Prediction by reservoir porosity using micro-seismic attribute analysis by machine learning algorithms in an Iraqi Oil Field.

Mohammed Wahab Raheem[a], Ph.D abdulbasit Khadim Shuker[b]

[a] Department of Computer Science, Diyala University, Baquba, Diyala, Iraq.
[b]Department of Computer Science, Diyala University, Baquba, Diyala, Iraq.

Email [1:]mohammed_wahab2002@yahoo.com, Email [2:]dr.abdulbasit@sciences.uodiyala.edu.iq

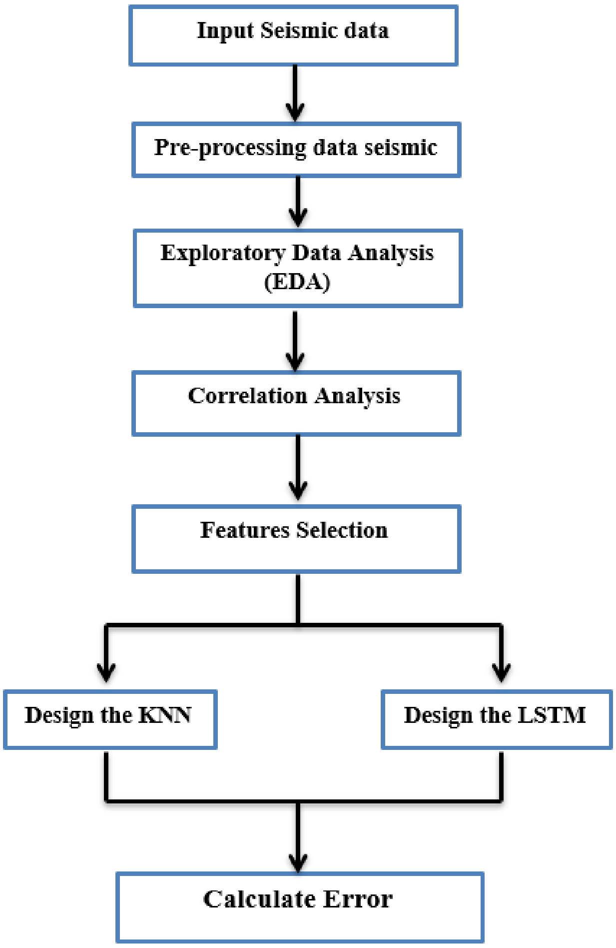
Article History: *Received: 10 June 2021; Revised: 12 July 2021; Accepted: 27 August 2021*

Abstract: The current work deals with an intelligent application that uses machine learning to analyze and attribute the resulting seismic data to improve and predict the locations of exploration, drilling, and production operations in petrophysical oil. Statistical analyzes of exploratory data analysis (EDA) were used to extract seismic features. This follows the application of two intelligent approaches of Recurrent Neural Networks (RNN) and K-Nearest Neighbors (KNN) to predict porosity. The parameters of seismic data with different seismic attributes were detected at the site by the seismic time-series method. The long-term memory (LSTM) algorithm is the most appropriate way to handle serial data. LSTM has a high capacity for data structure manipulation, which is applied for porosity prediction. Which depends a lot on choosing the best attributes? The two approaches evaluate by absolute error (MAE) and root means square error (RMSE). The results for both models used showed that using (LSTM) is more effective than using (KNN) in predicting porosity through seismic data, where the mean absolute error was obtained. (MAE) 0.017, while with KNN the mean absolute error (MAE) is 0.260 and the results showed that the model used can predict porosity very effectively.

Keywords: Long short-term memory (LSTM), K-Nearest Neighbors Regression, Recurrent Neural Networks (RNN), predict porosity, seismic data.

