, 32: 14899-14908. <https://doi.org/10.1007/s00521-020-04845-3>
- [6] Xu, Q. (2014). A comparison study of extreme learning machine and least squares support vector machine for structural impact localization. *Mathematical Problems in Engineering*, 2014: 906732. <https://doi.org/10.1155/2014/906732>
- [7] Ahmad, I., Basher, M., Iqbal, M.J., Rahim, A. (2018). Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE Access*, 6: 33789-33795.

- <https://doi.org/10.1109/ACCESS.2018.2841987>
- [8] Škrjanc, Ž., Vulić, M. (2016). Comparison of the directional survey calculation methods applied on real well data. *Measurement*, 94: 239-244. <https://doi.org/10.1016/j.measurement.2016.08.002>
- [9] Mahajan, N.H., Khataniar, S., Patil, S.L., Dandekar, A.Y., Fatnani, A.K. (2018). Anti-collision risk management guidelines for Alaska North Slope directional wells. *Journal of Petroleum Science and Engineering*, 166: 650-657. <https://doi.org/10.1016/j.petrol.2018.03.069>
- [10] Gul, S. (2021). Machine learning applications in drilling fluid engineering: A review. In *International Conference on Offshore Mechanics and Arctic Engineering*, 85208: V010T11A007. <https://doi.org/10.1115/OMAE2021-63094>
- [11] Wang, B., Sharma, J., Chen, J., Persaud, P. (2021). Ensemble machine learning assisted reservoir characterization using field production data—An offshore field case study. *Energies*, 14(4): 1052. <https://doi.org/10.3390/en14041052>
- [12] Liang, H., Zou, J., Liang, W. (2019). An early intelligent diagnosis model for drilling overflow based on GA-BP algorithm. *Cluster Computing*, 22: 10649-10668. <https://doi.org/10.1007/s10586-017-1152-5>
- [13] Rahman, A.U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Mosavi, A. (2022). Rainfall prediction system using machine learning fusion for smart cities. *Sensors*, 22(9): 3504. <https://doi.org/10.3390/s22093504>
- [14] Ghazal, T.M., Al Hamadi, H., Umar Nasir, M., Gollapalli, M., Zubair, M., Adnan Khan, M., Yeob Yeun, C. (2022). Supervised machine learning empowered multifactorial genetic inheritance disorder prediction. *Computational Intelligence and Neuroscience*, 2022: 1051388. <https://doi.org/10.1155/2022/1051388>
- [15] Alhaidari, F., Shaib, N.A., Alsafi, M., Alharbi, H., Alawami, M., Aljindan, R., Zagrouba, R. (2022). ZeVigilante: Detecting Zero-Day malware using machine learning and sandboxing analysis techniques. *Computational Intelligence and Neuroscience*, 2022: 1615528. <https://doi.org/10.1155/2022/1615528>
- [16] Ibrahim, N.M., Gabr, D.G.I., Rahman, A.U., Dash, S., Nayyar, A. (2022). A deep learning approach to intelligent fruit identification and family classification. *Multimedia Tools and Applications*, 81(19): 27783-27798. <https://doi.org/10.1007/s11042-022-12942-9>
- [17] Khan, M.B.S., Nawaz, M.S., Ahmed, R., Khan, M.A., Mosavi, A. (2022). Intelligent breast cancer diagnostic system empowered by deep extreme gradient descent optimization. *Mathematical Biosciences and Engineering*, 19(8): 7978-8002. <https://doi.org/10.3934/mbe.2022373>
- [18] Nasir, M.U., Khan, S., Mehmood, S., Khan, M.A., Rahman, A.U., Hwang, S.O. (2022). IoMT-based osteosarcoma cancer detection in histopathology images using transfer learning empowered with blockchain, fog computing, and edge computing. *Sensors*, 22(14): 5444. <https://doi.org/10.3390/s22145444>
- [19] Gollapalli, M., Musleh, D., Ibrahim, N., Khan, M.A., Abbas, S., Atta, A., Omer, A. (2022). A neuro-fuzzy approach to road traffic congestion prediction. *Computers, Materials and Continua*, 72(3): 295-310.
- [20] Musleh, D., Alotaibi, M., Alhaidari, F., Mohammad, R.M. (2023). Intrusion Detection System Using Feature Extraction with Machine Learning Algorithms in IoT. *Journal of Sensors and Actuator Networks*, 12(2): 29. <https://doi.org/10.3390/jsan12020029>
- [21] Alqarni, A., Rahman, A. (2023). Arabic Tweets-based sentiment analysis to investigate the impact of COVID-19 in KSA: A deep learning approach. *Big Data and Cognitive Computing*, 7(1): 16. <https://doi.org/10.3390/bdcc7010016>
- [22] Zaman, G., Mahdin, H., Hussain, K., Abawajy, J., Mostafa, S.A. (2021). An ontological framework for information extraction from diverse scientific sources. *IEEE Access*, 9: 42111-42124. <https://doi.org/10.1109/ACCESS.2021.3063181>
- [23] Alhaidari, F., Almotiri, S.H., Al Ghamdi, M.A., Khan, M.A., Rehman, A., Abbas, S., Khan, K.M. (2021). Intelligent software-defined network for cognitive routing optimization using deep extreme learning machine approach. *Computers, Materials & Continua*, 67(1): 1269-1285.
