Petroleum Science 20 (2023) 3712–3722

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Bottom hole pressure prediction based on hybrid neural networks and Bayesian optimization

Chengkai Zhang [a, b, c], Rui Zhang [b], Zhaopeng Zhu [a, c], Xianzhi Song [a, b, c, *], Yinao Su [d], Gensheng Li [a, b, c], Liang Han [a, c]

[a] College of Petroleum Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
[b] College of Artificial Intelligence, China University of Petroleum (Beijing), Beijing, 102249, China
[c] National Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
[d] CNPC Engineering Technology R&D Company Limited, Beijing, 102206, China

ARTICLE INFO

Article history:
Received 24 October 2022
Received in revised form
29 March 2023
Accepted 10 July 2023
Available online 21 July 2023

Edited by Jia-Jia Fei

Keywords:
Bottom hole pressure
Spatial-temporal information
Improved GRU
Hybrid neural networks
Bayesian optimization

ABSTRACT

Many scholars have focused on applying machine learning models in bottom hole pressure (BHP) prediction. However, the complex and uncertain conditions in deep wells make it difficult to capture spatial and temporal correlations of measurement while drilling (MWD) data with traditional intelligent models. In this work, we develop a novel hybrid neural network, which integrates the Convolution Neural Network (CNN) and the Gate Recurrent Unit (GRU) for predicting BHP fluctuations more accurately. The CNN structure is used to analyze spatial local dependency patterns and the GRU structure is used to discover depth variation trends of MWD data. To further improve the prediction accuracy, we explore two types of GRU-based structure: skip-GRU and attention-GRU, which can capture more long-term potential periodic correlation in drilling data. Then, the different model structures tuned by the Bayesian optimization (BO) algorithm are compared and analyzed. Results indicate that the hybrid models can extract spatial-temporal information of data effectively and predict more accurately than random forests, extreme gradient boosting, back propagation neural network, CNN and GRU. The CNN-attention-GRU model with BO algorithm shows great superiority in prediction accuracy and robustness due to the hybrid network structure and attention mechanism, having the lowest mean absolute percentage error of 0.025%. This study provides a reference for solving the problem of extracting spatial and temporal characteristics and guidance for managed pressure drilling in complex formations.

1. Introduction

Accurately predicting the bottom hole pressure (BHP) is critical for managed pressure drilling (MPD), which is an essential technology for safe and efficient drilling. Real-time BHP calculation can significantly reduce incidents such as wellbore influx, wellbore instability, well control issues, and uncontrolled flow (He and Yun, 2018; Sami, 2021). Traditional BHP estimation methods include mechanistic and empirical models and downhole sensor acquisition, but they have limitations (Marfo et al., 2021; Nait Amar and Zeraibi, 2019). The empirical models (Aziz and Govier, 1972; Duns and Ros, 1963; Hagedorn and Brown, 1965; Orkiszewski, 1967) and mechanistic models (Ansari et al., 1994; Chokshi et al., 1996; Gomez et al., 2000) require assumptions based on certain correction factors and empirical coefficients, and the iterative solution time is long, making it difficult to achieve the necessary accuracy and real-time efficiency required in drilling operations. Downhole sensors are costly and face the challenge of failure under extreme high-temperature and high-pressure conditions, leading to the collection of invalid data from deep wells.

The oil and gas industry has recently seen a surge in the application of intelligent technologies, including machine learning algorithms, to solve complex nonlinear problems. Artificial neural networks (ANNs) have been extensively used to improve the accuracy of bottom hole pressure (BHP) prediction, with various researchers proposing ANN models with different structures (Chen et al., 2017; Jahanandish et al., 2011; Mask et al., 2019;

* Corresponding author. College of Petroleum Engineering, China University of Petroleum (Beijing), Beijing, 102249, China.
E-mail address: songxz@cup.edu.cn (X. Song).

https://doi.org/10.1016/j.petsci.2023.07.009

Mohammadpoor et al., 2010; Okoro et al., 2021; Sami and Ibrahim, 2021; Spesivtsev et al., 2018). Al Shehri et al. (2020) explored the use of functional neural networks (FNN) and long short-term memory (LSTM) in addition to ANN for BHP prediction. Marfo et al. (2021) developed M5 prime decision tree method to predict flowing bottom hole pressure and compared it with other basic models to demonstrate its superior performance. Zhu et al. (2022) proposed a hybrid parallel network model that combines convolution neural network (CNN) and LSTM (CNN-LSTM) as well as backpropagation neural network (BPNN) and LSTM (BPNN-LSTM) to extract temporal features in MPD data and enhance prediction accuracy. However, complex data-driven models often have a large number of parameters, and manual experimentation is time-consuming and may not achieve the global optimal solutions. Therefore, researchers have integrated optimization algorithms into the model to further improve the computational efficiency and accuracy of the model. Nait Amar et al. (2018) demonstrated the superiority of the hybridization ANN and grey wolves optimization (ANN-GWO) over other hybridizations such as genetic algorithm (GA) and particle swarms optimization (PSO) or the BPNN alone. Nait Amar and Zeraibi (2019) proposed a hybrid model based on the combination of support vector regression and firefly optimization algorithm (SVR-FFA), which realizes the prediction of BHP in vertical wells of multiphase flow. The superior robustness and accuracy of SVR-FFA is demonstrated through a comparison with SVR-GA. Tariq et al. (2019) and Liang et al. (2020) indicated the use of hybrid models such as PSO-ANN and GA-BPNN, which can perform better in terms of prediction accuracy and convergence speed.

MPD data contains both the local spatial information between features and contextual information related to depth variation. Although the above models have shown improvement in prediction accuracy, in terms of extracting information, they only focus on point-to-point mapping information or variation information in a single dimension (e.g., temporal or spatial) and may not be able to capture spatial-temporal information simultaneously, leading to the loss of critical information.

In this paper, we propose a hybrid neural network based on the CNN and Gate Recurrent Unit (GRU). The CNN is utilized to extract the spatial local dependence features of drilling data, while the GRU is used to capture the context information of the data changing with depth. Additionally, considering the long-term potential periodic characteristics of MWD data, improved GRU structures with skip connections and attention mechanism are proposed, respectively. Finally, to obtain the optimal prediction model and further improve the prediction accuracy of the BHP, the Bayesian optimization algorithm is used to tune the hyper-parameters of the model. The main contribution of this paper can be further defined as follows:

(1) We propose a novel CNN-GRU hybrid neural network, which combines the strengths of traditional intelligent models and effectively captures spatial local correlation and depth-dependent contextual information.
(2) We enhance the GRU component by adding skip connections and an attention mechanism, which improves the model's ability to extract long-term potential periodic information.
(3) To improve the prediction accuracy of BHP and computational efficiency, this paper adopts the Bayesian Optimization (BO) algorithm to automatically tune model hyper-parameters and avoid time-consuming manual experiments.
(4) The prediction results of the proposed model and the existing basic model are compared and analyzed, and its superiority is validated.

The remainder of this paper is organized as follows. Section 2 illustrates related approach for processing data, the structure of the proposed models and the principle of BO algorithm. Section 3 presents comparison and results, which compares the performance of three different model structures, the impact of skip periods on the prediction results and the superiority of the proposed model over other basic models. We end with some conclusions and expectation in Section 4.

2. Methodology

Fig. 1 illustrates the workflow of this paper for predicting BHP using the proposed model. The workflow comprises four main steps: data preprocessing, model establishment, tuning, and application. The details of each step will be discussed in the next sections.

2.1. Data preprocessing

2.1.1. Data composition and description

The original dataset used in this study is the 100,000 sets of MWD data from the surface monitoring in the Tarim Oilfield located in Xinjiang, China. The dataset includes 70 distinct features, such as logging data, measurement while drilling data, and drilling fluid data, with a maximum measured depth of 6705 m and a maximum vertical depth of 4942 m. To ensure the quality of the data and mitigate the impact of environmental noise, the 3 σ-rule (Lehmann, 2013) was applied to eliminate fluctuating BHP anomalies, which is a simple and widely utilized method for detecting outliers. Following data cleaning, a total of 80,000 data points were available for training, validation, and testing the model. The characteristics of the cleaned BHP are presented in Table 1.

