

Data-driven S-wave velocity prediction method via a deep-learning-based deep convolutional gated recurrent unit fusion network

Jun Wang¹ and Junxing Cao¹

ABSTRACT

The S-wave velocity (V_S) is a fundamental parameter in geophysical analysis, prestack seismic inversion, and reservoir prediction. For various reasons, the availability of directly measured values of V_S is low, especially in old wells. Therefore, indirect estimations of V_S data on the basis of available reservoir information are important, and the development of a high-efficiency and low-cost prediction method is necessary. We have developed a novel prediction method that combines the convolutional neural network (CNN) and the gated recurrent unit (GRU) algorithms, based on a deep convolutional GRU (DCGRU) approach. More specifically, a CNN structure is used to identify and memorize the complex relationship between V_S and well-log data, whereas a GRU network is introduced to extract key features of the data series in the depth direction. Owing to its structure, the DCGRU approach can seamlessly account for data trends with depth, local correlations across data series, and the actual depth accumulation effect. This approach is tested on data sets from an actual reservoir; it provides more reliable and accurate V_S predictions not only compared with empirical models but also compared with the CNN and GRU algorithms applied separately. Our approach has potential for accurately estimating V_S from log data.

INTRODUCTION

The S-wave velocity (V_S) is a fundamental parameter in a variety of geophysical models, such as the seismic wave forward modeling, prestack seismic data inversion, and reservoir prediction (Smith and Gidlow, 2000; Downton, 2005; Rajabi et al., 2010; Aranibar et al., 2013; Akhundi et al., 2014; Fattahi et al., 2016; Vernik et al., 2017).

Because of the difficulty and high cost of performing S-wave logging, V_S data for reservoir prediction are often lacking (Rezaee et al., 2007; Tan et al., 2015; Xu et al., 2017). Therefore, the development of a fast, cheap, and reliable alternative method for V_S prediction is necessary.

Traditional V_S prediction approaches are mainly classified into two groups: empirical models and theoretical petrophysical models. The former are based on laboratory or logging data and allow the elucidation of linear or simple nonlinear relationships between V_S and the P-wave velocity (V_P) or other parameters such as density or gamma-ray values (Carroll, 1969; Han et al., 1986; Krief et al., 1990; Miller and Stewart, 1990; Dvorkin et al., 1999; Greenberg and Castagna, 1992; Castagna et al., 1993; Gal et al., 1998; Jørstad et al., 2001; Brocher, 2005; Mabrouk and Pennington, 2009; Hossain et al., 2012; Ojha and Sain, 2014; Oloruntobia et al., 2019). A variety of empirical models are available depending on the characteristics of the reservoir. These models can offer useful predictions; however, they are interval based and depend on the rock type. Moreover, their calibration is time-consuming and requires human expertise.

Theoretical petrophysical models rely on the wave-propagation process in the medium. Depending on the structure and mineralogy of the rock, a mathematical formula can be constructed to characterize the relationship between microscopic and elastic parameters, which can then be used to predict V_S (Xu and White, 1995, 1996; Pride et al., 2004; Lee, 2006; Xu and Payne, 2009; Luo et al., 2016). The accuracy of such a prediction is high, but the algorithm is complex and requires many parameters, resulting in low computational efficiency. Furthermore, assumptions on the pore shape — which add further complexity and indeterminacy — are necessary in most models.

With the rapid development of artificial intelligence (AI) strategies and improvements in hardware, the interest around AI-based technologies has been increasing in the field of oil and gas exploration and development (Maiti et al., 2007; Fu et al., 2018; Lin et al., 2018; Ray and Myer, 2019; Zhao et al., 2019b; Feng, 2020;

Manuscript received by the Editor 11 December 2020; revised manuscript received 11 June 2021; published ahead of production 20 July 2021; published online 24 September 2021.

¹State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation of Chengdu University of Technology, Chengdu 610059, China and Chengdu University of Technology, School of Geophysics, Chengdu 610059, China. E-mail: wjlwy@stu.cduto.edu.cn; caojx@cduto.edu.cn (corresponding author). © 2021 Society of Exploration Geophysicists. All rights reserved.

Meshalkin et al., 2020; Wang et al., 2020b; Xi and Huang, 2020).

In recent years, with machine learning (ML) applications becoming increasingly common in science and engineering, many researchers have attempted to use ML methods for V_S prediction. Support vector machines, fuzzy logic models, and artificial neural networks have been used (Eskandari et al., 2004; Rezaee et al., 2007; Rajabi et al., 2010; Akhundi et al., 2014; Maleki et al., 2014; Anemangely et al., 2019; Mergini et al., 2019; Zhu et al., 2020a; Olayiwola and Sanuade, 2021). However, these ML methods only conduct point-to-point mapping. In other words, they can only relate V_S predictions at a certain depth to well-log data at the same depth, whereas they cannot account for spatial correlations (along depth and with nearby wells). Because this contrasts with geologic experience, the V_S values estimated by these methods have a questionable accuracy.

Deep learning (DL) is a recently developed class of ML algorithms for AI applications (LeCun et al., 2015). It includes many popular models, such as the convolutional neural network (CNN) (Nebauer, 1998) and the recurrent neural network (RNN) (Schuster and Paliwal, 1997). CNN models have gained increasing attention among geophysicists owing to their ability to automate the interpretation and inversion of seismic and well-log data (Li and Harris, 2018; Xiong et al., 2018; Yuan et al., 2018; Zhang et al., 2018b; Wu et al., 2019; Zhong et al., 2019; Feng et al., 2020; Zhu et al., 2020b). However, even CNNs only rely on well-log data at the same depth for the prediction of V_S . Therefore, they are unable to extract key features along the depth direction. Conversely, rock properties often exhibit trends with depth that are relevant in geologic studies. Therefore, the use of RNNs would be preferable, as such algorithms use an internal input from the previous step (e.g., the trend) together with external inputs (other available log types).

The long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) network is an advanced type of RNN. Because gate structures are constructed in self-looped neuron cells, the LSTM algorithm can preserve long-term previous information for future use without any further modification, thereby solving the vanishing gradient problem of the RNN. Using this method, the V_S is estimated from a series of data inputs while considering their internal relationships and trends (Zhang et al., 2018a, 2020; Chen et al., 2020; Pham et al., 2020; Wang et al., 2020a). The LSTM approach performs well with hyphen data; however, its internal structure is complex and it requires a long training time. To overcome these issues, a gated recurrent unit (GRU) was proposed (Graves et al., 2013), based on the LSTM algorithm. A GRU-based V_S prediction method can effectively mine contextual information from logging data and reduce the prediction multiplicity. However, it does not

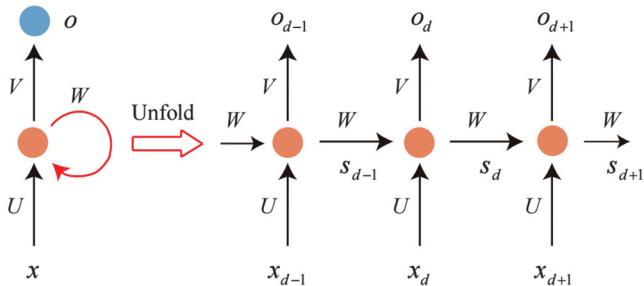


Figure 1. Structure of an RNN and its unfolded network.

consider the local correlation among logs; thus, it is unable to fully identify the complex relationship between the well-log data and V_S . To overcome this limitation, this method should be combined with another network algorithm. Ensemble learning is a popular approach that can “gather the strengths” of all of its constituent methods (i.e., various ML algorithms).

The purpose of this study was to construct an effective AI system to achieve an exhaustive prediction of V_S . Based on the advantages of CNN and GRU networks in data mining and the idea of ensemble learning, a deep convolutional GRU (DCGRU) fusion network algorithm was developed. The root-mean-square error (rms error) was adopted as the loss function. Dropout, ReduceLROnPlateau, and early stopping methods were used to avoid overfitting. The optimal DCGRU network was obtained by progressively reducing the loss function. The V_S predictions from log data of two exploration areas confirmed the high prediction accuracy of the proposed method.

THEORY

Convolutional neural network

The CNN (Bengio, 2009; LeCun et al., 2015) is inspired by the structure of the visual system, which can extract features from input data by establishing multiple filters. With an increasing network depth, deeper features can be extracted. CNN is one of the most successful DL methods. Its network structures include 1D, 2D, and 3D CNNs (Zhao et al., 2019a). One-dimensional CNNs are mainly used for sequential data processing (Abdeljaber et al., 2017; Kiranyaz et al., 2019), whereas 2D CNNs are often used for image and text recognition, and 3D CNNs are mainly used for medical imaging and video data recognition. Considering that V_S is predicted by solving a multisequence input problem, and coupling exists between different curves, the convolution kernel in the CNN network used in this study requires a 1D structure.

Logging data feature local correlations, which makes the use of a CNN suitable. In addition, compared with the traditional multilayer perceptron and other neural networks, the CNN reduces the complexity of the network model and achieves better generalization by using local connections and by sharing the weight of the convolution layer. Over the years, CNNs have been widely used for image classification, facial recognition, and other applications. CNNs abstract the input information into multilayer features and retain the spatial topology of the original data.