1. Introduction

The analysis of seismic and micro-seismic data in the oil industry has evolved into a massively data-intensive industry[1]. It is mainly involved from the knowledge of reservoir properties, which has attracted considerable interest due to its ability to monitor the development of unusual oil fields[2]. It is important to review some key ingredients of micro-seismic data processing, such as data acquisition, characteristics of micro-seismic events, events such as that concerning event location and relocation as well as the source of mechanism analysis. The uncertainties and issues of the current processing workflow must be included, which is critical to integrate the micro-seismicity into reservoir sensors [3]. Information storage takes a long time due to frequent backflow which is mostly due to insufficient measurement of error of signals flowing backward. Hochreiter briefly reviews the analysis of this problem, by addressing it by introducing a novel, efficient, gradient-based method called Long Short-Term Memory (LSTM) [1][3]. It can minimize the time of steps exceeding one thousand by enforcing constant error flow at a constant time unit. It is local in space and time; its computational complexity per time step and weight is negligible[4]. In this study, our experiments with local data include standardized, distributed, and real-valued representations. Our results are crystallized through the use and comparison of machine learning algorithms. Since regression algorithms are one of the main features of supervised learning algorithms as they model dependencies and relationships between target output and input features to predict the value of new data. Deep learning algorithms such as LSTM are an important example of Sequential Recurrent Neural Network (RNN) [5]. Because it's faster learning and performs many Successful operations. LSTM also solves complex, artificial and time-consuming tasks that were not solved by previous traditional neural network algorithms[6]. There are two types of methods of obtaining data for the oil location, directly and indirectly, through the well information record and indirectly through seismic surveys, and both methods must be reliable considering the values of porosity and permeability and the number of available fluids, which are considered the most important petrophysical properties of the reservoir[7][8]. The physical properties are the main identity of the reservoir, through which we can predict the porosity and thus measure the specifications of the reservoir or well [9]. When dealing with the characteristics of the oil reservoir, we must realize the importance of building any machine learning model that depends on the values of those properties, as they represent the inputs to this model, such as neural networks and other algorithms. The outputs are porosity prediction [10]. Some wells in southern Iraq do not have a (base record) or well record. These data include improving reservoir characterization and simulation, reducing drilling time and increasing drilling safety, optimizing production pump performance, and improving petrochemical asset management[11], so this research aims to predict porosity through data of seismic characteristics resulting from seismic surveys of reservoir sites through seismic teams deployed in wells locations[12]. Two models types of algorithms were used, the first is linear regression algorithms, where the characteristics of the seismic data were applied to a model that was designed with an algorithm such as the nearest neighbor regression for what is known to be a convergent, consistent, and somewhat accurate algorithm for the data [10][13]. The second type is the use of the Recurrent Neural Network model RNN, and one of its most famous algorithms that have been dealt with is the LSTM from the data [14]. In this study, an attempt was made to explore the capabilities of these techniques to predict porosity in the field of southern Iraq, and also included a comparison between the protection of these techniques and the selection the best as below flowchart (1).



Flowchart (1): approach used in the work

2.Methodology

1. Data collection and seismic data generation

The process of interpreting the interior of the Earth and the information it contains, especially the physical properties of rocks and layers of the Earth or sediments resulting from accumulations of fluids (reservoirs), requires a very special type of discovery, which is the generation of seismic waves utilizing special devices, and these waves are returned to the Earth to be received by acoustic sensors that measure [15]. These sensors reflect waves as well as energy and time. In addition, there are time frequencies of wave arrival, which generate a huge amount of data and a realistic representation of the wave as Fig (1,2) [16].

