

RESEARCH ARTICLE

Crude Oil Price Time Series Forecasting: A Novel Approach Based on Variational Mode Decomposition, Time-Series Imaging, and Deep Learning

ZI-JIAN PENG, CHUAN ZHANG^{ID}, AND YU-XIN TIAN^{ID}

School of Business Administration, Northeastern University, Shenyang 110169, China

Corresponding author: Chuan Zhang (czhang@mail.neu.edu.cn)

This work was supported in part by the National Social Science Fund of China under Grant 19BGL229.

ABSTRACT Accurately forecasting crude oil prices is challenging due to market noise and non-stationarity. To address these challenges, we propose a forecasting framework that incorporates variational mode decomposition (VMD), time-series imaging, and bidirectional gated recurrent unit network (BGRU). Our approach eliminates additional assumptions and auxiliary data. First, the raw data are preprocessed through normalization, followed by decomposing multiple stationary sub-series through VMD. Subsequently, three time-series imaging techniques, recurrence plot (RP), Gramian angular field (GAF), and Markov transition field (MTF), are employed respectively to transform the sub-series into two-dimensional images. A convolutional neural network (CNN) is then used to extract features from these images. Finally, the extracted features are fed into BGRU for prediction and the Adam optimizer is used to train models. Experimental evaluations are conducted using a dataset from the U.S. Energy Information Administration (EIA), consisting of the weekly spot prices free on board (FOB) of the West Texas Intermediate (WTI) crude oil, spanning from June 19, 1998, to May 12, 2023. Results demonstrate that all three models constructed following our approach outperform benchmark methods. Specifically, our VMD-RP-BGRU model achieves the best forecasting performance with MAE=2.429, MSE=10.94, MAPE=2.94%, and R-squared=0.9418. The model exhibits reductions of 21.64%, 23.15%, and 36.46% in MAE, MSE, and MAPE, respectively, compared to the seasonal autoregressive integrated moving average (SARIMA) model, and reductions of 21.18%, 22.70%, and 36.08% compared to the Holt-Winters exponential smoothing (HWES) model. Our study contributes to the advancement of crude oil price forecasting techniques and supports informed decision-making in the energy sector.

INDEX TERMS Crude oil price forecasting, deep learning, variational mode decomposition, time-series imaging, bidirectional gated recurrent unit.

I. INTRODUCTION

Crude oil is one of the essential energy sources for modern industrial production and transportation, often referred to as the “lifeblood of industry.” With the globalization of the economy and financial markets, fluctuations in crude oil prices impact global economic stability and development.

The associate editor coordinating the review of this manuscript and approving it for publication was Akin Tascikaraoglu^{ID}.

According to the International Energy Agency (IEA), crude oil constituted 30.9% of the global total energy supply in 2019, establishing it as the largest primary energy source [1]. Fluctuations in crude oil prices can significantly affect global economic stability. For instance, the 2008 global financial crisis was exacerbated by the sharp increase in crude oil prices, which reached a peak of \$145.31 per barrel in June 2008 [2]. The surge in oil prices contributed to the economic downturn by reducing consumer spending and increasing production

costs. Conversely, the drastic year-over-year decline of 31.9% in crude oil prices in 2014 imposed notable economic hardships on oil-exporting nations, thereby impeding their fiscal stability and prospects for development [3]. Therefore, accurate forecasting of crude oil prices is of significant importance for energy policy formulation, risk management, and investment decision-making [4], [5].

The main characteristics of crude oil price time series data are non-stationarity, non-linearity, and high volatility [6]. Various unstable factors, including economic conditions, political events, and trader expectations, exert influence on oil prices [7]. All of these factors make accurate price prediction challenging and pose significant difficulties in forecasting. Traditional methodologies for time series analysis, such as moving average (MA), exponential smoothing (ES), and autoregressive integrated moving average (ARIMA) [8], have proven inadequate in achieving satisfactory predictions. With advancements in computer technology, deep learning methods have been successfully applied to forecasting crude oil price time series, particularly recurrent neural networks (RNN) [9], long short-term memory (LSTM) [10], [11], [12], [13], [14], [15], and gated recurrent unit (GRU) [16], [17]. Leveraging their strong capabilities in handling substantial volumes of data and capturing non-linear relationships, deep learning models outperform their linear counterparts in predicting crude oil prices. However, the persistent non-stationarity exhibited by crude oil prices continues to negatively impact the predictive performance of deep learning models, indicating the need for further enhancement. Furthermore, historical time series encompass a multitude of implicit features, and recurrent neural network models face certain limitations in uncovering these latent features, thus necessitating the exploration of more effective methodologies.

To overcome these challenges and improve the accuracy of crude oil price forecasting, we propose a novel forecasting framework that combines variational mode decomposition (VMD), time series imaging, and deep learning. Our approach is entirely data-driven, without any assumption or additional search for auxiliary data potentially related to crude oil prices. Specifically, our framework consists of the following five steps: 1) Data preprocessing: The original data are standardized by division to reduce the influence of scale variance in the time series. 2) Decomposing subsequences by the VMD: The non-stationary time series is decomposed into simple, stationary intrinsic mode function (IMF) series with certain periodic patterns. 3) Time series image generation: We employ three time series imaging techniques, Recurrence Plot (RP), Gramian Angular Field (GAF), and Markov Transition Field (MTF), respectively, to transform the one-dimensional IMF series into two-dimensional images, which reveal more features of these series and facilitate subsequent processing by Convolutional Neural Network (CNN). 4) Image feature extraction using CNN: By leveraging the powerful computational vision capabilities of CNN, we extract and utilize

the hidden features embedded in the time series, avoiding the need for manually collecting additional auxiliary data. 5) Forecasting through Bidirectional Gated Recurrent Unit Network (BGRU): We employ the BGRU module to extract information from both the forward and backward directions, train the model, and obtain the final prediction results.

Our primary contributions can be summarized as follows:

(1) The originality of this study lies in the proposal of three new hybrid forecasting models, namely VMD-RP-BGRU, VMD-GAF-BGRU, and VMD-MTF-BGRU, based on different time series imaging techniques. These models integrate the advantages of VMD, time series imaging, CNN, and BGRU. We applied the proposed methods to publicly available West Texas Intermediate (WTI) crude oil price data to evaluate their performance. Comparative experimental results demonstrate that our proposed VMD-RP-BGRU model achieves the best prediction performance, followed by VMD-MTF-BGRU and VMD-GAF-BGRU. Specifically, VMD-GAF-BGRU, VMD-MTF-BGRU, and VMD-RP-BGRU exhibit a MAE reduction of 21.64%, 23.15%, and 36.46% compared to the SARIMA model, and 21.18%, 22.70%, and 36.08% compared to the HWES model, respectively.

(2) We have, for the first time, employed the time series imaging technique to analyze the time series of crude oil prices. By transforming one-dimensional subseries into two-dimensional image matrices, additional feature information embedded in the sequence is revealed, thereby enhancing the performance of crude oil price forecasting. Ablation experimental results demonstrate a significant improvement in the predictive performance of the crude oil price time series through the use of Recurrence Plot (RP) time series imaging compared to other ablation experiment benchmarks. Specifically, after employing the RP time series imaging method, there was a reduction of 27.51% in MAE, 41.62% in MSE, and 24.29% in MAPE, while the coefficient of determination R^2 increased by 4.61%.

(3) Our study expands the research on crude oil price forecasting and enhances the theoretical framework of deep learning-based time series prediction methods, thereby contributing to the field. Moreover, our approach proves to be effective and holds the potential to provide valuable guidance to energy policy-makers, macroeconomic regulators, and investors.

While our approach shows promising results, it is important to acknowledge its limitations. Currently, our methodology solely relies on univariate analysis and does not consider the influence of other significant factors or exogenous variables. However, to further enhance the accuracy of predictions, future research endeavors can explore the incorporation of additional variables, such as news media and search engine data.

The subsequent sections of this paper are structured as follows. Section II provides an overview of the related work in crude oil price forecasting. Section III presents the

methodology proposed in this study. Section IV substantiates the effectiveness and superiority of our proposed approach by means of comparative analyses with existing methods and ablation experiments. The last section provides a summary and conclusion for the paper.

II. RELATED WORK

Developing an accurate prediction method to forecast future price fluctuations in the oil market has consistently been a focal point for investors. Traditional statistical techniques offer various models for time series analysis, such as the autoregressive moving average (ARMA) model [8] and the autoregressive integrated moving average (ARIMA) model (see [8], [18], [19]). For example, [20] applied the ARIMA technique to forecast oil prices and observed reasonable predictions for the annual average oil prices worldwide. However, these statistical methods rely on assumptions of stationarity and linearity, rendering their performance inadequate when applied to non-stationary and chaotic forecasting of crude oil prices.

To address the non-stationarity, extreme irregularity, and multiscale variability of crude oil prices, an increasing number of studies have adopted a sequence decomposition approach. This approach decomposes the original time series data into several fixed components using various decomposition techniques, which often exhibit certain stationarity and regularity. The main decomposition methods include empirical mode decomposition (EMD) [21] and its variants [22], [23], [24], wavelet analysis [25], and VMD [26]. Reference [27] utilized the ensemble empirical mode decomposition (EEMD) combined with extreme learning machines (EELM) to predict spot prices of crude oil. Using a dataset of crude oil spot prices of West Texas Intermediate (WTI), their proposed EEMD-EELM-ADD model demonstrated the highest prediction accuracy, which was statistically confirmed by the Diebold-Mariano (DM) test with a 90% confidence level. In another study, [28] constructed a hybrid model consisting of complete ensemble empirical mode decomposition (CEEMD), support vector machines (SVM), particle swarm optimization (PSO), and Markov-switching generalized autoregressive conditional heteroskedasticity to more effectively capture the nonlinearity and volatility of crude oil price time series. Using daily closing prices of WTI crude oil, their hybrid model achieved a MAE of 0.042, RMSE of 0.053, and MAPE of 0.057, outperforming SVM-PSO and MS-GARCH models. Reference [29] introduced a new forecasting approach for oil price prediction by combining VMD, LSTM networks, and a moving window strategy, known as the VMD-LSTM-MW model. Using monthly WTI data, their proposed model exhibited significantly better MAPE compared to the VMD-GA-SVM model, EEMD-GA-SVM model, GA-SVM model, and ARIMA model, with relative improvements of 35.39%, 40.16%, 72.06%, and 67.23%, respectively. Additionally, [30] utilized an improved complete ensemble empirical mode decomposition with adaptive

noise (ICEEMDAN) method to decompose the original crude oil futures price series into a set of subseries and reconstructed them into high-frequency, low-frequency, and trend components using permutation entropy (PE). Their proposed ICEEMDAN-PE-EMD-PSR-CSSA-KELM model achieved significant reductions of 52%, 49.8%, and 44.8% in MAE, MAPE, and RMSE, respectively, in one-step ahead forecasting for Brent crude oil futures prices compared to the PSR-KELM model. Furthermore, [31] proposed two hybrid predictors based on RNN, employing VMD, sample entropy (SE), and GRUs for oil price prediction. Using the WTI dataset, their VMD-SE-GRU framework exhibited a RMSE of 0.6735, MAE of 0.4585, MAPE of 0.8059, and an R^2 value of 0.9272.

