**CHAPTER TWO**

**LITERATURE REVIEW**

* 1. **Conventional Petroleum Accumulation**

Petroleum is made up of hydrocarbons that are the byproducts of fossilized organic matter, mostly plankton, which has been buried under silt to create source rocks. Liquids and gases have formed as a result of the high temperature at great depths. Petroleum was discharged from the source beds because it was less dense than the surrounding water. It then moved upward through porous rock, including some sandstone and limestone, until it was eventually blocked by nonporous rock, such shale or solid limestone. The bending, faulting, and erosion of the Earth's crust led to the formation of geologic structures that eventually contained petroleum reserves.

Petroleum can exist either separately or in combination in gaseous, liquid, or near-solid forms. The more solid phase may be referred to as bitumen, tar, pitch, or asphalt, whereas the liquid phase is frequently referred to as crude oil. When both phases coexist, the liquid often covers the more-solid phase while the gas typically rests on top of the liquid.

New sources of rock oil were actively sought once it was learned in the 19th century that rock oil would provide a distilled product (kerosene) suited for lanterns. It is now widely accepted that Edwin Laurentine Drake's 1859 well at Titusville, Pennsylvania, United States, was the first one to be dug particularly for oil. The success of this well, which was dug close to an oil seep, inspired more drilling nearby and eventually brought about similar exploration elsewhere. By the turn of the century, other governments and nations had begun digging their own oil wells due to the rising global demand for petroleum goods.

Around the world, 150 million barrels of crude oil were produced in 1900. The majority (80%) of the remaining was manufactured in the United States, with half of it being generated in Russia.

The second decade of the 20th century saw the introduction and expansion of vehicle use, which increased demand for petroleum-based products. In 1925 and 1940, the annual production was more than one billion barrels. By the final decade of the 20th century, more than 20 billion barrels were being produced annually by almost a million wells spread across more than 100 nations.

* 1. **Petroleum Prospecting and Exploration**

The original Drake well was constructed not far from a known crude oil surface seepage. Such seepages were the sole verifiable signs of subsurface oil and gas production for many years. New techniques for assessing the potential of subsurface rock formations were developed, nevertheless, as demand increased. To effectively explore for oil today, data from seismic surveys, geologic framing, geochemistry, petrophysics, Geographic Information Systems (GIS) data collection, geostatistics, drilling, reservoir engineering, and other surface and subsurface investigative techniques must be integrated. The main technique for searching for petroleum is geophysical exploration, which includes seismic analysis. In more complicated and difficult exploration conditions, such sub-salt structures and deep water, gravity and magnetic field approaches are also historically trustworthy evaluation techniques.

Beginning with GIS, gravity, magnetic, and seismic surveys enable geoscientists to efficiently narrow their search for target assets to investigate, decreasing the risks involved with exploration drilling. There are three primary types of exploratory methods: (1) surface approaches, such as geologic feature mapping, made possible by GIS, (2) area surveys of gravity and magnetic fields, and (3) seismographic methods. These approaches reveal the existence or absence of subsurface characteristics that are favorable for petroleum accumulations. There is still no technique to anticipate the presence of profitable subterranean oil reserves with 100 percent accuracy.

Another strategy relies on surface signs of potential subsurface rock formations. Subsurface folds and faults in rock formations are sometimes replicated in surface structures. A modest protrusion in an otherwise level ground surface, for example, may suggest the existence of subsurface salt domes. Oil and gas accumulations are frequently caused by uplift and faulting in the geological formations surrounding these domes.

* + 1. **Surface Techniques**

Crude oil seeps can appear as a tar-like deposit in a low terrain, such as the oil springs mentioned by Marco Polo near Baku, Azerbaijan, on the Caspian Sea. They usually appear as a thin skim of oil on little streams that run through a region. This latter phenomena is what gave rise to the name Oil Creek in Pennsylvania, where Drake's well was dug. Natural gas seeps are seldom visible, but sensors may detect natural gas concentrations in air as low as 1 part in 100,000. Similar devices have been used to detect traces of gas in saltwater. These geochemical surface prospecting approaches are inapplicable to the vast majority of petroleum reservoirs with no surface leakage.

* + 1. **Gravity and Magnetic Survey**

Although gravity is essentially constant at the Earth's surface, it is somewhat higher where solid rock formations are close to the surface. As a result, gravitational force rises over the tops of anticlinal (arch-shaped) folds while decreasing over the tops of salt domes. A sensitive sensor called as a gravimeter may detect very slight variations in gravitational force. Measurements are taken on a precise grid across a vast region, and the findings are plotted and evaluated to determine the existence of prospective oil or gas reserves.

Magnetic surveys make advantage of the magnetic characteristics of particular types of rock, which impact the Earth's normal magnetic field when they are close to the surface. Sensitive devices are now again utilized to map abnormalities across broad regions. Surveys are frequently conducted from aircraft over land and from ocean-going boats across continental shelves. Magnetotellurics (MT), a related technology, measures the natural electromagnetic field at the Earth's surface. Anomalies are caused by the various electrical resistivities of rock formations, which are then mapped and interpreted to represent subterranean geologic characteristics. Before more expensive and time-consuming seismic surveys, MT is becoming a more cost-effective filter for identifying a petroleum play (a group of oil fields or petroleum deposits with comparable geology features). MT is sensitive to what is present within the strata of the Earth.