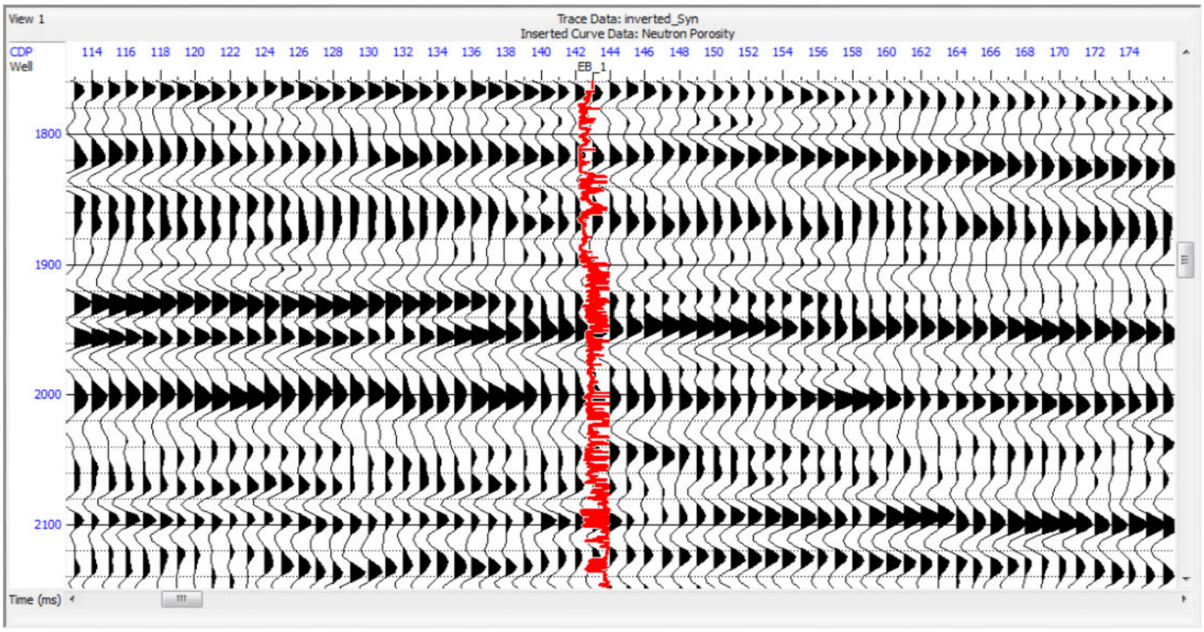


Fig (1): Seismic section

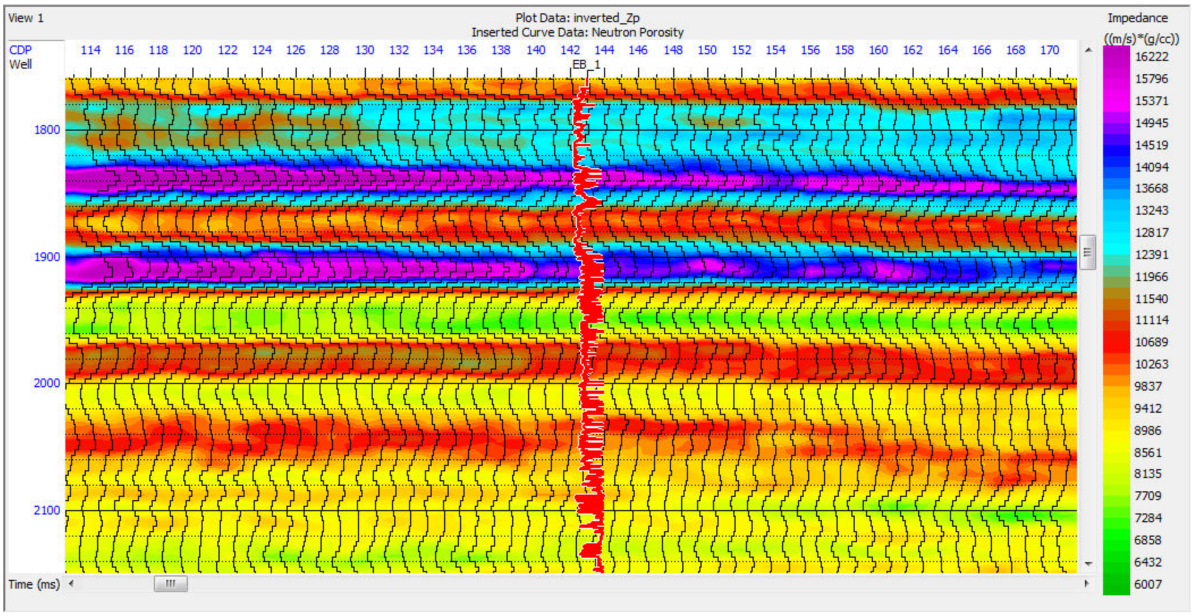


Fig (2): Acoustic impedance seismic attribute

2. Descriptive statistics and building Selected Data

Modern management of reservoir engineering relies heavily on reservoir characterization and plays a major role in making decisions, improving reliability, and predicting methods. Descriptive statistics are used to give a general description of characteristics or a data set, such as the mean of a variable, standard deviation, or frequency[17]. The average of the first, second or third quartile Also, inferential statistics can help us understand the collective characteristics of the elements of the data sample.In this work, seismic data containing features were presented. These features have varying correlation rates. We try to reduce the characteristics by removing the values that have a very high correlation coefficient.This work uses data analyses to build the relationship between seismic data and petrophysical attributes [18]. Many different types of seismic features can be elicited from seismic data and list the common types of attributes that can be generated from seismic data Table 1 [19].