With the development of artificial intelligence (AI), machine learning models are increasingly being used for time series forecasting. The main machine learning models include LSTM (see [12], [29], [32]) and GRU (see [31], [33]). Compared to traditional statistical techniques, AI-based models do not require any assumptions about the data distribution. They possess good generalization and self-learning abilities, leading to a significant improvement in forecasting accuracy. For example, [34] proposed a hybrid method combining LSTM and GRU architectures for crude oil price forecasting and addressed challenges such as long-term dependencies, overfitting, and hyperparameter tuning. Experimental results using daily crude oil price data demonstrated that their hybrid model slightly outperforms the individual LSTM and GRU models. Reference [35] proposed a new hybrid forecasting technique based on local mean decomposition (LMD), ARIMA and LSTM models to improve the accuracy of crude oil price forecasting. Experimental results using WTI crude oil price data showed that their LMD-SD-ARIMA-LSTM model achieves an MAE of 0.106 and MAPE of 1.124. Reference [36] proposed a text-driven crude oil price forecasting method based on CNN and LSTM, which uses text information from various sources such as social media, news and expert comments to capture the impact of market sentiment and expectations on crude oil price fluctuations. Results indicated that the SVR model incorporating text and financial features outperforms the model using only financial features, with improvements of 6.67% and 2.5% in MAE and RMSE, respectively. Reference [37] proposed a learning paradigm that combines the trajectory similarity (TS) method, VMD, and sample entropy (SE) to forecast nonlinear and highly volatile crude oil price series, and achieved better performance than benchmarks by using LSTM networks and ARIMA models. For the Brent dataset, the MAPE of their TS model was reduced by 45.66%, 59.55%, 55.12%, 44.08%, and 30.07% compared to Naive, SVR, ELM, ANN, and LSTM, respectively. Unlike previous deep learning models that assumed unidirectional relationships in time series data, the BGRU method enhances model performance by extracting information from both forward and backward directions [38]. For example, [39] integrated

news sentiment analysis, variational mode decomposition, attention mechanism, and BGRU for oil price forecasting, improving forecasting performance by capturing both qualitative and quantitative information. Experimental results using daily WTI crude oil futures price data demonstrated that their FinBERT-VMD-Att-BiGRU model achieves an RMSE of 0.0077 and MAE of 0.0049 for one-step ahead prediction.

In recent years, there have been significant advancements in computer vision technologies, which are gradually finding applications in the field of forecasting. One emerging area of research involves utilizing deep learning architectures with imaging techniques for time series forecasting. However, the utilization of time series imaging methods in the domain of crude oil price forecasting remains relatively unexplored. Time series imaging has proven to be an effective approach for analyzing time-dependent data and has demonstrated its efficacy in various domains, including tourism demand forecasting [40], drought prediction [41], and residential electricity consumption forecasting [42]. These studies have provided evidence of the ability of time series imaging to enhance prediction accuracy. Specifically, combining time series imaging with CNN can fully leverage the advantages of CNN in computer vision processing and enhance the accuracy of time series forecasting.

In view of this, this study takes into account the non-stationarity and complex variations of crude oil price fluctuations. By integrating VMD, time series imaging techniques, CNN, and BGRU, this study aims to fully leverage the strengths of VMD in non-stationary sequence decomposition, time series imaging in feature representation, CNN in image processing, and BGRU in sequence data handling. Through comparative experiments, the superiority of the proposed forecasting framework is demonstrated.

III. METHODOLOGY

A. METHODOLOGICAL FRAMEWORK AND PRINCIPLE

Fig. 1 illustrates the comprehensive framework of the methodology proposed in this research. The methodology comprises five components: 1) Data preprocessing; 2) Decomposing subsequences by the VMD; 3) Time series image generation; 4) Image feature extraction using CNN; 5) Forecasting through BGRU.

In the data preprocessing stage, the raw time series data are transformed into the required format for subsequent modeling using a normalization method, eliminating the prediction challenges caused by scale variance. Since crude oil prices exhibit continuous fluctuation and are non-stationary time series, the VMD component is used to decompose multiple subsequences used for forecasting the crude oil price in the next period. Each subsequence is decomposed into a set of IMF series that exhibit certain periodic patterns.

In the time series image generation module, the one-dimensional IMF sequence is transformed into two-dimensional image data using time series imaging techniques. This transformation is crucial for revealing the relationship between observations with different time lags in the IMF.

It captures the temporal correlations and patterns of various components of crude oil prices series in response to environmental factors, which are helpful for forecasting future price changes.

In the next section, leveraging the powerful computer vision processing capability of CNN, features from each time series image are gradually extracted using convolutional kernels, capturing local feature information. Subsequently, the dimensionality of the results from the convolutional layers is reduced through pooling layers. This aims to retain essential information while reducing computational complexity. Additionally, it helps prevent overfitting and improves the model's generalization ability. Finally, the features extracted by the CNN module are transformed into the input format required by the BGRU network through fully connected layers.

The last module is the BGRU network, which consists of a two-layer GRU network that is trained simultaneously in both forward and backward directions along the time sequence. The outputs from both directions are then concatenated and fed into an output layer. As proposed by [43], time series data are bidirectional, meaning that the current state reflects the past state and forms the basis for the future state. Therefore, BGRU is more advantageous in capturing the long-term correlations of time series observations as a whole. The CNN-BGRU model is trained using the Adam optimizer with the objective of minimizing MSE to obtain the predictive model for forecasting future crude oil prices.

In summary, in our proposed method, the strengths of VMD in non-stationary sequence decomposition, time series imaging in feature representation, CNN in image processing, and bidirectional GRU in sequence data handling are fully leveraged.

B. THE SPECIFIC PROCESS OF THE METHOD

The detailed expansion of the steps shown in Fig. 1 is described below.

1) DATA PREPROCESSING

To accurately forecast crude oil prices, our prediction model must capture both the global and local variations inherent in the time series data. The global variation primarily arises from scale (or level) variance, while the local variation is characterized by local nonlinear trends that encompass recurring patterns and temporary relationships. Deep learning models have strong nonlinear fitting capabilities and can effectively capture and model local variations in time series. However, they often encounter challenges when dealing with scale variance. By eliminating scale variation, deep learning models can fully utilize their nonlinear fitting capabilities, outputting improved forecasts.

In this paper, we employ a division-based method known as scale normalization for preprocessing the raw data. Scale normalization involves dividing the current value by the previous value, as denoted by the equation:

$$x_t = \frac{y_t}{y_{t-1}}, \quad (1)$$

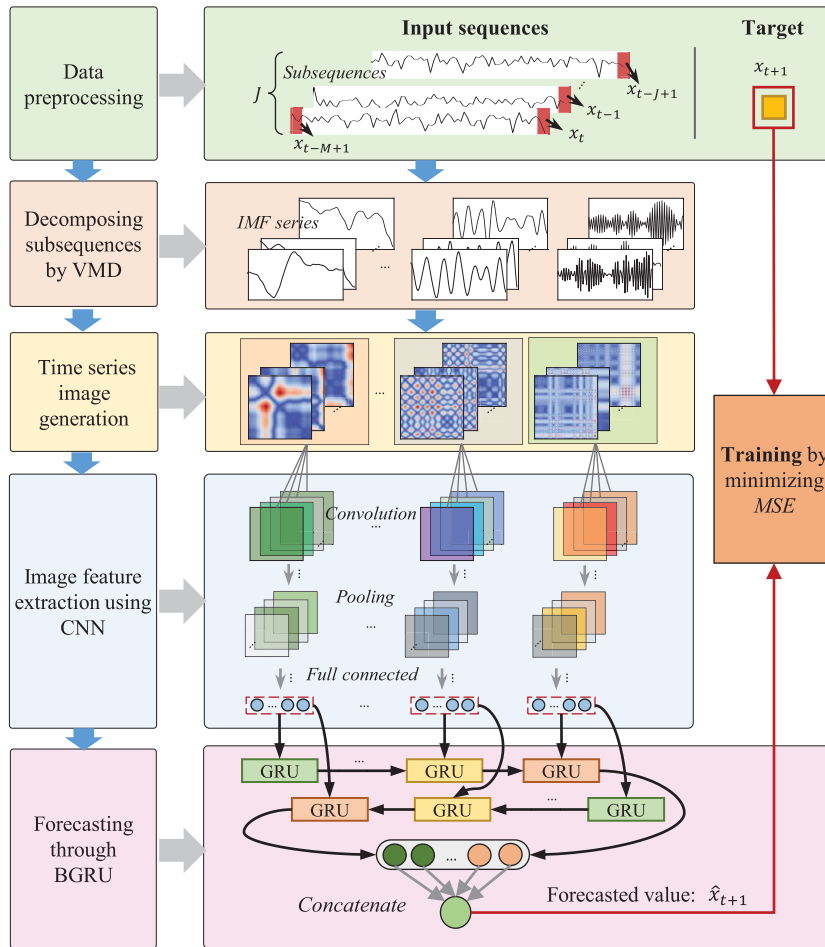


FIGURE 1. Methodology framework.

where y_t represents the crude oil price at the current time, y_{t-1} represents the crude oil price at the previous time, and x_t represents the normalized value of the time series at the current time. The normalized series of crude oil prices can be represented as a vector

$$\mathbf{x} = (x_1, x_2, \dots, x_t). \quad (2)$$

We forecast the future value x_{t+1} using multiple historical subsequences in \mathbf{x} . Each subsequence is defined with a length referred to as the “image size” (denoted as M). A subsequence used for predicting x_{t+1} can be represented as:

$$\mathbf{s}_t = (x_{t-M+1}, \dots, x_{t-1}, x_t). \quad (3)$$

2) DECOMPOSING SUBSEQUENCES BY VMD

Let $f(t)$ represent an original sequence that undergoes decomposition using the VMD module, generating K decomposed signals, i.e., IMF series. VMD is an adaptive and fully non-recursive method for mode variation and signal processing. This technique has the advantage of determining the number of mode decompositions. Its adaptability is demonstrated by determining the appropriate number of mode decompositions based on the specific situation of the

given sequence. Subsequently, in the search and solving process, it can adaptively match the optimal center frequency and finite bandwidth for each mode and effectively separate the IMFs, thus obtaining the effective decomposition components of the given signal. For a detailed explanation of the principles of VMD, please refer to [44].