2.1.2. Feature engineering

To explore the relationship between the characteristic variable and the target variable BHP, reduce the noise effect of irrelevant variables on BHP prediction, and enhance the robustness of the model. In this part, we use the distance correlation coefficient (Székely and Bakirov, 2007) to measure the correlation, which can describe the linear and nonlinear relationship between variables at the same time. The calculation formula is defined as follows:

$$dCor_n^2(X,Y)=\begin{cases}\dfrac{dCov_n^2(X,Y)}{\sqrt{dCov_n^2(X)dCov_n^2(Y)}} & dCov_n^2(X)dCov_n^2(Y)>0\\ 0\; dCov_n^2(X)dCov_n^2(Y)=0\end{cases} \tag{1}$$

where $dCor_n^2(X,Y)$ is the distance correlation coefficient of X and Y, $dCov_n^2(X,Y)$ is the distance covariance of X and Y, $dCov_n^2(X)$ is the standard deviation of the distance of X, and $dCov_n^2(Y)$ is the standard deviation of the distance of Y.

The correlation coefficient indicates the strength of the relationship between two variables. A larger coefficient implies a stronger correlation, while a coefficient of 0 indicates independence. Table 2 depicts the medium-strong correlations ($\geq$0.5) between features and their distance correlation coefficients. These correlations align with physical drilling laws, such as the strong correlation between BHP and well depth and true vertical depth, in line with the momentum conservation equation in the annular multiphase flow model. The response mechanism between BHP and engineering parameters is also evident, such as BHP increasing with the back-pressure pump flow rate. Meanwhile, considering that the rheological parameters of drilling fluid play a major part in predicting BHP (Zhu et al., 2022), some stable features in the field

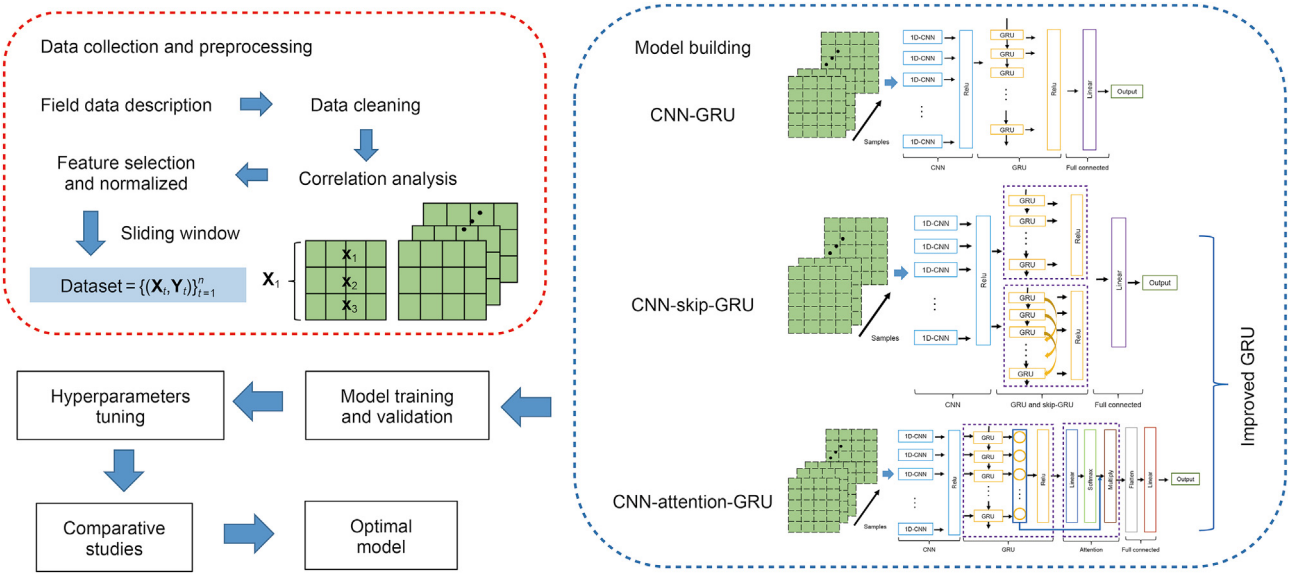

Fig. 1. The overall workflow of this study.

Table 1
Data description of BHP.

Evaluation standard	Value
Count	80,000
Mean, MPa	56.61
Standard deviation, MPa	0.40
Min, MPa	55.37
Max, MPa	57.58
First quartile, MPa	56.26
Second quartile, MPa	56.70
Third quartile, MPa	56.70

are difficult to reveal the intrinsic relation by statistical methods alone. For instance, BHP will increase with the drilling fluid density and viscosity. Consequently, this paper selectively supplements the performance parameters of drilling fluid, including funnel viscosity, sand content and drilling fluid density. Finally, 12 characteristic parameters are selected as the input features, as shown in Table 3, and their statistical description is presented in Table 4.

Data normalization is a common practice in machine learning that scales the data to a range of [0, 1]. This improves the efficiency of the gradient descent algorithm and prevents numerical issues during training. Before training the model in this study, all input parameters are normalized, as shown in the following equation:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

where $\tilde{x}$ is the normalized value of x, $x_{\min}$ and $x_{\max}$ are the maximum and minimum values of the variable x.

After comparing the effects of different dataset division methods, the experimental dataset is divided into training, validation, and test sets in a 7:1:2 ratio, with 70% of the data used for model training, 10% of the data is used as the validation set to obtain the optimal hyper-parameters, and the remaining 20% for evaluating the model.

Table 3
Input features of prediction of BHP.

Number	Features	Parameter category
1	Well depth	Others
2	Rotary speed	Engineering
3	Riser pressure	Engineering, Hydraulics
4	Inlet flow rate	Engineering
5	Outlet flow rate	Engineering
6	True vertical depth	Engineering
7	Back-pressure pump flow rate	Engineering
8	Outlet density	Drilling fluid
9	Total pool volume	Engineering
10	Drilling fluid density	Drilling fluid
11	Sand content	Drilling fluid
12	Funnel viscosity	Drilling fluid

2.1.3. Sample construction

To extract sequence information, the sliding window method (Bao et al., 2020) is employed, which involves moving a fixed-sized window over the MWD data sequence. Fig. 2 illustrates the sliding window process, where the input window gradually shifts downward over measured depth to construct the dataset for the model training. Window size is one of the hyper-parameters that needs to be optimized, and it determines the length of the historical MWD data sequence used as input for the model.

2.2. Model building

In the analysis of measurement while drilling (MWD) data, depth is commonly used as a proxy for time, treating the data as a

Table 2
The distance correlation coefficient of features and BHP.

Input parameters	Well depth	Rotary speed	Riser pressure	Inlet flow rate	Outlet flow rate	True vertical depth	Back-pressure pump flow rate	Outlet density	Total pool volume
Coefficient	0.64	0.51	0.50	0.52	0.51	0.62	0.57	0.52	0.51

Table 4
The statistical description of input parameters.