- [24] Gollapalli, M., Alqahtani, T.A., Alhamed, D.H., Alnassar, M.R., Alajmi, A.M., Alali, Y.H., Abdulqader, M.M., Saadeldin, A. (2023). Intelligent modelling techniques for predicting used cars prices in Saudi Arabia. *Mathematical Modelling of Engineering Problems*, 10(1): 139-148. <https://doi.org/10.18280/mmep.100115>
- [25] Olatunji, S.O., Alansari, A., Alkhorasani, H., Alsubaii, M., Sakloua, R., Alzahrani, R., Alsaleem, Y., Alassaf, R., Farooqui, M., Ahmed, M.I.B., Alhiyafi, J. (2022). Preemptive diagnosis of Alzheimer's disease in the Eastern Province of Saudi Arabia Using Computational Intelligence Techniques. *Computational Intelligence and Neuroscience*, vol. 2022, Article ID: 5476714, 2022. <https://doi.org/10.1155/2022/5476714>
- [26] Alotaibi, S.M., Basheer, M.I., Khan, M.A. (2021). Ensemble machine learning based identification of pediatric epilepsy. *Computers, Materials & Continua*, 68(1): 149-165.
- [27] Rahman, A. (2020). GRBF-NN based ambient aware realtime adaptive communication in DVB-S2. *Journal of Ambient Intelligence and Humanized Computing*, 1-11. <https://doi.org/10.1007/s12652-020-02174-w>
- [28] Rahman, A.U., Dash, S., Luhach, A.K. (2021). Dynamic MODCOD and power allocation in DVB-S2: A hybrid intelligent approach. *Telecommunication Systems*, 76: 49-61. <https://doi.org/10.1007/s11235-020-00700-x>
- [29] Mahmud, M., Lee, M., Choi, J.Y. (2020). Evolutionary-based image encryption using RNA codons truth table. *Optics & Laser Technology*, 121: 105818. <https://doi.org/10.1016/j.optlastec.2019.105818>
- [30] Dash, S., Luhach, A.K., Chilamkurti, N., Baek, S., Nam, Y. (2019). A Neuro-fuzzy approach for user behaviour classification and prediction. *Journal of Cloud Computing*, 8(1): 1-15. <https://doi.org/10.1186/s13677-019-0144-9>
- [31] Gollapalli, M., Alamoudi, A., Aldossary, A., Alqarni, A., Alwarthan, S., AlMunsour, Y.Z., Abdulqader, M.M., Mohammad, R.M., Chabani, S. (2022). Modeling algorithms for task scheduling in cloud computing using CloudSim. *Mathematical Modelling of Engineering Problems*, 9(5): 1201-1209. <https://doi.org/10.18280/mmep.090506>
- [32] Olatunji, S.O., Alansari, A., Alkhorasani, H., Alsubaii, M., Sakloua, R., Alzahrani, R., Alsaleem, Y., Almutairi, M., Alhamad, N., Alyami, Albandari., Alshobbar, Z.,



- Alassaf, Reem., Farooqui, M., Ahmed, M.I.B. (2022). A novel ensemble-based technique for the preemptive diagnosis of rheumatoid arthritis disease in the Eastern Province of Saudi Arabia Using Clinical Data. *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 2339546. <https://doi.org/10.1155/2022/2339546>
- [33] Musleh, D., Aldhafferi, N., Alqahtani, A., Alfifi, H. (2018). Adaptive communication for capacity enhancement: A hybrid intelligent approach. *Journal of Computational and Theoretical Nanoscience*, 15(4): 1182-1191. <http://dx.doi.org/10.1166/jctn.2018.7191>
- [34] Khan, M.A., Abbas, S., Atta, A., Ditta, A., Alquhayz, H., Khan, M.F., Naqvi, R.A. (2020). Intelligent cloud-based heart disease prediction system empowered with supervised machine learning. *Computers, Materials & Continua*, 65(1): 139-151.
- [35] Naseem, M.T., Qureshi, I.M., Muzaffar, M.Z. (2020). Robust and fragile watermarking for medical images using redundant residue number system and chaos. *Neural Network World*, 30(3): 177-192.
- [36] Gollapalli, M. (2022). Ensemble machine learning model to predict the waterborne syndrome. *Algorithms*, 15(3): 93. <https://doi.org/10.3390/a15030093>
- [37] Gollapalli, M.A., Chabani, S. (2022). Modeling and verification of aircraft takeoff through novel quantum nets. *Computers, Materials and Continua*, 72(2): 3331-3348.