GRU neural network

RNNs are a specialized class of deep neural networks (DNNs) for processing sequential data, such as audio signals and text data (Pasa et al., 2015). Unlike feed-forward neural networks, RNNs allow information to cycle through a feedback loop that takes the current input and what was learned from previous time steps into consideration. Such loops enable the algorithm to capture long temporal dependencies in sequential data and permit an exhaustive analysis of hyphen data, such as well-log data. The typical structure of RNNs is shown in Figure 1. Similar to the structure of a fully connected network, it also contains input, hidden, and output layers. However, the hidden layer has a different structure. In fact, the nodes in the hidden layers are connected to each other, and the output at the current time step is affected not only by the current input but also by the input from previous steps. The structure of an RNN is

particularly suited for dealing with sequential data (such as well logs). RNNs are theoretically capable of learning long-term dependencies from hidden states; however, in practical applications, they might fail when training with long sequences because of numerical issues with gradient calculation (Bengio et al., 1994).

To solve this limitation, a more effective solution would be to use gated RNN architectures, such as LSTM or GRU. LSTM is specifically designed to account for long-term dependencies. The memory cell and transfer belt structure in the LSTM algorithm can hold information for later use and prevent early signals from gradually disappearing during the processing. However, the training of LSTM networks usually takes a long time owing to its complex internal structure. The GRU is a simplified version of the LSTM and has a more concise model structure (Wang et al., 2018). The GRU combines the forget and input gates into an update gate, so that information is selectively transmitted in the hidden layer. This effectively alleviates the problem of gradient disappearance while keeping important information in memory. Because its performance is similar to that of the LSTM for various tasks, the GRU was considered to be more suitable for this study because of its simpler structure, fewer parameters, stronger prevention of overfitting, and faster convergence speed.

The GRU network is used as an example to illustrate the log data feature extraction process (Figure 2). The reset gate information is calculated using the output layer and the hidden layer state of the forward GRU network at a previous depth. The reset gate is used to control the importance of information at a previous depth, which can effectively reduce the risk of the vanishing gradient. The terms r_d and x_d are the reset gate and the output layer at depth d , respectively, and h_{d-1} is the hidden layer state of the forward GRU network at depth $d-1$. The reset gate that contains the previous depth log information is stored in new memory \tilde{h}_d . The update gate is used to control the degree of combination between current and previous depth information. A larger update gate indicates that more state information of the previous depth is retained. Finally, we obtain hidden layer state h_d , which is the output of the forward GRU network. It is related to the depth information retained by the current layer and previous memory. In Figure 2, $\sigma(x)$ represents the activation function; \circ represents the dot product; and U_d , W_d , and b_d are the weights and biases of the forward GRU network.

Deep DCGRU fusion network

The temporal relationship between shifted sequences captured by the GRU architecture can identify a geologic trend. To account for the local correlation across log data series, which depends on the depositional facies (Cant, 1994; Pham et al., 2020), we extracted the characteristics of the data using CNNs. In other words, the data were first input into CNNs to extract different characteristics, which were then received by the GRU to calculate the V_s . Therefore, we represented each temporal point by a 1D CNN to encode local correlation information around a point and applied the DCGRU to capture spatio-temporal correlations (Shi et al., 2015; Zhao and Itti, 2018). The DCGRU calculates the future

state of a certain cell in a spatial grid using the inputs and past states of its local neighbors.

METHODOLOGY

To predict the V_s using a deep neural network, we define a function f as

$$Y_d = f(\mathbf{X}, d), \quad (1)$$

where \mathbf{X} is the well-log data and Y_d is the V_s at depth d . The function f is generally highly nonlinear and difficult to obtain, so it is appropriate to approximate it using a DNN. Neural networks can estimate Y_d by considering the trend of well-log data with depth using the following mapping:

$$f_\theta : \mathbf{X}, \mathbf{d} \rightarrow \mathbf{Y}, \quad (2)$$

where θ represents the parameters of the deep neural network, which can be divided into two groups: network weights and biases. In addition, if only the input effects at the same depth are considered, the mapping in equation 2 can be simplified as

$$f_\theta : \mathbf{X} \rightarrow \mathbf{Y}. \quad (3)$$

Equations 2 and 3 state that, given a training set, a neural network can be trained to perform an approximate mapping to V_s . The training is an optimization process that yields network parameters.

For a well-log data series, the defined depth sequence matrix \mathbf{X} can be expressed in equation 4. The set $\mathbf{X}_d = (x_d^1, x_d^2, \dots, x_d^m)$ contains the values of m log attributes at depth d . The set $\mathbf{X}^m = (x_1^m, x_2^m, \dots, x_d^m, \dots, x_D^m)$ contains the values of a well-log attribute along depth, such as V_p , gamma ray (GR), neutron porosity (NPHI), and density (RHOB). The term $\mathbf{Y} = (y_{a1}, y_{a2}, \dots, y_{ad}, \dots, y_{aD})$ shows the historical information

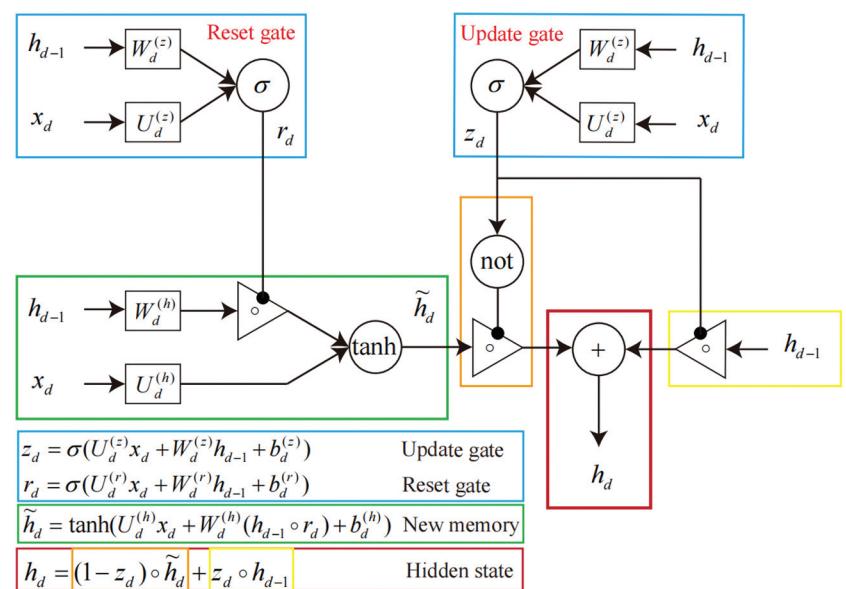


Figure 2. Internal structure of a GRU cell. A GRU network consists of a reset gate and an update gate and makes predictions at a given depth by using a gate mechanism to control the input and deep memory information.

of the predicted V_S :

$$\mathbf{X} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^m \\ x_2^1 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \vdots \\ x_d^1 & x_d^2 & \cdots & x_d^m \\ \vdots & \vdots & \vdots & \vdots \\ x_D^1 & x_D^2 & \cdots & x_D^m \end{bmatrix}. \quad (4)$$

Network structure

The DCGRU network structure is illustrated in Figure 3. A notable feature of CNNs is the local connection. In the field of computer vision, in a certain area of an image, the correlation between pixels within a relatively small distance is typically high, whereas the correlation is low when the distance becomes large (Li et al., 2020). The nature of the image conforms to the idea of a local connection, and thus CNNs are widely used in image processing. For our well-log data, the information carried at similar depths will be similar. Therefore, we chose CNNs to extract the characteristics of the data for prediction. As the knowledge of V_S in shallow layers may help the network predict values in deeper layers, the V_S values of shallow layers are calculated first, and their information is preserved by the GRU to calculate the V_S values of deeper layers.

The DCGRU network in this study is composed of an input module, a CNN module, a GRU module, and a fully connected layer. The input module uses conventional logging data as its input. The CNN module captures the deep spatial features of the input logging data sequence. The GRU module extracts key features from the well-log data along the depth direction and further captures internal correlations. Finally, in the fully connected layer, the final V_S prediction is obtained based on the extracted features.

Prediction framework

The prediction framework based on the deep DCGRU fusion network model is shown in Figure 4. First, a sensitivity analysis for

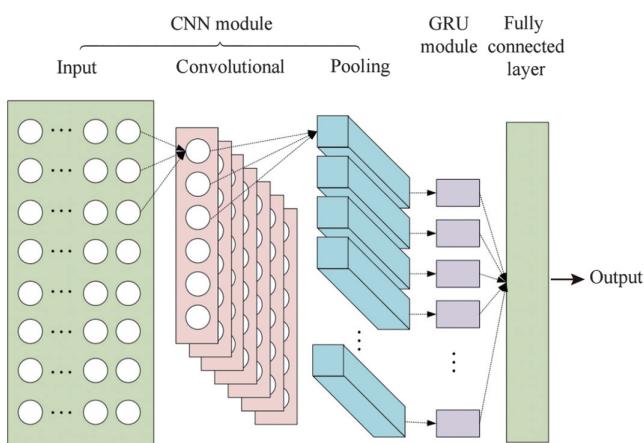


Figure 3. The DCGRU fusion network structure. A DCGRU network is mainly composed of an input module, a CNN module, a GRU module, and a fully connected layer.

conventional well logs and V_S is performed, which aims at choosing well logs that are relevant for V_S prediction. Then, the well-log data are divided into a training set and a test set, which is used to test the model accuracy. The output of the DCGRU model is connected to the fully connected layer, which outputs the V_S prediction (Zhang et al., 2018a; Song et al., 2020). By calculating the rms error between the predicted and actual V_S values, the parameters of the DCGRU model can be updated. Through continuous training, an optimal DCGRU network model can be obtained. Finally, the test set is input into the optimal DCGRU network to evaluate V_S , and the prediction performance of the model is measured by the rms error, the mean absolute error (MAE), and the mean absolute percentage error (MAPE).