	Well depth, m	Rotary speed, rpm	Riser pressure, MPa	Inlet flow rate, L/s	Outlet flow rate, L/s	True vertical depth, m
Mean	6087.99	11.91	19.37	14.06	14.06	4935.39
Min	5832.82	5.26	17.24	13.14	13.50	4930.15
Max	6703.78	23.60	23.23	15.47	15.57	4942.07

	Back-pressure pump flow rate, L/s	Outlet density, g/cm^3	Total pool volume, m^3	Drilling fluid density, g/cm^3	Funnel viscosity, s	Sand content, %
Mean	8.35	1.19	146.44	1.1	44	0.2
Min	6.78	1.15	106.68	1.1	43	0.2
Max	10.78	1.26	175.22	1.2	45	0.2

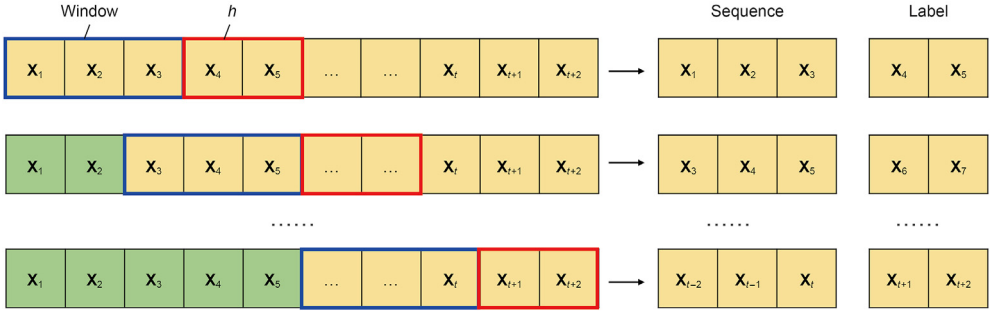

Fig. 2. Sliding window scheme for training. Where $\mathbf{X}_t$ is the input MWD data sequence in depth t, the label is predicted BHP, the blue color box is the input sequence samples, and the red color box is the corresponding predicted label. Figure takes the window size is 3, which means the input set is 0.3 m and predicted outcome is 0.2 m.

time series that varies with depth. Consequently, predicting BHP can be transformed into a time series forecasting problem, which typically involves a combination of short-term and long-term change patterns. Recurrent neural networks (RNNs) have been intensively studied in energy prediction due to their ability to capture the long-term dependencies from the historical information (Asala et al., 2019; Huang et al., 2022; Lu et al., 2021; Wei et al., 2019). Two commonly used RNN variants are LSTM (Hochreiter and Schmidhuber, 1997) and GRU (Chung et al., 2014). In this study, we opted for GRU as it is computationally cheaper and performs similarly to LSTM (Group, 2017). On the other hand, CNN models have shown powerful performance in identifying sudden changes at various levels of granularity in the data (Kanwal et al., 2022; Ye et al., 2022). However, they are not suitable for the long-term memory tasks since they operate only on the current frame for forecasting.

2.2.1. CNN-GRU

In this section, the hybrid model CNN-GRU is proposed to enhance the accuracy prediction of BHP. The CNN component is used to uncover the internal connections and reduce the scale and complexity of MWD data. Meanwhile, the temporal memory capacity of the GRU neural network is employed to learn the internal dynamic change patterns of the data. The hybrid neural network CNN-GRU comprises the CNN component, the GRU component and the full connected component, as depicted in Fig. 3. The modeling process can be divided into the following parts:

(1) The CNN component is a convolutional network (Bezdan and Bacanin, 2019) without a pooling layer, which is designed to extract local feature information and dependencies to a greater extent. The output is then mapped non-linearly using the Relu activation function. The convolutional layer operation flow of CNN is illustrated in Fig. 4. The output results of the CNN component can be obtained in this step:

$$\mathbf{x}_k = \mathrm{Relu}(\mathbf{W}_k * \mathbf{x} + \mathbf{b}_k) \tag{3}$$

where $\mathbf{x}$ is the input vector, $\mathbf{W}_k$ is the weight matrice, * is the convolution operation and $\mathbf{x}_k$ is the outputs, and Relu is an activation function.

(2) After the feature extraction by the CNN component, the extracted features are transmitted to the GRU which includes two gate structures, namely update gate and reset gate. The GRU selectively memorizes and forgets the information passed to seek pre and post correlation between BHP and other parameters. A single GRU and its detailed schematic are shown in Fig. 5. The input vector $\mathbf{x}_t$ of the GRU at time t is given by:

$$\mathbf{r}_t = \sigma(\mathbf{x}_t\mathbf{W}_{xr} + \mathbf{h}_{t-1}\mathbf{W}_{hr} + \mathbf{b}_r) \tag{4}$$

$$\mathbf{u}_t = \sigma(\mathbf{x}_t\mathbf{W}_{xu} + \mathbf{h}_{t-1}\mathbf{W}_{hu} + \mathbf{b}_u) \tag{5}$$

$$\mathbf{c}_t = \mathrm{Relu}(\mathbf{x}_t\mathbf{W}_{xc} + \mathbf{r}_t \odot (\mathbf{h}_{t-1}\mathbf{W}_{hc}) + \mathbf{b}_c) \tag{6}$$

$$\mathbf{h}_t = (1 - \mathbf{u}_t) \odot \mathbf{h}_{t-1} + \mathbf{u}_t \odot \mathbf{c}_t \tag{7}$$

where σ is the sigmoid function, $\mathbf{x}_t$ is the input at time t, $\mathbf{W}_{xr}$, $\mathbf{W}_{hr}$, $\mathbf{W}_{xu}$, $\mathbf{W}_{xc}$ are the weight matrices, $\mathbf{b}_r$, $\mathbf{b}_u$, $\mathbf{b}_c$ are offsets, $\mathbf{h}_{t-1}$ is the output at the previous time, $\mathbf{h}_t$ is the current output, $\mathbf{r}_t$ is the output of the reset gate neuron, $\mathbf{u}_t$ is the output of the update gate, $\mathbf{c}_t$ is the hidden state calculated based on reset gate, and $\odot$ is the element-wise product.

(3) Following the GRU component, the data flow into the full connected component, which is a full connected layer that integrates the output features to obtain the final BHP prediction results. The output of the full connected is calculated as:

$$\hat{\mathbf{Y}}_t = \mathbf{W}^T\mathbf{h}_t + \mathbf{b} \tag{8}$$

where $\hat{\mathbf{Y}}_t$ is the model's prediction results at time t.

2.2.2. Improved GRU components

However, GRU often loses some crucial information in this process due to the gradient vanishing or exploding, which makes it difficult to discover periodic signals. To overcome this issue and

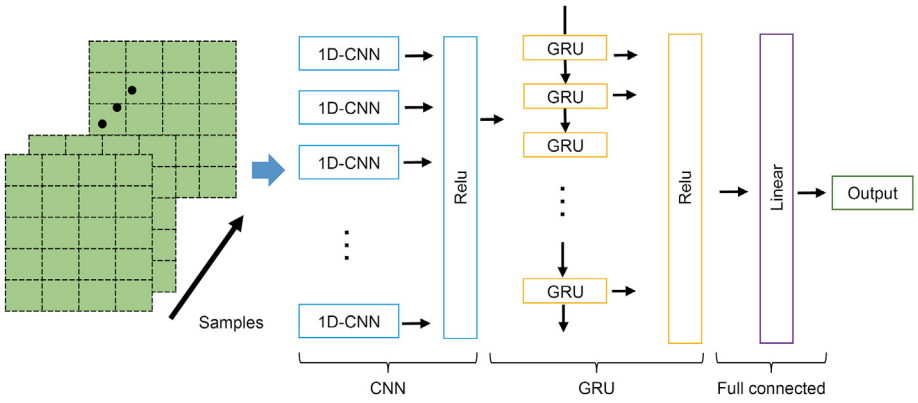

Fig. 3. Structure of the hybrid neural network CNN-GRU.

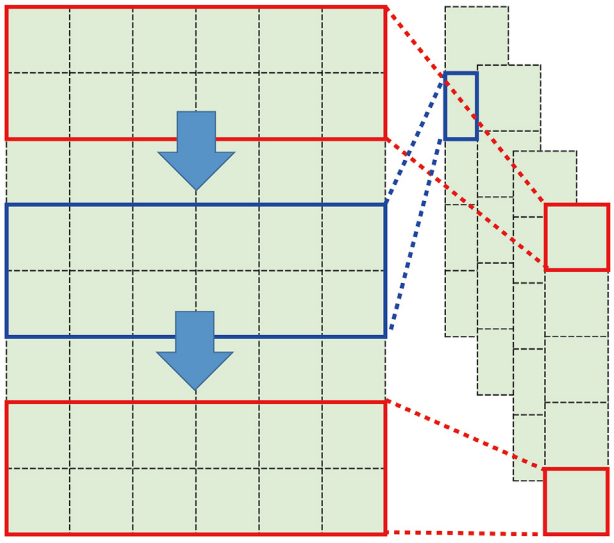

Fig. 4. The operation flow of convolution layer for one sample.