Subsalts (salts whose bases have not been totally neutralized by acid) are particularly resistant to electromagnetic waves, whereas porous rocks are frequently conductive due to the saltwater and brines trapped inside them. Anomalies such as salt domes are used by petroleum geologists to identify possible stratigraphic traps for petroleum.

* + 1. **Seismographic techniques**

Even if there are no surface indications of the presence of large geologic anomalies such as anticlines (arch-shaped folds in subterranean layers of rock), fault blocks (sections of rock layers separated by a fracture or break), and salt domes, the survey methods described above can reveal their presence. They cannot, however, be depended on to detect smaller, less visible traps and unconformities (gaps) in the stratigraphic arrangement of rock strata that may contain petroleum reserves. Seismic surveying, which takes use of the sound-transmitting and sound-reflecting qualities of subsurface rock formations, may identify and locate these. Seismic waves move at varying speeds through different types of rock formations and are reflected by interfaces between them. Typically, the sound-wave source is a tiny explosion in a shallow drilled hole. To pick up and record the transmitted and reflected sound-wave arrivals, microphones are set at varying distances and orientations from the explosive site.

The operation is done at regular intervals throughout a large region. An skilled seismologist can then evaluate the collected data to determine the outlines of the subsurface formations.

The primary difference between offshore and land-based seismic data collecting is the manner of setup. One of the most crucial components of offshore seismic surveys is knowing where the ship and receivers are at all times, which is made possible by transferring real-time global positioning system (GPS) information from satellites to GPS reference and monitoring stations, and subsequently to the ship. Real-time readings have become an integral aspect of the seismic sound-wave capture, data processing, and analysis process.

**2.3. Oil Production Prediction**

Hydrocarbon exploration and production is a global sector that is both technically difficult and financially risky. Because well performance prediction is crucial for profit making, it begins in the early phases of production to anticipate future recovery. If a well's expected output does not meet a certain level, it might be remedied or even shut down to prevent additional loss.

For future production prediction, it is usual practice to compare early stage production data from a new well with historical production data from surrounding wells. Type curve matching is a well-known method for estimating cumulative production. Analogous wells are initially categorized based on the similarity of their cumulative production curves, and the average production profiles for each group of wells are determined as the type curves.  (Mohaghegh & Gaskari, 2009; Sproule, 2015)

The early cumulative output of a new well is then matched to the various type curves, and the closest curve is chosen by engineers to estimate the new well's future well performance. Matching type curves is efficient but subjective. It is difficult to replicate the research because various engineers may produce different estimates for a certain well based on their visual interpretations of the type curves. Another difficulty is that cumulative production curves tend to appear quite similar to one another over time, making it difficult to differentiate performance variances in individual wells.

According to the research literature, data mining techniques are gaining popularity in the oil industry for production data analysis because they are objective, resistant to poor data quality, and more accurate than statistical approaches (Mohaghegh et al., 2008; Ma et al., 2015; Zhong et al. 2015). Different data mining methods, such as Support Vector Machines, Random Forests, and Boosted Regression Trees, have shown to be efficient and useful tools for understanding oil and gas operation and production (LaFollette et al. 2012; Esmaili et al. 2013; Zhong et al. 2015). However, the time-series production data in these research were either averaged or represented by their maximum or cumulative values.

These statistical measurements oversimplify production curves and do not reflect each well's whole production pattern. As a result, an effective and understandable data-driven prediction technique based on time-series production data will be investigated.

**2.4 Artificial Neural Network**

A complicated non-linear relationship between input and target variables can be identified or approximated using an artificial neural network (ANN), a technique in artificial intelligence. Artificial neural networks are computational models inspired by biological systems that are capable of machine learning and pattern recognition. They are based on human central nervous systems, particularly the brain, and are used in computer science and related disciplines. Over the past few decades, artificial neural networks have attracted a lot of attention and are effectively being used in a variety of scientific fields, including physics, engineering, economics, and geology. These problems are more about classification, control and complex behavior prediction (Nikravesh, et al. 2003; Graupe 2007).

The neurons (nodes) in ANN's hidden layers use nonlinear activation functions to process a weighted sum of input data. An artificial neural network (ANN) is created by teaching a network to reflect the inherent connections contained in data. The network is trained using a data set that contains both input (or predicting) and target variables. The unknown network parameters (weights and biases) are estimated using an inverse problem procedure that minimizes the discrepancy between the network output and the known values of the target variables.