Feature selection

Feature selection is an important step in DL methods. It involves selecting appropriate features that contribute the most to the prediction of the unknown parameter. The number and type of features can affect the prediction accuracy. The V_S prediction based on DCGRU is in complete accordance with well logs. Therefore, well logs are significant for the prediction results. Well logs contain diverse information, such as mineral compositions, sedimentary sequences, and pore characteristics. Thus, it is necessary to preliminarily analyze the relationships between well logs and V_S before making the actual prediction. In this study, the Pearson correlation coefficient (PCC) (Pearson, 1896) was used to quantitatively analyze the degree of correlation between V_S and input logs.

Data normalization

Data normalization eliminates the systematic errors induced by different measurement tools, maximizes the reflection of geologic information in log data, and causes all log data to follow the same standard as much as possible. Furthermore, data normalization has the effect of speeding up the training process and prevents the exploding gradient issues (Hrynaszkiewicz, 2010). In this study, we used the MinMaxScaler normalization method to scale the logging data into the range [0,1].

Evaluation metrics

We used the rms error, MAE, and MAPE to evaluate the model performance:

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)|, \quad (5)$$

$$\text{rms error} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}, \quad (6)$$

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \quad (7)$$

where m is the number of samples, y_i is the actual value, and \hat{y}_i is the predicted value.

Model parameter setting

In the training process, the parameters, including the number of previous time steps used to perform the prediction at the current time (i.e., the step size) and the number of training datapoints for each

training (i.e., the batch size), are crucial for the performance of the training model (Zhang et al., 2018a; Song et al., 2020). The parameters were obtained according to previous experience and by comparative experiments. The structures of the DCGRU, deep GRU (DGRU), deep LSTM (DLSTM), and deep CNN (DCNN) DL models are listed in Table 1. In addition, the Adam algorithm (Kingma and Ba, 2015) was used as the optimization method. The adaptive learning rate adjustment (ReduceLROnPlateau) (set patience = 5, i.e., the number of epochs with no improvement after which the learning rate will be reduced), early stopping (set patience = 10, i.e., the number of epochs with no improvement after which the training will be stopped), and dropout (Srivastava et al., 2014) were used to avoid overfitting. These hyperparameters can be tuned to maximize the performance. However, when comparing the performance of different algorithms for a specific task, it is difficult to choose completely fair hyperparameters. A fair comparison depends on several factors (Chen et al., 2020). Therefore, to minimize the unfairness, we set all parameters as equal except for the network type.

Model training

The error variations of different models in the model training process are shown in Figure 5. It can be seen that the training-phase loss values decreased significantly for all models, demonstrating that the learning rate was appropriate and that the gradient descent process was correctly implemented. After a certain amount of learning, the loss value stabilized at a minimum value. Figure 5 shows that all models reached their optimal state; however, the training error of the DCGRU was lower and more stable than that of the other models. This indicates that the

Figure 4. The V_s prediction framework of the DCGRU fusion network model. First, a sensitivity analysis for conventional well-log data and V_s is performed, which aims at identifying the data series that can best constrain V_s . Then, the well-log data are divided into a training set used to train the DCGRU model and a testing set used to test the model accuracy. The training set is input into the neural network structure. The output of the DCGRU model is connected to the fully connected layer to yield the V_s prediction.

Table 1. Structures of the developed DL models.

Model	Structure
DCGRU	1 Convolutional layer (64 filters + 4 kernel_size + Relu activation + padding ‘same’ + 1 stride) + MaxPooling (1 pooling size) + 1 Convolutional layer (64 filters + 4 kernel_size + Relu activation + padding ‘same’ + 1 stride) + MaxPooling (2 pooling size) + Flatten + 2 GRU layer (40 neurons + Tanh activation + Sigmoid recurrent_activation) + 1 Dense layer (1 neuron, linear activation) + Compile (root mean_squared_error loss, Adam optimizer, ReduceLROnPlateau and early stopping strategy, maximum epochs is 1000, time steps is 5, batch size is 10, dropout is 0.1).
DGRU	2 GRU layer (64 neurons + Tanh activation + Sigmoid recurrent_activation) + 2 GRU layer (40 neurons + Tanh activation + Sigmoid recurrent_activation) + 1 Dropout layer (0.1) + 1 Dense layer (1 neuron, linear activation) + Compile (root mean_squared_error loss, Adam optimizer, ReduceLROnPlateau and early stopping strategy, maximum epochs is 1000, time steps is 5, batch size is 10, dropout is 0.1).
DLSTM	2 LSTM layer (64 neurons + Tanh activation + Sigmoid recurrent_activation) + 2 LSTM layer (40 neurons + Tanh activation + Sigmoid recurrent_activation) + 1 Dropout layer (0.1) + 1 Dense layer (1 neuron, linear activation) + Compile (root mean_squared_error loss, Adam optimizer, ReduceLROnPlateau and early stopping strategy, maximum epochs is 1000, time steps is 5, batch size is 10, dropout is 0.1).
DCNN	1 Convolutional layer (64 filters + 4 kernel_size + Relu activation + padding ‘same’ + 1 stride) + MaxPooling (1 pooling size) + 1 Convolutional layer (64 filters + 4 kernel_size + Relu activation + padding ‘same’ + 1 stride) + MaxPooling (2 pooling size) + Flatten + 1 Dropout layer (0.1) + 1 Dense layer (1 neuron, linear activation) + Compile (root mean_squared_error loss, Adam optimizer, ReduceLROnPlateau and early stopping strategy, maximum epochs is 1000, time steps is 5, batch size is 10, dropout is 0.1).

ability of the DCGRU to account for the local correlation across logs and their “temporal” characteristics (i.e., the trends along the depth direction) results in a better elucidation of the relationship between the logging data and V_S .

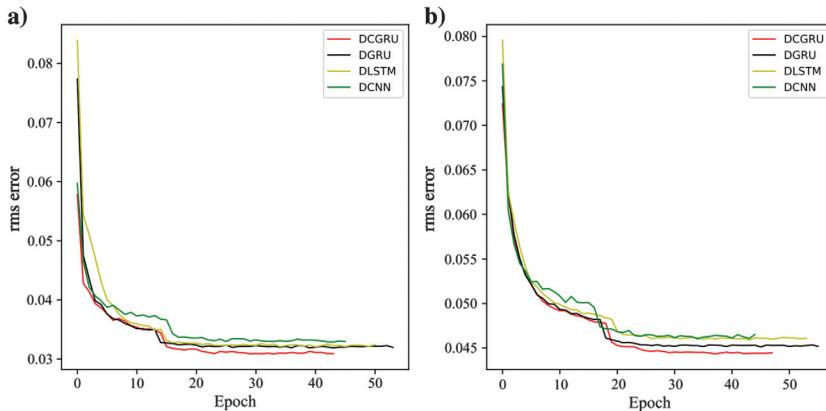


Figure 5. Training error curves of the DCGRU, DGRU, and DCNN models: (a) case I and (b) case II.