Let $\{u_k\} = \{u_1, u_2, \dots, u_K\}$ and $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_K\}$ represent the k -th decomposed signal and the center frequency of the signal, respectively. λ^n denotes the noise tolerance in the n -th iteration calculation, which is used to satisfy the fidelity requirement of signal decomposition. $\hat{u}_k^n(\omega)$, $\hat{f}(\omega)$ and $\hat{\lambda}^n(\omega)$ correspond to the Fourier transforms of $u_k^n(\omega)$, $f(\omega)$ and $\lambda^n(\omega)$ in the n -th iteration calculation, respectively. The main steps of VMD are as follows:

Step 1. Let $n = 0$. Initialize $\{\hat{u}_k^1\}$, $\{\omega_k^1\}$, $\hat{\lambda}^1$ and the max number of iterations N .

Step 2. For each mode, update \hat{u}_k^n and ω_k^n using the following equations:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i < k} \hat{u}_i^{n+1}(\omega) - \sum_{i > k} \hat{u}_i^n(\omega) + \hat{\lambda}^n(\omega)/2}{1 + 2\alpha(\omega - \omega_k^n)^2}, \quad (4)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega} \quad (5)$$

Step 3. For each iteration, update λ^n by the following equation:

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left[\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right]. \quad (6)$$

Step 4. If $\sum_k \left(\left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 / \left\| \hat{u}_k^n \right\|_2^2 \right) < \varepsilon$ and $n < N$ are not satisfied, return to **Step 2**; otherwise, complete the iteration and output the final $\{u_k\}$ and $\{\omega_k\}$.

In this paper, K IMF series obtained from the decomposition of the subsequence s_t can be represented as:

$$V(s_t, K) = \begin{pmatrix} u_1^{t,1} & u_2^{t,1} & \dots & u_M^{t,1} \\ u_1^{t,2} & u_2^{t,2} & \dots & u_M^{t,2} \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{t,K} & u_2^{t,K} & \dots & u_M^{t,K} \end{pmatrix} = \begin{pmatrix} v_{t,1} \\ v_{t,2} \\ \vdots \\ v_{t,K} \end{pmatrix}. \quad (7)$$

3) TIME SERIES IMAGE GENERATION

The CNN model has demonstrated significant advantages in computer vision. However, when dealing with one-dimensional IMF series, it becomes necessary to employ time series imaging techniques to convert the one-dimensional series into two-dimensional images. This conversion allows for the mapping of temporal correlations and patterns between each observation within the lag order to pixels in the resulting images. Consequently, CNNs are better equipped to recognize and learn these extracted features, facilitating improved performance in analysis tasks.

There are three commonly used time series imaging techniques: recurrence plot (RP) (see [45], [46]), Gramian angular field (GAF) [47], and Markov transition field (MTF) [47]. These techniques are capable of preserving the temporal and spatial relationships inherent in the time series data, as well as capturing essential characteristics such as periodicity and trend. By leveraging these transformation methods, we can effectively convert a complex time series analysis problem into a more manageable image classification problem. This conversion not only enhances the prediction accuracy but also improves computational efficiency [40].

Let the vector $\mathbf{X} = (x_1, x_2, \dots, x_n)$ denote the time series to be processed. The computational processes of these three techniques are as follows.

a: RECURRENCE PLOT (RP)

RP technique is widely used to visualize the periodicity of trajectories in phase space, making it a valuable tool for analyzing the periodicity, non-stationarity, and chaos present in time series data. RP effectively captures the similarity and stability of the internal structure of time series, making it particularly suitable for analyzing short-term time series. The output values of each pixel in the original RP method are only

0 and 1. The process is as follows: First, the time-domain space of the time series is transformed into the phase space, where each point x_i in the time domain is transformed into the corresponding state in the phase space:

$$s_i = (x_i, x_{i+\tau}, \dots, x_{i+(\mu-1)\tau}), i = 1, 2, \dots, n - (\mu - 1)\tau, \quad (8)$$

where μ represents the dimension of the extracted trajectory, and τ denotes the time delay. Next, the distance (vector norm) between each pair of states is calculated, and threshold binarization is performed to obtain an $n \times n$ recursive plot, where the calculation formula for each point in the image is:

$$g_{i_1, i_2}^{RP} = \Theta(\varepsilon - \|s_{i_1} - s_{i_2}\|), i_1, i_2 = 1, \dots, n, \quad (9)$$

where ε is the distance threshold that makes $g_{i_1, i_2}^{RP} \in \{0, 1\}$, and $\Theta(\cdot)$ represents the Heaviside function.

b: GRAMIAN ANGULAR FIELD (GAF)

GAF technique utilizes the Gramian matrix and polar coordinate system to transform time series data into visual representations in the form of images. This encoding process enables the images to effectively capture the temporal correlation present within different time intervals. The calculation process for GAF is as follows: First, normalize \mathbf{X} to the interval $[-1, 1]$ to obtain $\tilde{\mathbf{X}}$, where the formula for calculating the i -th element is:

$$\tilde{x}_i = \frac{2x_i - \max(\mathbf{X}) - \min(\mathbf{X})}{\max(\mathbf{X}) - \min(\mathbf{X})}. \quad (10)$$

Then, $\tilde{\mathbf{X}}$ is represented in polar coordinates based on the following formulas:

$$\begin{cases} \varphi_i = \arccos(\tilde{x}_i) \\ r_i = \frac{i}{C} \end{cases}, i = 1, 2, \dots, n, \quad (11)$$

where C is the constant factor used to regularize the span of the polar coordinates. Finally, the sequence represented in polar coordinates is transformed into a two-dimensional image of size $n \times n$, where the calculation formula for each point in the image is:

$$g_{i_1, i_2}^{GAF} = \sin(\varphi_{i_1} - \varphi_{i_2}), i_1, i_2 = 1, 2, \dots, n. \quad (12)$$

The resulting image effectively captures the temporal correlations within different time intervals.

c: MARKOV TRANSITION FIELD (MTF)

The fundamental concept behind the MTF technique is to convert time series data into image format using a Markov state transition matrix based on quantile bins. By preserving the temporal information within the sequence, the Markov transition probability captures the changes in the quantile bin to which the observed values belong. The calculation process

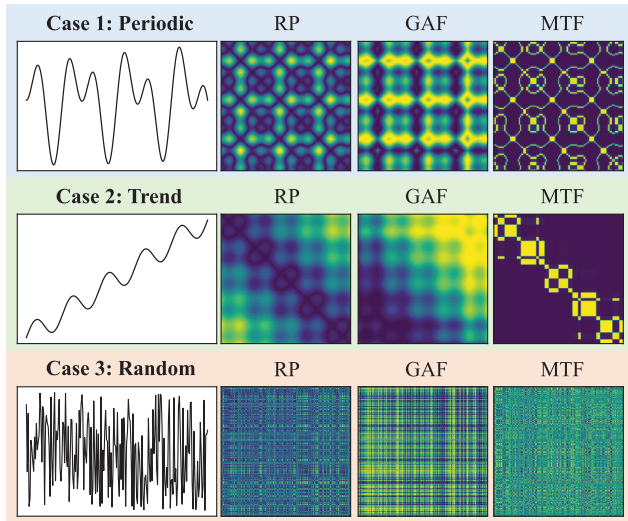


FIGURE 2. Representative cases of RP, GAF, and MTF images.

for MTF is as follows: First, divide \mathbf{X} into Q quantile bins to obtain B_1, B_2, \dots, B_Q . Then, calculate

$$\begin{cases} w_{q_1, q_2} = \mathbb{P}(x_{i-1} \in b_{q_1} | x_i \in b_{q_2}) \\ \sum_{q_2} w_{q_1, q_2} = 1, \\ q_1, q_2 = 1, 2, \dots, Q, \end{cases} \quad (13)$$

to obtain a weighted adjacency matrix $W = (w_{q_1, q_2})_{Q \times Q}$. Finally, a two-dimensional image of size $n \times n$ is obtained, where the value of each point in the image is:

$$g_{i_1, i_2}^{MTF} = w_{q_1, q_2} (x_{i_1} \in B_{q_1}, x_{i_2} \in B_{q_2}), \quad i_1, i_2 = 1, 2, \dots, n; \quad q_1, q_2 = 1, 2, \dots, Q. \quad (14)$$

By employing time series imaging techniques, the IMF series from (7) can be transformed into images. Taking the k -th IMF series as an example, the expression is as follows:

$$I_{\#}(\mathbf{v}_{t,k}) = \begin{pmatrix} g_{1,1}^{t,k,\#} & \dots & g_{1,m}^{t,k,\#} & \dots & g_{1,M}^{t,k,\#} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{m,1}^{t,k,\#} & \dots & g_{m,m}^{t,k,\#} & \dots & g_{m,M}^{t,k,\#} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{M,1}^{t,k,\#} & \dots & g_{M,m}^{t,k,\#} & \dots & g_{M,M}^{t,k,\#} \end{pmatrix}, \quad (15)$$

where $\#$ represents a time series imaging method, $\# \in \{RP, GAF, MTF\}$.

To provide a more intuitive representation of these three types of images, we present typical examples of RP, GAF, and MTF images in Fig. 2. In these images, lighter regions correspond to higher values. It is apparent that applying the same method to different types of curves or employing different methods on the same curve leads to distinct images. In Case 1, the time series exhibits periodic trends, and the distribution of values in the three images clearly demonstrates periodicity. The regularity of value distribution is more pronounced in the GAF and RP images compared to the MTF image, and there is a certain level of complementarity

between the GAF and RP images. In Case 2, the time series exhibits an increasing trend with periodicity. The distribution of values in the RP image demonstrates a periodic increasing trend along the main diagonal axis, while the values in the GAF image cyclically increase from the bottom-left to the top-right. Conversely, the values in the MTF image exhibit a periodic decreasing trend along the main diagonal axis. In Case 3, the time series lacks any discernible pattern, and the corresponding distributions of values in the three types of images do not exhibit any regularity.

4) IMAGE FEATURE EXTRACTION USING CNN

We utilize CNN to extract features from the time series images. The CNN model utilized in our approach consists of only one convolutional layer and one pooling layer. A single convolutional layer effectively detects features such as edges and shapes in the images, while multiple convolutional layers can introduce excessive complexity and redundancy, increasing computational costs and the risk of overfitting. A single pooling layer efficiently reduces the spatial dimensions of the feature maps while preserving the most prominent features. This helps to reduce the number of parameters and computations in the network and prevents overfitting by providing a more abstract representation of the features. The use of multiple pooling layers may result in the loss of important information and excessive reduction in the resolution of the feature maps. Therefore, employing a single pooling layer is more appropriate for the current task.

