Journal of Petroleum Science and Engineering 203 (2021) 108602

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: http://www.elsevier.com/locate/petrol

Machine learning - A novel approach of well logs similarity based on synchronization measures to predict shear sonic logs

Muhammad Ali [a], Ren Jiang [b], Ma Huolin [a,*], Heping Pan [a], Khizar Abbas [c], Umar Ashraf [d], Jar Ullah [a]

[a] Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, PR China
[b] Research Institute of Petroleum Exploration and Development, Petro-China Company Limited, Beijing, PR China
[c] School of Economics and Management, China University of Geosciences, Wuhan, PR China
[d] Institute for Ecological Research and Pollution Control of Plateau Lakes, School of Ecology and Environmental Science, Yunnan University, Kunming 650500, China

ARTICLE INFO

Keywords:
Machine learning
Geophysical similarity analysis
Multidimensional scaling
Jaccard and Overlap similarities
Well logs

ABSTRACT

This study proposes a novel approach to predict missing shear sonic log responses more precisely and accurately using similarity patterns of various wells with similar geophysical properties, which is important in decision making and planning of hydrocarbon exploration. Deep Neural Network (DNN) along with the similarity metrics such as Jaccard and Overlap similarities are employed to examine the relationship between the wells. Further, dimensionality reduction techniques including Multi-Dimensional Scaling (MDS) and well-ranking process are applied to extract common geophysical responses of the wells. A higher response indicates the existence of a strong similarity. This can also be verified by the superimposed of well log data. The potential benefits of our novel method are following; (a) it does not follow the zone-by-zone prediction of the missing logs such as rock physics methods, (b) it outputs the uncertainties facilitated that is by the least-squares method. Having the potential of demonstrating shear sonic log prediction in hydrocarbon-bearing zones, which cannot be precisely predicted by the Greenberg-Castagna method that only works in brine-saturated rocks, this approach will provide improved accuracy, where shear sonic logs are missing and need to be predicted for geomechanics, rock physics, and other applications.

1. Introduction

Contemporary computational techniques have proven quite beneficial in estimating unknown parameters such as facies (Ashraf et al., 2019; Bestagini et al., 2017), faults and fractures (Ashraf et al., 2020a; Wu et al., 2019), and petrophysical logs using rock physics models (Ali et al., 2020; Li et al., 2020; Ali et al., 2019). Amongst them, deep machine learning has attained substantial popularity for its capability to accurately estimate the shear wave velocity using the petrophysical logs (Anemangely et al., 2019; Weijun et al., 2017). It is of supreme importance to precisely measure the shear wave velocity because it plays an essential role in predicting the geology of an area through various ways (Chen et al., 2018; Tong et al., 2013; Tonni and Simonini, 2013). It is prevalent in the oil and gas industry during the logs recording procedure that different kinds of logs could be missing in some wells, this could be due to many reasons such as tool failure during drillings and wellbore instability (Geng and Wang, 2020). It could be costly to run logging tools in those wells to get these missing log responses again, such as density, shear sonic, and sonic logs. In such cases, the correlate similarity approach plays the most crucial role in predicting missing logs accurately because it follows the patterns of the reference log from the surrounding well logs to account for stratigraphic variations to predict the log of interest and geology of an area.

Different empirical methods have been employed to estimate the missing well logs: Gardner et al. (1974) provided a reasonable empirical data drive equation related to density and sonic for brine-saturated rock types. Smith (2007) and Faust (1953) methods provided an interval empirical base about the relations between sonic and resistivity logs. Greenberg and Castagna (1992) and Castagna et al. (1985) proposed empirical relationships for different minerals to estimate the sonic shear curve from the compressional sonic curve.

Likewise, the idea of Machine Learning (ML) to deal with these problems by predicting missing well log data was also employed by some researchers. Verma et al. (2014) proposed a Visibility Graph

* Corresponding author.
E-mail addresses: muhammad_ali@cug.edu.cn (M. Ali), mhl70@163.com (M. Huolin), umarash2010@hotmail.com (U. Ashraf).

https://doi.org/10.1016/j.petrol.2021.108602
Received 29 October 2020; Received in revised form 3 January 2021; Accepted 25 February 2021
Available online 9 March 2021

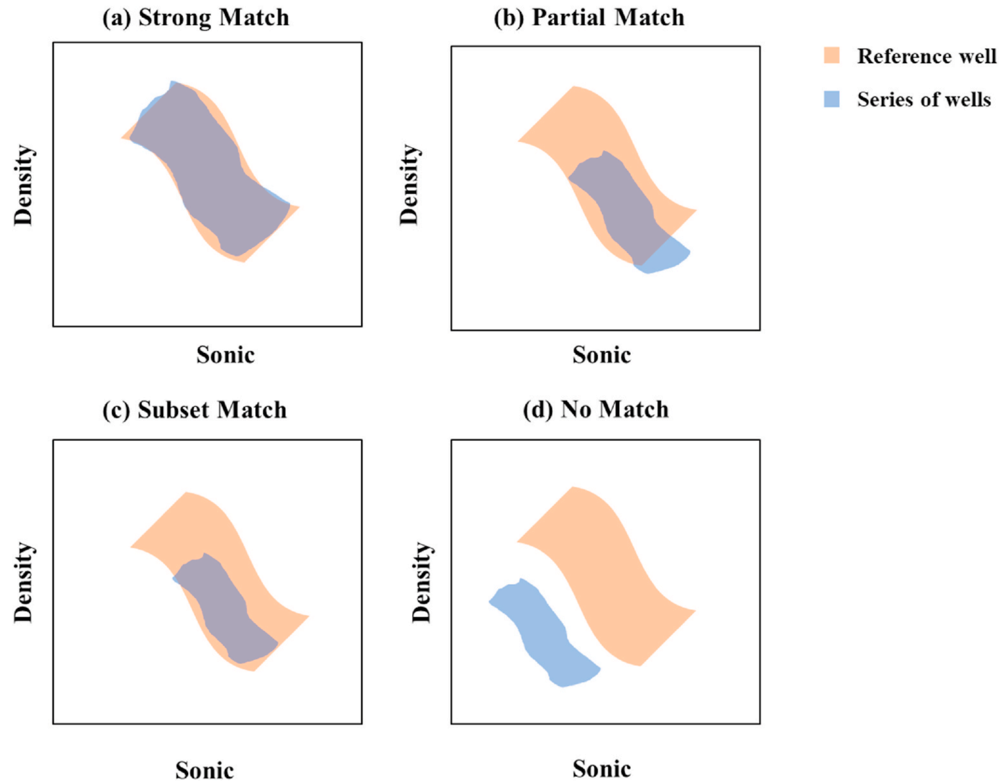


Fig. 1. Example of the similarity among the pairs of the wells. The reference well footmark displayed in orange has verified the footmark against the series of wells in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Similarity (VGS) and Synchronization Likelihood (SL) methods taking porosity and gamma-ray (GR) log responses as a reference to measure the correlation similarity among the diverse wells. Akkurt et al. (2018) designed an unsupervised outlier detection algorithm that could identify the outliers in density and sonic logs and determine the footmark of a well from an arbitrary number of logs. Similarity metrics can be used to compare wells geometry based on their geophysical footmark, and rebuilt density and sonic logs with uncertainty estimates. Bader et al. (2018) proposed a technique to predict missing logs by correlating a similar petrophysical log response from surrounding wells. Freire et al. (2002) established a semiautomatic algorithm for the correlation of statistical stratigraphic of well logs. It comprises of a combined estimation of the correlation between shrink or stretch and shift, located in the selected window, and cannot compare a set of logs at the same time. Al-Anazi and Gates (2010) predicted permeability distributions and classified electrofacies in highly heterogeneous sandstone reservoirs using the nonlinear Support Vector Machine (SVM) technique. There are plenty of other rock physics-based models according to the unconventional and conventional settings of the reservoir to produce a suitable prediction of logs, nevertheless, they are all interval to interval-based, dependent on the lithological unit, and require human time and expertise in their calibration.