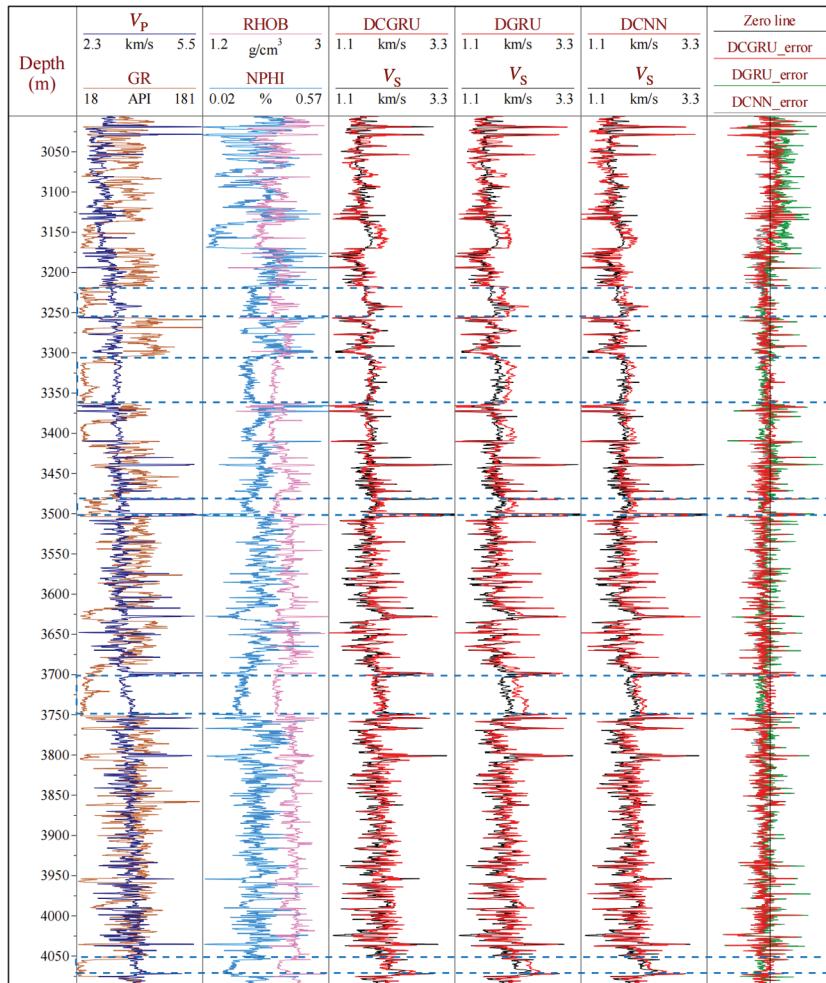


Figure 6. Prediction results of the DCGRU, DGRU, and DCNN models for case I.

NUMERICAL EXPERIMENTS

To verify the performance of the proposed method, the results of the DGRU, DLSTM, and DCNN methods and those of empirical equations (Castagna et al., 1993; Eskandari et al., 2004; Brocher, 2005) were compared with the output of the DCGRU. In particular, two cases were considered. In case I, training was performed using data from the Satyr-5 well and testing was done using data from the Callirhoe-1 well. In case II, training was performed using data from the 15-9-F-1A and 15-9-F-11T2 wells and testing was done using data from the 15-9-F-11A well. The computer configuration used in this experiment was i7-9750H processor, Windows 10 64-bit operating system, 2.6 GHz clock frequency, and 128 GB memory. We chose TensorFlow, a Python-based DL library to implement the models.

Case I

Satyr-5 and Callirhoe-1 are appraisal and exploration wells, respectively, located in the Northern Carnarvon Basin offshore of Western Australia. The well-log data include GR, NPHI, RHOB, and V_P measurements. In Callirhoe-1, log data at depths from 2690 to 4128 m were used. In Satyr-5, data at depths from 2100 to 4100 m were considered. In total, 7219 and 13,140 data points were available for the Callirhoe-1 and Satyr-5 observation wells, respectively. The detailed geology of the sites can be obtained from the literature (Jeong et al., 2020).

Because the data set of Satyr-5 is much larger than that of Callirhoe-1, we chose the former for training and the latter for testing. To ensure that the input data contained as much useful information as possible, we exploited the logs of all four parameters (GR, NPHI, RHOB, and V_P). The experimental results are shown in Figures 6 and 7, respectively. The rightmost column in these figures shows the relative error. The rms error, MAE, and MAPE values of the experimental results are shown in Tables 2 and 3.

Comparison of prediction performance of DL models

Table 2 shows that the prediction error of the DCGRU model is lower than that of the DGRU, DLSTM, and DCNN models. A lower error indicates higher prediction accuracy. In particular, the rms error, MAE, and MAPE are lower for the DCGRU model compared with those in the DGRU, DLSTM, and DCNN models, indicating that the DCGRU can effectively combine the local perception capability of a CNN and the long-term memory of the GRU. In other words, by taking the spatiotemporal features of the input

data into account, the DCGRU fusion network can achieve better prediction accuracy.

Figure 6 shows a comparison of the predictions of the DCGRU, DGRU, and DCNN models with the actual values. Specifically, it is shown in the dashed blue box in Callirhoe-1 that the values of the V_s have a step increase in this interval. The DGRU and DCNN fail to predict this step change. However, thanks to its ability to account for the trend of well-log data with depth, the local correlation information

of different log series, and the actual depth accumulation effect, the DCGRU can predict the step increase. Different log sequences and data at different depths (approximately 15–25 m) have different effects on the accuracy of the predicted values.

It can also be seen from the figure that the performance of the DCNN and DGRU networks is slightly worse than that of the DCGRU. This is a consequence of the fact that the former two are only sensitive to spatial or temporal characteristics. Therefore, it is difficult to obtain high-precision prediction results by only relying on a single spatial or temporal feature. The DCGRU overcomes this issue by combining the advantages of the CNN and GRU networks; thus, it can mine more comprehensive feature information from the log data and can achieve better fitting and generalization. These results further show that the DCGRU fusion network can effectively exploit the advantages of CNNs in spatial feature extraction as well as those of GRUs in temporal feature extraction.

To make a more comprehensive comparison of the prediction performance of the DCGRU model, the running times (the time required for model training and prediction) of each model are listed in Table 2. It can be seen that the DCNN network has the shortest running time. The LSTM introduces a gated memory unit, which increases the number of parameters to be trained and results in slower convergence and longer running times. The GRU represents an improvement of the LSTM. It retains the good learning performance of the LSTM, with fewer parameters and faster convergence. The running time of the DCGRU network is longer than that of the DCNN and shorter than that of the DGRU; however, as noted previously, it retains the advantages of the CNN and the GRU. It can be seen that the DCGRU fusion network can effectively identify correlations between well logging and V_s data, which not only improves the prediction accuracy of the model but also reduces its running time. By comprehensively comparing the prediction error and running time of each

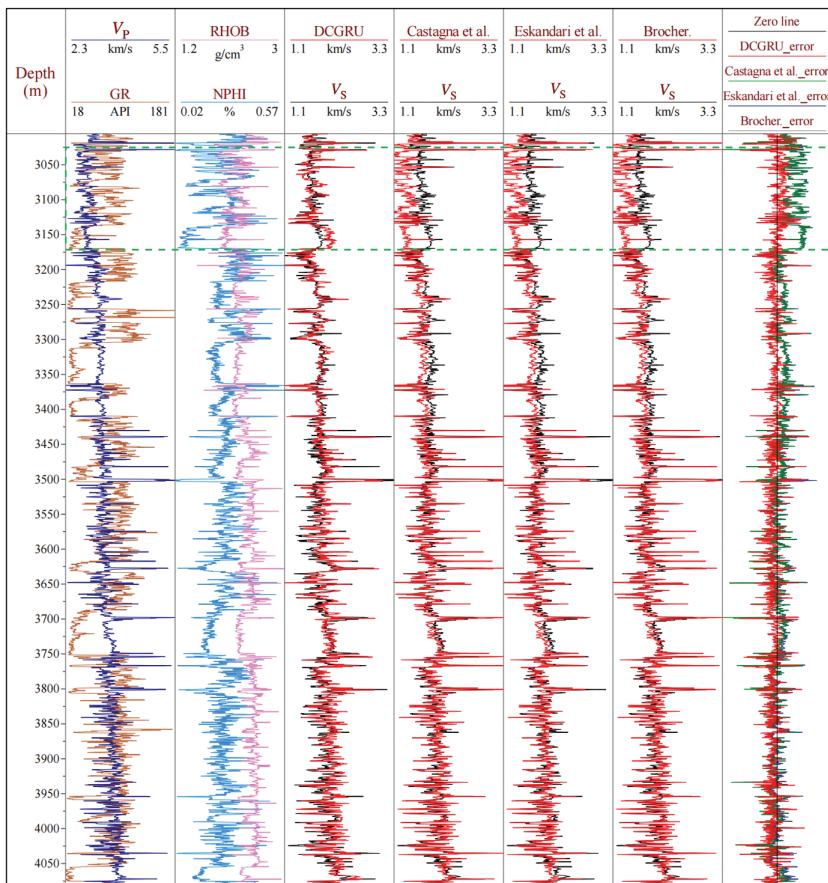


Figure 7. Prediction results of the DCGRU and three empirical equations for case I.

Table 2. Comparison of prediction results from different models (DCGRU, DGRU, DLSTM, and DCNN).

Well names	Models	Rms error	MAE	MAPE	Time (s)
Callirhoe-1	DCGRU	0.1498	0.1142	6.4325	119.18
	DGRU	0.1697	0.1310	7.3293	259.64
	DLSTM	0.1747	0.1316	7.3814	282.87
	DCNN	0.1626	0.1307	7.3537	48.47
15-9-F-11A	DCGRU	0.0852	0.0594	2.4603	153.79
	DGRU	0.087	0.0606	2.5044	376.68
	DLSTM	0.0876	0.0608	2.5085	494.12
	DCNN	0.0919	0.0625	2.5804	101.13

Table 3. Comparison of prediction results using different models (DCGRU and three empirical equations).

Well names	Models	Rms error	MAE	MAPE
Callirhoe-1	DCGRU	0.1498	0.1142	6.4325
	Castagna et al.	0.1986	0.1468	8.1801
	Eskandari et al.	0.1927	0.1427	7.9701
	Brocher	0.1952	0.1452	8.1356
15-9-F-11A	DCGRU	0.0852	0.0594	2.4603
	Castagna et al.	0.2830	0.2534	10.0723
	Eskandari et al.	0.1122	0.0776	3.0820
	Brocher	0.2206	0.2010	8.1098

Table 4. Common empirical equations for V_S estimation.