Let $S_{\#}(\mathbf{v}_{t,k})$ represent a submatrix of $I_{\#}(\mathbf{v}_{t,k})$ with a shape of $f \times f$, where f denotes the size of the filter matrix in the convolutional layer. We use grid search to determine the optimal value of the filter size f . Specifically, we have:

$$S_{\#}(\mathbf{v}_{t,k})_{f_1, f_2} = \begin{pmatrix} g_{f_1, f_2}^{t,k,\#} & \dots & g_{f_1, f_2 + f - 1}^{t,k,\#} \\ \vdots & \ddots & \vdots \\ g_{(f_1 + f - 1), f_2}^{t,k,\#} & \dots & g_{(f_1 + f - 1), (f_2 + f - 1)}^{t,k,\#} \end{pmatrix}, \quad f, f_1, f_2 \in \mathbb{N}^+, f + f_1, f + f_2 \leq M + 1. \quad (16)$$

The feature extraction process by the convolution kernel weight matrix $W_{Cov} \in \mathbb{R}^{f \times f}$ (i.e., the filter) yields a feature $c_{f_1, f_2}^{t,k,\#}$, which is expressed as follows:

$$c_{f_1, f_2}^{t,k,\#} = \Phi \left[W_{Cov} * S_{\#}(\mathbf{v}_{t,k})_{f_1, f_2} + b_{Cov} \right], \quad (17)$$

where $\Phi(\cdot)$ is a nonlinear activation function, which is usually the rectified linear unit (ReLU) function defined as $\text{Relu}(x) = \max(x, 0)$. We use the ReLU activation function because it can increase the nonlinearity of the neural network, avoid the gradient vanishing problem that occurs when using sigmoid or tanh activation functions, and lower the computational cost of the neural network by only judging whether the input is greater than zero without performing exponential operations [48]. The asterisk symbol $*$ denotes the convolution operation, and b_{Cov} represents the bias term corresponding to each layer. Through (17), the convolution

feature matrix $C_{\#}(\mathbf{v}_{t,k})$ is calculated as follows:

$$C_{\#}(\mathbf{v}_{t,k}) = \begin{pmatrix} c_{1,1}^{t,k,\#} & \cdots & c_{1,f_2}^{t,k,\#} & \cdots & c_{1,(M-f+1)}^{t,k,\#} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{f_1,1}^{t,k,\#} & \cdots & c_{f_1,f_2}^{t,k,\#} & \cdots & c_{f_1,(M-f+1)}^{t,k,\#} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{(M-f+1),1}^{t,k,\#} & \cdots & c_{(M-f+1),f_2}^{t,k,\#} & \cdots & c_{(M-f+1),(M-f+1)}^{t,k,\#} \end{pmatrix} \times \begin{pmatrix} c_{1,1}^{t,k,\#} & \cdots & c_{1,f_2}^{t,k,\#} & \cdots & c_{1,(M-f+1)}^{t,k,\#} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{f_1,1}^{t,k,\#} & \cdots & c_{f_1,f_2}^{t,k,\#} & \cdots & c_{f_1,(M-f+1)}^{t,k,\#} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{(M-f+1),1}^{t,k,\#} & \cdots & c_{(M-f+1),f_2}^{t,k,\#} & \cdots & c_{(M-f+1),(M-f+1)}^{t,k,\#} \end{pmatrix}. \quad (18)$$

Then, we apply pooling operation to the convolution feature matrix to further reduce the dimensionality and extract salient features. Let the filter matrix size of the pooling layer be denoted as $p \times p$. The stride of the filter matrix in the pooling layer is given by s , and the size of the feature matrix after the pooling operation is $D \times D$. We can express this matrix as:

$$P_{\#}(\mathbf{v}_{t,k}) = \begin{pmatrix} z_{1,1}^{t,k,\#} & \cdots & z_{1,d_2}^{t,k,\#} & \cdots & z_{1,D}^{t,k,\#} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{d_1,1}^{t,k,\#} & \cdots & z_{d_1,d_2}^{t,k,\#} & \cdots & z_{d_1,D}^{t,k,\#} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{D,1}^{t,k,\#} & \cdots & z_{D,d_2}^{t,k,\#} & \cdots & z_{D,D}^{t,k,\#} \end{pmatrix}. \quad (19)$$

Two calculation methods exist for performing the pooling operation: average pooling and maximum pooling. The corresponding formulas are:

$$z_{d_1,d_2}^{t,k,\#} = \sum \left\{ c_{[1+(d_1-1)s]:[p+(d_1-1)s], [1+(d_2-1)s]:[p+(d_2-1)s]}^{t,k,\#} \right\} / p^2, \quad (20)$$

$$d_1, d_2 = 1, 2, \dots, D, \text{ and}$$

$$z_{d_1,d_2}^{t,k,\#} = \max \left\{ c_{[1+(d_1-1)s]:[p+(d_1-1)s], [1+(d_2-1)s]:[p+(d_2-1)s]}^{t,k,\#} \right\}, \quad (21)$$

$$d_1, d_2 = 1, 2, \dots, D.$$

In this paper, we implement the maximum pooling method. We use the maximum pooling method because it can reduce the dimensionality of images by extracting only the most salient features of the data, increase the translation invariance of the representation, lower the computational cost by reducing the number of parameters to learn, and prevent overfitting by providing an abstracted form of the representation.

Flatten the obtained $P_{\#}(\mathbf{v}_{t,k})$, and then obtain a one-dimensional vector composed of $D \times D$ elements: $\mathbf{Z}_{t,k}^{\#} = (z_{1,1}^{t,k,\#}, \dots, z_{1,D}^{t,k,\#}, \dots, z_{D,1}^{t,k,\#}, \dots, z_{D,D}^{t,k,\#})$. The one-dimensional vectors obtained from above time series imaging and CNN processing for each IMF series are spliced together to form a fully connected layer. Hence, a new one-dimensional vector composed of $D \times D \times K$ elements is generated as follows:

$$\mathbf{Z}_t^{\#} = (z_{1,1}^{t,1,\#}, \dots, z_{D,D}^{t,1,\#}, \dots, z_{1,1}^{t,2,\#}, \dots, z_{D,D}^{t,2,\#}, \dots, z_{1,1}^{t,K,\#}, \dots, z_{D,D}^{t,K,\#}). \quad (22)$$

5) FORECASTING THROUGH BGRU

The number of subsequences used for training is defined as the ‘‘Image Lag Period’’ and denoted by J . The outputs of the fully connected layer in the previous step $\mathbf{Z}_{t-j+1}^{\#}, \dots, \mathbf{Z}_{t-1}^{\#}, \mathbf{Z}_t^{\#}$ for $j = 1, 2, \dots, J$, are passed to a BGRU network [33], with a structure illustrated in Fig. 3(a), and a single cell of the network presented in Fig. 3(b). The network consists of three parts: a reset gate, an update gate, and a candidate hidden state. The computation formulas for each component in Fig. 3(b) are as follows:

$$\mathbf{r}_t = \sigma(\mathbf{W}_{h,r}\mathbf{h}_{t-1} + \mathbf{W}_{i,r}\mathbf{I}_t + \mathbf{b}_r), \quad (23)$$

$$\mathbf{u}_t = \sigma(\mathbf{W}_{h,u}\mathbf{h}_{t-1} + \mathbf{W}_{i,u}\mathbf{I}_t + \mathbf{b}_u), \quad (24)$$

$$\tilde{\mathbf{h}}_t = \tanh[\mathbf{W}_{r,c}(\mathbf{r}_t \otimes \mathbf{h}_{t-1}) + \mathbf{W}_{i,c}\mathbf{I}_t + \mathbf{b}_c], \quad (25)$$

$$\mathbf{h}_t = (1 - \mathbf{u}_t) \otimes \mathbf{h}_{t-1} + \mathbf{u}_t \otimes \tilde{\mathbf{h}}_t, \quad (26)$$

where \mathbf{r}_t , \mathbf{u}_t , and $\tilde{\mathbf{h}}_t$ represent the reset gate, update gate, and candidate hidden state, respectively. \mathbf{h}_t represents the hidden state. The symbol \otimes denotes element-wise multiplication. \mathbf{W} represents the weight vectors for each component, and \mathbf{b} represents the bias terms for each component. The function $\sigma(x) = 1/[1 + \exp(-x)]$ is the sigmoid function, and $\tanh(x) = [\exp(x) - \exp(-x)]/[\exp(x) + \exp(-x)]$ is the hyperbolic tangent function. Moreover, the term $(1 - \mathbf{u}_t) \otimes \mathbf{h}_{t-1}$ determines the information to be retained in the previous hidden state, and $\mathbf{u}_t \otimes \tilde{\mathbf{h}}_t$ determines the information to be retained in the candidate hidden state.

From Fig. 3(a), it can be observed that the BGRU connects the output state values of the hidden layer from the GRU in two different directions, and the expression is given by:

$$\mathbf{H}_t = \overrightarrow{\mathbf{h}}_t \oplus \overleftarrow{\mathbf{h}}_t, \quad (27)$$

where $\overrightarrow{\mathbf{h}}_t$ and $\overleftarrow{\mathbf{h}}_t$ represent the forward and backward states of the hidden layer in the GRU at the time t , the symbol \oplus represents the concatenation operation, and \mathbf{H}_t is the final output value of the BGRU. Finally, after passing through the fully connected layer, the output value can be obtained as:

$$\hat{x}_{t+1} = \mathbf{W}_x \mathbf{H}_t + b_x, \quad (28)$$

where \mathbf{W}_x and b_x denote the weight and the bias term, respectively, of the final fully connected layer.

Since the model is trained on normalized data, inverse normalization operation is required to obtain the forecasted value of the crude oil price at the next period, \hat{y}_{t+1} . The formula is given as:

$$\hat{y}_{t+1} = \hat{x}_{t+1} \times y_t, \quad (29)$$

where y_t represents the actual crude oil price at the period t .