The concept of a neural network dates back to 1943, when mathematician Walter Pitts and neurophysiologist Warren McCulloch published a paper on the potential functions of neurons (McCulloch and Pitts. 1943). To describe how neurons works in the brain worked, they created a simple neural network model using electrical circuits. It was only in the 1950s that it became feasible to simulate a fictitious neural network, thanks to developments in computer science and technology. It has been noted in past decades that neural networks made up of numerous neurons can be more effective than those made up of only one neuron. The gradient descent-based learning principles for large neural networks were created and popularized by McClelland and James in 1986. Recently, scientists have made enormous advances in neural networks and created cutting-edge methods like self-organizing maps and associative memories (Mehrotra, et al. 1997). Numerous sources, including Mehrotra et al. (1997), Poulton (2001), and Suh (2012), give thorough background material on the development of artificial neural networks.

Any neural network is taught using a learning algorithm and a training data set. Unsupervised learning and supervised learning are the two main categories of neural network learning methods. The supervised learning is used to uncover a hidden structure in unlabeled data.

The goal is to classify, find features or regularities in the training data, or both. A cluster analysis is the most common use of unsupervised learning. In contrast, the supervised learning method requires that target values be provided. The input vector requires a training dataset, and will generate the rules according to the desired output by adjusting the weights.

The weights are then used to process the test data set's inputs. The weights will be adjusted after providing the desired output to the net to match the model to the desired goal. The learning process iteration will be continued till the intended goal is realized.

**2.5 Artificial Neural Networks Architecture**

An artificial neural network is formed of several artificial neurons that are coupled together according to a certain network architecture. The two most prevalent forms of network architectures are Single Layer Perceptron (SLP) and Multilayer Perceptron (MLP) which comprises of one input and output layer with any number of hidden layers, as seen in Fig. 2.1.

The number of layers and neurons (nodes) in each layer must be chosen when designing a neural network. There is a trade-off between accuracy and overfitting of the data: with insufficient neurons, the mismatch between network predictions and actual values of the target variables could not be minimized, while with excessive neurons, the overfitting of the network parameters could occur.

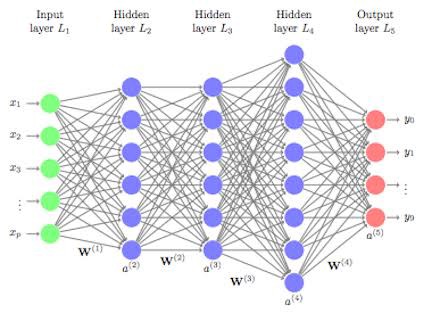


Fig.2.1: Multilayer Perceptron (MLP) neural network configuration

**2.6 Application of Artificial Neural Networks in Petroleum Engineering**

Petroleum engineering has used neural networks for a variety of purposes, including classification (Stundner and Al-Thuwaini. 2001), reservoir characterization or property prediction (Tang, et al. 2011; Raeesi, et al. 2012; Aminian, et al. 2003; Chaki, et al. 2014), proxy for recovery performance prediction (Aw), and more (Mohaghegh. 2000; Maschio and Schiozer. 2014), as well as the design or optimization of industrial processes and well trajectory. (Stoisits, et al. 1999; Ayala H and Ertekin. 2007; Artun, et al. 2012; Yeten, et al. 2002; Oberwinkler, et al. 2004; Malallah and Nashawi. 2005; Zangl, et al. 2006).

Specifically, neural networks have been used in recent years as a proxy model to perform EOR (enhanced oil recovery) screening, predict heavy oil recoveries, characterize reservoir properties in unconventional reservoirs, and more (Queipo, et al. 2002; Popa, et al. 2011; Popa and Cassidy. 2012; Ahmadloo, et al. 2010; Fedutenko, et al. 2012; Fedutenko, et al 2014; Mohammadpoor, et al. 2012).

The overall size of the training data set and the input vector dimension continue to influence how many free parameters (i.e., weights and bias connections) are present in a hidden layer.

There are several correlations between training data set size and user-defined error parameters in the literature (Haykin, 2001), and extrapolation detection methods are used in some of these relationships. Several design challenges pertaining to the creation of ANNs have recently been reviewed, and can be found in  Al-Bulushi et al. (2012).

Most studies on network architecture design adhere to a few broad guidelines and arrive at the ideal network architecture by trial and error, however others have suggested generalizing training to include learning an acceptable architecture through Bayesian inference, where Poulton (2001) discusses probabilistic networks as an approach.

The usage of functional neural networks, where the activation functions and weight connections connected with the neurons might fluctuate and be approximated through training, was also recently advocated in the literature (El-Sebakhy, et al. 2012). The majority of projects make use of back-propagation or other gradient-based optimization techniques as the learning algorithm, only a few authors have considered the application of more generic global optimization algorithms like a genetic algorithm optimization approaches like the learning algorithm (Saemi, et al. 2007).

Numerous studies also stress the significance of data pre-processing, which includes normalization and outlier detection for input and target data utilized in training (Tang, et al. 2011). Additionally, experimental design is frequently carried out in applications where answers from thorough flow simulations are utilized to train a network that would act as a stand-in for reservoir performance prediction. This reduces redundant training data and cuts down on computation time (Queipo, et al. 2002).

In order to account for heterogeneity and uncertainty in reservoir parameters, several authors have also suggested include indices in the set of predictive variables, such as a Dykstra Parsons coefficient and the permeability ratio of neighboring strata (Ahmadloo, et al. 2010).