Model	Equation
Castagna et al.	$V_S = -0.05509*V_P^2 + 1.0168*V_P - 1.0305$
Eskandari et al.	$V_S = -0.1236*V_P^2 + 1.612*V_P - 2.0357$
Brocher	$V_S = 0.0064*V_P^4 - 0.1238*V_P^3 + 0.7949*V_P^2 - 1.2344*V_P + 0.7858$

Table 5. Correlation coefficients between well logs and the V_S .

Well names	CALI	V_P	GR	NPHI	PEF	RHOB	RT
15-9-F-1A	0.5570	0.9492	0.5858	0.8351	-0.3459	-0.5154	-0.7578
15-9-F-11T2	0.4603	0.9121	0.4924	0.8699	-0.1512	-0.7080	-0.8055
15-9-F-11A	0.2358	0.9120	0.6598	0.8552	-0.3700	-0.5048	-0.6952

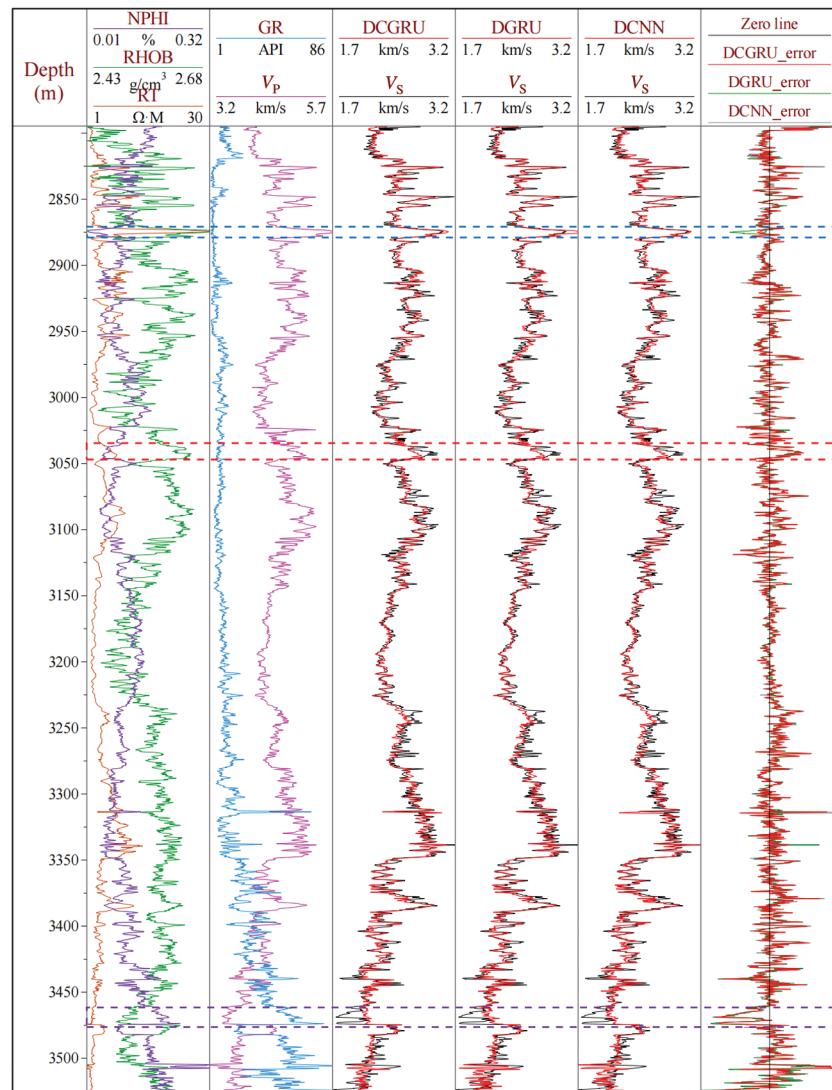


Figure 8. Prediction results of the DCGRU, DGRU, and DCNN models for case II.

model, it can be concluded that the proposed DCGRU model achieves a high performance-to-price ratio, thereby offering obvious advantages.

Comparison of the prediction performance of empirical equations

In current research, various empirical relations are used to predict V_S . Three classical and commonly used relations were developed by Castagna et al. (1993), Eskandari et al. (2004), and Brocher (2005), and they are shown in Table 4. The experimental results and errors resulting from the use of these relations are shown in Figure 7 and Table 3, respectively. Figure 7 shows that the predicted values of the DCGRU are the closest to the actual values. Especially in the dashed green box in Figure 7, the values predicted by the empirical equations are significantly lower than the actual values, and they exhibit trends that deviate from the actual one. Table 3 shows the obvious limitations of the empirical equations compared with the DCGRU. Indeed, even though they are usually calibrated with large amounts of data, empirical equations are not universally applicable. Conversely, the DL DCGRU model overcomes the limitations arising from the use of a linear regression and achieves a higher prediction accuracy.

Case II

To further verify the performance of the proposed method, we used logging data from the Volve field located in the southern part of the Norwegian North Sea (Equinor 2018), where oil accumulation was discovered in the Middle Jurassic Hugin sandstone formations. The detailed geology of the Norwegian continental shelf can be found in the literature (Brekke et al., 2001; Martinsen and Dreyer, 2001; Faleide et al., 2015). In this research area, 21 wells are available, but V_S data could not be obtained in most cases. We selected three wells (15-9-F-1A, 15-9-F-11T2, and 15-9-F-11A) for which V_S curves were available. In particular, data from 15-9-F-1A and 15-9-F-11T2 (15,818 sample points) were chosen as the training set, and data from 15-9-F-11A (7301 samples) were used for validation.

For the observation wells, the following logging data were available: the photoelectric absorption cross-section index (PEF), true formation resistivity (RT), borehole diameter (CAL1), V_P , GR, NPHI, and RHOB. Different input curves can have different impacts on the output data. To select the well-logging data with the best correlation with V_S , the PCC was used. The results are shown in Table 5, from which it can be concluded that V_S is primarily correlated with V_P , which explains why researchers have developed empirical correlations based on this

parameter. The correlation between V_S and NPHI, RHOB, GR, and RT was also higher than that exhibited by other logging data. Accordingly, we recommend using the V_P , NPHI, RHOB, GR, and RT logging data to train a DCGRU for estimating the V_S .

The model results are shown in Figures 8 and 9. The calculation errors are listed in Tables 2 and 3. It can be clearly seen from Figure 8 and Table 2 that the DCGRU model has a good predictive power. The curve fitting degree in Figure 8 shows that the DCGRU exhibits a higher correlation compared with DGRU and DCNN. In the region where V_S is the highest (the dashed blue box), the DCGRU offers the best prediction. However, in the dashed red box, the predicted values of the three models are lower than the actual values, and the predicted values of the three models are higher than the actual values in the dashed purple box. In other words, in these regions, DCGRU, DGRU, and DCNN all failed to accurately predict the actual trend of V_S .

DL models can effectively clarify the intrinsic relationships between logging data and V_S . This is especially true for the DCGRU fusion network. In general, if the DCGRU correctly partitions the reservoir depth features, it can yield a high-precision prediction even if the amount of data is small. However, if there is an error in the raw data, then the error in the DCGRU memory cell will obviously increase.

In terms of computational burden, as shown in Table 2, DCGRU requires a longer time than DCNN, but a much shorter time than DGRU. It can be concluded that the DCGRU offers a good compromise in terms of the cost-benefit ratio.

It can be seen from Figure 9 and Table 3 that, consistent with the results of case I, the performance of the empirical equations is not ideal. Among these equations, the model proposed by Eskandari et al. (2004) exhibits the lowest error, yet it underestimates the V_S values. In contrast, the models proposed by Castagna et al. (1993) and Brocher (2005) overestimate V_S considerably. This again confirms that the DCGRU model can offer the best prediction ability across the tested DL and empirical approaches.

DISCUSSION

A comparison of the prediction performance of the DL models showed that all four DL models yielded relatively good estimates, thereby demonstrating the ability of DL models in pattern identification. The prediction errors of DCGRU were lower and more stable than those of DGRU, DLSTM, and DCNN. Moreover, the DCGRU fusion network model could deal better with multi-point and multidimensional sparse feature data, and its prediction accuracy was the highest. With the same parameters, DCGRU achieves a lower prediction error because the information is stored and filtered by the internal structure. The prediction performance of the DCNN prediction model was higher than that of the DGRU model for case I, indicating that the spatial features extracted by the DCNN are better than those extracted by DGRU. However, the prediction performance

of the DGRU prediction model for case II was higher than that of the DCNN model, indicating that the temporal features extracted by DGRU are better than those extracted by CNN. This example shows that it is difficult to obtain high-precision prediction results by relying solely on a single spatial or temporal feature. A comparison of the prediction performance of empirical equations shows that the DCGRU fusion network is very powerful for estimating V_S . The DCGRU was able to evaluate the complex causal relationship between V_S and petrophysical parameters. The DCGRU can not only transmit reservoir features within the same layer but also correct the extracted features between layers. Using this method, V_S can be predicted from a series of input log data by considering variation trends and contextual information with depth. Although the prediction ability of DCGRU is not higher than that of DGRU and DCNN at all locations, DCGRU achieves better overall performance in evaluating the V_S trend with depth.