C. TRAINING OF THE CNN-BGRU MODEL

In this study, the CNN-BGRU model is optimized using the Adam algorithm. The Adam algorithm is selected due to its faster convergence and superior learning performance when compared to other adaptive learning rate algorithms. Moreover, the Adam algorithm overcomes the limitations observed

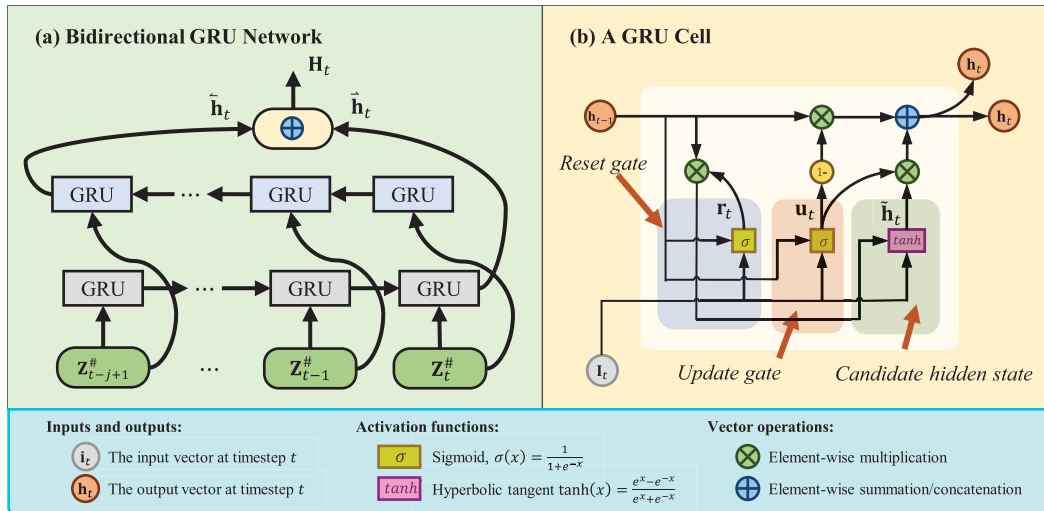


FIGURE 3. BGRU and its unit structure.

in alternative optimization techniques, such as unstable learning rates, slow convergence, and large fluctuations in the loss function resulting from high-variable parameter updates. For a comprehensive understanding of the Adam algorithm, we refer readers to the work of Kingma and Ba [49].

We employ grid search to determine an optimal initial learning rate, and use exponential decay to multiply the learning rate by 0.9 every 20 iterations. This approach accounts for the changes in the training process, as a fixed learning rate may not effectively adapt. As training progresses, a smaller learning rate becomes necessary to ensure convergence to a good local optimum and prevent oscillation. Exponential decay, a widely used method, allows the learning rate to decrease exponentially with the number of iterations, enabling smooth adaptation to the evolving training process.

We divide the training data into small batches of size 32 and calculate the loss function and gradient on each batch. The choice of this batch size balances model performance improvement and computational cost reduction. A smaller batch size can increase the generalization ability of the model, avoid falling into local optima, and accelerates each iteration. Conversely, a larger batch size may cause the model to converge to sharp minima, which compromises generalization. By selecting 32 as the batch size, we satisfy the memory requirements of most CPUs and GPUs while maintaining a suitable balance.

We set a maximum number of iterations to 800, which is an empirical value that can achieve a balance between ensuring model convergence and avoiding overfitting. Insufficient iterations may result in underfitting, where the model fails to fully capture the data’s underlying features and patterns. Conversely, an excessive number of iterations may lead to overfitting, causing the model to struggle with adapting to new data. To address overfitting, we employ an early stopping strategy. If the validation set’s loss function fails to improve for 40 consecutive iterations, we halt the training process.

This adaptive approach dynamically adjusts the number of iterations based on the model’s actual performance, preventing overfitting and unnecessary computations.

D. EVALUATION METRICS

To assess and compare the forecast performance of various models, this study employs four evaluation metrics to gauge the accuracy of the forecasts: absolute error (AE_t), mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and determination coefficient R^2 . Smaller values of AE_t , MAE, RMSE, and MAPE indicate better forecast accuracy, with predicted values being closer to actual values, and reflecting the high forecast performance of the model. Conversely, a larger value of the determination coefficient R^2 suggests a better training result of the model. The calculation formulas for these metrics are provided below:

$$AE_t = |\hat{y}_t - y_t|, \tag{30}$$

$$MAE = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t|, \tag{31}$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2}, \tag{32}$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{y}_t - y_t}{y_t} \right|, \tag{33}$$

$$R^2 = 1 - \frac{\sum (\hat{y}_t - y_t)^2}{\sum (y_t - \bar{y})^2}, \tag{34}$$

where \hat{y}_t , y_t and \bar{y} represent the forecasted value, the actual value, and the mean actual value, respectively.

To assess whether the proposed method exhibits superior forecasting accuracy compared to a benchmark model from a statistical perspective, we employ Diebold-Mariano (DM)

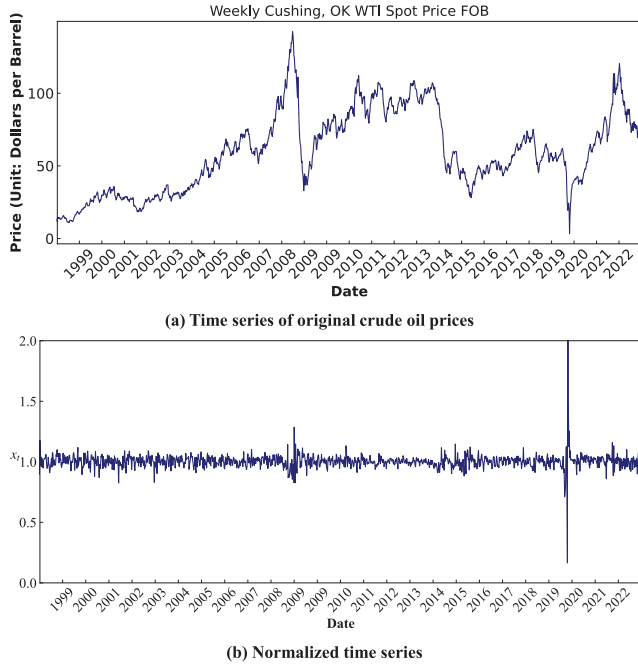


FIGURE 4. Time series data of crude oil price.

test [50]. The null and alternative hypotheses for this test are formulated as follows:

$$H_0 : \mathbb{E} [L(e_t^1)] = \mathbb{E} [L(e_t^2)], \quad (35)$$

$$H_1 : \mathbb{E} [L(e_t^1)] \neq \mathbb{E} [L(e_t^2)], \quad (36)$$

where $L(\cdot)$ is loss function of forecasting, e_t^1 and e_t^2 are the forecasting errors of the two models. The null hypothesis H_0 posits that model 1 and model 2 have approximately the same forecast performance, whereas the alternative hypothesis H_1 suggests that there is a significant difference in forecasting accuracy between model 1 and model 2. The DM test is calculated by

$$DM = \frac{\frac{1}{T} \sum_{t=1}^T [L(e_t^1) - L(e_t^2)]}{\sqrt{S^2/T}} S^2, \quad (37)$$

where S^2 is an estimator of the variance $[L(e_t^1) - L(e_t^2)]$. If $DM < 0$, the model 1 is superior to the model 2; otherwise, the model 2 is superior to the model 1.

IV. EXPERIMENTS AND ANALYSIS

A. DATA

The time series data of international crude oil prices is obtained from the U.S. Energy Information Administration (EIA) website (https://www.eia.gov/dnav/pet/pet_pri_spt_s1_w.htm). The dataset comprises weekly spot prices free on board (FOB) of the Oklahoma Cushing (Cushing, OK) WTI crude oil, spanning from June 19, 1998, to May 12, 2023. Table 1 presents the descriptive statistics of the dataset, which consists of 1300 data points with a mean of 59.89 and

TABLE 1. Descriptive statistic.

Item	Value
count	1300
mean	59.89
std	27.48
min	3.32
25%	36.73
50%	57.98
75%	80.77
max	142.52

a variance of 27.48. The range of values spans from 3.32 to 142.52. Fig. 4 (a) illustrates the line graph of the crude oil price time series, while Fig. 4 (b) visualizes the normalized series calculated using (1).

B. EXPERIMENTAL ENVIRONMENT AND EXPERIMENTAL DESIGN

The hardware environment for this experiment consists of an Intel Xeon Platinum 8255C @ 2.50GHz CPU with 12 cores, an NVIDIA GeForce RTX 3090 GPU, and an ubuntu20.04 system with 43GB of memory. The development environment includes Python 3.8, Cuda 11.2, and TensorFlow 2.9.0.

The dataset from September 18, 1998, to March 8, 2019, containing 1069 weeks, is chosen as the training set for model training. The dataset from March 15, 2019, to June 11, 2021, containing 118 weeks, is selected as the validation set to determine the optimal hyperparameters. The dataset from June 18, 2021, to May 12, 2023, containing 100 weeks, is used as the testing set to evaluate the performance of the model. During training, the maximum number of training epochs is set to 800, batch size to 32, and the loss function to MSE.

C. SELECTION OF HYPERPARAMETERS

The proposed models involve several hyperparameters, including K , image size, image lag period, filter number, kernel size, GRU units, and initial learning rate. In this study, we utilize random grid search to determine the optimal hyperparameters by conducting multiple experiments and comparing the performance of the validation set. The selected hyperparameters for the proposed forecasting models are presented in Table 2.

D. RESULTS AND DISCUSSION

This section is divided into comparative experiments and ablation experiments.

1) COMPARISON WITH EXISTING METHODS

The experiments conducted in this section compare the proposed method with the following benchmark models. MA: Moving Average model, ARIMA: Autoregressive Integrated Moving Average model, SARIMA: Seasonal Autoregressive Integrated Moving Average model, and HWES: Holt-Winters Exponential Smoothing model.

TABLE 2. The architecture and the training hyperparameters of VMD-GAF-BGRU, VMD-MTF-BGRU and VMD-RP-BGRU proposed in this paper.

Hyperparameter	Search space	VMD-GAF-BGRU	VMD-MTF-BGRU	VMD-RP-BGRU
K	[2,3,4,5,6]	2	2	2
Image size	[16,18,20,22,24,26,30,40,50,60,80,100,200]	40	80	100
Image lag period	[1,2,3,4]	1	1	1
Filter number	[4,5,10,50,100]	100	50	50
Kernel size	[2,3,4,5]	5	5	5
GRU units	[4,8,16,32]	4	32	16
Initial learning rate	[0.01,0.005,0.001,0.0005]	0.005	0.01	0.01

TABLE 3. Comparison with existing time series forecasting methods.