**Neural Network (NNet)**

Jia (Jia and Zhang 2016) forecasted Barnett Shale production using neural networks. Jia preprocessed the data in his research by giving each individual point a distinct weight in order to reduce the impact of outlier data. He provided the bottom hole pressure and time series history matching findings in the prediction section. In a comparison of the final prediction findings from the neural network model with those from the exponential, harmonic, and hyperbolic models, it was found that the neural network model produced far more precise forecasts than did the other three traditional empirical models. Fig. 2 depicts the straightforward neural network architecture used by Jia (Jia and Zhang 2016).

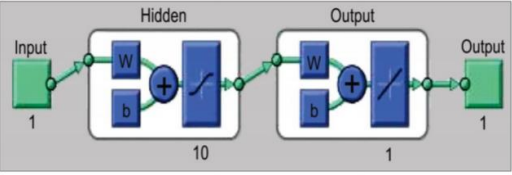


Fig. 2— Simple neural network structure (adapted from Jia and Zhang 2016)

**Random Forest (RF)**

According to Aulia (Aulia et al. 2014), a subset of the bottom-hole pressure (BHP) can contribute to the oil recovery factor in the field, therefore he used Latin Hypercube Monte Carlo (LHMC) and random forest (RF) to discover such subsets. Aulia further mentioned that because RF can determine the relevance of RF's independent variables based on a collection of uncertainty runs, LHMC and RF coupled can be a worldwide method to sensitivity analysis. Hedge (Hedge et al. 2015) demonstrated the feasibility of predicting rate of penetration during real-time operations utilizing statistical learning approaches such as trees, bagging, and random forest.

Hedges also predicted that random forest and bagging approaches might be used to evaluate the relative relevance of input parameters, providing useful information for drilling parameter forecasts and optimizing on-the-fly rig floor modifications.

**Support Vector Machine (SVM)**

Zhao (Zhao et al. 2006) used a water saturation curve computed from density and resistivity logs of a gas well in the Gulf of Mexico to train a -intensive support vector machine (SVM) to execute the regression of water saturation from seismic data. This method estimates the water saturation away from the wellbore and can be used to discriminate between commercial and low saturation gas. Zhao (Zhao et al. 2014) demonstrated the application of proximal SVM (PSVM) to classify lithofacies; particularly, Zhao employed PSVM to distinguish limestone from shale in a Barnett Shale gas field. Zhao's solution was based on two PSVM applications, one for waveform classification and the other for well data classification.

 The PSVM classifier's validity in binary classification was shown by the promising findings in both seismic and well log data. Artificial neural networks (ANN) and SVM were both used in Anifowose's paper (Anifowose et al. 2012) to estimate porosity and permeability of oil and gas reservoirs with carbonate platforms. The findings suggest that SVM outperforms ANN. Anifowose's solution uses stratified sampling rather than the traditional static technique of data division, with six datasets used. The SVM method was supposed to help petroleum exploration engineers estimate various attributes more accurately, resulting in less exploration time and more output.

With the use of prior information (supervised learning), several researchers in the oil sector are using machine learning and deep learning to anticipate production of water, gas, and oil as well as to identify patterns in well test data. This section will discuss work that uses machine learning to forecast oil, water, and gas output as well as the well pressure and well oil/gas/water ratio. With the currently employed techniques, such as simulation, the well production forecast process is time-consuming and expensive in terms of computer resources. Due to the fact that our dataset is a continuous (time-series) problem type, we had to thoroughly investigate and investigate the likelihood of getting good results with our dataset (Abdullayeva, F. 2019)

The methods now in use have a bottleneck effect when creating new wells. Numerous ML and DL strategies have been applied to improve the performance of production prediction as technology has advanced. Additionally, the preliminary search and comprehension of the pertinent machine learning research addressing well-performed classification and prediction served as the foundation for the organization of this part. Additionally, the research were carefully examined, and numerous concepts connected to the core theme were gathered from several papers. The literature research was then divided into various sub-areas of oil production that were pertinent to our topic.

**2.1. Oil Flowrate Prediction**

According to operating parameters and choke size, the volume of oil flowing was determined in a recent artificial intelligence research. For critical and subcritical flow, the study employed two distinct datasets: choke size, upstream pressure, temperature, gas/oil ratio, and water cut for critical flow. Along with downstream pressure, the subcritical dataset included the same data as the critical flow dataset. Then, a range of AI technologies, including artificial neural networks, fuzzy logic, and functional networks, as well as widely used techniques like Gilbert correlation, were put to the test. The ANN system demonstrated notable accuracy with a correlation coefficient of 0.89 when compared to current correlation techniques, which have a maximum correlation value of 0.30 (al Ajmi, et. Al 2015).