Table 1: common types from seismic data attributes.

Amplitude weighted cosine phase AWP	Average frequency AF	Instantaneous phase IP
Amplitude envelop AE	Filters F	Instantaneous frequency IF
Amplitude weighted frequency AWF	Impedance I	Dominant frequency DF

3. Correlation Analysis for seismic data attributes.

The coefficients of the physical properties of the Earth's layers are recorded by seismic reflections when all sample data are used in creating the model that predicts porosity. This leads to accuracy in predicting physical or porosity properties.

4. Recurrent neural network (RNN)

Recurrent Neural Network (RNN). A recurrent Neural Network (RNN) is a type of neural network that is used to process sequential data. In this work, RNN was used because the data used in building the model is seismic data of different periods[20]. The RNN shares weight with the outputs as the hidden layers communicate. This feature makes RNN different from the rest of the data-related neural networks between them in a circular way, which leads to the output communication with all the hidden layers, and this helps to reduce the number of parameters. The study uses a deep recurrent neural network to find the relationship between the properties of seismic (petrophysical) data to predict porosity as input the network with data as input to the output to predict the porosity expected from these features. This training is based on moderated training data. (RNN) is a neural network that simulates a discrete-time dynamical system that has an input x_t, an output $y_{t,}$ and a hidden state h_t. In our notation the subscript t .

$$H_t = f h(x_t, h_{t-1})$$

$$Y_t = f_0(h_t)$$

where f_h and f_o are a state transition function and an output function, respectively. Each function is parameterized by a set of parameters; θ_h and θ_o

represent time [21]. And we can show the Structural-RNN architectures by fig (3)

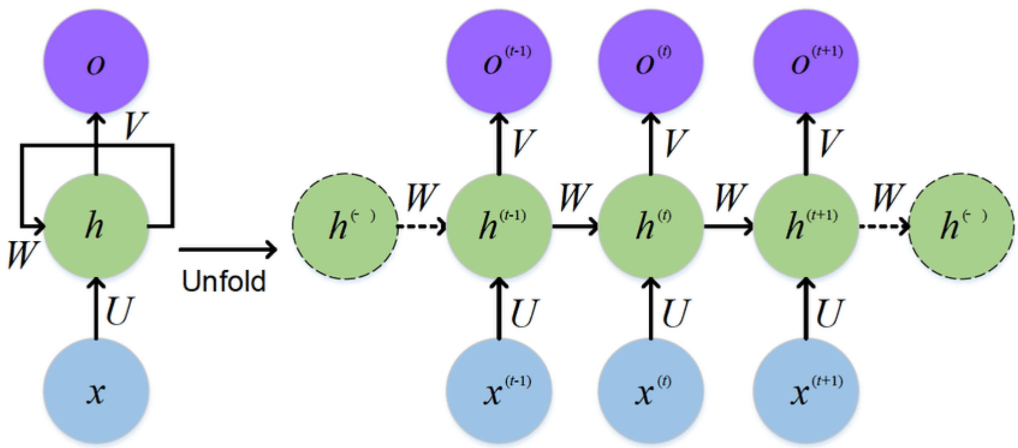


Fig (3): Structural-RNN architectures

5. Long short-term memory (LSTM)

The Long and Short Term Memory (LSTM) network is one of the RNN optimization tasks. It can effectively solve the problems of data using cascading sequences, N-1, N, N+1, and scaling, and makes the network have stronger memory capacity. In addition, LSTM can also recall longer data, which has an internal "LSTM cell" (self-rotation) [20]. Before that, LSTM simply enforces element-by-element nonlinearity It is similar to the common recycling network, and each module not only has the same input and output structure but also has a gate control system with more parameters and information flow control [22]. The structure of the LSTM hidden layer is shown in $C_{(t-1)}$ is the node state of the previous sequence of hidden layers, $h_{(t-1)}$ is the output of the previous sequence of nodes of the hidden layer, $x_{(t)}$ is the input of the node of the hidden layer node of the current sequence, C_t is the node state of the hidden layer of the current sequence, $h_{(t)}$ is the output hidden node The class for the current sequenceσ is non-linear activation function sigmoid, and tanh is the hyperbolic tangent function. Compared with RNN, LSTM is better at learning long-term dependence between sequence data, while LSTM has a complex structure, multiple parameters, and slow convergence speed fig (4).