This paper introduces a novel approach to help predict missing shear sonic log by using Machine Learning (ML) based similarity algorithms and Deep Neural Network (DNN). The novel approach uses similarity patterns of various wells with similar geophysical properties to predict missing log responses precisely and accurately. This approach analyzes the prediction of missing logs and will produce shear sonic logs prediction accurately from essential logs. We provide an example from the benchmark dataset where a complete shear sonic log is predicted and compared with both the original shear sonic log and the outcome of a conventional method for estimating missing shear sonic logs.

This study uses data of five wells of Lower Goru Formation in the Middle Indus Basin, Pakistan: an area composed of a continuous formation of thin shale and sand layer intercalations (Ashraf et al., 2019). The identification of the mineralogy is quite challenging due to the heterogeneous nature of the geological formations in the Lower Goru Formation (Ashraf et al., 2020a; Ehsan and Gu, 2020). Several authors have addressed the issue of reservoir heterogeneity by utilizing modified approaches within the Lower Goru Formation (Ashraf et al., 2020b; Ehsan et al, 2018, 2019). All logging data obeys the depositional environment similarity which is acquired from seismic and well logs data. The recording of well logs data at equivalent intervals of depth can also be deemed as time samples, because of minimal heterogeneities in the relationship of age-depth.

2. Methods

2.1. Well similarity analysis

The concept of characterization of petrophysical response for each well by their similarity to match the pairs of wells recognizes that which wells have similar log responses. This kind of procedure is called wells similarity analysis. It is completely automatic and fundamental for data Quality Control (QC) and selecting wells to be included in the Machine Learning training dataset. We describe an ML automatic procedure for analyzing wells similarity of different wells where the visual examination is not effective and possible for us. We adopt that idea from a video of Fred Jason, a product manager, and senior petrophysicist at CCG company (Jason, 2019), and Rıdvan Akkurt, a petrophysics advisor in the innovation team at Schlumberger (Akkurt et al., 2018). We utilized the idea of similarity metrics e.g., Jaccard similarity and Overlap similarity to measure the correlation among well footmarks, reduction techniques including a Multi-Dimensional Scaling (MDS), and novel well-ranking process to extract information based on wells having common petrophysical responses.

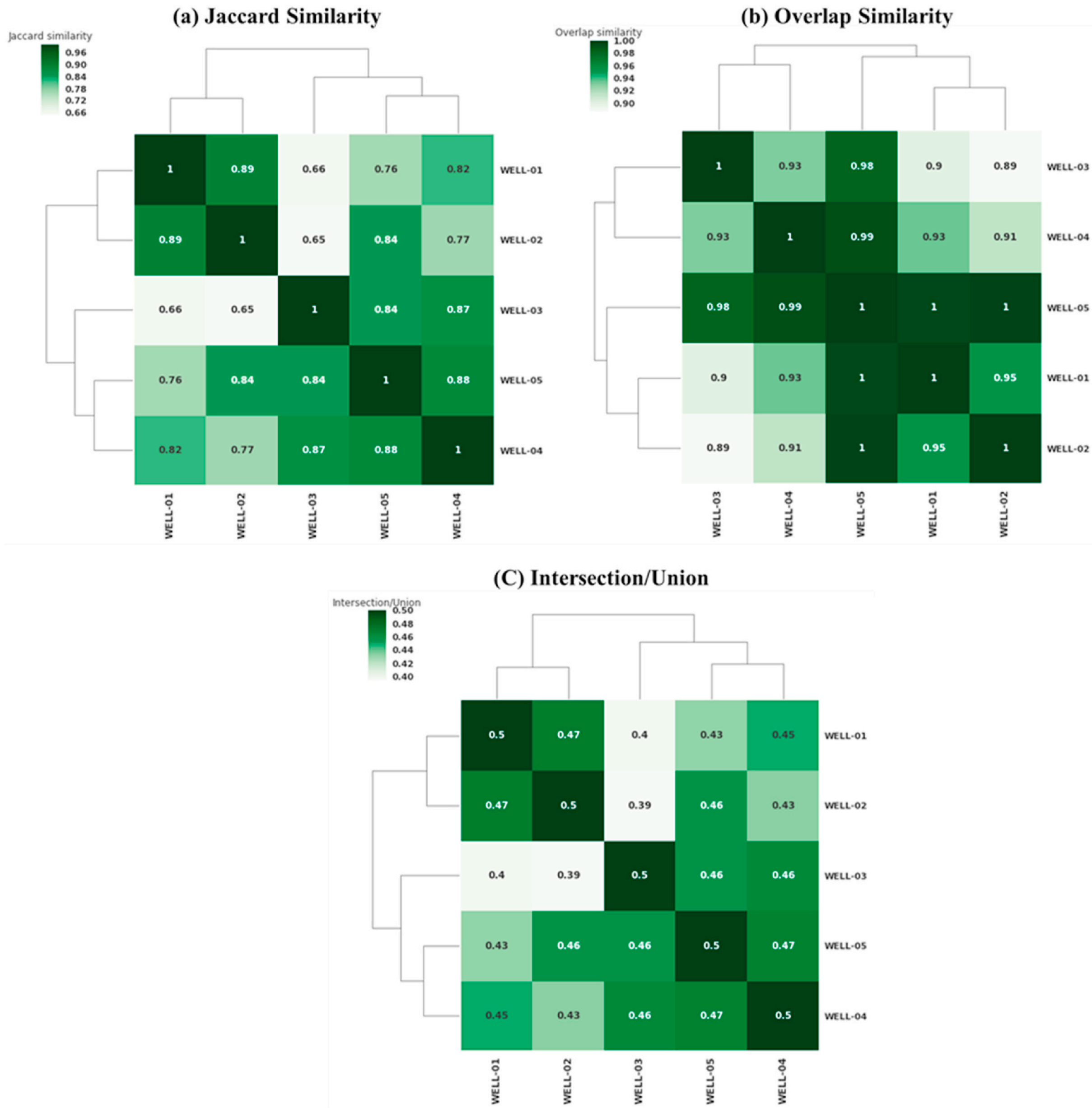


Fig. 2. Matrix plots of similarity metrics for the set of data. (a) Jaccard similarity, (b) Overlap similarity, and (c) intersection/union.

2.2. Similarity metrics

Fig. 1 demonstrates a sequence of schematic patterns to exemplify the broad spectrum of scenarios that can arise between two wells by comparing their footmarks as shown in Fig. 1 (a)–(d). The similarity of a reference well (shaded color background in orange) is compared against the similarity of other wells of the dataset (shaded color background in blue) that reveals gradual mismatch against the reference well. The similarity metrics described in this section are just two from a toolkit of several that we use to quantify the type and strength of the footmark relationships in cases like those illustrated.

2.3. Jaccard similarity

Jaccard similarity proposed by Jaccard (1912) is utilized to compare the footmark geometry from a pair of wells i and j. The size of the intersection to the size of the union of the footmarks is the Jaccard similarity ratio. To avoid computing these quantities by numerical integration, we utilized the combined set of data from the well i and j based on the computation as follows:

$$Jaccard_{(i,j)} = \left(\frac{N_{ij}}{N_i - N_j + N_{ij}}\right) \tag{1}$$

whereas i presents even numbers of wells, j refers to odd numbers of wells, N_i denotes the number of samples fall inside the well i footmark, N_j are the samples fall inside the well j footmark, and N_{ij} infers the numbers samples fall inside both wells' footmarks.