An orifice flow meter's oil flow rate was recently predicted using machine learning algorithms, according to a study. From a dataset of 1037 data records, inputs for pressure, temperature, viscosity, square root of differential pressure, and oil-specific gravity were taken. The employed machine learning (ML) techniques, including the adaptive neuro-fuzzy inference system (ANFIS), least squares support vector machine (LSSVM), radial basis function (RBF), multilayer perceptron (MLP), and gene-expression programming (GEP), all produced results with high levels of accuracy, with correlation coefficients varying from 0.90 to 0.99 (Ghorbani, H. 2020).

According to the authors, tight gas fields constitute a significant source of hydrocarbons for the energy sector. Large amounts of information are produced by the process. As a consequence, data may be assessed, forecasted, and utilized to spot patterns between dependent and independent variables using machine learning techniques. In this study, the generalized linear model (GLM) and the artificial neural network (ANN) were both used. To determine how well-planned new wells will recover is the study's main objective. A dataset of 224 wells was evaluated for production data. These models' predictions were contrasted with the actual initial gas production rate of the wells. A GLM model had a mean square error of 1.57 while an ANN model had a mean square error of 1.24, according to research.

Additionally, the reservoir thickness was shown to be responsible for 36.5 percent of the initial gas output, followed by flowback rate (29%), according to the ANN model's performance index. As a consequence, the ANN model beat the GLM model when it came to projecting gas production and looking back (Ghorbani, H 2020).

The authors discussed how to anticipate the flow rate of oil production in circumstances when direct measurement is challenging, which is a challenge encountered by petroleum engineers across the world. In order to provide a novel methodology for calculating oil flow rate in two-phase oil and gas flow through wellhead chokes, this study uses an artificial neural network approach.

As a function of the choke upstream pressure, choke size, and generating gas-to-oil ratio, the oil flow rate is estimated using the artificial neural networks (ANNs) model. The accuracy of the proposed model was tested to many well-known empirical correlations, and the findings showed that the projected oil flow rates of the new model are quite near to actual observed data. Additionally, this model estimates the oil flow rate more quickly and accurately than empirical correlations with more than three input parameters, with an accuracy rate of about 87% (Mirzaei-Paiaman, A.; Salavati, S 2012)

**2.2. Production of Wells**

A research has been proposed that describes a hybrid model for time series prediction of oil output based on a combination of CNN and LSTM networks. The model's CNN layer is applied to the current time window first, extracting the features, and then the LSTM is used to forecast the relationship between the time windows. The goal of this study is to use deep neural networks to build a system that can estimate oil output accurately based on well debt data. They conclude that the CNN + LSTM model's root mean squared logarithmic error (RMSLE) value is 0.186891 (Abdullayeva, F. 2019).

The authors presented their study as not only constructing but also evaluating a proxy model utilizing field data from 1239 horizontal wells. As a consequence of this investigation, the following conclusions were reached. Initially, the datasets were investigated using the exploratory data analysis (EDA) approach, which comprised outlier analysis, categorical and numerical variable analysis, and correlation analysis. From a total of 1239 wells, 1150 were used as training data using EDA. Second, principal component analysis (PCA) was used to minimize the size of the input variable. Finally, VIA employed the RF, GBM, and XGBoost algorithms to find independent factors with a significant influence on cumulative gas production during a 12-month projection (CGP12). Using this strategy, the average relevance ranking of the independent variables was calculated.

Finally, hyperparameter sensitivity analysis is utilized to build more predictive DNN models (Han, D.; Kwon, S. 2021)

The author reviewed a previous paper that described the use of RNN-LSTM to predict oil, gas, and water outputs of wells based on injection patterns in a time-series way. The RNN-LSTM model estimated oil, water, and gas output with a first-year accuracy of more than 90% and production values with a 73.63 percent accuracy for up to 5 years (Pal, M. 2021).

The authors described their work by proposing an automated method for assisting technical teams in increasing data quality in production data analysis through the use of machine learning techniques, improving reliable production forecasting, lowering operating costs, and optimizing drilling schedules.

Input characteristics for this study included reservoir pressure, water cut, wellhead pressure, choke size, and rate. The following machine learning techniques were effectively employed to remove noisy data and outliers. For instance, the K-means clustering technique is employed to identify outliers or anomalies. Second, at the analysis stage, support vector regression (SVR) is a potent method for reducing noise from data. The aforementioned techniques can increase the precision of production forecasts (Doan, T.T.; van Vo, M. 2020)

The study examined how supporters predicted oil, water, and gas output in reservoirs with water injection using artificial neural networks. The researchers used the Bayesian regularization method to train the ANN model.