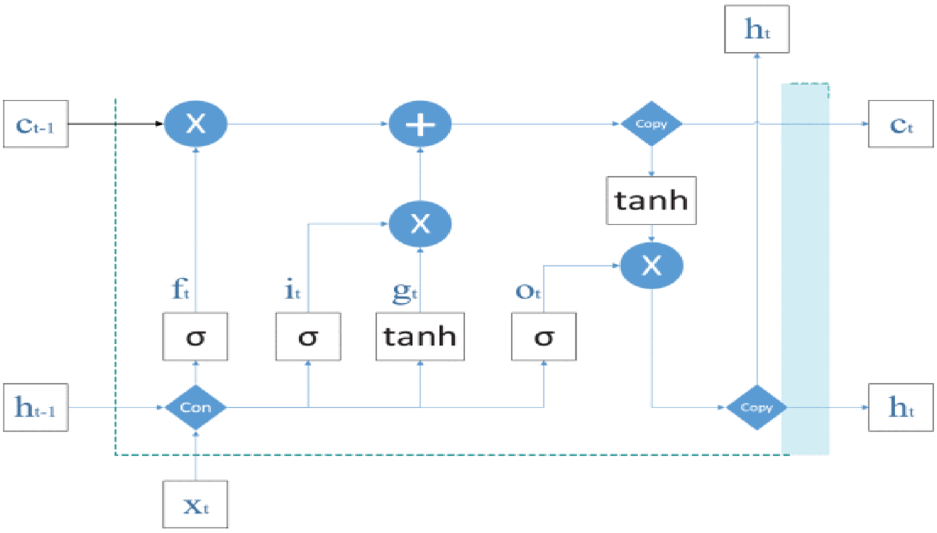


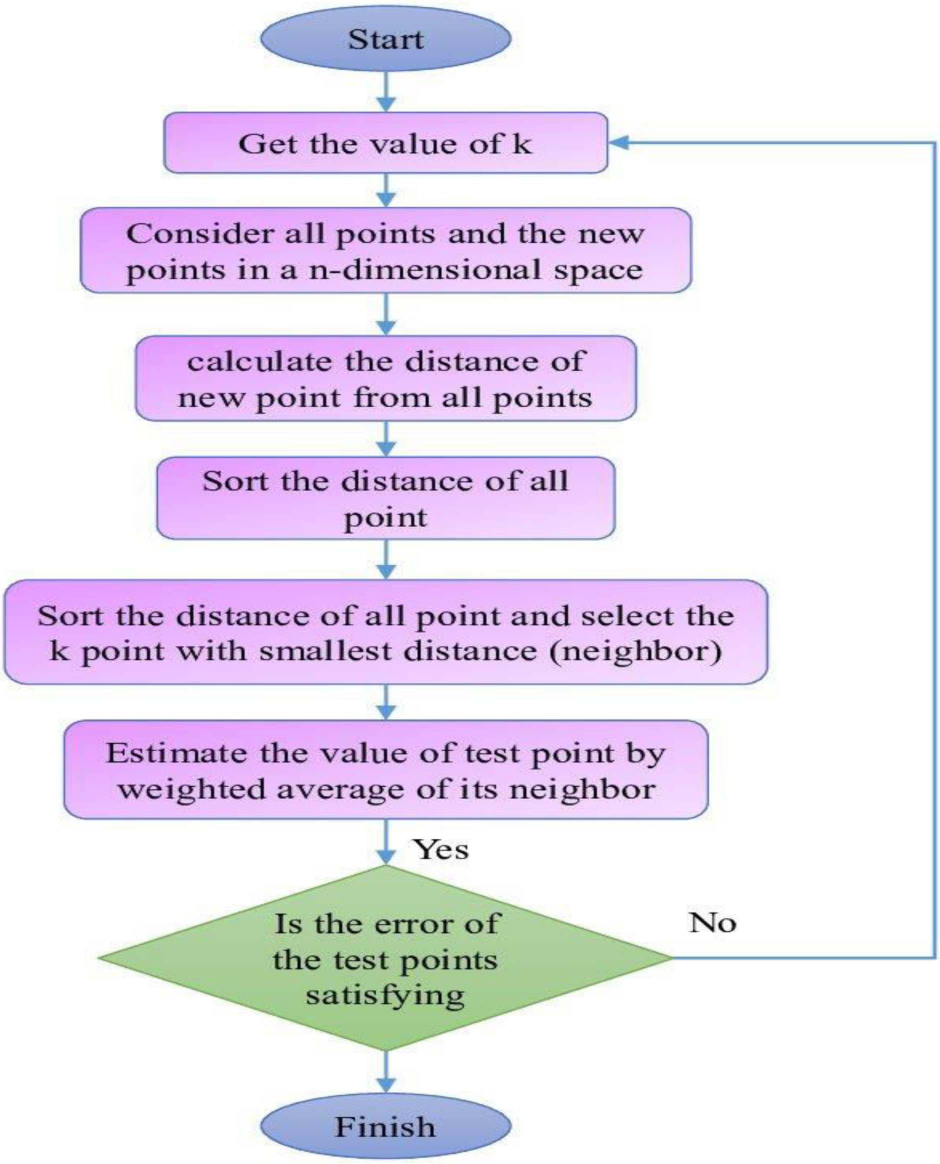
Fig (4): Long short-term memory (LSTM) architectures