The interpretation of Jaccard similarity is uncomplicated as the probability from the combined set of data is within the footmark of well i and j, hence, it is a valuable measurement of similarity between two footmarks, and it ranges as follows:

$$0 \leq Jaccard_{(i,j)} \leq 1 \tag{2}$$

According to the above equation, if Jaccard similarity is close to one, it means the footmarks of the wells i and j strongly match with each other and both wells are similar. However, if the footmarks do not match, it means the Jaccard similarity is equal to zero. Jaccard similarity

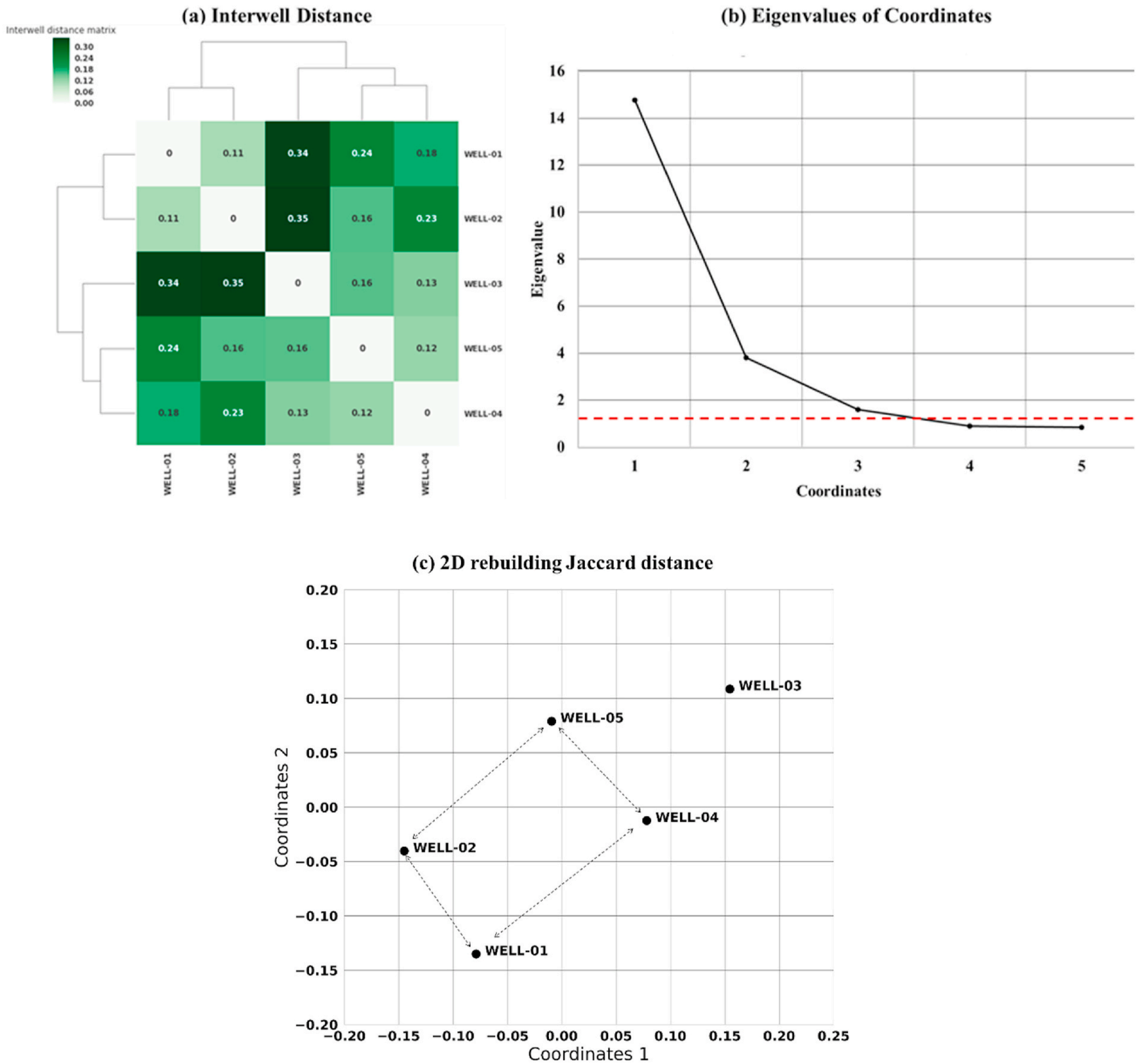


Fig. 3. MDS map of the Jaccard similarity matrix of the well dataset. (a) Jaccard distance, (b) eigenvalues of the MDS coordinates, and (c) 2D rebuilding of the Jaccard distances.

does not depend on the order of the wells because it is symmetric:

$$Jaccard_{(i,j)} = Jaccard_{(j,i)} \tag{3}$$

2.4. Overlap similarity

Jaccard similarity in some cases is not sensitive, where the footmark of the well is a subgroup of the reference well. In this case, Jaccard similarity will decrease, even though the footmark of the subgroup well is completely falling inside the footmark of the reference well. It is important to classify these conditions because the subgroup well is a strong factor to construct a predictive model to predict missing logs in the reference well. Such observation can be identified by Overlap similarity (Vijaymeena and Kavitha, 2016) and can be calculated by employing the following equation:

$$\text{Overlap}_{(i,j)} = \frac{N_{ij}}{\min(N_i, N_j)} \tag{4}$$

Both similarity indices do not rely on the order of the well such as $\text{Overlap}_{(i,j)} = \text{Overlap}_{(i,j)}$, and has a range between 0 and 1. On the other hand, Jaccard and Overlap similarities are always equivalent to 1 when the footmark of the reference well is completely a subgroup of footmark well or vice versa. Overlap similarity is unable to show which well is the subgroup, *i* or *j* but it is unimportant to acquire this information as the subgroup well is the one that is set to a minimum denominator of the function of overlap.

The result of similarities is calculated from both similarity index for every pair of wells in the set of data in Fig. 2, which shows similarity in matrix form where every pixel's similarities are colored with strength among the pair of wells marked on the column and row organized through unsupervised hierarchical clustering. Note that each matrix is symmetric and has ones on the leading diagonal where a well is compared to itself. Well-03 is distinctive in its mismatch with well-01, well-02, well-4, and well-05.

2.5. Jaccard distance and multidimensional scaling

This method is used to visualize the matrix similarity or dissimilarity structure displayed in the above portion. It utilizes the Jaccard distance (inter-well distance) matrix that is complementary to the Jaccard similarity (Jaccard index) and can be computed as:

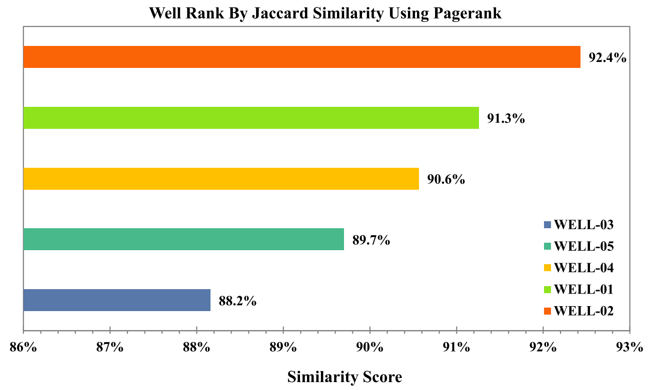


Fig. 4. Well similarity scores derived by the implementation of the PageRank algorithm to the well set of data.

$$d_{(i,j)} = 1 - \left(\frac{N_{ij}}{N_i - N_j + N_{ij}}\right) \tag{5}$$

The (MDS) method places every well in its low dimensional space so that Jaccard distance "d" could be extracted as a result of plausible approximation (Akkurt et al., 2018). The implementation of (MDS) to the set of data is displayed in Fig. 3. We transformed the Jaccard similarity matrix (Fig. 2a) into a Jaccard distance as displayed in Fig. 3 (a). Eigen-decomposition of the Jaccard distance matrix produces a set of coordinates for each well in a reduced number of dimensions. The eigenvalues are shown in Fig. 3 (b) is a diagnostic indicating that the Jaccard distance can be reduced to just three important coordinates that have higher than 1 eigenvalue.