The ANN model attained a coefficient of determination of higher than 0.9. The delay in the time-step term is a crucial component that might enhance the model's prediction, which the researchers also found since other models disregard it (Negash, B.M.; Yaw, A.D. 2020)

The idea that various factors, including geology and completion, play a role in gas output has been the subject of a study proposal. Using machine learning techniques, we can develop a production prediction model and identify the key factor impacting production. There were 159 horizontal wells that were aimed towards the Duvernay formation. Then, the crucial variables were discovered using Pearson correlation and grey-connection analysis. Finally, three statistical models were produced using multiple linear regression (MLR), support vector regression (SVR), and Gaussian process regression (GPRFor the first six months of production, total oil and gas output was predicted (Guo, Z.; Wang et. Al 2021)

A study that uses the attention mechanism and combines it with the convolutional neural network (CNN) and the long short-term memory neural network to create a hybrid model has been proposed (attention-CNN-LSTM). Two wells in a southern Chinese oilfield provided daily average data for this investigation. Each day of the 23 years that the T1 well data were gathered represents a unique record, while the T2 well data were only collected for 17.5 years. Support vector regression (SVR), back-propagation neural network (BP), standard CNN LSTM, and other techniques have also been investigated. With an RMSE (root mean square error) score of 0.315 and 0.402, an MAE score of 0.218 and 0.303, and a MAPE (mean absolute percentage error) score of 0.008 and 0.005 for T1 and T2 wells, respectively, attention-CNN-LSTM appears to have the lowest score uncontested among all other models, though, according to three different comparison measures (Jabbari, M.; et. Al 2021).

The suggested approach for calculating oil production from a single well is discussed in a recent paper. They applied the LSTM technique. To get a better outcome, the model has been altered and adjusted. Several optimization techniques, such the batch normalization layer, which increases prediction accuracy, can speed up training and improve the model's generalization skills. From January 1973 to May 1995, 1275 wells are included in the dataset. The utilized dataset also included 43 characteristics. Additionally, the author devised a data-labeling technique based on the various water cut phases. Each well must be a part of at least one of the four stages. The program also provides other labeling options, such well type. They chose 65 wells from the dataset, which had been split into 90/10, 1409 samples for the training, after the tool had preprocessed the data. For the test set, the remaining 157 samples were taken into account. The author employed RF and SVM in addition to the LSTM to compare them. RMSE, MAS, and R2 are the employed metrics. With an RMSE score of 383.33, the LSTM has the lowest value, indicating higher performance. In addition, LSTM's value in MAE is 285.34, which is the lowest. The LSTM approach has the greatest R2 of all of the techniques, at 0.786 (Xia, L.; Shun et. al 2020).

The study states that the hydraulic fracture design parameters, fracture network features, such as fracture spacing and fracture conductivity, and simulation data from 370 reservoir simulations were used to build the three ANN architectures. The output of shale gas is greatly impacted by these variables. Based on an additional set of 92 simulations, the testing outcomes confirmed a strong connection between input and goal functions, with an R2 better than 0.86. Additionally, throughout the course of 1, 5, 10, and 15 years, there was good agreement between observed and forecast total gas output, with an R2 > 0.94 and error rates < 15%. For wells that have been producing for a brief period of time, using the peak production rate can increase forecast accuracy (Nguyen-Le, V.; Shin, H. 2020).

**2.3. Prediction of Well Production Enhancement**

A research that examines a special deep recurrent neural network built for use by petroleum engineers to examine the characteristics of active oil wells has been offered. Additionally, this study examines the most widely used time-series models using the deep structure of a GRU recurrent neural network, which serves as the model's foundation. The purpose of this study is to evaluate and contrast the layered DGRU model's robustness and accuracy to the most widely used time-series models. They end up with a collection of 21,000 observations of US oil production data spanning 90 years after pre-processing the information. Training 40%, Validation 20%, and Testing 20% make up the three subgroups that comprise the whole dataset collection. Additionally, this method can be used to assess the effectiveness of both the trained model and the stacked deep GRU model using the testing dataset (unknown new data). The DGRU model also assessed the performance and accuracy of other time-series models. Multi-RNN, single GRU, Multi-GRU/ANN, and Multi-LSTM models are among those used in this work. Using R2 measurement and a 10-step prediction process, the total accuracy performance results were as follows:

• GRU: 55.078%

• GRU/ANN Multi: 72%

51% for Multi LSTM

• DGR Stacked: 70%

They discovered that the model's design is a great deal more straightforward than that of LSTM and RNN. Our DGRU model performs better than other traditional models due to its low level of complexity and capacity to handle long interval time-series information. As a consequence, our DGRU model may be utilized to predict long-term relationships in a complex time series dataset (Al-Shabandar et. Al 2021).

**Prediction of Pressure Gradients**

The usage of a multilayer feed-forward ANN was described in order to estimate the horizontal oil-water flow pressure gradient utilizing five inputs: oil superficial velocity, water superficial velocity, pipe diameter, pipe roughness, and oil viscosity. The data from 765 experiments were separated into three sections: training (60%), validation (20%), and testing (20%).

When compared to the other techniques investigated (two-fluid model, homogeneous model, and correlation), the ANN model had an APE of 0.30 percent, an AAPE of 2.9 percent, and an SD of 7.6 percent, which is extremely low (Al-Wahaibi, T.; Mjalli, F.S. 2014).