6. Approach K-Nearest Neighbors Regression(KNN)

This algorithm works on the principle that points with similar inputs have the same output, where the points are classified based on their similar properties in the n-dimensional chromosome. n is the number of input parameters[23]. KNN is a statistical algorithm used to estimate an unknown or missing value based on its nearest neighbor. The principle that KNN works on is (similarity of features) [24]. That is, in the sense of predicting new data values, this means that a value is assigned to the new point based on how similar it is to other points in the several exercises. Usually, points that are the closest distance are determined as the closest neighbors to the point. The unknowns of their surroundings. The techniques used to measure the distance between the nearest neighbor are numerous and often range from the most complex to the simplest, but the simplest equation can be taken as follows:

$$h(\mathrm{x},\mathrm{y}) = \|x - y\| \sqrt{\sum_{i=1}^{n} (Xi, Yi)^2}$$

Where a new point is (x) and an existing point is (y) and It is used for categorical parameters. If the value (x) and the value (y) are the same, the distance h will be equal to 0. Otherwise h=1[23][24].



Flowchart (2): flowchart for the k-nearest neighbor regression

7. Model Evaluation Criteria

To properly evaluate the performance of the model, it is necessary to choose good evaluation criteria for the model. There is no single efficiency criterion that can provide a complete description of the model's performance, so we chose to apply two criteria for this model:

- Mean Absolute Error (MAE) measures how far predicted values are away from observed values. We can calculate MAE through the following equation

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|x_i - y_i|$$

- Root Mean Square Error (RMSE) is a measure of how concentrated the data is around the line of best fitWe can calculate MAE through the following equation.

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(x_{\text{predicted}} - y_{\text{observed}})^2}{n}}$$

In general, RMSE is used to accurately and very effectively reflect the accuracy of prediction results, and the reason is due to its sensitivity in that the closer it is to 0, the more accurate the prediction results.

In this work, both criteria were used on two different models of algorithms, but on the same data set.

8. Results and discussion

In this work, we presented a model for two approaches based on time-series data describing a type of seismic data for petrophysical properties. By using correlation coefficients, features are narrowed down, and features with high correlation are excluded in between, so that we can choose the typical features to train the model on to get high and perfect prediction accuracy, the data is divided into two groups and randomly, one group to train with 80% of the total data set and another set to test 20 % of the total data as shown in Figure (7). Indeed, the results of the STM algorithm were very good, as it achieved a lower error rate than what was obtained by the KNN algorithm, as shown in table (2). This indicates that the proposed model can give a good prediction of time-series data and is ideal for oil exploration.