In Fig. 3 (c) the two essential coordinates of MDS are utilized to generate a 2D view of the wells, where the plotted Jaccard distances are an approximate rebuilding of the Jaccard distance matrix in Fig. 3 (a). Because Jaccard distance is the complement to the footmark similarity. In Fig. 3 (c), the nearer wells have more similar footmarks than the wells away from other wells. The main wells of the dataset, well-01, well-02, well-04, and well-05, plot furthest apart at the vertices of the square indicating that these four wells have the most similar footmarks, whereas well-03 have the least similar footmark.

2.6. Well similarity ranking

PageRank (Brin and Page, 1998) is a famous link-based ranking algorithm. The primary concept of the PageRank algorithm is that the significance score of a document equals the sum of those propagated from its in-link neighbors. We employed the PageRank algorithm (Brin and Page, 1998) to arrange the wells according to their significance in the similarity matrix. It measures the strength of footmarks similarity and compares it with all other wells. It ranks the wells based on the strength of footmarks similarity showing the petrophysical responses of the whole set of data.

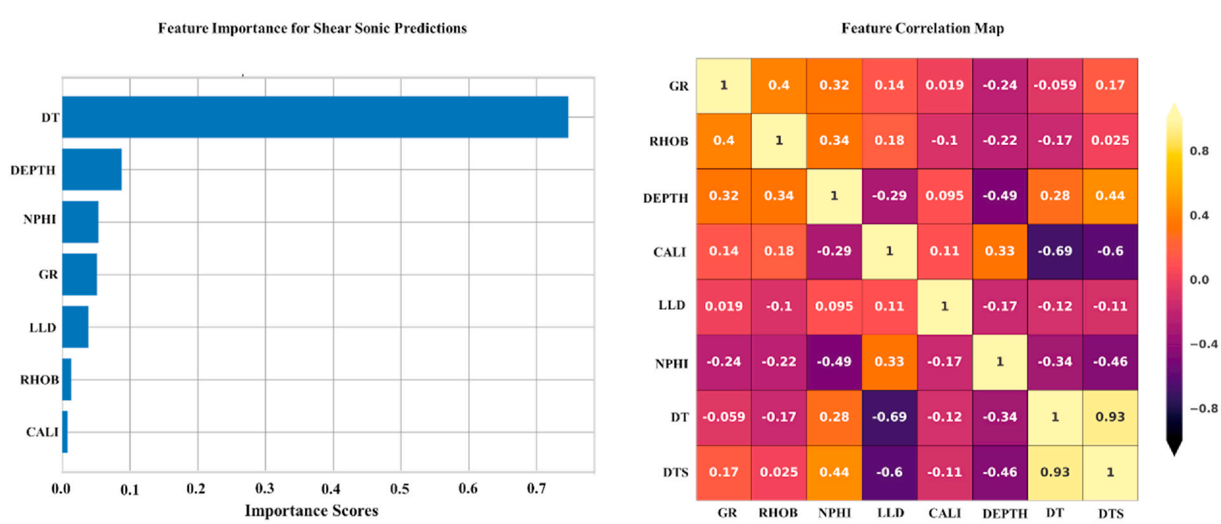
$$PR_{(P_i)} = \frac{1-d}{N} + d \sum_{P_j \in M_{(P_i)}} \frac{PR_{P_j}}{L_{(P_j)}} \tag{6}$$

where $P_1, P_2 \ldots\ldots\ldots P_N$ are the Jaccard similarities of wells in the dataset, N is the total number of wells, $M_{(P_i)}$ is the set of wells that link to P_i , P_j is the rank of well 'j' and $L_{(P_j)}$ is a number of outgoing edges of well j, and d is a damping factor (usually, the favorable value for "d" is 0.85).

Fig. 4 illustrates the similarity score of the wells derived from the Jaccard similarity matrix displayed in Fig. 2 (a). The similarity score is shown in descending order of significance and is consistent with the components they were constructed. The most important wells are well-01 to well-05 which contains high similarity scores. The next well, well-03, of the dataset has the lowest scores but it is acceptable.

In practice, we utilized both techniques, multidimensional scaling and well scoring (ranking), to generate understandable summaries about the well that is the most demonstrative of the petrophysical responses of a dataset and the wells having a small part of those responses. The wells that have the highest similarity scores in a project make good factors for existing well logs in order to generate new (synthetic) well logs for wells that are missing the shear sonic log.

A critical part of our missing log prediction approach is to optimize the selection of the wells providing the training data. The motivation for our approach is that any predictive model tends to perform poorly in situations where the values of the predictors are outside the range seen in the training data. The petrophysical footmark described in the previous section helps to avoid these situations by identifying wells that have similar petrophysical footmarks as the target well, with the result that the predictive model training is specialized for and adapted to the predictive features seen in the target well. The Deep Neural Network is the most commonly used technique to generate models for missing logs prediction that extracts the required information and learns from



(a) (b)

Fig. 5. (a) Weight feature importance scores and (b) heatmap of correlation features for the shear sonic log.