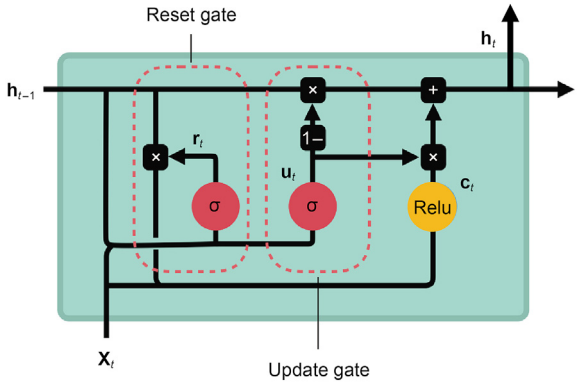

Fig. 5. Structure of a single GRU.

effectively use the periodic pattern in MWD data, we propose a novel skip-GRU component. The overall CNN-skip-GRU architecture is shown in Fig. 6, which utilizes skip connections to connect the hidden cells with a certain interval to transmit periodic signals effectively. The calculation updating process of the skip connection can be computed as:

$$\mathbf{r}_t = \sigma(\mathbf{x}_t\mathbf{W}_{xr} + \mathbf{h}_{t-s}\mathbf{W}_{hr} + \mathbf{b}_r) \tag{9}$$

$$\mathbf{u}_t = \sigma(\mathbf{x}_t\mathbf{W}_{xu} + \mathbf{h}_{t-s}\mathbf{W}_{hu} + \mathbf{b}_u) \tag{10}$$

$$\mathbf{c}_t = \mathrm{Relu}(\mathbf{x}_t\mathbf{W}_{xc} + \mathbf{r}_t \odot (\mathbf{h}_{t-s}\mathbf{W}_{hc}) + \mathbf{b}_c) \tag{11}$$

$$\mathbf{h}_t = (1 - \mathbf{u}_t) \odot \mathbf{h}_{t-s} + \mathbf{u}_t \odot \mathbf{c}_t \tag{12}$$

where s is the number of hidden cells skipped and has to be defined manually. In this part, based on human experience and pressure measurement data while drilling, we initially define s to 10.

However, the skip connection needs to manually define the period interval of the skip, which is not applicable to the sequence whose period length is dynamic over time. In order to further enhance the accuracy of the model and better adapt to the bottom hole pressure prediction task, an attention mechanism (Bahdanau et al., 2014) is proposed in this part. The attention mechanism learns the weighted combination at each hidden window of the input matrix. Fig. 7 illustrates the architecture of the attention mechanism. Finally, the output of attention layer can be calculated as follows:

$$\boldsymbol{\alpha}_t = \mathrm{Softmax}(f(\mathbf{H}_t, \mathbf{h}_{t-1})) \tag{13}$$

$$\widehat{\mathbf{Y}}_t = \mathbf{W}[\boldsymbol{\alpha}_t\mathbf{H}_t; \mathbf{h}_{t-1}] + \mathbf{b} \tag{14}$$

where $\mathbf{H}_t = [\mathbf{h}_{t-w}, \mathbf{h}_{t-w+1}, \ldots, \mathbf{h}_{t-1}]$ is a matrix stacking the hidden state of GRU column-wisely, w is the window size, f is a scoring function cosine, $\boldsymbol{\alpha}_t$ is the attention weights at time t, and Softmax is an activation function.

2.2.3. Parameter tuning using Bayesian optimization

The process of tuning hyper-parameters is crucial for developing an optimal model. In the proposed hybrid model, there are several hyper-parameters required to be pre-defined, such as window size, number of epochs, number of neurons in CNN and number of neurons in GRU, etc. Traditional population optimization algorithms like GA and PSO require enough sample points for initialization, resulting in low optimization efficiency (Asante-Okyere et al., 2022; Otaki et al., 2022). Grid search, random search, and Bayesian optimization are other common methods for hyper-parameter tuning (Snoek et al., 2012). Grid search compares the performance of all the hyperparameter permutation and combination but computationally expensive, while random search may not yield satisfactory results due to its randomness. Compared with these methods, BO algorithm does not easily fall into a local

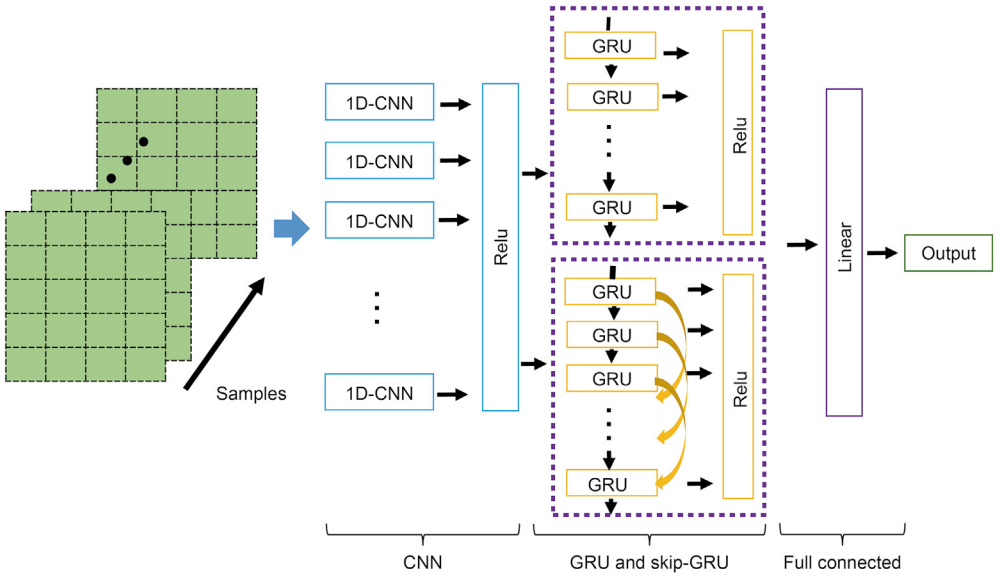

Fig. 6. Structure of the hybrid neural network CNN-skip-GRU.

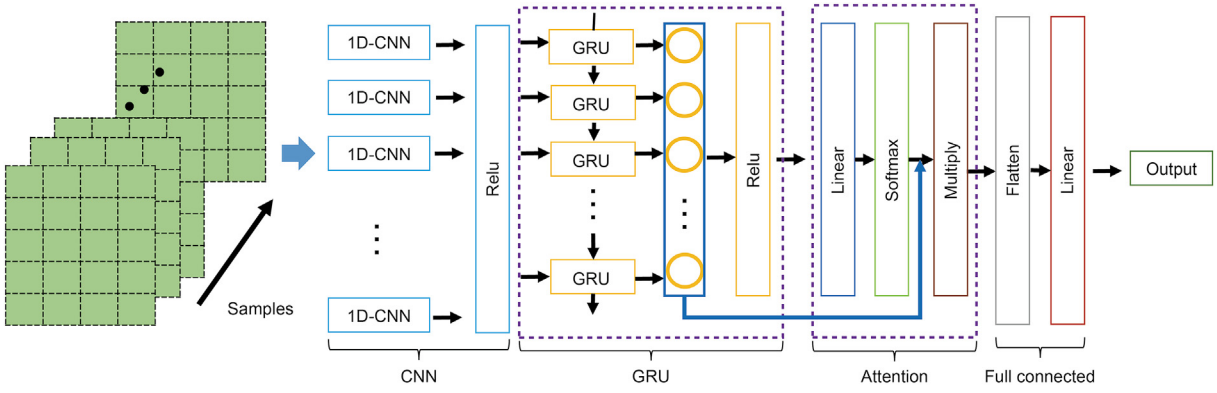

Fig. 7. Structure of the hybrid neural network CNN-attention-GRU.

solution, and it is a reliable faster way to machine learning for hyperparameter tuning when the objective functions are expensive to evaluate or potentially intractable (Bao et al., 2020; Frazier, 2018).