Table 2—MAE, RMSE, for two model

Criteria	KNN	LSTM
MAE	**0.260**	**0.086**
RMSE	**0.295**	**0.092**

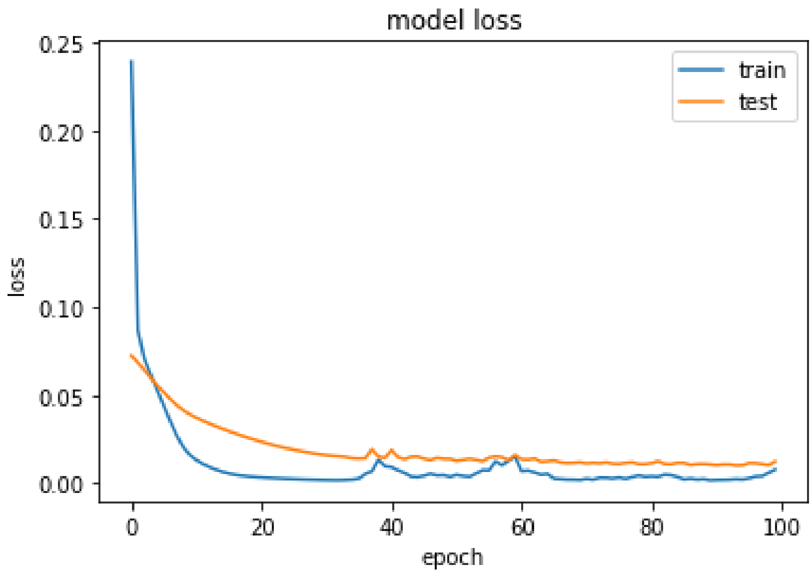


Figure (5): Performance validation plot for training and testing (LSTM)

Conclusions

The main application area of this approach is reservoir characterization. This work introduced the RNN approach to predicting porosity. We tried another method that uses KNN regression algorithms to see the appropriateness of characterizing reservoir properties using seismic data. The results showed that the prediction error rate was in line with the training and testing of the LSTM model, and it is much better than the results achieved using the KNN regression model.RNN and through one of its algorithms, LSTM, the latter providing better results in predicting porosity. This good accuracy of porosity prediction helped to address the uncertainty and uncertainty of drilling, especially in the absence of well recording data other than seismic data, enhances the confidence of oil industry professionals.

- The LSTM approach and the KNN were tested on a small amount of data. These models can use a large amount of data in specific reservoir characterization to improve the ideal environment in the reservoir characterization.

- Machine learning has tackled the major challenges that complexity professionals have faced in dealing with large amounts of data and overcome mistrust in exploration and production.
- Statistical analysis also contributed to describing and analyzing the data and giving a logical picture of the type and size of the data and whether it was linear or non-linear. It also gives us the correlation coefficient between them to choose the characteristics that can be involved in training.

All of these improvements that the approach provides help boost the confidence of oil exploration, exploration, and production professionals. Also, the method used for forecasting gives us high accuracy, so we suggest using it with all seismic data and it is computationally inexpensive.

References

[1] S. Hochreiter and J. Schmidhuber, "Bridging long time lags by weight guessing and 'Long Short-Term Memory,'" *Spat. Model. Biol. Artif. Syst.*, no. January 1996, pp. 65–72, 1996, [Online]. Available: http://www.bioinf.jku.at/publications/older/3104.pdf.

[2] X. Q. Pang, C. Z. Jia, and W. Y. Wang, "Petroleum geology features and research developments of hydrocarbon accumulation in deep petroliferous basins," *Pet. Sci.*, vol. 12, no. 1, pp. 1–53, 2015, doi: 10.1007/s12182-015-0014-0.

[3] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014.

[4] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent Advances in Recurrent Neural Networks," pp. 1–21, 2017, [Online]. Available: http://arxiv.org/abs/1801.01078.

[5] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[6] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, 2019.

[7] J. R. Fanchi and R. L. Christiansen, *Introduction to petroleum engineering*. 2016.

[8] G. Zhang, Z. Wang, H. Li, Y. Sun, Q. Zhang, and W. Chen, "Permeability prediction of isolated channel sands using machine learning," *J. Appl. Geophys.*, vol. 159, pp. 605–615, 2018.