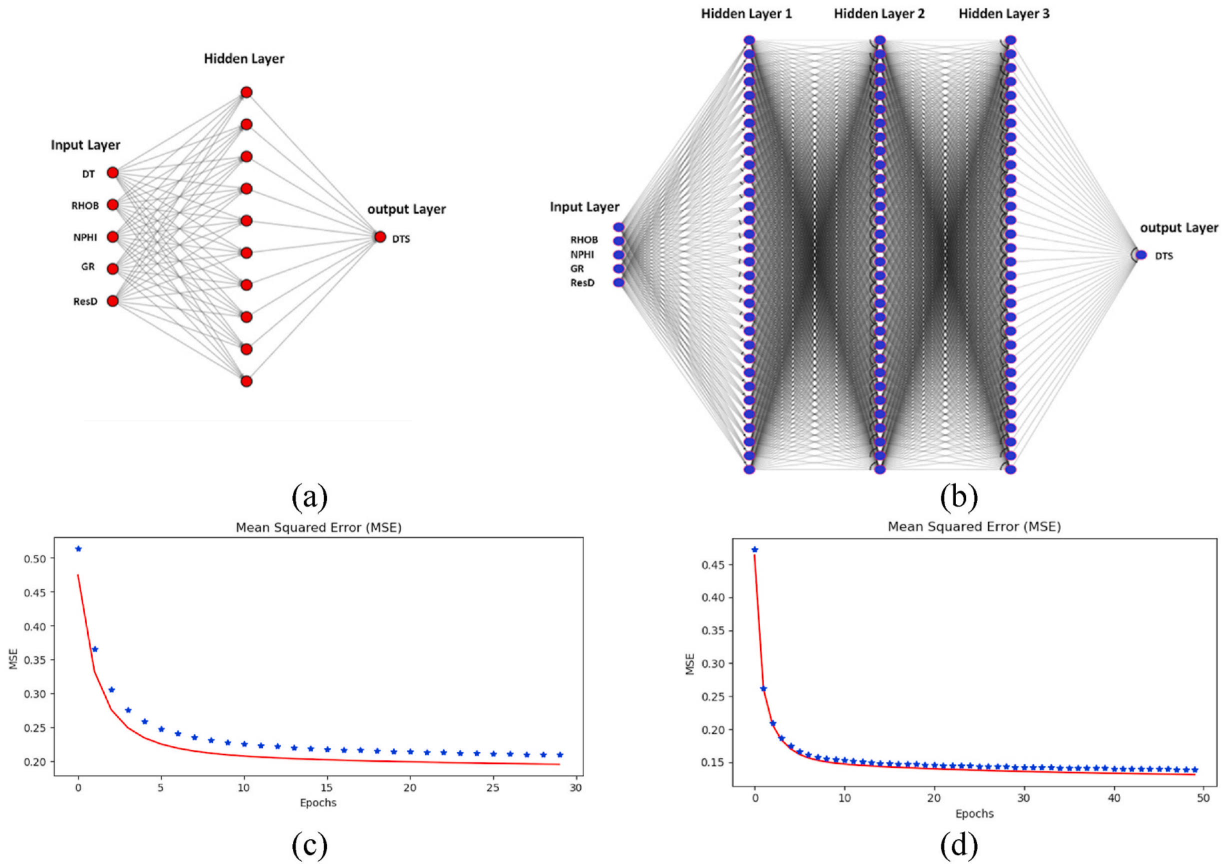


Fig. 6. (a) Shallow network diagram. (b) Deep network diagram. (c) Training error (blue curve) and test error (red curve) of the shallow network. (d) Training error (blue curve) and test error (red curve) of deep network. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

existing well logs dataset to generate new (synthetic) logs for wells with missing logs.

2.7. Optimizing the importance of rank feature

Before building a deep neural network model for the prediction of the shear sonic curve. It is important to study the existing log data about which log curve is the most suitable or highly correlated with shear sonic curve prediction. Therefore, the feature selection process is used to identify the logs with relatively higher relevancy with shear sonic. Being a key preprocessing step in ML and data mining, feature selection is the process of selecting effective features from the original feature curves to reduce the dimension of the dataset (Anifowose et al., 2014; Tao et al., 2019). In the case of high-dimensional datasets, the use of the feature curve owing to the low correlation with the target curve will result in a low-quality model. There are several conventional log curves, it is necessary to select the feature curve possessing a higher correlation with the shear sonic curve before training the model. I Incorporating the XGBoost algorithm (Chen and Guestrin, 2016), the importance was ranked according to its gain value in all boosting decision trees. The importance of predicting the shear velocity is increasing with the growth of the importance score. As shown in Fig. 5(a), the order of importance is DT < DEPTH < NPHI < GR < LLD < RHOB < CALI. Similarly, Fig. 5(b) also presents the significance of the correlation between logs using Pearson correlation heat map. Most of the logs found to have a significant positive correlation with DTS, such as the density log (RHOB), gamma-ray (GR), and sonic log (DT), being the highly correlated log, had a significant positive relation with shear sonic (DTS). However, caliper (CALI), resistivity (LLD), and NPHI were negatively correlated with the shear sonic log (DTS). Therefore, the top 5 feature curves had been selected according to their importance scores presented in Fig. 5 (a).

2.8. Deep Neural Network (DNN)

Neural Network (NN) is a very robust tool of supervised machine learning, which is characterized by its neurons (processing units of NN) and activation functions (determine an output of a neuron in the result of constraining the summation to finite value), biases (shift/process the input of activation function that is determined by its range), and equivalent weights (an input signal to a neuron) (Guresen and Kayakutlu, 2011; Haykin, 2011). When a NN has multiply fully connected layers, it is called a deep learning mechanism or deep neural network (DNN). DNN can be applied to perform either regression or classification tasks. This study uses a regression task performed by the DNN, as the regression technique can compute the predictive value of the target well logs. The activations function enforces non-linearity and discriminates DNN from linear regression techniques (Guresen and Kayakutlu, 2011; Haykin, 2011). The structure of the neural network is designed as displayed in Fig. 6. Since the purpose is to predict missing logs from the existing well-logs dataset to generate new (synthetic) logs, the regression task is carried out by the network. A DNN has been constructed with 5 nodes in the input layer (this layer inputs the initial data) whose labels are identical with the selected feature curves (Fig. 5a) and 3 hidden layers train the input data. The number of neurons in each hidden layer are 32, the output layer has 1 node (this layer produces the results based on input data), and all layers are fully connected. The activation function of each layer is rectified linear unit (ReLU). A ReLU is a non-linear function of supervised machine learning that allows error back-propagation throughout the multiple layers. In order to compare the

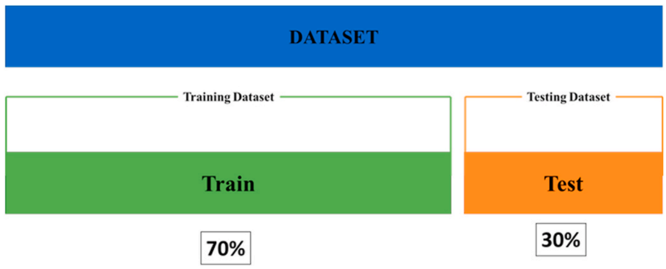


Fig. 7. The dataset is divided into two sets.- Set 1 is the training dataset and set 2 is the testing dataset.

result, a shallow network has been constructed in the same way, with only one hidden layer, 10 neurons in this layer, and the output layer has 1 node. Comparing the training and test error curves of these two networks, the training and testing errors of the deep network are smaller than those of the shallow network, and the error curves are more stable (Fig. 6c and d).

The study uses the mean square error (MSE) loss function because loss function is used to estimate the error between the predicted value ($\hat{y}_t$) and the real value (y_i). It is a non-negative real-valued function, which is usually expressed by L (Y, f(x)). The small value of the loss function indicates more robustness of the model. The loss functions used for regression are mainly mean absolute error (MAE) and MSE. The relationship between MAE loss and the absolute error is linear while the relation between MSE loss and error is square. So, when the error is