The authors presented their work by developing and implementing machine learning techniques. The pressure gradient in a liquid-liquid flow was calculated using six different methods, including support vector machine (SVM), Gaussian process (GP), random forest (RF), artificial neural network (ANN), k-nearest neighbor (kNN), and a fusion model. Using SVM, seven predictor values were picked as the best collection of predictors with the fewest errors. The seven predictors include oil and water velocities, FP, input diameter, oil and water density, and oil viscosity. With the exception of the ML-fusion model, which had a (p 0.05), the GP model had the best prediction accuracy in compared to other models (Wahid, M.F. et. al 2021)

**2.5 Fault Prediction**

The early development of a machine learning (ML) model for early fault prediction of a centrifugal pump in the oil and gas sector that is straightforward and deployable was studied. Process and equipment sensors on the chosen machinery were utilized to review past data in real-time. Raw sensor data, mostly from temperature, pressure, and vibration probes, was denoised, pre-processed, and sequentially coded in order to train the model.

In this study, support vector machine (SVM) and multilayer perceptron are two techniques that are compared (MLP). High accuracy was recorded by the SVM algorithm at 98.1%. The MLP method outperformed the SVM algorithm in terms of total accuracy, recording 98.2%. The main goal of this work is not to build very accurate ML models, but rather to show that it is feasible to make accurate predictions using a straightforward and understandable ML approach. Overall, the findings imply that the suggested algorithms deliver good overall classification performance and are effective in determining the health condition of the monitored machine (Orrù, P.F. et. al 2020).

A research has been suggested that addresses an unique hybrid LSTM-SAE learning approach that will be utilized to improve fault detection accuracy while addressing the drawbacks of RNN training and the usage of a single methodology alone. All of the vibration indications are provided by the DCS of the electrical generator. Approximately 2000 samples were divided into a training dataset (80%) and a testing dataset (20%) before being input into the suggested DL architecture. These samples are separated into two groups: training and testing datasets with errors and datasets without errors. The sample data were selected and utilized to train and test RNN-LSTM in preparation for use in the fault detection stage of the suggested DL architecture. The following key contributions to the present work were intended for this study:

• A unique DL fault detection framework is developed based on the RNN-LSTM, SAE, and particle swarm optimization (PSO) techniques to balance the three processes of parameter optimization, fault feature extraction, and fault detection.

• The suggested DL technique can not only adapt the essential features, but also materialize patterns without preserving the past sequence inputs, thanks to an unique hybrid mathematical approach that addresses RNN training restrictions such as fading error, deficit, gradient vanishing, and backflow.

roaches, with a 99.67% detection accuracy. Furthermore, the findings show that the proposed deep learning system can detect errors in industrial data without using labels. This can help data engineers properly extract characteristics and avoid relying on human expertise based on the unsurprised fault detection technique. Furthermore, according to this study, the suggested framework has the greatest accuracy when compared to other techniques, which are RNN and ANN, which have accuracy of (69.0%) and (49.5%), respectively. When compared to previous techniques, the proposed DL framework is more time efficient and has a quicker detection time (0.17 s) (Alrifaey, M. et. al 2021).

**Prediction of Bottom-Hole Pressure**

It is necessary to compute the bottom hole pressure in order to determine the pressure drop in wells (BHP). To estimate the BHP, the author used ANN with a hybrid genetic algorithm and particle swarm optimization (HGAPSO). Oil flow rate, gas flow rate, water flow rate, oil API, depth, tubing diameter, surface temperature, wellhead pressure, and bottom-hole temperature are the nine inputs that are employed (Ahmadi, M.A. 2019).

Results from testing this model on a population of 150 had a maximum inaccuracy of 10%. Although other approaches, including GA-ANN and PSO-ANN, have been tried, the hybrid approach HGAPSO-ANN has demonstrated the highest success rate. The article (Khamehchi, E.; Bemani 2021) demonstrates a different way to compute BHP. The author employed extreme learning machine (ELM) and gradient tree boosting (GTB) (ELM). Eight variables made up the dataset that was used: actual vertical depth, average deviation, average angle, wellhead pressure, oil flow rate, water flow rate, and gas flow rate. The combination of these two techniques produced a mean relative error that was less than 4%. A recent research also examines the use of ANN, KNN, and random forests to forecast the flowing bottom hole pressure (FBHP). An error rate of 2.5%, 3.6%, and 4% for ANN, random forest, and KNN, respectively, was obtained using these methods with the following nine inputs: flowing oil rate, flowing gas rate, flowing water rate, production tubing internal diameter, well perforation depth, oil gravity, surface temperature, well bottom hole temperature, and wellhead pressure (Sami, N.A.; Ibrahim, D.S. 2021)

**2.8. Summary of Related Work**

In conclusion, complicated non-linear situations cannot be handled by machine learning. However, deep learning is best used for complicated non-linear problems, thus as a first step, our system will utilize deep learning to estimate the rates of production for oil, water, and gas. After completing the initial stage, we intend to enhance the system to provide more deep-learning-supported solutions for forecasting critical factors in oil extraction, such as the oil/water ratio, well pressure, and flow rate. We intend to investigate the use of ANN because it is ideal for our issue type. Other approaches, like as RNN and CNN, are less feasible; CNN excels in image recognition and may be used with RNN to enhance RNN module results. However, additional experimentation with other ways will be required to improve the system's efficiency.