Benchmark	MAE	MSE	MAPE	R^2
MA	4.454	31.49	0.05185	0.8336
ARIMA	4.003	26.12	0.04659	0.8620
SARIMA	3.822	24.68	0.04458	0.8696
HWES	3.800	23.46	0.04439	0.8761
VMD-GAF-BGRU	2.995	12.66	0.03497	0.9326
VMD-MTF-BGRU	2.937	13.46	0.03428	0.9284
VMD-RP-BGRU	2.429	10.94	0.02938	0.9418

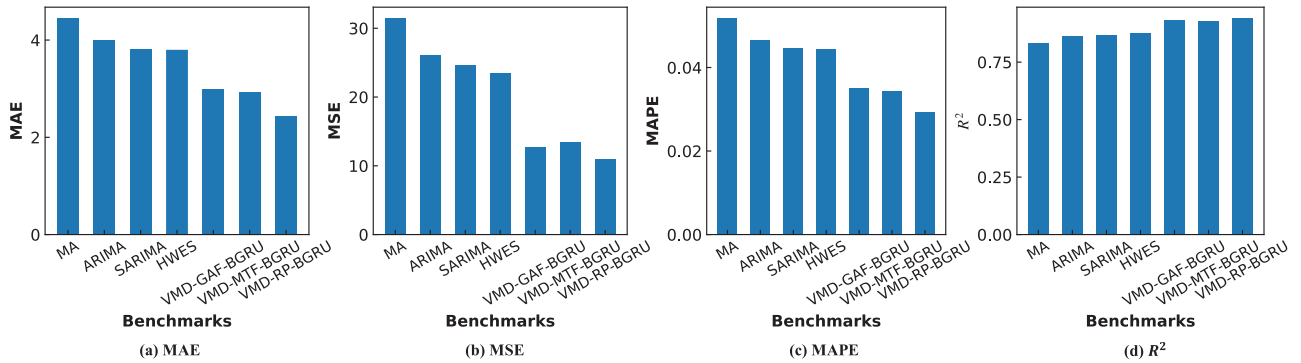


FIGURE 5. Evaluation of performance for different models.

- MA: The MA model is a widely employed technique for smoothing time series data and reducing random fluctuations. It calculates the moving average based on past data to derive the forecasted value.
- ARIMA: The ARIMA model is a general and flexible approach that can capture various patterns and structures in time series data. By integrating autoregression (AR) and differencing, the ARIMA model extends the capabilities of the MA model, enabling the elimination of seasonality and trends present in the data.
- SARIMA: The SARIMA model is suitable for data that exhibits periodic or cyclical patterns over time. It is an extension of the ARIMA model that takes into account seasonal factors.
- HWES: The HWES model is a popular and robust method that can handle data with both additive and multiplicative seasonality. By applying exponential smoothing to historical data, the HWES model generates a weighted average that effectively captures changes in trends.

According to different time series imaging techniques used, we propose three models: VMD-GAF-BGRU,

VMD-MTF-BGRU, and VMD-RP-BGRU, based on our methodology. These models have the same components of VMD and deep learning, but differ in their utilization of GAF, MTF, and RP algorithms for generating time series images. The experimental results, highlighting the best-performing values in bold, are presented in Table 3. To facilitate an easier comparison between the models, we provide Fig. 5. From Table 3 and Fig. 5, the following can be observed:

In terms of MAE and MAPE metrics, the forecasting performance ranking is: VMD-RP-BGRU > VMD-MTF-BGRU > VMD-GAF-BGRU > HWES > SARIMA > ARIMA > MA. In terms of MSE and R^2 metrics, the forecasting performance ranking is: VMD-RP-BGRU > VMD-GAF-BGRU > VMD-MTF-BGRU > HWES > SARIMA > ARIMA > MA. The symbol $A > B$ indicates that method A outperforms method B. In summary, based on any of the evaluation metrics, our proposed VMD-RP-BGRU model demonstrates the best predictive performance, followed by the VMD-MTF-BGRU and VMD-GAF-BGRU models.

There are several possible reasons for the superior performance of our proposed method, which combines VMD, time-series imaging, and BGRU: 1) The nonlinearity and

TABLE 4. The DM test results of the comparative experimental results.

DM Test	DM value	p-value	Result
VMD-GAF-BGRU v.s. MA	-5.47	0.000***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-GAF-BGRU v.s. ARIMA	-2.83	0.005***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-GAF-BGRU v.s. SARIMA	-2.22	0.026**	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-GAF-BGRU v.s. HWES	-2.40	0.016**	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-MTF-BGRU v.s. MA	-5.12	0.000***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-MTF-BGRU v.s. ARIMA	-2.47	0.014**	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-MTF-BGRU v.s. SARIMA	-1.91	0.057*	Accept H_0 , decline H_1 (Statistically insignificant difference.)
VMD-MTF-BGRU v.s. HWES	-2.03	0.042**	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-RP-BGRU v.s. MA	-4.35	0.000***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-RP-BGRU v.s. ARIMA	-3.20	0.001***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-RP-BGRU v.s. SARIMA	-2.76	0.006***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-RP-BGRU v.s. HWES	-2.74	0.006***	Accept H_1 , decline H_0 (Statistically significant difference.)

Note: The mark ***, **, and * indicate the 1%, 5% and 10% significance levels, respectively.

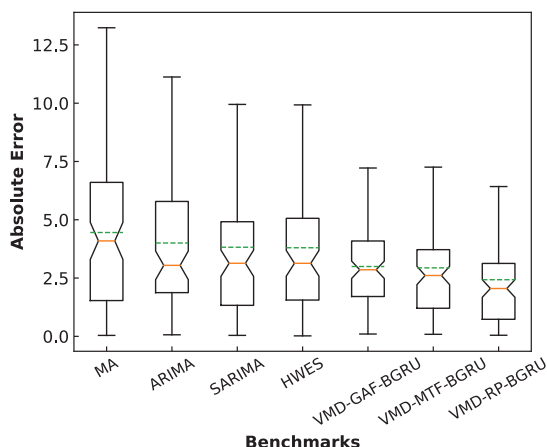


FIGURE 6. Boxplot of comparative experimental results.

complexity of crude oil price data can be effectively addressed by VMD, which decomposes the complex non-linear signal into a set of basic mode functions representing different frequency oscillatory components in the signal. By decomposing the signal, local features and patterns can be better captured, thereby improving the representation capability of the signal. 2) Time-series imaging leverages patterns and trends from historical data for prediction. By combining time-series imaging with VMD, modeling and prediction can be performed on the basic mode functions obtained from the decomposition, thereby capturing long-term dependencies and trends of the signal more accurately. 3) Bidirectional GRU utilizes both past and future context information, enabling a more comprehensive capture of the nonlinear relationships in the time series data. By applying bidirectional GRU to the prediction of the basic mode functions, it combines the local features obtained from the decomposition with the global dependencies, leading to improved prediction accuracy.

HWES, SARIMA, ARIMA, and MA are commonly used methods in time series forecasting. The possible reasons for the differences in their forecasting performance are as follows: 1) HWES and SARIMA exhibit stronger modeling capabilities for trends and seasonality compared to ARIMA

and MA. HWES, through exponential smoothing, effectively captures trend and seasonal changes in the data. It adapts the smoothing parameter adaptively, better accommodating trend and seasonal patterns in the data. In contrast, while SARIMA can handle seasonal data, its modeling capability may be limited when dealing with complex or irregular seasonal patterns. 2) HWES smoothes the data through exponential smoothing, eliminating some noise and reducing data volatility. This helps extract trend and seasonal components from the data, resulting in smoother prediction outcomes. In contrast, SARIMA requires differencing operations when dealing with noise and non-stationary data, and larger fluctuations or noise may affect its forecasting performance. ARIMA and MA are more suitable for modeling stationary time series, and their forecasting performance may be relatively poorer for non-stationary data.

We calculate the absolute errors of each model at different points in the test set to facilitate statistical comparison. We present boxplots of the absolute errors for each benchmark, as shown in Fig. 6. The solid orange line on each box represents the median absolute error of the corresponding model's test points, while the dashed green line represents MAE. Further statistical analysis reveals the following findings:

Firstly, VMD-GAF-BGRU, VMD-MTF-BGRU, and VMD-RP-BGRU achieve reductions in MAE of 21.64%, 23.15%, and 36.46%, respectively, compared to SARIMA. They also outperform HWES with reductions of 21.18%, 22.70%, and 36.08% in MAE, respectively. Moreover, these three models demonstrate superiority over SARIMA in 61%, 62%, and 65% of cases, and outperform HWES in 58%, 64%, and 65% of cases.

Secondly, all three of our proposed forecasting models exhibit exceptional predictive performance. This is attributed to the effective utilization of VMD's advantages in handling non-linear and non-stationary signals, the integration of time series imaging and convolutional neural network feature extraction, and the ability of BGRU to capture local non-linear patterns in time series. Among the existing time series prediction methods, SARIMA and HWES perform relatively well. This may be attributed to their ability to capture both the scale variations and seasonal patterns in time series.

TABLE 5. Comparison of ablation experiments.

Benchmark	MAE	MSE	MAPE	R^2
BGRU	3.709	26.38	0.04347	0.8597
VMD-RP	3.269	18.68	0.03771	0.9006
VMD-BGRU	3.350	18.73	0.03880	0.9004
RP-BGRU	3.201	18.81	0.03692	0.9000
VMD-RP-GRU	2.986	14.93	0.03490	0.9206
VMD-RP-BGRU	2.429	10.94	0.02938	0.9418

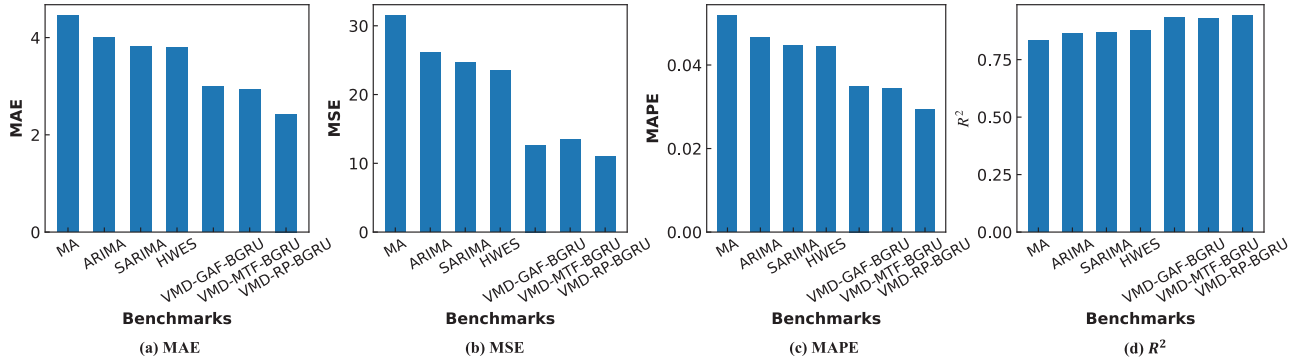


FIGURE 7. Evaluation of performance for different ablation benchmarks.