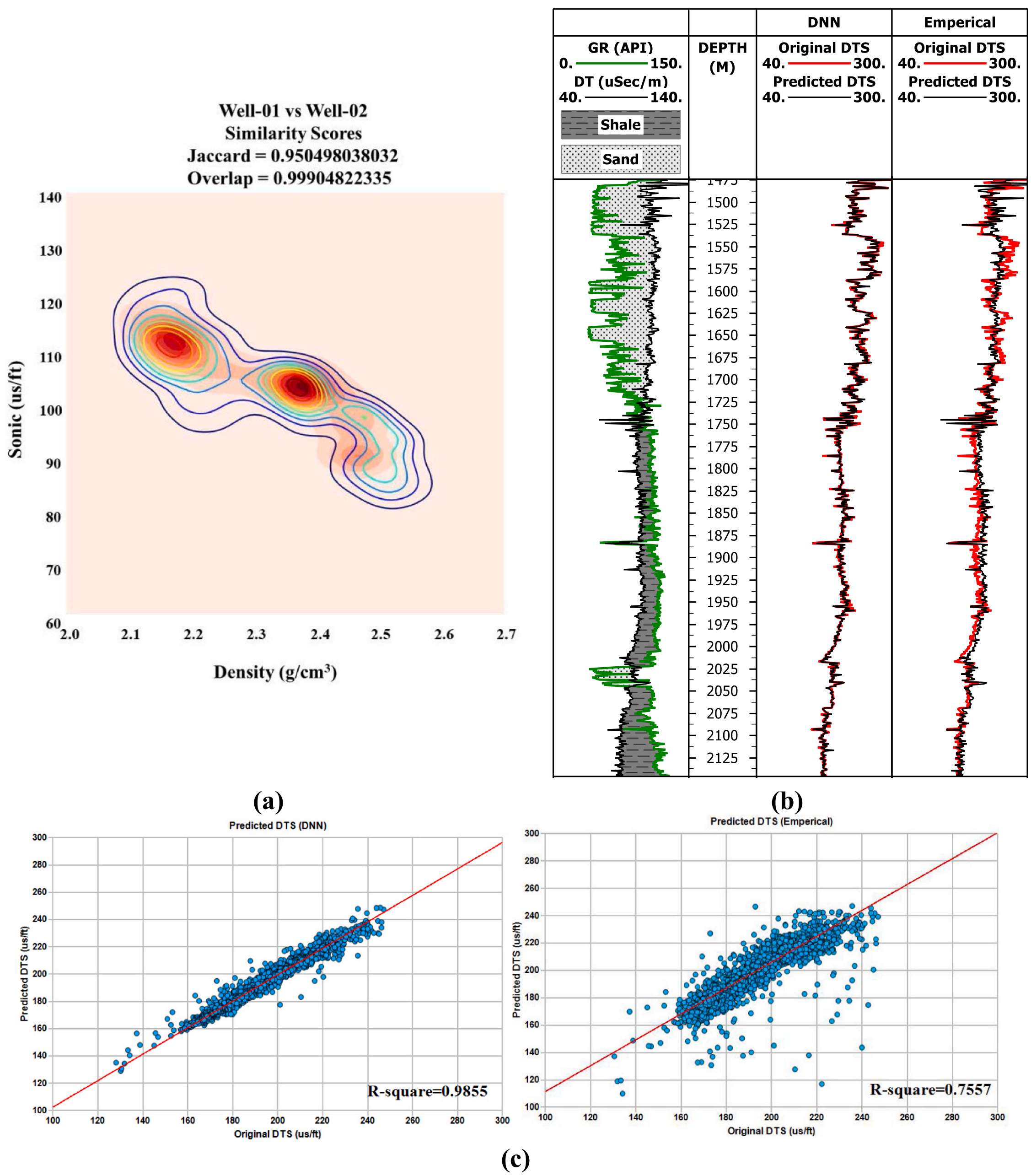


Fig. 8. (a) Similarity map shows the similarity scores between the wells. (b) Comparison between measured log and predicted log by the DNN and empirical estimation for DTS in the testing well: in track-1 GR and sonic log, in track-2 original DTS vs predicted DTS by DNN, and in the last track original DTS vs predicted DTS using the empirical estimation. (c) The cross-plots show the correlation between original DTS vs predicted DTS (empirical) and original DTS vs predicted DTS (DNN).

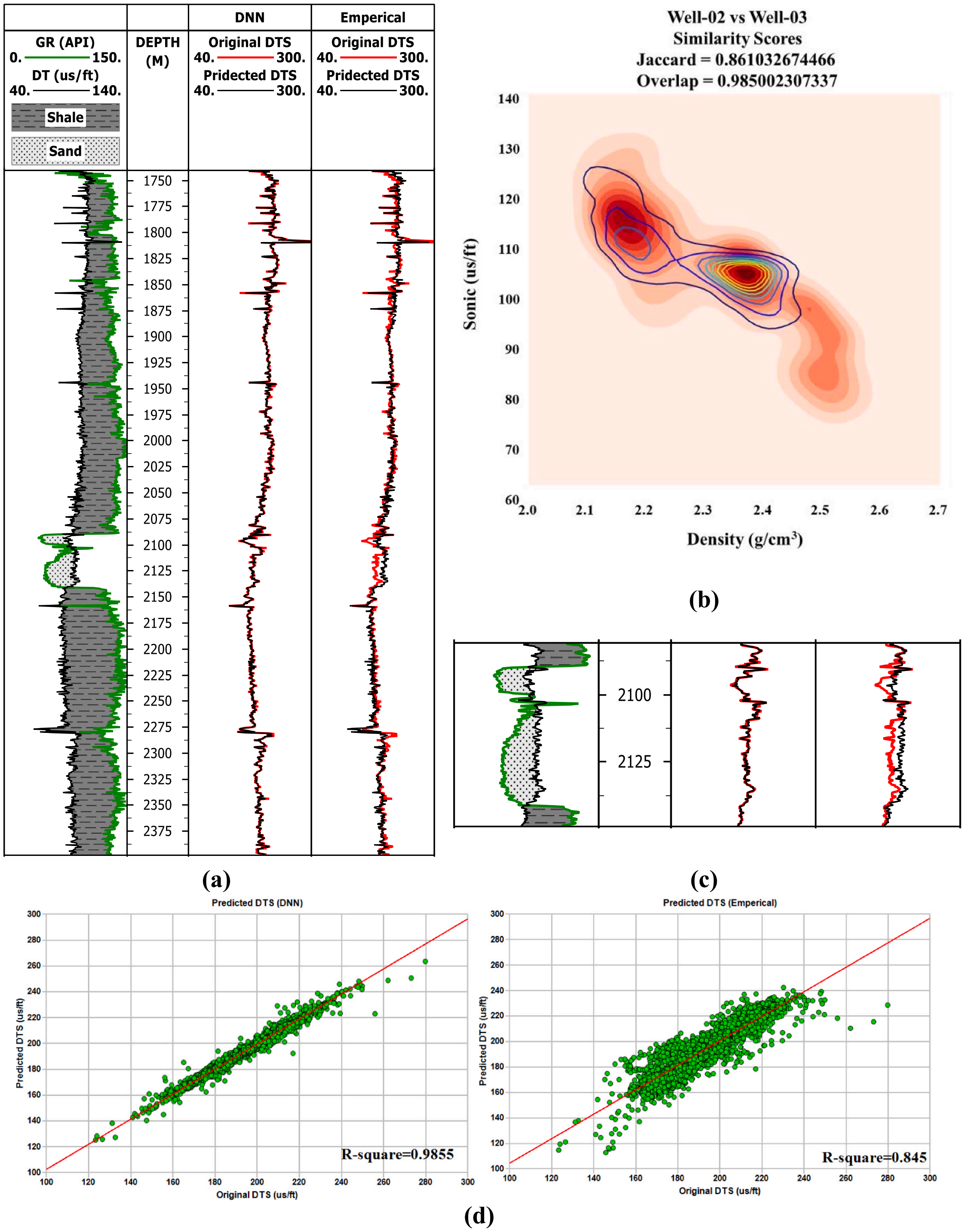


Fig. 9. (a) Similarity map shows the similarity scores between the wells. (b) Comparison between measured log and predicted log by the DNN and empirical estimation for DTS in the testing well: in track-1 GR and sonic log, in track-2 original DTS vs predicted DTS by DNN, and in the last track original DTS vs predicted DTS using the empirical estimation. (c) Enlarged image that visualizes the result of clear prediction for both methods in the hydrocarbon bearing zone. (d) The cross-plots show the correlation between original DTS vs predicted DTS (empirical) and original DTS vs predicted DTS (DNN).

large, MAE loss will be far greater than MSE loss and when an abnormal value with a large error appears in the data, MAE will produce a very large loss, which will have an adverse influence on the model training (Al-Farisi et al., 2002; Theys et al., 2014). Therefore, the MSE loss function is adopted in this study.

$$J_{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y}_i| \tag{7}$$

Further, to control the quality of DNN performance, a powerful cross-validation approach was applied consisting of two datasets as shown in Fig. 7. Set 1 (well-01, well-02, and well-4 have 70% of the data selected as the training set according to well similarity score) will have all kinds of the well logs data and will be utilized as a training dataset to detect the minimization performance during the training process and avoid overfitting. The remaining 30% computes set 2 is used as the testing set according to well similarity score that will have few numbers of wells with missing logs and will be utilized afterward training procedure ends to predict the missing logs.