In summary, the results showed that the proposed DCGRU method can offer more accurate V_S prediction than other DL and empirical methods. The V_S prediction can be automated and implemented in real time using a trained prediction model, which can allow geologists to perform time-efficient and cost-effective field interpretation. As such, the proposed method can be an effective tool for improving the efficiency of geologic formation interpretation for field development.

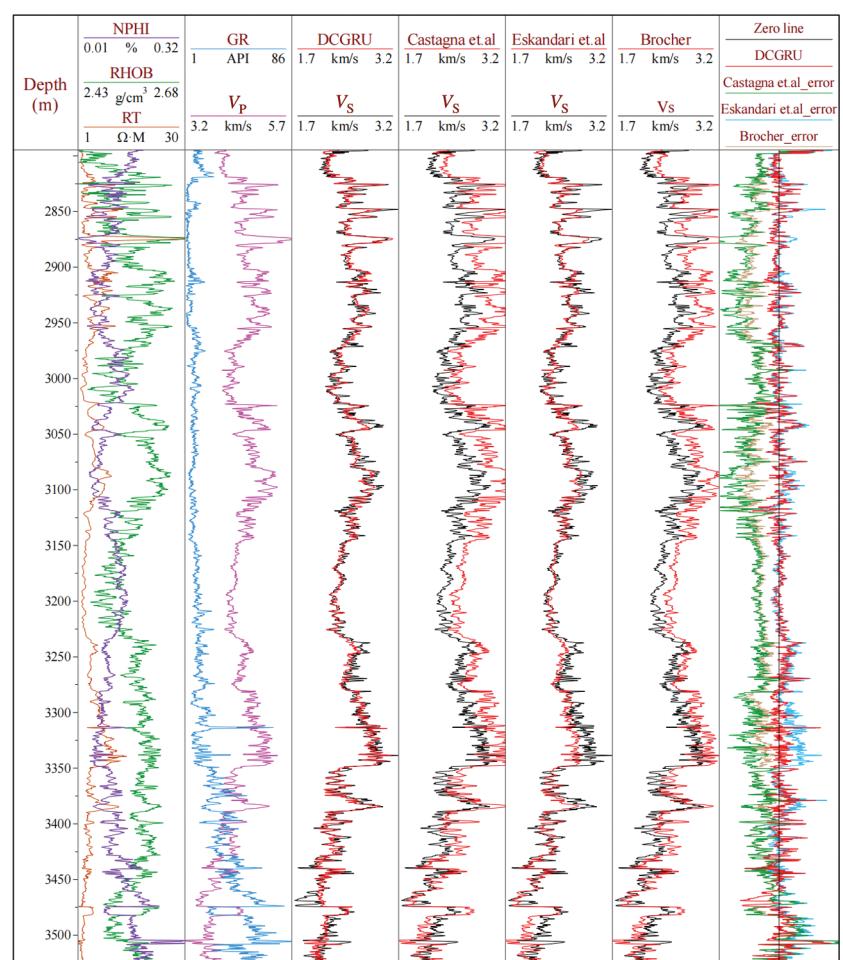


Figure 9. Prediction results of the DCGRU and three empirical equations for case II.

The scale of the data used in this study was relatively small. In future research, we expect to perform V_s prediction by accounting for more characteristics and including more data dimensions. In addition, we plan to perform further validations in more heterogeneous geologic settings to ensure the general applicability of the proposed method.

CONCLUSION

The S-wave velocity is one of the most important parameters for geophysical, geomechanical, and reservoir characterization. In this study, we proposed a novel approach — the DCGRU fusion network model — to predict V_s from logging data. With respect to the sedimentary continuity of the reservoir, logging data are considered as vertical sequences. The model takes the geologic trend and local correlation across logs into account, offering an improved ability to extract information on reservoir features. In fact, the DCGRU fusion network can correctly identify reservoir features in the depth direction.

To evaluate the proposed method, three empirical relations (Castagna, Eskandari, and Brocher) and three DL models (DGRU, LSTM, and DCNN) were implemented. Based on experimental results from two case studies, it was concluded that the DCGRU fusion network provides more accurate, efficient, reliable, and robust V_s predictions than other evaluated models. Furthermore, it offers an enhanced calculation speed and is less subjective compared with the empirical methods because it does not rely on expert judgement to ensure the accuracy of the prediction results. However, it is worth noting that the DCGRU method — similarly to other methods — cannot yield a completely reliable estimation of V_s . In addition, the effects of the geologic environment, well spacing, geologic events, and other factors on the prediction results have not been assessed in this study, and further in-depth and detailed research is still warranted.

ACKNOWLEDGMENTS

The authors would like to thank the editors and reviewers for their thorough reading and many constructive comments on this paper. This work was financially supported by the National Natural Science Foundation of China (grant nos. 42030812, 42042046, 41974160, and 41430323) and the Sinopec Science and Technology Department project (grant no. P20055-6).

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

REFERENCES

- Abdeljaber, O., O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, 2017, Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks: *Journal of Sound & Vibration*, **388**, 154–170, doi: [10.1016/j.jsv.2016.10.043](https://doi.org/10.1016/j.jsv.2016.10.043).
- Akhundi, H., M. Ghafoori, and G. R. Lashkaripour, 2014, Prediction of shear wave velocity using artificial neural network technique, multiple regression and petrophysical data: A case study in Asmari Reservoir (SW Iran): *Open Journal of Geology*, **4**, 303–313, doi: [10.4236/ojg.2014.47023](https://doi.org/10.4236/ojg.2014.47023).
- Anemangely, M., A. Ramezanzadeh, H. Amiri, and S. A. Hoseinpour, 2019, Machine learning technique for the prediction of shear wave velocity using petrophysical logs: *Journal of Petroleum Science and Engineering*, **174**, 306–327, doi: [10.1016/j.petrol.2018.11.032](https://doi.org/10.1016/j.petrol.2018.11.032).
- Aranibar, A., M. Saneifar, and Z. Heidari, 2013, Petrophysical rock typing in organic-rich source rocks using well logs: *Unconventional Resources Technology Conference*, 1154–1162.
- Bengio, Y., 2009, Learning deep architectures for AI: *Foundations & Trends in Machine Learning*, **2**, 1–127, doi: [10.1561/2200000006](https://doi.org/10.1561/2200000006).
- Bengio, Y., P. Simard, and P. Frasconi, 1994, Learning long-term dependencies with gradient descent is difficult: *IEEE Transactions on Neural Networks*, **5**, 157–166, doi: [10.1109/72.279181](https://doi.org/10.1109/72.279181).
- Brekke, H., H. I. Sjulstad, C. Magnus, and R. W. William, 2001, Sedimentary environments offshore Norway — An overview: *Norwegian Petroleum Society Special Publications*, **10**, 7–37.
- Brocher, T. M., 2005, Empirical relations between elastic wavespeeds and density in the earth's crust: *Bulletin of the Seismological Society of America*, **95**, 2081–2092, doi: [10.1785/0120050077](https://doi.org/10.1785/0120050077).
- Cant, D., 1994, Facies models: Response to sea level change: *Geological Journal*, **29**, 27–45.
- Carroll, R. D., 1969, The determination of the acoustic parameters of volcanic rocks from compressional velocity measurements: *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, **6**, 557–579, doi: [10.1016/0148-9062\(69\)90022-9](https://doi.org/10.1016/0148-9062(69)90022-9).
- Castagna, J. P., M. L. Batzle, and T. K. Kan, 1993, Rock physics: The link between rock properties and AVO response, Offset-dependent reflectivity — Theory and practice of AVO analysis: *SEG*, 135–171.
- Chen, W., L. Yang, B. Zha, M. Zhang, and Y. Chen, 2020, Deep learning reservoir porosity prediction based on multilayer long short-term memory network: *Geophysics*, **85**, no. 4, WA213–WA225, doi: [10.1190/geo2019-0261.1](https://doi.org/10.1190/geo2019-0261.1).
- Downton, J. E., 2005, Seismic parameter estimation from AVO inversion: *University of Calgary*.
- Dvorkin, J., M. Prasad, A. Sakai, and D. Lavoi, 1999, Elasticity of marine sediments: Rock physics modeling: *Geophysical Research Letters*, **26**, 1781–1784, doi: [10.1029/1999GL900332](https://doi.org/10.1029/1999GL900332).
- Equinor, 2018, Volve data village, <https://data.equinor.com/dataset/Volve>, accessed 18 September 2020.
- Eskandari, H., M. R. Rezaee, and M. Mohammadnia, 2004, Application of multiple regression and artificial neural network techniques to predict shear wave velocity from wireline log data for a carbonate reservoir, South-West Iran: *CSEG Recorder*, **42**, 40–48.
- Faleide, J. I., K. Bjørlykke, and R. H. Gabrielsen, 2015, Geology of the Norwegian continental shelf, in K. Bjørlykke, ed., *Petroleum geoscience*: Springer, 603–637, doi: [10.1007/978-3-642-34132-8_25](https://doi.org/10.1007/978-3-642-34132-8_25).
- Fattah, H., H. Nazar, and A. Monaghan, 2016, Hybrid ANFIS with ant colony optimization algorithm for prediction of V_s from a carbonate reservoir in Iran: *International Journal of Mining and Geo-Engineering*, **50**, 231–238.
- Feng, R., 2020, Lithofacies classification based on a hybrid system of artificial neural networks and hidden Markov models: *Geophysical Journal International*, **221**, 1484–1498, doi: [10.1093/gji/ggaa083](https://doi.org/10.1093/gji/ggaa083).
- Feng, R., T. M. Hansen, D. Grana, and N. Balling, 2020, An unsupervised deep-learning method for porosity estimation based on post-stack seismic data: *Geophysics*, **85**, no. 6, M97–M105, doi: [10.1190/geo2020-0121.1](https://doi.org/10.1190/geo2020-0121.1).
- Fu, C., N. Lin, D. Zhang, B. Wen, Q. Wei, and K. Zhang, 2018, Prediction of reservoirs using multi-component seismic data and the deep learning method: *Chinese Journal of Geophysics*, **61**, 293–303.
- Gal, D., J. Dvorkin, and A. Nur, 1998, A physical model for porosity reduction in sandstones: *Geophysics*, **63**, 454–459, doi: [10.1190/1.1444346](https://doi.org/10.1190/1.1444346).
- Graves, A., A. R. Mohamed, and G. Hinton, 2013, Speech recognition with deep recurrent neural networks: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6645–6649.
- Greenberg, M. L., and J. P. Castagna, 1992, Shear-wave velocity estimation in porous rocks: Theoretical formulation, preliminary verification and applications: *Geophysical Prospecting*, **40**, 195–209, doi: [10.1111/j.1365-2478.1992.tb00371.x](https://doi.org/10.1111/j.1365-2478.1992.tb00371.x).
- Han, D. H., A. Nur, and D. Morgan, 1986, Effects of porosity and clay content on wave velocities in sandstones: *Geophysics*, **51**, 2093–2107, doi: [10.1190/1.1442062](https://doi.org/10.1190/1.1442062).
- Hochreiter, S., and J. Schmidhuber, 1997, Long short-term memory: *Neural Computation*, **9**, 1735–1780, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- Hossain, Z., T. Mukerji, and I. L. Fabricius, 2012, Vp-Vs relationship and amplitude variation with offset modelling of glauconitic greensand: *Geophysical Prospecting*, **60**, 117–137, doi: [10.1111/j.1365-2478.2011.00968.x](https://doi.org/10.1111/j.1365-2478.2011.00968.x).
- Hrynaszkiewicz, I., 2010, A call for BMC research notes contributions promoting best practice in data standardization, sharing and publication: *BMC Research Notes*, **3**, 235–237, doi: [10.1186/1756-0500-3-235](https://doi.org/10.1186/1756-0500-3-235).
- Jeong, J., E. Park, I. Emelyanova, M. Pervukhina, L. Esteban, and S. T. Yun, 2020, Application of conditional generative model for sonic log estimation considering measurement uncertainty: *Journal of Petroleum Science and Engineering*, **196**, 108028, doi: [10.1016/j.petrol.2020.108028](https://doi.org/10.1016/j.petrol.2020.108028).