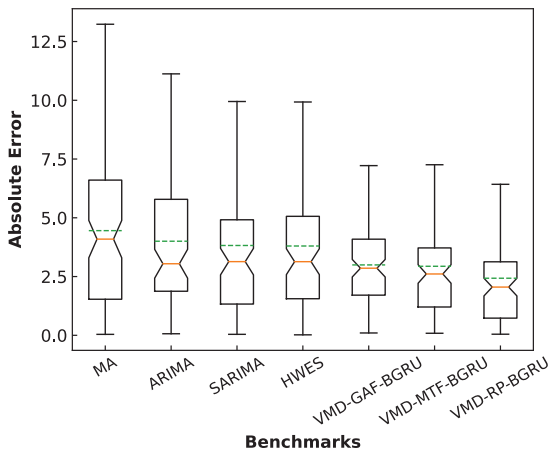


FIGURE 8. Boxplot of ablation experimental results.

To evaluate the superiority of our methodology, we conduct DM tests to determine whether the forecasting accuracy of our proposed models is significantly superior to that of benchmarks [51], [52], [53]. Table 4 presents the results of the DM tests conducted on existing forecasting models. When the p -value is greater than or equal to 0.05, the null hypothesis is accepted (Accept H_0 , decline H_1), indicating that the two models have similar performances. When the p -value is less than 0.05, the null hypothesis is rejected (Accept H_1 , decline H_0), indicating that the two models have significantly different performances. It can be observed that, except for the “VMD-MTF-BGRU v.s. SARIMA” comparison, which does not meet the conditions to reject the null hypothesis (at a 10% confidence level), all other DM tests exhibit significance at the 1% or 5% level. These results can strongly indicate

the significant superiority of our proposed methodology over existing methods in general. Thus, our findings underscore the effectiveness and advancement of our proposed methodology for forecasting crude oil prices.

2) ABLATION EXPERIMENTS

To illustrate the contributions of each component of the proposed forecasting framework, we conduct ablation experiments by removing certain parts of the best-performing model, VMD-RP-BGRU. The constructed ablation experiment benchmarks are as follows:

- VMD-RP-GRU: This model replaces the BGRU network in VMD-RP-BGRU with a unidirectional GRU network.
- RP-BGRU: This model directly performs time series imaging without the VMD process, followed by CNN and BGRU networks to generate the predictions.
- VMD-BGRU: This model decomposes the original subsequences using VMD into multiple IMF series but skips the time series imaging step, directly inputting the IMF series into the BGRU network to generate the predictions.
- VMD-RP: This model performs VMD, time series imaging, and CNN feature extraction operations, and then directly outputs the predictions through a fully connected layer.
- This model omits the VMD process, time series imaging, and CNN feature extraction operations. Instead, it directly trains the BGRU network on the original subsequences to generate the predictions.

Table 5 presents the comparison results of the ablation experiments, with the best-performing results highlighted

TABLE 6. The DM test results of the ablation experimental results.

DM Test	DM value	<i>p</i> -value	Result
VMD-RP-BGRU v.s. BGRU	-3.97	0.000***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-RP-BGRU v.s. VMD-RP	-2.62	0.009***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-RP-BGRU v.s. VMD-BGRU	-2.79	0.005***	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-RP-BGRU v.s. RP-BGRU	-2.34	0.019**	Accept H_1 , decline H_0 (Statistically significant difference.)
VMD-RP-BGRU v.s. VMD-RP-GRU	-1.90	0.057*	Accept H_0 , decline H_1 (Statistically insignificant difference.)

Note: The mark ***, **, and * indicate the 1%, 5% and 10% significance levels, respectively.

in bold. Fig. 7 provides a more intuitive visualization of the predictive performance of different ablation benchmarks. Similarly, we have plotted the boxplot of the absolute errors of the ablation benchmarks at different points in the test set in Fig. 8 for a more in-depth analysis.

From Table 5 and Fig. 7, it can be observed that removing any component of our method leads to a decrease in predictive performance. Through analysis of Fig. 8, the following findings are obtained: (1) The MAE, MSE, and MAPE of VMD-RP-BGRU are 27.51%, 41.62%, and 24.29% lower than those of VMD-BGRU, respectively, and the R^2 of VMD-RP-BGRU is 4.61% higher than that of VMD-BGRU. This indicates that the use of time series imaging (RP) significantly improves the predictive performance of crude oil price time series. (2) The MAE, MSE, and MAPE of VMD-RP-BGRU are 24.13%, 41.85%, and 20.43% lower than those of RP-BGRU, respectively, and the R^2 of RP-BGRU is 4.65% higher than that of VMD-BGRU. This indicates that the use of VMD also significantly improves the predictive performance of crude oil price time series. (3) The ablation benchmark without using VMD or RP time series imaging, i.e., BGRU, exhibits the worst predictive performance. (4) Introducing the GRU network, particularly the BGRU network, significantly improves the prediction performance. Specifically, using a GRU network (i.e., VMD-RP-GRU) reduces the MAE, MSE, and MAPE of the model's predictions by 8.66%, 20.09%, and 7.47%, respectively, with an R^2 improvement of 2.22%, compared to VMD-RP. The adoption of BGRU network (i.e., VMD-RP-BGRU) further reduces the MAE, MSE, and MAPE of the model compared to using the GRU network (i.e., VMD-RP-GRU) by 18.67%, 26.77%, and 15.8%, respectively, and increases the R^2 by 2.31%.

Based on Fig. 8 and further statistical analysis, we can observe that in terms of absolute value errors, VMD-RP-BGRU outperforms BGRU, VMD-RP, VMD-BGRU, RP-BGRU, and VMD-RP-GRU in 85%, 85%, 90%, 82%, and 82% of cases, respectively. This result further confirms the necessity and rationality of the different components constructed in our proposed methodology.

In summary, from the perspective of MAE and MAPE, the contribution of each component to improving predictive ability can be ranked as follows: time series imaging + CNN, BGRU, and VMD. While in terms of MSE and R^2 , the rank is: VMD, time series imaging + CNN, and BGRU. These results may arise from the varying sensitivity of different evaluation metrics to the nature and biases of the predictions, leading to different performances of the components across different

metrics. MAE and MAPE are absolute error metrics that are more sensitive to the differences between predicted values and true values. MSE and R^2 are squared error metrics that are more sensitive to the deviations and goodness of fit between predicted values and true values. Specifically:

1) **Time series imaging + CNN**: By transforming time series data into images and applying CNN for feature extraction and modeling, the patterns and trends in the time series data can be better captured. This helps to reduce the absolute error between predicted values and true values, thus lowering MAE and MAPE.

2) **BGRU**: BGRU can capture the long-term dependencies and contextual information in time series data. By utilizing both past and future context, BGRU provides a more comprehensive data representation, improving the accuracy of predictions and reducing MAE and MAPE. However, BGRU may not perform as well as VMD and time series imaging + CNN in terms of goodness of fit, resulting in a comparatively inferior ranking in MSE and R^2 .

3) **VMD**: VMD decomposes the signal into a set of basic mode functions, extracting the main components of the signal. Although VMD performs well in capturing local features (for MSE and R^2), it may be less sensitive to absolute error metrics such as MAE and MAPE, resulting in a lower ranking in these metrics.

Table 6 presents the results of DM tests conducted on the forecast errors of ablation benchmarks in the test set. It can be observed that, except for the comparison between "VMD-RP-BGRU v.s. VMD-RP-GRU" where the *p*-value does not meet the criteria for rejecting the null hypothesis, all other model comparisons exhibit statistically significant differences.

V. CONCLUSION

This paper proposes a novel approach for crude oil price forecasting based on VMD, time series imaging, and deep learning. The main purpose is to address the limitations of traditional forecasting methods in dealing with non-stationary, non-linear, and highly volatile crude oil price time series data. The original data are preprocessed through normalization, and then VMD is used to decompose the subsequences used for prediction into multiple stationary IMF series. These IMF series are converted into two-dimensional images using three different time series imaging techniques: RP, GAF, and MTF. CNN is employed to extract features from the images, and BGRU is used for prediction. Experimental evaluations are conducted on publicly available crude oil price

datasets, comparing the proposed method with commonly used forecasting methods such as SARIMA and HWSE. The results demonstrate significant advantages of our proposed approach, particularly the VMD-RP-BGRU model, which achieves superior forecasting performance across multiple evaluation metrics including MAE, MSE, MAPE, and R^2 . Specifically, our approach outperforms the benchmark methods, with the VMD-RP-BGRU model achieving the best forecasting performance, exhibiting MAE=2.429, MSE=10.94, MAPE=2.94%, and R^2 =0.9418. It showcases substantial improvements of 21.64%, 23.15%, and 36.46% in MAE, MSE, and MAPE, respectively, compared to the SARIMA model, as well as reductions of 21.18%, 22.70%, and 36.08% compared to the HWSE model.

The proposed approach offers several advantages and notable contributions that can be summarized in three key aspects: (1) The application of time series imaging techniques in crude oil price forecasting is introduced for the first time. By transforming one-dimensional series into two-dimensional image matrices, the approach uncovers additional inherent feature information, thereby enhancing the accuracy of crude oil price forecasting. (2) Three novel hybrid models, namely VMD-RP-BGRU, VMD-GAF-BGRU, and VMD-MTF-BGRU, are proposed, incorporating different time series imaging methods. These models effectively combine the strengths of VMD, time series imaging, CNN, and BGRU. (3) This study contributes to the expansion of research in crude oil price prediction theory and enriches the repertoire of deep learning-based time series forecasting methods, thus contributing to theoretical advancements in the field.

However, this study has limitations that can serve as directions for future work. The proposed approach is univariate and does not consider other exogenous variables. Future research can incorporate several exogenous variables, such as news media and search engine data, to further enhance the accuracy of predictions.

This study holds significant implications for various stakeholders including investors, crude oil producers, consumers, and policy-makers. For investors, accurate forecasts can inform investment decisions and risk management strategies. Crude oil producers can utilize reliable price forecasts to optimize production plans and resource allocation. Consumers, such as industries heavily reliant on crude oil, can better anticipate price fluctuations and adjust their operations accordingly. Additionally, policy-makers can benefit from accurate forecasts when formulating energy policies, regulating market activities, and promoting economic stability. Overall, this study contributes to advancing the field of time series forecasting and offers valuable tools for decision-making in the energy sector.

ACKNOWLEDGMENT

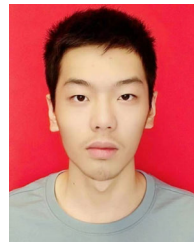
The authors would like to thank the editor, the associate editor, and the reviewers for their constructive comments,

which have led to a dramatic improvement of the earlier version of this article.