BO involves four components: objective function, domain space, optimization algorithm and historical data of the results (Snoek et al., 2012). In this study, the objective function is to minimize the mean absolute percentage error (MAPE) of training and validation set, namely the objective function f of the BO is:

$$f = \min(\text{training } MAPE + \text{validation } MAPE) \tag{15}$$

In the domain space part, we define different spaces according to different hyper-parameters, such as window size, number of neurons in CNN and GRU networks, and number of epochs. The core part of BO is the selection of search algorithm, and in this paper we use Tree-structured Parzen Estimator (TPE) (Bergstra et al., 2011). TPE sequentially constructs models to approximate the performance of hyper-parameters based on historical measurements H, and subsequently selects new hyper-parameters to test based on this model. Finally, the optimal solution can be obtained by iterative update as shown in Table 5 (Frazier, 2018).

2.2.4. Evaluation criteria

To assess and compared the predictive performances of the proposed models, three performance evaluation metrics are selected, which are MAPE, root mean square error (RMSE) and mean absolute error (MAE). MAPE represents the average absolute prediction error of m samples. RMSE is the square root of the mean squared errors of m samples. MAE represents the average absolute prediction error of m samples. They are mathematically defined as:

$$\text{MAPE} = \frac{1}{m}\sum_{i=1}^{m}\frac{\left|y_i - y_{\text{pre}}\right|}{y_i} \times 100 \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - y_{\text{pre}}\right)^2} \tag{17}$$

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}\left|y_i - y_{\text{pre}}\right| \tag{18}$$

where y_i is the observed BHP values and y_{pre} is the calculated BHP values which is predicted by the developed models.

Table 5
The BO for hyper-parameters tuning.

1 Initialize the probabilistic surrogate with n0 points, $\mathbf{H}_0 = \{(x_1, \mathrm{f}_1), (x_2, \mathrm{f}_2), \cdots, (x_{n_0}, \mathrm{f}_{n_0})\}$. Where x is the different hyper-parameter combination.
2 for $n = 0, 1, \ldots, N$:
3 Select new query point x_{n+1} by optimizing acquisition function AC
$x_{n+1} = \mathrm{argmax}\ \mathrm{AC}(x; \mathbf{H}_n)$
4 Evaluate $\mathrm{f}(x_{n+1})$
5 Augment data $\mathbf{H}_{n+1} = \mathbf{H}_n \cup \{(x_{n+1}, \mathrm{f}(x_{n+1}))\}$
6 Update the probabilistic model

3. Results and discussion

3.1. GRU structure

For predicting bottom hole pressure, different model structures exhibit varying prediction effects due to their different abilities to extract and learn feature information. Firstly, we define the ranges of the model parameters based on manual experiments and the problem scale. In order to minimize the impact of experimental randomness, we conduct multiple sets of comparative experiments to ensure consistency among the parameters of the three models. Moreover, to mitigate overfitting, the dropout probability is set to 0.2 and early stopping strategy is utilized in model validation stage which monitors whether the validation set error continues to increase. The parameters of window size and number of epochs are set as 30 and 40, respectively. Table 6 compares the test results of different GRU structures. It is observed that the proposed CNN-skip-GRU and CNN-attention-GRU models exhibit superior performance. With an increase in the number of neurons, the prediction accuracy slightly declines, indicating that more complex models learn less meaningful information when the models are trained on a consistent amount of data. Among them, the MAPE of CNN-skip-GRU is lower than CNN-GRU. CNN-Attention-GRU shows the best performance on the test set, with a MAPE decrease to 0.029%. Hence, we choose CNN-attention-GRU as the baseline model.

3.2. CNN-attention-GRU with BO

Based on the above experiment, it is evident that different hyperparameter settings can result in different results. Therefore, it is necessary to optimize the hyperparameters and select a group of optimal hyperparameters for enhancing model performance. In this section, we highlight the benefits of employing BO to find optimal model. As shown in Table 7, different parameter ranges are defined for BO. Then, by setting the number of trials for BO as 100, the performance of the CNN-attention-GRU model on the validation set is compared and the optimal model is retained in time for predicting the final test set. Fig. 8 illustrates the errors and results of conducting 100 hyperparameter search experiments. The optimal hyperparameters found by the BO algorithm are a window size of 45, 60 epochs, 55 CNN neurons, and 70 GRU neurons, with a minimum MAPE of 0.025%. Using the BO algorithm instead of manual experience, we effectively reduce the original MAPE to 0.025% and further improve the BHP prediction accuracy.

Table 7
Range of hyper-parameter values for BO.

Parameters	Value range and step
Window size	(30, 50, 5)
Number of epochs	(50, 80, 5)
Number of neurons in CNN	(50, 80, 5)
Number of neurons in GRU	(50, 100, 5)

In previous tests, the training data are considered to be noise-free. However, in practice, field data may contain noise from sensors and other equipment, which can affect the stability of the predicted model. Therefore, it is important to evaluate the performance of the model when predicting with noisy data. In this section, training data with noise ratio of 10% and 20% are added to each input features respectively, the evaluation metrics of the model are shown in Table 8. For instance, we added noise to the real input drilling fluid density while keeping the other input features as ground truth to demonstrate the robustness of the model when drilling fluid density is noisy. Additionally, an extreme scenario is considered where all the input features are mixed with noise.

Table 8 demonstrates that the performance of proposed model is almost unaffected by the noise. Compared to other input features, the evaluation metrices of drilling fluid density and back-pressure pump flow rate showed obvious changes when noise was added, but the model still gave reliable predictions. The MAPE of drilling fluid density increases from 0.025% (clean data) to 0.069% (with 10% noise) and 0.088% (with 20% noise). This may be caused by a lack of diversity in the collected drilling fluid density data. The drilling fluid density is usually slightly changed in the field for drilling safety, resulting in a narrow range of collected field data. For train set, the mean value of drilling fluid density is 1.1 and the standard deviation closes to 0.1, which means that most of the drilling fluid density changes in the scope of 1.0–1.2. Meanwhile, the prediction model is more sensitive to the fluctuation of BHP in the test set, and the MAPE increases to 0.072% (with 10% noise) and 0.092% (with 20% noise), indicating that the back-pressure pump flow rate as the most direct and efficient BHP control parameter. In fact, the BHP

Table 6
Comparison of test results with different model parameters. The optimal parameters and results are highlighted in bold font.

Model	Neurons in CNN	Neurons in GRU	MAPE, %	MAE, MPa	RMSE, MPa
CNN-GRU	50	50	0.040	0.037	0.092
CNN-GRU	60	60	0.039	0.035	0.077
CNN-GRU	70	70	0.041	0.040	0.087
CNN-skip-GRU	50	50	0.033	0.024	0.062
CNN-skip-GRU	60	60	0.035	0.026	0.085
CNN-skip-GRU	70	70	0.035	0.026	0.075
CNN-attention-GRU	50	50	0.030	0.019	0.049
CNN-attention-GRU	**60**	**60**	**0.029**	**0.018**	**0.045**
CNN-attention-GRU	70	70	0.031	0.020	0.047