- Jørstad, A., T. Mukerji, and G. Mavko, 2001, Model-based shear-wave velocity estimation versus empirical regressions: Geophysical Prospecting, **47**, 785–797, doi: [10.1046/j.1365-2478.1999.00154.x](https://doi.org/10.1046/j.1365-2478.1999.00154.x).
- Kingma, D. P., and J. Ba, 2015, Adam: A method for stochastic optimization: Preprint at the arXiv: 1412.6980.
- Kiranyaz, S., O. Avci, O. Abdeljaber, T. Ince, and D. J. Inman, 2019, 1D convolutional neural networks and applications: A survey: Electrical Engineering and Systems Science, arXiv: 1905.03554.
- Krief, M., J. Garat, J. Stellingwerf, and J. Ventre, 1990, A petrophysical interpretation using the velocities of P and S waves (full-waveform sonic): Log Analyst, **31**, 355–369.
- LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–442, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- Lee, M. W., 2006, A simple method of predicting Vs: Geophysics, **71**, no. 6, F161–F164, doi: [10.1190/1.2357833](https://doi.org/10.1190/1.2357833).
- Li, D., and J. M. Harris, 2018, Full waveform inversion with nonlocal similarity and model-derivative domain adaptive sparsity-promoting regularization: Geophysical Journal International, **215**, 1841–1864, doi: [10.1093/gji/ggy380](https://doi.org/10.1093/gji/ggy380).
- Li, J., Y. Liu, C. Yin, X. Ren, and Y. Su, 2020, Fast imaging of time-domain airborne EM data using deep learning technology: Geophysics, **85**, no. 5, E163–E170, doi: [10.1190/geo2019-0015.1](https://doi.org/10.1190/geo2019-0015.1).
- Lin, N., D. Zhang, K. Zhang, S. Wang, C. Fu, J. Zhang, and C. Zhang, 2018, Predicting distribution of hydrocarbon reservoirs with seismic data based on learning of the small-sample convolution neural network: Chinese Journal of Geophysics, **61**, 4110–4125.
- Luo, S., P. Yang, G. Hu, and S. Liu, 2016, S-wave velocity prediction based on the modified P-L model and matrix equation iteration: Chinese Journal of Geophysics, **59**, 1839–1848.
- Mabrouk, W. M., and W. D. Pennington, 2009, Compressional and shear wave velocity in terms of petrophysical parameters in clean formations: Journal of Petroleum Science and Engineering, **65**, 62–66, doi: [10.1016/j.petrol.2008.12.024](https://doi.org/10.1016/j.petrol.2008.12.024).
- Maiti, S., R. K. Tiwari, and H. J. Kümpel, 2007, Neural network modelling and classification of lithofacies using well log data: A case study from KTB borehole site: Geophysical Journal International, **169**, 733–746, doi: [10.1111/j.1365-246X.2007.03342.x](https://doi.org/10.1111/j.1365-246X.2007.03342.x).
- Maleki, S., A. Moradzadeh, R. G. Rabii, R. Gholami, and F. Sadeghzadeh, 2014, Prediction of Vs using empirical correlations and artificial intelligence methods: NRIAG Journal of Astronomy and Geophysics, **3**, 70–81, doi: [10.1016/j.nrjag.2014.05.001](https://doi.org/10.1016/j.nrjag.2014.05.001).
- Martinsen, O. J., and T. Dreyer, 2001, Sedimentary environments offshore Norway — Palaeozoic to recent: An introduction, in O. J. Martinsen and T. Dreyer, eds., Sedimentary environments offshore Norway — Palaeozoic to recent: Volume 10: Norwegian Petroleum Society Special Publications, Elsevier, 1–5.
- Mergini, B., H. Izadi, and H. Memarian, 2019, Shear wave velocity prediction using Elman artificial neural network: Carbonates and Evaporites, **34**, 1281–1291.
- Meshalkin, Y., A. Shakirov, E. Popov, D. Koroteev, and I. Gurbatova, 2020, Robust well-log based determination of rock thermal conductivity through machine learning: Geophysical Journal International, **222**, 978–988, doi: [10.1093/gji/ggaa209](https://doi.org/10.1093/gji/ggaa209).
- Miller, S. L. M., and R. R. Stewart, 1990, Effects of lithology, porosity and shaliness on P- and S-wave velocities from sonic logs: Canadian Journal of Exploration Geophysics, **26**, 94–103.
- Nebauer, C., 1998, Evaluation of convolutional neural networks for visual recognition: IEEE Transactions on Neural Networks, **9**, 685–696, doi: [10.1109/72.701181](https://doi.org/10.1109/72.701181).
- Ojha, M., and K. Sain, 2014, Velocity-porosity and velocity-density relationship for shallow sediments in the Kerala-Konkan basin of western Indian margin: Journal of the Geological Society of India, **84**, 187–191, doi: [10.1007/s12594-014-0122-2](https://doi.org/10.1007/s12594-014-0122-2).
- Olayiwola, T., and O. A. Sanuade, 2021, A data-driven approach to predict compressional and shear wave velocities in reservoir rocks: Petroleum, **7**, 199–208, doi: [10.1016/j.petlm.2020.07.008](https://doi.org/10.1016/j.petlm.2020.07.008).
- Oloruntobia, O., D. Onalo, S. Adedigbe, L. James, R. Chunduru, and S. Butt, 2019, Data-driven shear wave velocity prediction model for siliciclastic rocks: Journal of Petroleum Science and Engineering, **183**, 106293.
- Pasa, L., A. Testolin, and A. Sperduti, 2015, Neural networks for sequential data: A pre-training approach based on hidden Markov models: Neurocomputing, **169**, 323–333, doi: [10.1016/j.neucom.2014.11.081](https://doi.org/10.1016/j.neucom.2014.11.081).
- Pearson, K., 1896, Mathematical contributions to the theory of evolution — 3: Regression, heredity, and panmixia: Philosophical Transactions of the Royal Society of London, **187**, 253–318.
- Pham, N., X. Wu, and E. Z. Naeini, 2020, Missing well log prediction using convolutional long short-term memory network: Geophysics, **85**, no. 4, WA159–WA171, doi: [10.1190/geo2019-0282.1](https://doi.org/10.1190/geo2019-0282.1).
- Pride, S. R., J. G. Berryman, and J. M. Harris, 2004, Seismic attenuation due to wave-induced flow: Journal of Geophysical Research, Solid Earth, **109**, 247–278, doi: [10.1029/2003JB002639](https://doi.org/10.1029/2003JB002639).
- Rajabi, M., B. Bohlolli, and E. G. Ahangar, 2010, Intelligent approaches for prediction of compressional, shear and Stoneley wave velocities from conventional well log data: A case study from the Sarvak carbonate reservoir in the Abadan Plain (Southwestern Iran): Computers & Geosciences, **36**, 647–664, doi: [10.1016/j.cageo.2009.09.008](https://doi.org/10.1016/j.cageo.2009.09.008).
- Ray, A., and D. Myer, 2019, Bayesian geophysical inversion with trans-dimensional Gaussian process machine learning: Geophysical Journal International, **217**, 1706–1726, doi: [10.1093/gji/ggz111](https://doi.org/10.1093/gji/ggz111).
- Rezaee, M. R., A. Kadkhodaie-Ilkchhi, and A. Barabasi, 2007, Prediction of Vs from petrophysical data utilizing intelligent systems: An example from a sandstone reservoir of Carnarvon Basin, Australia: Journal of Petroleum Science and Engineering, **55**, 201–212, doi: [10.1016/j.petrol.2006.08.008](https://doi.org/10.1016/j.petrol.2006.08.008).
- Schuster, M., and K. K. Paliwal, 1997, Bidirectional recurrent neural networks: IEEE Transactions on Signal Processing, **45**, 2673–2681, doi: [10.1109/78.650093](https://doi.org/10.1109/78.650093).
- Shi, X., Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, 2015, Convolutional LSTM network: A machine learning approach for precipitation nowcasting: Advances in Neural Information Processing Systems, Curran Associates, Inc., 802–810.
- Smith, G. C., and M. Gidlow, 2000, A comparison of the fluid factor with λ and μ in AVO analysis: 70th Annual International Meeting, SEG, Expanded Abstracts, 122–125, doi: [10.1190/1.1815615](https://doi.org/10.1190/1.1815615).
- Song, S., J. Hou, L. Dou, Z. Song, and S. Sun, 2020, Geologist-level wireline log shape identification with recurrent neural networks: Computers & Geosciences, **134**, 104313, doi: [10.1016/j.cageo.2019.104313](https://doi.org/10.1016/j.cageo.2019.104313).
- Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 2014, Dropout: A simple way to prevent neural networks from overfitting: Journal of Machine Learning Research, **15**, 1929–1958.
- Tan, M., X. Peng, H. Cao, H. Wang, and Y. Yuan, 2015, Estimation of shear wave velocity from wireline logs in gas-bearing shale: Journal of Petroleum Science and Engineering, **133**, 352–366, doi: [10.1016/j.petrol.2015.05.020](https://doi.org/10.1016/j.petrol.2015.05.020).
- Vernik, L., J. Castagna, and S. J. Movie, 2017, Shear-wave velocity prediction in unconventional shale reservoirs: Geophysics, **83**, no. 1, MR35–MR45, doi: [10.1190/geo2017-0349.1](https://doi.org/10.1190/geo2017-0349.1).
- Wang, J., J. Cao, and S. Yuan, 2020a, Shear wave velocity prediction based on adaptive particle swarm optimization optimized recurrent neural network: Journal of Petroleum Science and Engineering, **194**, 107466, doi: [10.1016/j.petrol.2020.107466](https://doi.org/10.1016/j.petrol.2020.107466).
- Wang, K., Q. Huang, and W. Wu, 2020b, Application of long short-term memory neural network in geoelectric field data processing: Chinese Journal of Geophysics, **63**, 3015–3024.
- Wang, Y., W. Liao, and Y. Chang, 2018, Gated recurrent unit network-based short-term photovoltaic forecasting: Energies, **11**, 2163, doi: [10.3390/en11082163](https://doi.org/10.3390/en11082163).
- Wu, X., L. Liang, Y. Shi, Z. Geng, and S. Fomel, 2019, Multitask learning for local seismic image processing: Fault detection, structure-oriented smoothing with edge-preserving, and seismic normal estimation by using a single convolutional neural network: Geophysical Journal International, **219**, 2097–2109, doi: [10.1093/gji/ggj418](https://doi.org/10.1093/gji/ggj418).
- Xi, X., and J. Huang, 2020, Location and imaging of scatterers in seismic migration profiles based on convolution neural network: Chinese Journal of Geophysics, **63**, 687–714.
- Xiong, W., X. Ji, and Y. Ma, 2018, Seismic fault detection with convolutional neural network: Geophysics, **83**, no. 5, O97–O103, doi: [10.1190/geo2017-0666.1](https://doi.org/10.1190/geo2017-0666.1).
- Xu, J., M. Tan, X. Wang, and C. Wu, 2017, Predicting acoustic-wave velocities and fluid sensitivity to elastic properties in fractured carbonate formation: Interpretation, **5**, no. 1, SB69–SB80, doi: [10.1190/INT-2016-0067.1](https://doi.org/10.1190/INT-2016-0067.1).
- Xu, S., and M. A. Payne, 2009, Modeling elastic properties in carbonate rocks: The Leading Edge, **28**, 66–74, doi: [10.1190/1.3064148](https://doi.org/10.1190/1.3064148).
- Xu, S., and R. E. White, 1995, A new velocity model for clay-sand mixture: Geophysical Prospecting, **43**, 91–118, doi: [10.1111/j.1365-2478.1995.tb00126.x](https://doi.org/10.1111/j.1365-2478.1995.tb00126.x).
- Xu, S. Y., and R. E. White, 1996, A physical model for Vs predicting: Geophysical Prospecting, **44**, 687–717, doi: [10.1111/j.1365-2478.1996.tb00170.x](https://doi.org/10.1111/j.1365-2478.1996.tb00170.x).
- Yuan, S., J. Liu, and S. Wang, 2018, Seismic waveform classification and first-break picking using convolution neural networks: IEEE Geoscience and Remote Sensing Letters, **15**, 272–276, doi: [10.1109/LGRS.2017.2785834](https://doi.org/10.1109/LGRS.2017.2785834).
- Zhang, D., Y. Chen, and J. Meng, 2018a, Synthetic well logs generation via recurrent neural networks: Petroleum Exploration and Development, **45**, 598–607.
- Zhang, G., Z. Wang, and Y. Chen, 2018b, Deep learning for seismic lithology prediction: Geophysical Journal International, **215**, 1368–1387, doi: [10.1093/gji/ggy344](https://doi.org/10.1093/gji/ggy344).
- Zhang, Y., H. Zhong, Z. Wu, H. Zhou, and Q. Ma, 2020, Improvement of petrophysical workflow for Vs prediction based on machine learning methods for complex carbonate reservoirs: Journal of Petroleum Science and Engineering, **192**, 107234, doi: [10.1016/j.petrol.2020.107234](https://doi.org/10.1016/j.petrol.2020.107234).