REFERENCES

- [1] IEA. (Sep. 10, 2021). *Key World Energy Statistics 2021*. International Energy Agency, Paris, France. [Online]. Available: <https://www.iea.org/reports/key-world-energy-statistics-2021>
- [2] K. Joo, J. H. Suh, D. Lee, and K. Ahn, "Impact of the global financial crisis on the crude oil market," *Energy Strategy Rev.*, vol. 30, Jul. 2020, Art. no. 100516.
- [3] Bureau of Labor Statistics. (May 2015). *The 2014 Plunge in Import Petroleum Prices: What Happened?* U.S. Dept. Labor, Washington, DC, USA. [Online]. Available: <https://www.bls.gov/opub/btn/volume-4/pdf/the-2014-plunge-in-import-petroleum-prices-what-happened.pdf>
- [4] J. Yuan, J. Li, and J. Hao, "A dynamic clustering ensemble learning approach for crude oil price forecasting," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106408.
- [5] J. Guo, Z. Zhao, J. Sun, and S. Sun, "Multi-perspective crude oil price forecasting with a new decomposition-ensemble framework," *Resour. Policy*, vol. 77, Aug. 2022, Art. no. 102737.
- [6] Y. Guo, F. Ma, H. Li, and X. Lai, "Oil price volatility predictability based on global economic conditions," *Int. Rev. Financial Anal.*, vol. 82, Jul. 2022, Art. no. 102195.
- [7] B. Sévi, "Forecasting the volatility of crude oil futures using intraday data," *Eur. J. Oper. Res.*, vol. 235, no. 3, pp. 643–659, Jun. 2014.
- [8] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, VIC, Australia: OTexts, 2018.
- [9] X.-Y. Jiang, Y.-F. Cheng, W.-Q. Zhu, Y.-H. Bai, L.-Z.-J. Xu, X.-Q. Wu, and R.-C. Jin, "Effect of chromium on granule-based anammox processes," *Bioresource Technol.*, vol. 260, pp. 1–8, Jul. 2018.
- [10] Y. Baek and H. Y. Kim, "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module," *Exp. Syst. Appl.*, vol. 113, pp. 457–480, Dec. 2018.
- [11] J. Cao, Z. Li, and J. Li, "Financial time series forecasting model based on CEEMDAN and LSTM," *Phys. A, Stat. Mech. Appl.*, vol. 519, pp. 127–139, Apr. 2019.
- [12] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018.
- [13] S. Borovkova and I. Tsiamas, "An ensemble of LSTM neural networks for high-frequency stock market classification," *J Forecast.*, vol. 38, no. 6, pp. 600–619, Mar. 2019.
- [14] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Exp. Syst. Appl.*, vol. 178, Sep. 2021, Art. no. 115019.
- [15] S. W. Lee and H. Y. Kim, "Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation," *Exp. Syst. Appl.*, vol. 161, Dec. 2020, Art. no. 113704.
- [16] T. B. Shahi, A. Shrestha, A. Neupane, and W. Guo, "Stock price forecasting with deep learning: A comparative study," *Mathematics*, vol. 8, no. 9, p. 1441, Aug. 2020.
- [17] X. Li and P. Tang, "Stock index prediction based on wavelet transform and FCD-MLGRU," *J. Forecasting*, vol. 39, no. 8, pp. 1229–1237, Dec. 2020.
- [18] Y. Xiang and X. H. Zhuang, "Application of ARIMA model in short-term prediction of international crude oil price," *Adv. Mater. Res.*, vols. 798–799, pp. 979–982, Sep. 2013.
- [19] S. C. Hillmer and G. C. Tiao, "An ARIMA-model-based approach to seasonal adjustment," *J. Amer. Stat. Assoc.*, vol. 77, no. 377, pp. 63–70, Mar. 1982.
- [20] C.-L. Zhao and B. Wang, "Forecasting crude oil price with an autoregressive integrated moving average (ARIMA) model," in *Proc. Fuzzy Inf. Eng. Operations Res. Manag.*, Berlin, Germany, 2014, pp. 275–286.
- [21] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London, Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
- [22] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, pp. 1–41, Jan. 2009.

- [23] F. Zhou, Z. Huang, and C. Zhang, "Carbon price forecasting based on CEEMDAN and LSTM," *Appl. Energy*, vol. 311, Apr. 2022, Art. no. 118601.
- [24] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147, doi: [10.1109/ICASSP.2011.5947265](https://doi.org/10.1109/ICASSP.2011.5947265).
- [25] J. Wang, X. Li, T. Hong, and S. Wang, "A semi-heterogeneous approach to combining crude oil price forecasts," *Inf. Sci.*, vols. 460–461, pp. 279–292, Sep. 2018.
- [26] T. Zhang, Z. Tang, J. Wu, X. Du, and K. Chen, "Multi-step-ahead crude oil price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm," *Energy*, vol. 229, Aug. 2021, Art. no. 120797.
- [27] L. Yu, W. Dai, and L. Tang, "A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting," *Eng. Appl. Artif. Intell.*, vol. 47, pp. 110–121, Jan. 2016.
- [28] H. Abdollahi, "A novel hybrid model for forecasting crude oil price based on time series decomposition," *Appl. Energy*, vol. 267, Jun. 2020, Art. no. 115035.
- [29] Y. Huang and Y. Deng, "A new crude oil price forecasting model based on variational mode decomposition," *Knowl.-Based Syst.*, vol. 213, Feb. 2021, Art. no. 106669.
- [30] J. Sun, P. Zhao, and S. Sun, "A new secondary decomposition-reconstruction-ensemble approach for crude oil price forecasting," *Resour. Policy*, vol. 77, Aug. 2022, Art. no. 102762.
- [31] S. Zhang, J. Luo, S. Wang, and F. Liu, "Oil price forecasting: A hybrid GRU neural network based on decomposition-reconstruction methods," *Exp. Syst. Appl.*, vol. 218, May 2023, Art. no. 119617.
- [32] K. Zheng, P. Li, S. Zhou, W. Zhang, S. Li, L. Zeng, and Y. Zhang, "A multi-scale electricity consumption prediction algorithm based on time-frequency variational autoencoder," *IEEE Access*, vol. 9, pp. 90937–90946, 2021.
- [33] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734.
- [34] G. R. Govind and A. S. Babu, "Hybrid deep learning model to forecast crude oil price," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, Lalitpur, Nepal, Apr. 2023, pp. 19–23.
- [35] J. Nasir, M. Aamir, Z. U. Haq, S. Khan, M. Y. Amin, and M. Naem, "A new approach for forecasting crude oil prices based on stochastic and deterministic influences of LMD using ARIMA and LSTM models," *IEEE Access*, vol. 11, pp. 14322–14339, 2023.
- [36] X. Li, W. Shang, and S. Wang, "Text-based crude oil price forecasting: A deep learning approach," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1548–1560, Oct. 2019.
- [37] M. Li, Z. Cheng, W. Lin, Y. Wei, and S. Wang, "What can be learned from the historical trend of crude oil prices? An ensemble approach for crude oil price forecasting," *Energy Econ.*, vol. 123, Jul. 2023, Art. no. 106736.
- [38] Q. Zhu, F. Zhang, S. Liu, Y. Wu, and L. Wang, "A hybrid VMD-BiGRU model for rubber futures time series forecasting," *Appl. Soft Comput.*, vol. 84, Nov. 2019, Art. no. 105739.
- [39] Y. Fang, W. Wang, P. Wu, and Y. Zhao, "A sentiment-enhanced hybrid model for crude oil price forecasting," *Exp. Syst. Appl.*, vol. 215, Apr. 2023, Art. no. 119329.
- [40] J.-W. Bi, H. Li, and Z.-P. Fan, "Tourism demand forecasting with time series imaging: A deep learning model," *Ann. Tourism Res.*, vol. 90, Sep. 2021, Art. no. 103255.
- [41] W. Tian, J. Wu, H. Cui, and T. Hu, "Drought prediction based on feature-based transfer learning and time series imaging," *IEEE Access*, vol. 9, pp. 101454–101468, 2021.
- [42] G. Zhang and J. Guo, "A novel ensemble method for hourly residential electricity consumption forecasting by imaging time series," *Energy*, vol. 203, Jul. 2020, Art. no. 117858.
- [43] R. Kadari, Y. Zhang, W. Zhang, and T. Liu, "CCG supertagging via bidirectional LSTM-CRF neural architecture," *Neurocomputing*, vol. 283, pp. 31–37, Mar. 2018.
- [44] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- [45] X. Li, Y. Kang, and F. Li, "Forecasting with time series imaging," *Exp. Syst. Appl.*, vol. 160, Dec. 2020, Art. no. 113680.
- [46] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhys. Lett. (EPL)*, vol. 4, no. 9, pp. 973–977, Nov. 1987.
- [47] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proc. 24th Int. Conf. Artif. Intell.*, 2015, pp. 3939–3945.
- [48] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
- [49] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.
- [50] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 13, no. 3, pp. 253–263, Jul. 1995.
- [51] E. Zhao, P. Du, and S. Sun, "Historical pattern recognition with trajectory similarity for daily tourist arrivals forecasting," *Exp. Syst. Appl.*, vol. 203, Oct. 2022, Art. no. 117427.
- [52] K. Ijaz, Z. Hussain, J. Ahmad, S. F. Ali, M. Adnan, and I. Khosa, "A novel temporal feature selection based LSTM model for electrical short-term load forecasting," *IEEE Access*, vol. 10, pp. 82596–82613, 2022.
- [53] T. Mutavhatsindi, C. Sigauke, and R. Mbuva, "Forecasting hourly global horizontal solar irradiance in South Africa using machine learning models," *IEEE Access*, vol. 8, pp. 198872–198885, 2020.



ZI-JIAN PENG is currently pursuing the bachelor's degree in industrial engineering with Northeastern University, China. His research interests are forecasting, supply chain management, and revenue management.



CHUAN ZHANG received the Ph.D. degree in control theory and control engineering from Northeastern University (NEU), Shenyang, China, in 2002. He is currently a Professor with the School of Business Administration, NEU. He has published over 30 articles in leading journals, including *International Journal of Forecasting*, *Annals of Operations Research*, *Information Sciences*, *Journal of Cleaner Production*, and *Expert Systems with Applications*. His current research interests include data-driven operations management, big data mining and forecasting, sustainable operations management, green supply chain management, and platform operations management.



YU-XIN TIAN is currently pursuing the Ph.D. degree in management science and engineering with Northeastern University, China. He has published over 15 articles in leading journals, including *International Journal of Forecasting*, *Annals of Operations Research*, *Journal of Cleaner Production*, and *Expert Systems with Applications*. His research interests are data-driven operational research, artificial intelligence, machine learning, forecasting, sustainable operations management, and green supply chain management.