3. Results and discussion

Once the similarity score of the wells is acquired from the novel approach, we can start with the final phase of the research which is to combine the similarity approaches with DNN to predict the missing shear sonic log precisely and accurately, where well similarity score is higher or reasonable. Then, its results are compared with both the measured shear sonic log and the outcomes of an empirical estimation from Greenberg and Castagna (1992) technique.

A well having a higher similarity score and complete well logs dataset is employed as a reference well for the prediction of missing shear sonic log. For instance, Fig. 8 (a) shows the highest similarity score between well-01 (shaded color background in orange) and well-02 (contour line), which makes a good factor for the prediction of missing shear sonic log. As we can see in Fig. 8 (b) that there was a precise accuracy of prediction of shear sonic log on the basis of similarity results using DNN technique: as compared to the conventional method of using Greenberg and Castagna (1992)'s technique. Another way of comparing these results is to cross-plot the original versus predicting shear sonic log data. In Fig. 8 (c), it is easy to visualize the improvement of the proposed approach over the conventional method for predicting a missing shear sonic log. Numerically, between 1500m and 2200m depth, the Greenberg and Castagna (1992)'s estimation compared to the original sonic log results in a correlation coefficient of 75%; the proposed approach achieves a correlation coefficient of 98% when compared against the original sonic log. Judging from these results, we conclude that the proposed method generates a reasonable approximation to the original shear sonic log.

From a detailed examination, we noticed that the quality of shear sonic log prediction depends significantly on the data availability. Nevertheless, crucially not all data input curves are of equal importance for the prediction of missing shear sonic log. Fig. 5 demonstrates the prediction importance rank for each standard input curve for shear sonic log prediction, as reported by the XGboost python library. It can be seen that by far the most important log is the compressional sonic log input curve. An interesting observation of Fig. 5 is that depth is the second important feature. This is very interesting because the petrophysicists use depth for providing context to the other curves. However, depth is just a symptom of the fact that they are actually concerned with the underlying geological formation. This makes a lot of sense because different geology will result in different input curve values that can mean the same thing.

Similarly, we also applied the proposed approach to another well with two different scenarios to check the consistent accuracy of the prediction of our model in the low similarity between the wells. For example, the first scenario is regarding the hydrocarbon-bearing zone, where the Greenberg-Castagna method is limited and only works in brine saturated rocks. Whereas there is a low well similarity score but acceptable for prediction in the second scenario. Jaccard similarity is not sensitive to the case shown in Fig. 9 (b) where the footmark of the well-03 (contour lines) is a subset of the reference well. In such cases, Jaccard similarity will be reduced, even if the footmark of the well-03 (contour lines) is completely contained within the footmark of the reference well. It is important that we are able to identify these situations since the well-03 is a strong candidate to build a predictive model to predict the missing shear sonic log in the well-03.

It is worth noting that Fig. 9 (a) showed the overall improvement in the accuracy of the prediction of shear sonic log on the basis of similarity results using DNN technique over the Greenberg and Castagna's method-accuracy was achieved in the hydrocarbon bearing zone whereas Greenberg and Castagna's method showed poor performance, as it is shown in Fig. 9 (c) that pictures enlarged and clear visuals of the reservoir of Fig. 9 (c). Thus, Fig. 9 (d) presents the rates of prediction accuracy of the shear sonic log using both conventional and DNN techniques with 80% and 95% correlation coefficient, respectively. It is easy to visualize the improvement of the proposed approach over the conventional method for predicting a missing shear sonic log.

Numerically, between 1700m and 2350m depth, the Greenberg and Castagna (1992)'s estimation compared to the original sonic log outcomes in a correlation coefficient of 84 %; the proposed approach achieves a correlation coefficient of 95 % when compared to the original sonic log. Judging from these results, we conclude that the proposed approach generates a reasonable approximation to the original shear sonic log in the hydrocarbon bearing zone.

4. Conclusions

This research proposed a novel approach to predict missing shear sonic log responses more precisely and accurately using similarity patterns of various wells with similar geophysical properties. These outcomes are confirmed in results by comparing the predicted logs with the original shear sonic logs. The prediction accuracy is shown in hydrocarbon bearing zones where Greenberg and Castagna's method is limited and only works in brine saturated rocks. The novel approach is very useful to identify the common geophysical log responses between wells, helpful to predict shear sonic log precisely and accurately, and allows the prediction of shear sonic log in the hydrocarbon bearing zone without having performed Gassmann fluid substitution. It does not follow the zone-by-zone prediction of the missing logs like rock physics methods do and it outputs the uncertainties facilitated by the least squares method. Having the potential of demonstrating shear sonic log prediction in hydrocarbon bearing zones, which cannot be precisely predicted by the Greenberg-Castagna method that only works in brine saturated rocks, this approach will provide improved accuracy where shear sonic logs are missing and need to be predicted for geomechanics, rock physics, and other applications. In this research, we only focus on the prediction of missing shear sonic log, nevertheless, this approach can also be extended to predict missing density and sonic logs in all well locations.

Credit author statement

Conceptualization, M.A; methodology, M.A; software, M.A and K.A; validation, M.A; formal analysis, M.A; investigation, M.A and K.A; resources, M.A; data correction, M.A; writing-original draft, M.A; and writing-review editing, M.A, M.H and K.A; visualization, supervision, M.H, U.A and R.J; project administration, funding acquisition, M.H.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I am thankful to the Directorate General of Petroleum Concessions (DGPC), Pakistan, for providing the sample and research data. This research was funded by financial support from the National Key Research and Development Program of China (2016YFC0600201), and the National Natural Science Foundation of China (Grant No's. 41630317 & 91858205).

References

Akkurt, R., Conroy, T.T., Psaila, D., Paxton, A., Low, J., Spaans, P., 2018. Accelerating and enhancing petrophysical analysis with machine learning: a case study of an automated system for well log outlier detection and reconstruction. In: SPWLA 59th Annu. Logging Symp. 2-6 June, London, UK.

Al-Anazi, A., Gates, I.D., 2010. A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. Eng. Geol. 114, 267–277. https://doi.org/10.1016/j.enggeo.2010.05.005.

Al-Farisi, O., Dajani, N., Boyd, D., Al-Felasi, A., 2002. Data management and quality control in the petrophysical environment. In: Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2002. ADIPEC 2002. https://doi.org/10.2523/78524-ms.

Ali, Muhammad, Khan, Muhammad Jahangir, Ali, Mubarik, Iftikhar, Shaista, 2019. Petrophysical analysis of well logs for reservoir evaluation: a case study of BKadanwariˆ gas field, middle Indus basin, Pakistan. Arab. J. Geosci. 12 (6), 1–12. https://doi.org/10.1007/s12517-019- 4389-x. In this issue.

Ali, M., Ma, H., Pan, H., Ashraf, U., Jiang, R., 2020. Building a rock physics model for the formation evaluation of the Lower Goru sand reservoir of the Southern Indus Basin in Pakistan. J. Petrol. Sci. Eng. https://doi.org/10.1016/j.petrol.2020.107461.