- [38] Abbas, T., Fatima, A., Shahzad, T., Shahzad, T., Rahman, A., Alissa, K., Ghazal, T.M., Al-Sakhnini, M.M., Abbas, S., Khan, M.A., Ahmed, A. (2023). Secure IoMT for disease prediction empowered with transfer learning in healthcare 5.0, the concept and case study. *IEEE Access*, 2023(1): 1-1. <https://doi.org/10.1109/ACCESS.2023.3266156>
- [39] Mahmud, M., Sultan, K., Aldhafferi, N., Alqahtani, A., Musleh, D. (2018). Medical image watermarking for fragility and robustness: A chaos, error correcting codes and redundant residue number system based approach. *Journal of Medical Imaging and Health Informatics*, 8(6): 1192-1200. <http://dx.doi.org/10.1166/jmih.2018.2431>
- [40] Min-Allah, Nasro, Alahmed, B.A., Albreek, E.M., Alghamdi, L.S., Alawad, D.A., Alharbi, A.S., Al-Akkas, N., Musleh, D., Alrashed, S. (2021). A survey of COVID-19 contact-tracing apps. *Computers in Biology and Medicine*, 137: 104787. <https://doi.org/10.1016/j.compbiomed.2021.104787>
- [41] Al Metrik M.A., Musleh, D.A. (2022). Machine learning empowered electricity consumption prediction. *Computers, Materials & Continua*, 72(1): 1427-1444. <https://doi.org/10.32604/cmc.2022.025722>
- [42] Alghamdi, A.S., Rahman, A. (2023). Data mining approach to predict success of secondary school students: A Saudi Arabian case study. *Education Science*, 13: 293. <https://doi.org/10.3390/educsci13030293>
- [43] Gollapalli, M., Alfaleh, A. (2022). An artificial intelligence approach for data modelling patients inheritance of sickle cell disease (SCD) in the eastern regions of Saudi Arabia. *Mathematical Modelling of Engineering Problems*, 9(4): 1079-1088. <https://doi.org/10.18280/mmep.090426>
- [44] Gollapalli, M., Alansari, A., Alkhorasani, H., Alsubaii, M., Sakloua, R., Alzahrani, R., Al-Hariri, M., Alfares, M., AlKhafaji, D., Al Argan, R., Albaker, W. (2022). A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM. *Computers in Biology and Medicine*, 147: 105757. <https://doi.org/10.1016/j.compbiomed.2022.105757>
- [45] Imran, M., Alsuhaibani, S.A. (2019). A neuro-fuzzy inference model for diabetic retinopathy classification. *Intelligent Data Analysis for Biomedical Applications*. Academic Press, pp. 147-172. <https://doi.org/10.1016/B978-0-12-815553-0.00007-0>
- [46] Alassaf, R.A., Alsulaim, K.A., Alroomi, N.Y., Alsharif, N.S., Aljubeir, M.F., Olatunji, S.O., Alahmadi, A.Y., Imran, M., Alzahrani, R.A., Alturayef, N.S. (2018). Preemptive diagnosis of diabetes mellitus using machine learning. In *2018 21st Saudi Computer Society National Computer Conference (NCC)*, pp. 1-5. <https://doi.org/10.1109/NCG.2018.8593201>
- [47] Olatunji, S.O., Alotaibi, S., Almutairi, E., Alrabae, Z., Almajid, Y., Altabee, R., Altassan, M., Ahmed, M.I.B., Farooqui, M., Alhiyafi, J. (2021). Early diagnosis of thyroid cancer diseases using computational intelligence techniques: A case study of a Saudi Arabian dataset. *Computers in Biology and Medicine*, 131: 104267. <https://doi.org/10.1016/j.compbiomed.2021.104267>
- [48] Sajid, N.A., Ahmad, M., Rahman, A.U., Zaman, G., Ahmed, M.S., Ibrahim, N.M., Ahmed, M.I.B., Krishna, G., Alzaher, R., Alkharraa, M., Alkhulaifi, D., Alqahtani, M., Salam, A.A., Saraireh, L., Gollapalli, M.A.S., Ahmed, R. (2023). A novel metadata based multi-label document classification technique. *Computer Systems Science and Engineering*, 46(2): 2195-2214. <http://dx.doi.org/10.32604/csse.2023.033844>
- [49] Musleh, D. Alkhales, T.A., Almakki, R.A., Alnajim, S.E., Almarshad S.K., Alhasaniah, R.S., Aljameel, S., Almuqhim, A.A. (2022). Twitter Arabic sentiment analysis to detect depression using machine learning. *Computers, Materials & Continua*, 71(2): 3463-3477. <http://dx.doi.org/10.32604/cmc.2022.022508>
- [50] Olatunji, S.O., Alsheikh, N., Alnajrani, L., Alanazy, A., Almusairii, M., Alshammasi, S., Alansari, A., Zaghdoud, R., Alahmadi, A., Basheer Ahmed, M.I., Ahmed, M.S., Alhiyafi, J. (2023). Comprehensive machine-learning-based models for the pre-emptive diagnosis of multiple sclerosis using clinical data: A retrospective study in the Eastern Province of Saudi Arabia. *International Journal of Environmental Research and Public Health*, 20(5): 4261. <https://doi.org/10.3390/ijerph20054261>