- Zhao, J., and L. Itti, 2018, shapeDTW: Shape dynamic time warping: Pattern Recognition, **74**, 171–184, doi: [10.1016/j.patcog.2017.09.020](https://doi.org/10.1016/j.patcog.2017.09.020).
- Zhao, J., X. Mao, and L. Chen, 2019a, Speech emotion recognition using deep 1D & 2D CNN LSTM networks: Biomedical Signal Processing and Control, **47**, 312–323, doi: [10.1016/j.bspc.2018.08.035](https://doi.org/10.1016/j.bspc.2018.08.035).
- Zhao, M., S. Chen, and Y. Dave, 2019b, Waveform classification and seismic recognition by convolution neural network: Chinese Journal of Geophysics, **62**, 374–382.
- Zhong, Z., T. C. Carr, X. Wu, and G. Wang, 2019, Application of a convolutional neural network in permeability prediction: A case study in the Jacksontburg-Stringtown oil field, West Virginia, USA: Geophysics, **84**, no. 6, B363–B373, doi: [10.1190/geo2018-0588.1](https://doi.org/10.1190/geo2018-0588.1).
- Zhu, H., C. Xu, P. Li, and C. Liu, 2020a, Structure analysis of shale and prediction of shear wave velocity based on petrophysical model and neural network: Global Geology, **23**, 155–165.
- Zhu, K., L. Wang, Y. Du, and C. Jiang, 2020b, DeepLog: Identify tight gas reservoir using multi-log signals by a fully convolutional network: IEEE Geoscience and Remote Sensing Letters, **17**, 568–571, doi: [10.1109/LGRS.2019.2930587](https://doi.org/10.1109/LGRS.2019.2930587).

Biographies and photographs of the authors are not available.