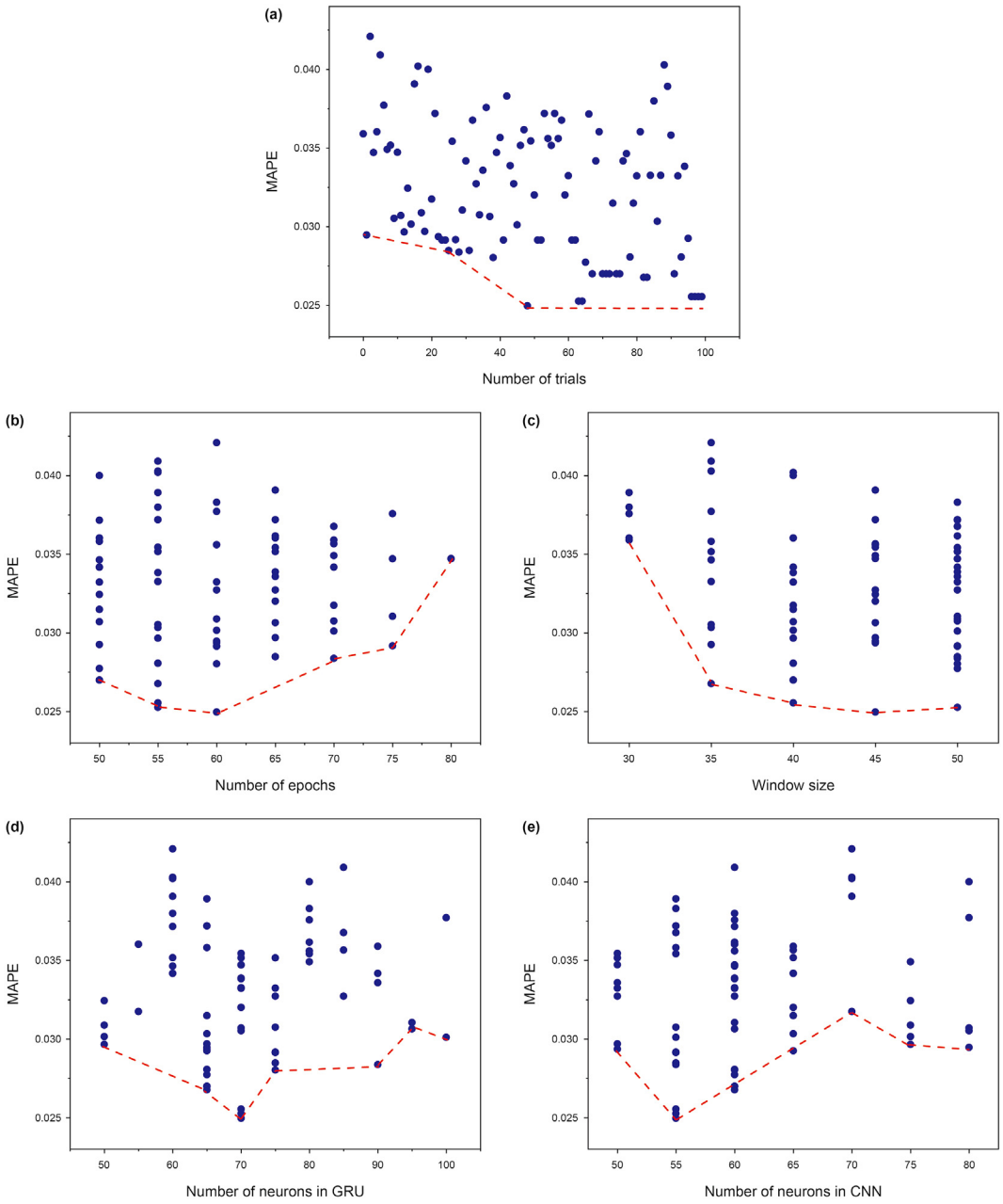

Fig. 8. The MAPE of the different parameters of CNN-GRU with 100 trials. (**a**) The MAPE of different trials. (**b**) The MAPE is low when number of epochs is concentrated at 50–60. (**c**) The MAPE is low when window size is concentrated at 40–50. (**d**) The MAPE is low when number of neurons in GRU is concentrated at 65–75. (**e**) The MAPE is low when number of neurons in CNN is concentrated at 55–60.

fluctuation of the test data is caused by the real-time adjustment of the back-pressure pump flow during MPD process. Finally, when all the features mixed with noise, it is found that the model is still powerful with a maximum MAPE of 0.229%. Therefore, the proposed CNN-attention-GRU model with BO algorithm is robust when the field data are affected by noise and is selected as the baseline model.

3.3. Impact of different skip intervals

In this section, we aim to select the optimal parameter skip in the CNN-skip-GRU model, analyzing the effect of different skip intervals, and comparing the performance of model. According to the evaluation metrics of Fig. 9, it shows that when the skip interval is set to 10, the model achieves the best performance. As the skip interval continues to increase, the performance of the model significantly degrades. The results indicate that the longer skip periods cause more loss of local information for MWD data, and the proposed CNN-skip-GRU model effectively captures the short-term autocorrelation of MWD data, making it critical to finding the suitable skip interval for accurately predicting BHP.

3.4. Comparison of basic and proposed model

To verify the reliability and superiority of the proposed hybrid models, the performance of the CNN-attention-GRU with BO model is compared with Random Forest (RF) (Breiman, 2001), Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), BPNN

Table 8
Model performance with adding noisy data.

Parameters	Noisy ratio = 10%			Noisy ratio = 20%		
	MAPE, %	MAE, MPa	RMSE, MPa	MAPE, %	MAE, MPa	RMSE, MPa
Well depth	0.033	0.018	0.078	0.038	0.021	0.091
Rotary speed	0.040	0.018	0.088	0.042	0.037	0.062
Standpipe pressure	0.039	0.022	0.086	0.039	0.022	0.086
Inlet flow rate	0.040	0.022	0.093	0.040	0.022	0.093
Sand content	0.043	0.024	0.093	0.038	0.021	0.090
Outlet flow rate	0.037	0.021	0.089	0.037	0.021	0.089
True vertical depth	0.039	0.020	0.089	0.039	0.020	0.089
Total pool volume	0.036	0.019	0.084	0.036	0.019	0.084
Back-pressure pump flow rate	0.072	0.062	0.120	0.092	0.073	0.132
Outlet density	0.041	0.023	0.093	0.041	0.023	0.093
Drilling fluid density	0.069	0.056	0.113	0.088	0.064	0.124
All features	0.153	0.086	0.146	0.229	0.128	0.164

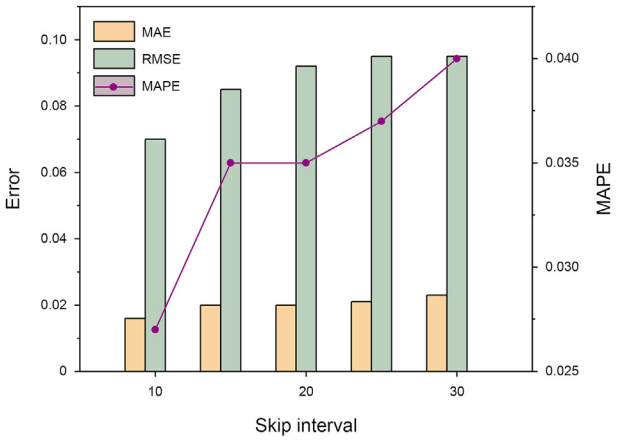

Fig. 9. Comparison of errors indicators with different skip intervals.

(Sheu and Choi, 1995), CNN, GRU, CNN-GRU, CNN-skip-GRU which are the most commonly applied intelligent models. The BO algorithm is also implemented in the hyperparameter tuning process of models.

Table 9 presents the prediction results of different models, showing that the proposed hybrid models outperform other basic models. In order to compares the prediction results more instinctively, Fig. 10(a)–(h) are provided, which depict the prediction results of different models. Fig. 10 clearly shows that traditional machine learning models such as RF and XGBoost fail to catch temporal relations, resulting in a significant gap between the real and predicted values. It demonstrates that traditional machine learning models show inadequate ability for time series prediction issue. BPNN in Fig. 10(c) also shows a certain deviation from the real value in the first 5000 points, indicating that it learns information with uniformity and hardly considers the correlation of temporal sequence. It shows great performance to match the point-to-point mapping. Although CNN in Fig. 10(d) has a high overall prediction accuracy, there are some abnormal fluctuations at the pressure sudden change. It indicates that CNN is sensitive to capturing local spatial fluctuations, and fits the pressure changing trends rapidly while it cannot follow the long-term pattern. In contrast, GRU in Fig. 10(e) has the ability to memorize long-term patterns, but fails to accurately predict the local maximum pressure variation point. Fig. 10(e) depicts that GRU shows strong capability to capture the correlation of long-term information and grasp global information accurately.

Table 9
Comparison of basic and proposed model. The optimal model and results are highlighted in bold font.