[9] M. B. Valentín *et al.*, "Estimation of permeability and effective porosity logs using deep autoencoders in borehole image logs from the brazilian pre-salt carbonate," *J. Pet. Sci. Eng.*, vol. 170, pp. 315–330, 2018, doi: https://doi.org/10.1016/j.petrol.2018.06.038.

[10] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, "Application of machine learning and artificial intelligence in oil and gas industry," *Pet. Res.*, 2021, doi: https://doi.org/10.1016/j.ptlrs.2021.05.009.

[11] M. Mohammadpoor and F. Torabi, "A new soft computing-based approach to predict oil production rate for vapour extraction (VAPEX) process in heavy oil reservoirs," *Can. J. Chem. Eng.*, vol. 96, no. 6, pp. 1273–1283, 2018, doi: 10.1002/cjce.23111.

[12] M. Mohammadpoor and F. Torabi, "Big Data analytics in oil and gas industry: An emerging trend," *Petroleum*, vol. 6, no. 4, pp. 321–328, 2020, doi: 10.1016/j.petlm.2018.11.001.

[13] F. I. Syed, A. AlShamsi, A. K. Dahaghi, and S. Neghabhan, "Application of ML & AI to model petrophysical and geo-mechanical properties of shale reservoirs – A systematic literature review," *Petroleum*, 2020, doi: 10.1016/j.petlm.2020.12.001.

[14] R. DiPietro and G. D. Hager, "Chapter 21 - Deep learning: RNNs and LSTM," in *The Elsevier and*

MICCAI Society Book Series, S. K. Zhou, D. Rueckert, and G. B. T.-H. of M. I. C. and C. A. I. Fichtinger, Eds. Academic Press, 2020, pp. 503–519.

[15] K. Schwarzer, "Chapter 7 - Geophysical prospection and sedimentological characteristics of subaquatic tsunami deposits," M. Engel, J. Pilarczyk, S. M. May, D. Brill, and E. B. T.-G. R. of T. and O. E. W. Garrett, Eds. Elsevier, 2020, pp. 115–142.

[16] W. Xiao, X. Yi, F. Pan, R. Li, and T. Xia, "Chapter 2 - Acoustic, Electromagnetic and Optical Sensing and Monitoring Methods," S. Pamukcu and L. B. T.-U. S. Cheng, Eds. Academic Press, 2018, pp. 43–139.

[17] Y. Zee Ma, "Uncertainty analysis in reservoir characterization and management: How much should we know about what we don't know?," *AAPG Mem.*, no. 96, pp. 1–15, 2011, doi: 10.1306/13301404M963458.

[18] P. J. R. Fitch, M. A. Lovell, S. J. Davies, T. Pritchard, and P. K. Harvey, "An integrated and quantitative approach to petrophysical heterogeneity," *Mar. Pet. Geol.*, vol. 63, pp. 82–96, 2015, doi: 10.1016/j.marpetgeo.2015.02.014.

[19] A. Contreras, C. Torres-Verdín, W. Chesters, K. Kvien, and T. Fasnacht, "Extrapolation of flow units away from wells with 3D pre-stack seismic amplitude data: Field example," *Petrophysics-The SPWLA J. Form. Eval. Reserv. Descr.*, vol. 47, no. 03, 2006.

[20] C. Tian and R. N. Horne, "Recurrent neural networks for permanent downhole gauge data analysis," *Proc. - SPE Annu. Tech. Conf. Exhib.*, vol. 0, no. October, 2017, doi: 10.2118/187181-ms.

[21] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.*, pp. 1–13, 2014.

[22] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *J. Mach. Learn. Res.*, vol. 3, no. 1, pp. 115–143, 2003, doi: 10.1162/153244303768966139.

[23] M. R. Mahdiani, E. Khamehchi, S. Hajirezaie, and A. Hemmati-Sarapardeh, "Modeling viscosity of crude oil using k-nearest neighbor algorithm," *Adv. Geo-Energy Res.*, vol. 4, no. 4, pp. 435–447, 2020, doi: 10.46690/ager.2020.04.08.

[24] J. S. Richman, "Chapter Thirteen - Multivariate Neighborhood Sample Entropy: A Method for Data Reduction and Prediction of Complex Data," in *Computer Methods, Part C*, vol. 487, M. L. Johnson and L. B. T.-M. in E. Brand, Eds. Academic Press, 2011, pp. 397–408.