Anemangely, M., Ramezanzadeh, A., Amiri, H., Hoseinpour, S.A., 2019. Machine learning technique for the prediction of shear wave velocity using petrophysical logs. J. Petrol. Sci. Eng. https://doi.org/10.1016/j.petrol.2018.11.032.

Anifowose, F.A., Labadin, J., Abdulraheem, A., 2014. Non-linear feature selection-based hybrid computational intelligence models for improved natural gas reservoir characterization. J. Nat. Gas Sci. Eng. https://doi.org/10.1016/j.jngse.2014.09.001.

Ashraf, Umar, Zhang, Hucai, Anees, Aqsa, Ali, Muhammad, Zhang, Xiaonan, Abbasi, Saiq Shakeel, Mangi, Hassan Nasir, 2020b. Controls on Reservoir Heterogeneity of a Shallow-Marine Reservoir in Sawan Gas Field, SE Pakistan: Implications for Reservoir Quality Prediction Using Acoustic Impedance Inversion. Water 12 (11), 2972. https://doi.org/10.3390/w12112972. In this issue.

Ashraf, U., Zhang, H., Anees, A., Mangi, H.N., Ali, M., Ullah, Z., Zhang, X., 2020a. Application of unconventional seismic attributes and unsupervised machine learning for the identification of fault and fracture network. Appl. Sci. https://doi.org/10.3390/app10113864.

Ashraf, U., Zhu, P., Yasin, Q., Anees, A., Imraz, M., Mangi, H.N., Shakeel, S., 2019. Classification of reservoir facies using well log and 3D seismic attributes for prospect evaluation and field development: a case study of Sawan gas field, Pakistan. J. Petrol. Sci. Eng. 175, 338–351. https://doi.org/10.1016/j.petrol.2018.12.060.

Bader, S., Wu, X., Fomel, S., 2018. Missing well log estimation by multiple well-log correlation. https://doi.org/10.3997/2214-4609.201800989.

Bestagini, P., Lipari, V., Tubaro, S., 2017. A machine learning approach to facies classification using well logs. https://doi.org/10.1190/segam2017-17729805.1.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine BT - computer Networks and ISDN Systems. Comput. Netw. ISDN Syst. 30, 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X.

Castagna, J., Batzle, M., Eastwood, R., 1985. Relationship between compressional and shear-wave velocities in classic silicate rocks. Geophysics 50, 571–581. https://doi.org/10.1190/1.1441933.

Chen, G., Zhu, J., Qiang, M., Gong, W., 2018. Three-dimensional site characterization with borehole data – a case study of Suzhou area. Eng. Geol. https://doi.org/10.1016/j.enggeo.2017.12.019.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Ehsan, M., Gu, H., 2020. An integrated approach for the identification of lithofacies and clay mineralogy through Neuro-Fuzzy, cross plot, and statistical analyses, from well log data. J. Earth Syst. Sci. https://doi.org/10.1007/s12040-020-1365-5.

Ehsan, M., Gu, H., Ahmad, Z., Akhtar, M.M., Abbasi, S.S., 2019. A modified approach for volumetric evaluation of shaly sand formations from conventional well logs: a case study from the talhar shale, Pakistan. Arabian J. Sci. Eng. https://doi.org/10.1007/s13369-018-3476-8.

Ehsan, M., Gu, H., Akhtar, M.M., Abbasi, S.S., Ullah, Z., 2018. Identification of hydrocarbon potential of talhar shale: member of lower Goru formation using well logs derived parameters, southern lower Indus Basin, Pakistan. J. Earth Sci. https://doi.org/10.1007/s12583-016-0910-2.

Faust, L.Y., 1953. A velocity function including lithologic variation. Geophysics 18, 271–288. https://doi.org/10.1190/1.1437869.

Freire, L., Roche, A., Mangin, J.F., 2002. What is the best similarity measure for motion correction in fMRI time series?. In: IEEE Transactions on Medical Imaging https://doi.org/10.1109/TMI.2002.1009383.

Gardner, G.H.F., Gardner, L.W., Gregory, A.R., 1974. Formation velocity and density - the diagnostic basics for stratigraphic traps. Geophysics. https://doi.org/10.1190/1.1440465.

Geng, Z., Wang, Y., 2020. Physics-guided deep learning for predicting geological drilling risk of wellbore instability using seismic attributes data. Eng. Geol. 105857. https://doi.org/10.1016/j.enggeo.2020.105857.

Greenberg, M.L., Castagna, J.P., 1992. SHEAR-WAVE velocity estimation IN porous rocks: theoretical formulation, preliminary verification and applications. Geophys. Prospect. https://doi.org/10.1111/j.1365-2478.1992.tb00371.x.

Guresen, E., Kayakutlu, G., 2011. Definition of artificial neural networks with comparison to other networks. Procedia Comput. Sci. 3, 426–433. https://doi.org/10.1016/j.procs.2010.12.071.

Haykin, S.O., 2011. Neural Networks and Learning Machines. Pearson Education.

Jaccard, P., 1912. The distribution OF the flora IN the alpine ZONE.1. New Phytol. 11, 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x.

Jason, F., 2019. Machine Learning Workflow for Automated Log Editing. https://register.gotowebinar.com/recording/1951884226031955458.

Li, Guoliang, Li, Guanfang, Wang, Y., Qi, S., Yang, J., 2020. A rock physics model for estimating elastic properties of upper Ordovician-lower Silurian mudrocks in the Sichuan Basin. China. Eng. Geol. https://doi.org/10.1016/j.enggeo.2019.105460.

Smith, J.H., 2007. A method for calculating pseudo sonics from E-logs in a clastic geologic setting. Gulf Coast Assoc. tion Geol. Soc. Trans. 57, 1–4.

Tao, Z., Huiling, L., Wenwen, W., Xia, Y., 2019. GA-SVM based feature selection and parameter optimization in hospitalization expense modeling. Appl. Soft Comput. J. https://doi.org/10.1016/j.asoc.2018.11.001.

Theys, P., Roque, T., Anadarko, Vik, M., Williams, J., Storey, M., 2014. Current status of well logging data deliverables and a vision forward. In: SPWLA 55th Annual Logging Symposium 2014.

Tong, L., Liu, L., Cai, G., Du, G., 2013. Assessing the coefficient of the earth pressure at rest from shear wave velocity and electrical resistivity measurements. Eng. Geol. https://doi.org/10.1016/j.enggeo.2013.05.012.

Tonni, L., Simonini, P., 2013. Shear wave velocity as function of cone penetration test measurements in sand and silt mixtures. Eng. Geol. https://doi.org/10.1016/j.enggeo.2013.06.005.

Verma, A.K., Routray, A., Mohanty, W.K., 2014. Assessment of similarity between well logs using synchronization measures. Geosci. Rem. Sens. Lett. IEEE 11, 2032–2036. https://doi.org/10.1109/LGRS.2014.2317498.

Vijaymeena, M.K., Kavitha, K., 2016. A survey on similarity measures in text mining. Mach. Learn. Appl. An Int. J. 3, 19–28.

Weijun, N., Qi, L., Tao, F., 2017. Prediction of shear wave velocity in shale reservoir based on logging data and machine learning. In: 2017 2nd International Conference on Knowledge Engineering and Applications, vol. 2017. ICKEA. https://doi.org/10.1109/ICKEA.2017.8169935.

Wu, X., Shi, Y., Fomel, S., Liang, L., Zhang, Q., Yusifov, A.Z., 2019. FaultNet3D: predicting fault probabilities, strikes, and dips with a single convolutional neural network. IEEE Trans. Geosci. Rem. Sens. https://doi.org/10.1109/TGRS.2019.2925003.