Model	MAPE, %	MAE, MPa	RMSE, MPa
RF	0.384	0.214	0.272
XGBoost	0.369	0.206	0.270
BPNN	0.148	0.087	0.113
CNN	0.042	0.037	0.062
GRU	0.062	0.048	0.066
CNN-GRU	0.039	0.033	0.050
CNN-skip-GRU	0.027	0.016	0.070
CNN-attention-GRU	**0.025**	**0.015**	**0.038**

The CNN-GRU model in Fig. 10(f) combines the respective strengths of CNN and GRU, and there is a slight improvement at the maximum pressure change, while they still have some gaps. It proves the ability of the proposed model to capture short-term and long-term patterns simultaneously, which the traditional prediction model is not equipped. With considering skip component, the CNN-skip-GRU model in Fig. 10(g) shows a significant improvement in responsiveness, indicating that the skip connection does help to alleviate the loss of information transmission in GRU subnetwork, and then prove the spatial and temporal characteristics of MWD data. It demonstrates that the temporal skip connections can extend the temporal span of the information flow, so that BHP at different depths help to interact in adjacent period. With considering attention mechanism, the CNN-attention-GRU model in Fig. 10(h) further improves the performance on basis of CNN-GRU and ponders both stability and rapid response at the inflection point, suggesting that the attention mechanism helps in focusing on global feature information. It also implies that attention mechanism is more suitable for handling sequences with dynamic periodic changes, and it is crucial in the prediction task of the field MWD data.

4. Conclusions

In this study, a novel hybrid model is proposed which is more robust, comprehensive, and accurate to deal with complex situations in the field. The proposed model can significantly improve the prediction accuracy of BHP and better response to real-time downhole fluctuations. At the same time, BO algorithm is used to tune the hyperparameters of all models, which can obtain optimal models for different structures faster and more reliable than manual experiments. The following conclusions can be drawn:

(1) The proposed hybrid neural networks show better performance than other basic neural networks in BHP fluctuations prediction. CNN-GRU combines the advantages of a single basic model and effectively capture the spatial and

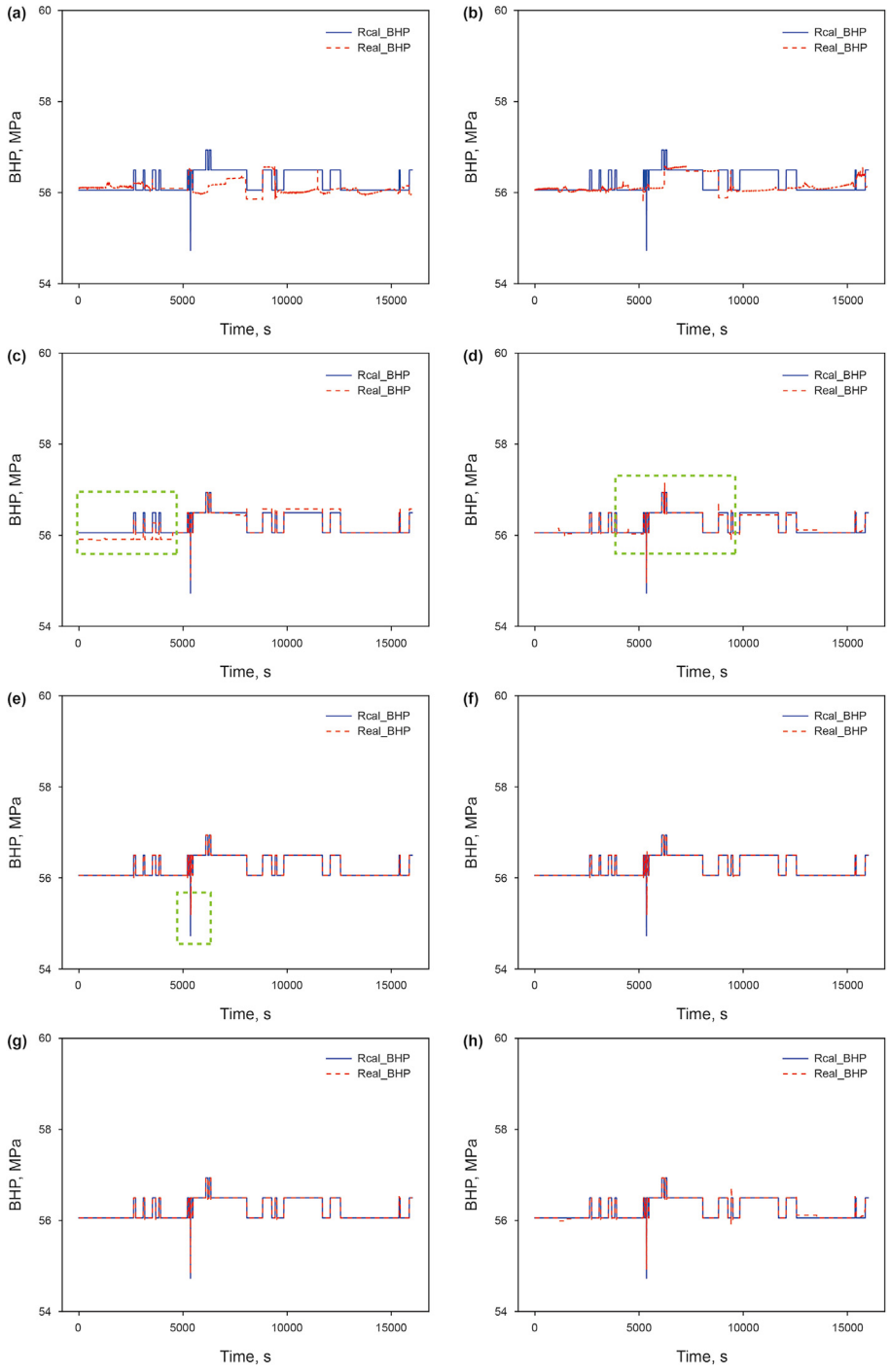

Fig. 10. Comparison of test results of different models. **(a)** XGBoost. **(b)** RF. **(c)** BPNN. **(d)** CNN. **(e)** GRU. **(f)** CNN-GRU. **(g)** CNN-attention-GRU. **(h)** CNN-skip-GRU.

contextual information of the MWD data. The GRU component with added skip connections can extract longer-term well depth information, and the selection of skip intervals plays a key role in the prediction results. The GRU component with the attention mechanism can automatically capture the presence of uncertain periodic signals.

(2) Compared with other models, CNN-attention-GRU achieves the highest prediction accuracy, and its MAPE, MAE, and RMSE are better than other models. The model also shows great robustness with 10% and 20% noise of input data.

(3) The MAPE of CNN-attention-GRU using the BO algorithm is decreased by 7.4% compared to the original, and MAE and RMSE is reduced by 6.2% and 9.5%, respectively.

This research can be applied to optimally solve multivariate time series problem, and provide a reference for precise pressure control in drilling. In the future, interpretable methods will be explored to capture the physical meaning behind the MWD data and further improve the transferability of the model for application in different new wells.

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.

Acknowledgements

The authors express their appreciation to National Key Research and Development Project “Key Scientific Issues of Revolutionary Technology” (2019YFA0708300), Strategic Cooperation Technology Projects of CNPC and CUPB (ZLZX2020-03), Distinguished Young Foundation of National Natural Science Foundation of China (52125401) and Science Foundation of China University of Petroleum, Beijing (2462022SZBH002).

References

Al Shehri, F.H., Gryzlov, A., Al, Tayyar, et al., 2020. Utilizing machine learning methods to estimate flowing bottom-hole pressure in unconventional gas condensate tight sand fractured wells in Saudi Arabia. In: SPE Russian Petroleum Technology Conference, Moscow, Russia.

Ansari, A.M., Sylvester, N.D., Sarica, C., et al., 1994. A comprehensive mechanistic model for upward two-phase flow in wellbores. SPE Prod. Facil. 9 (2), 143–151. https://doi.org/10.2118/20630-PA.

Asala, H.I., Chebeir, J.A., Manee, V., et al., 2019. An integrated machine-learning approach to shale-gas supply-chain optimization and refrac candidate identification. SPE Reservoir Eval. Eng. 22 (4), 1201–1224. https://doi.org/10.2118/187361-PA.

Asante-Okyere, S., Shen, C., Osei, H., 2022. Enhanced machine learning tree classifiers for lithology identification using Bayesian optimization. Applied Computing and Geosciences 16, 100100. https://doi.org/10.1016/j.acags.2022.100100.

Aziz, K., Govier, G.W., 1972. Pressure drop in wells producing oil and gas. J. Can. Petrol. Technol. 11 (3). PETSOC-72-03-04. https://doi.org/10.2118/72-03-04.

Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. Computer Science abs/1409, 0473. https://doi.org/10.48550/arXiv.1409.0473.

Bao, A., Gildin, E., Huang, J., et al., 2020. Data-driven end-to-end production prediction of oil reservoirs by EnKF-Enhanced recurrent neural networks. In: SPE Latin American and Caribbean Petroleum Engineering Conference. Bogota, Colombla.

Bergstra, J., Bardenet, R., Kégl, B., et al., 2011. Algorithms for Hyper-Parameter Optimization.

Bezdan, T., Bacanin, N., 2019. Convolutional Neural Network Layers and Architectures.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

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.

Chen, W., Di, Q., Ye, F., Zhang, J., et al., 2017. Flowing bottomhole pressure prediction for gas wells based on support vector machine and random samples selection. Int. J. Hydrogen Energy 42 (29), 18333–18342. https://doi.org/10.1016/j.ijhydene.2017.04.134.

Chokshi, R.N., Schmidt, Z., Doty, D.R., 1996. Experimental study and the development of a mechanistic model for two-phase flow through vertical tubing. In: SPE Western Regional Meeting. Anchorage, Alaska.

Chung, J., Gulcehre, C., Cho, K.H., et al., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.

Duns Jr., H., Ros, N.C.J., 1963. Vertical Flow of Gas and Liquid Mixtures in Wells. 6th World Petroleum Congress, Germany.

Frazier, P.I., 2018. A Tutorial on Bayesian Optimization.

Gomez, L.E., Shoham, O., Schmidt, Z., et al., 2000. Unified mechanistic model for steady-state two-phase flow: horizontal to vertical upward flow. SPE J. 5 (3), 339–350. https://doi.org/10.2118/65705-PA.

Group, N., 2017. R-NET: Machine Reading Comprehension with Self-Matching Networks.

Hagedorn, A.R., Brown, K.E., 1965. Experimental study of pressure gradients occurring during continuous two-phase flow in small-diameter vertical conduits. J. Petrol. Technol. 17 (4), 475–484. https://doi.org/10.2118/940-PA.

He, Z., Yun, T., 2018. Implement intelligent dynamic analysis of bottom-hole pressure with naive Bayesian models. Multimed. Tool. Appl. 78 (21), 29805–29821. https://doi.org/10.1007/s11042-018-6340-7.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Huang, R., Wei, C., Wang, B., et al., 2022. Well performance prediction based on Long Short-Term Memory (LSTM) neural network. J. Petrol. Sci. Eng. 208, 109686. https://doi.org/10.1016/j.petrol.2021.109686.

Jahanandish, I., Salimifard, B., Jalalifar, H., 2011. Predicting bottomhole pressure in vertical multiphase flowing wells using artificial neural networks. J. Petrol. Sci. Eng. 75 (3), 336–342. https://doi.org/10.1016/j.petrol.2010.11.019.

Kanwal, A., Lau, M.F., Ng, S.P.H., et al., 2022. BiCuDNNLSTM-1dCNN — a hybrid deep learning-based predictive model for stock price prediction. Expert Syst. Appl. 202, 117123. https://doi.org/10.1016/j.eswa.2022.117123.

Lehmann, R., 2013. 3σ-rule for outlier detection from the viewpoint of geodetic adjustment. J. Survey. Eng. 139 (4), 157–165. https://doi.org/10.1061/(ASCE)SU.1943-5428.0000112.

Liang, H., Wei, Q., Lu, D., et al., 2020. Application of GA-BP neural network algorithm in killing well control system. Neural Comput. Appl. 33 (3), 949–960. https://doi.org/10.1007/s00521-020-05298-4.

Lu, Q., Sun, S., Duan, H., et al., 2021. Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model. Energy Informatics 4 (2), 47. https://doi.org/10.1186/s42162-021-00166-4.

Marfo, S.A., Asante-Okyere, S., Ziggah, Y.Y., 2021. A new flowing bottom hole pressure prediction model using M5 prime decision tree approach. Modeling Earth Systems and Environment 8, 2065–2073. https://doi.org/10.1007/s40808-021-01211-7.

Mask, G., Wu, X., Ling, K., 2019. An improved model for gas-liquid flow pattern prediction based on machine learning. J. Petrol. Sci. Eng. 183, 106370. https://doi.org/10.1016/j.petrol.2019.106370.

Mohammadpoor, M., Shahbazi, K., Torabi, F., et al., 2010. A new methodology for prediction of bottomhole flowing pressure in vertical multiphase flow in Iranian oil fields using artificial neural networks (ANNs). In: SPE Latin American and Caribbean Petroleum Engineering Conference, Lima, Peru.

Nait Amar, M., Zeraibi, N., 2019. A combined support vector regression with firefly algorithm for prediction of bottom hole pressure. Appl. Sci. 2 (1), 23. https://doi.org/10.1007/s42452-019-1835-z.

Nait Amar, M., Zeraibi, N., Redouane, K., 2018. Bottom hole pressure estimation using hybridization neural networks and grey wolves optimization. Petroleum 4 (4), 419–429. https://doi.org/10.1016/j.petlm.2018.03.013.

Okoro, E.E., Obomanu, T., Sanni, S.E., et al., 2021. Application of artificial intelligence in predicting the dynamics of bottom hole pressure for under-balanced drilling: extra tree compared with feed forward neural network model. Petroleum 2405–6561. https://doi.org/10.1016/j.petlm.2021.03.001.

Orkiszewski, J., 1967. Predicting two-phase pressure drops in vertical pipe. J. Petrol. Technol. 19 (6), 829–838. https://doi.org/10.2118/1546-PA.

Otaki, D., Nonaka, H., Yamada, N., 2022. Thermal design optimization of electronic circuit board layout with transient heating chips by using Bayesian optimization and thermal network model. Int. J. Heat Mass Tran. 184, 122263. https://doi.org/10.1016/j.ijheatmasstransfer.2021.122263.

Sami, N.A., 2021. Application of machine learning algorithms to predict tubing pressure in intermittent gas lift wells. Petroleum Research 7 (2), 246–252. https://doi.org/10.1016/j.ptlrs.2021.09.006.

Sami, N.A., Ibrahim, D.S., 2021. Forecasting multiphase flowing bottom-hole pressure of vertical oil wells using three machine learning techniques. Petroleum Research 6 (4), 417–422. https://doi.org/10.1016/j.ptlrs.2021.05.004.

Sheu, B.J., Choi, J., 1995. Back-propagation neural networks. In: Neural Information Processing and VLSI, Boston, US, pp. 277–296.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. Adv. Neural Inf. Process. Syst. 25, 2960–2968. https://doi.org/10.48550/arXiv.1206.2944.

Spesivtsev, P., Sinkov, K., Sofronov, I., et al., 2018. Predictive model for bottomhole pressure based on machine learning. J. Petrol. Sci. Eng. 166, 825–841. https://doi.org/10.1016/j.petrol.2018.03.046.

Székely, G., Bakirov, R., 2007. Measuring and testing dependence by correlation of distances. Ann. Stat. 35 (6), 2769–2794. https://doi.org/10.1214/009053607000000505.

Tariq, Z., Mahmoud, M., Abdulraheem, A., 2019. Real-time prognosis of flowing bottom-hole pressure in a vertical well for a multiphase flow using computational intelligence techniques. J. Pet. Explor. Prod. Technol. 10 (4), 1411–1428. https://doi.org/10.1007/s13202-019-0728-4.

Wei, N., Li, C., Peng, X., et al., 2019. Daily natural gas consumption forecasting via the application of a novel hybrid model. Appl. Energy 250, 358–368. https://doi.org/10.1016/j.apenergy.2019.05.023.

Ye, W., Jiang, Z., Li, Q., et al., 2022. A hybrid model for pathological voice recognition of post-stroke dysarthria by using 1DCNN and double-LSTM networks. Appl. Acoust. 197, 108934. https://doi.org/10.1016/j.apacoust.2022.108934.

Zhu, Z., Song, X., Zhang, R., et al., 2022. A hybrid neural network model for predicting bottomhole pressure in managed pressure drilling. Appl. Sci. 12 (13), 6728. https://doi.org/10.3390/app12136728.