**References**

Abdullayeva, F.; Imamverdiyev, Y. Development of Oil Production Forecasting Method Based on Deep Learning. Optim. Inf. Comput. **2019**, 7, 826–839.

Ahmadi, M.A.; Chen, Z. Machine Learning Models to Predict Bottom Hole Pressure in Multi-Phase Flow in Vertical Oil Production Wells. Can. J. Chem. Eng. **2019**, 97, 2928–2940

al Ajmi, M.D.; Alarifi, S.A.; Mahsoon, A.H. Improving Multiphase Choke Performance Prediction and Well Production Test Validation Using Artificial Intelligence: A New Milestone. In Proceedings of the Society of Petroleum Engineers—SPE Digital Energy Conference and Exhibition 2015, The Woodlands, TX, USA, 3–5 March 2015.

Alrifaey, M.; Lim, W.H.; Ang, C.K. A Novel Deep Learning Framework Based RNN-SAE for Fault Detection of Electrical Gas Generator. IEEE Access **2021**, 9, 21433–21442.

Al-Shabandar, R.; Jaddoa, A.; Liatsis, P.; Hussain, A.J. A Deep Gated Recurrent Neural Network for Petroleum Production Forecasting. Mach. Learn. Appl. **2021**, 3, 100013.

Al-Wahaibi, T.; Mjalli, F.S. Prediction of Horizontal Oil-Water Flow Pressure Gradient Using Artificial Intelligence Techniques. Chem. Eng. Commun. **2014**, 201, 209–224

Amaechi, U.C.; Ikpeka, P.M.; Xianlin, M.; Ugwu, J.O. Application of Machine Learning Models in Predicting Initial Gas Production Rate from Tight Gas Reservoirs. Rud.-Geol.-Naft. Zb. **2019**, 34, 29–40.

Doan, T.T.; van Vo, M. Using Machine Learning Techniques for Enhancing Production Forecast in North Malay Basin. In Springer Series in Geomechanics and Geoengineering; Springer: Singapore, 2020; pp. 114–121

Ghorbani, H.; Wood, D.A.; Choubineh, A.; Tatar, A.; Abarghoyi, P.G.; Madani, M.; Mohamadian, N. Prediction of Oil Flow Rate through an Orifice Flow Meter: Artificial Intelligence Alternatives Compared. Petroleum **2020**, 6, 404–414.

Guo, Z.; Wang, H.; Kong, X.; Shen, L.; Jia, Y. Machine Learning-Based Production Prediction Model and Its Application in Duvernay Formation. Energies **2021**, 14, 5509.

Han, D.; Kwon, S. Application of Machine Learning Method of Data-driven Deep Learning Model to Predict Well Production Rate in the Shale Gas Reservoirs. Energies **2021**, 14, 3629

Jabbari, M.; Khushaba, R.; Nazarpour, K.; Han, S.; Zhong, X.; Shao, H.; Pan, S.; Wang, J.; Zhou, W. Prediction on Production of Oil Well with Attention-CNN-LSTM. J. Phys. Conf. Ser. **2021**, 2030, 012038

Khamehchi, E.; Bemani, A. Prediction of Pressure in Different Two-Phase Flow Conditions: Machine Learning Applications. Measurement **2021**, 173, 108665

Mirzaei-Paiaman, A.; Salavati, S. The Application of Artificial Neural Networks for the Prediction of Oil Production Flow Rate. Energy Sources **2012**, 34, 1834–1843

Negash, B.M.; Yaw, A.D. Artificial Neural Network Based Production Forecasting for a Hydrocarbon Reservoir under Water Injection. Pet. Explor. Dev. **2020**, 47, 383–392.

Nguyen-Le, V.; Shin, H. Artificial Neural Network Prediction Models for Montney Shale Gas Production Profile Based on Reservoir and Fracture Network Parameters. Energy **2022**, 244, 123150.

Orrù, P.F.; Zoccheddu, A.; Sassu, L.; Mattia, C.; Cozza, R.; Arena, S. Machine Learning Approach Using MLP and SVM Algorithms for the Fault Prediction of a Centrifugal Pump in the Oil and Gas Industry. Sustainability **2020**, 12, 4776.

Pal, M. On Application of Machine Learning Method for History Matching and Forecasting of Times Series Data from Hydrocarbon Recovery Process Using Water Flooding. Pet. Sci. Technol. **2021**, 39, 519–549.

Sami, N.A.; Ibrahim, D.S. Forecasting Multiphase Flowing Bottom-Hole Pressure of Vertical Oil Wells Using Three Machine Learning Techniques. Pet. Res. **2021**, 6, 417–422.

Wahid, M.F.; Tafreshi, R.; Khan, Z.; Retnanto, A. Prediction of Pressure Gradient for Oil-Water Flow: A Comprehensive Analysis on the Performance of Machine Learning Algorithms. J. Pet. Sci. Eng. **2021**, 208, 109265.

Xia, L.; Shun, X.; Jiewen, W.; Lan, M. Predicting Oil Production in Single Well Using Recurrent Neural Network. In Proceedings of the 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE, Fuzhou, China, 12–14 June 2020; pp. 423–430.