

RESEARCH

Open Access



Deep learning systems for forecasting the prices of crude oil and precious metals

Parisa Foroutan¹ and Salim Lahmiri^{1*}

*Correspondence:
salim.lahmiri@concordia.ca

¹ Department of Supply Chain and Business Technology Management, Concordia University, Montreal, Canada

Abstract

Commodity markets, such as crude oil and precious metals, play a strategic role in the economic development of nations, with crude oil prices influencing geopolitical relations and the global economy. Moreover, gold and silver are argued to hedge the stock and cryptocurrency markets during market downsides. Therefore, accurate forecasting of crude oil and precious metals prices is critical. Nevertheless, due to the nonlinear nature, substantial fluctuations, and irregular cycles of crude oil and precious metals, predicting their prices is a challenging task. Our study contributes to the commodity market price forecasting literature by implementing and comparing advanced deep-learning models. We address this gap by including silver alongside gold in our analysis, offering a more comprehensive understanding of the precious metal markets. This research expands existing knowledge and provides valuable insights into predicting commodity prices. In this study, we implemented 16 deep- and machine-learning models to forecast the daily price of the West Texas Intermediate (WTI), Brent, gold, and silver markets. The employed deep-learning models are long short-term memory (LSTM), BiLSTM, gated recurrent unit (GRU), bidirectional gated recurrent units (BiGRU), T2V-BiLSTM, T2V-BiGRU, convolutional neural networks (CNN), CNN-BiLSTM, CNN-BiGRU, temporal convolutional network (TCN), TCN-BiLSTM, and TCN-BiGRU. We compared the forecasting performance of deep-learning models with the baseline random forest, LightGBM, support vector regression, and k-nearest neighborhood models using mean absolute error (MAE), mean absolute percentage error, and root mean squared error as evaluation criteria. By considering different sliding window lengths, we examine the forecasting performance of our models. Our results reveal that the TCN model outperforms the others for WTI, Brent, and silver, achieving the lowest MAE values of 1.444, 1.295, and 0.346, respectively. The BiGRU model performs best for gold, with an MAE of 15.188 using a 30-day input sequence. Furthermore, LightGBM exhibits comparable performance to TCN and is the best-performing machine-learning model overall. These findings are critical for investors, policymakers, mining companies, and governmental agencies to effectively anticipate market trends, mitigate risk, manage uncertainty, and make timely decisions and strategies regarding crude oil, gold, and silver markets.

Keywords: Crude oil forecasting, Precious metal forecasting, Deep learning, Temporal convolutional networks, Time2Vector, LightGBM

Introduction

Nonrenewable commodities usually mined in certain countries can strongly impact their economies, policies, currencies, and international or political issues. Energy and precious metals markets, among other commodities, are well-known alternatives to stock markets (Pullen et al. 2014; Hussain Shahzad et al. 2017; Akbar et al. 2019; Adekoya et al. 2022; Phan et al. 2016; Sarwar et al. 2019). Their prices are critical indicators of economic health and crucial determinants for financial planning and decision making. In this regard, understanding the dynamics of such markets and forecasting their evolutions is crucial for portfolio optimization and management. Crude oil, a crucial energy commodity, is pivotal in global macroeconomics and influences the decisions made by policymakers like governments and central banks. Fluctuations in crude oil prices have profound implications for a country's political and economic security; therefore, accurate crude oil price forecasting is imperative. Crude oil market shocks in April 2020 and their impacts have increased interest in understanding oil price dynamics (Wang et al. 2021; Murshed and Tanha 2021; Balcilar et al. 2021; Zhang et al. 2022a, b; Enwereuzoh et al. 2021). Conversely, gold is important for investment portfolio diversification and hedging (ben Khelifa et al. 2021; Reboredo 2013; Baek 2019). Gold contributes a large portion of the commodity reserves of major economies. As of September 2022, the official United States (US) gold reserve was 8133.47 tons, approximately 66.6% of total US reserves.¹

Given these markets' multifaceted nature, forecasting the trajectories of these commodities is crucial in financial markets, serving as an essential tool for investors, policymakers, and analysts. For investors, anticipating price movements in crude oil and precious metals provides a strategic advantage in optimizing portfolio performance and risk management. A comprehensive understanding of potential price fluctuations allows investors to make informed decisions, allocate resources optimally, and ultimately enhance their overall financial returns (Bhowmik and Wang 2020). In contrast, policymakers rely on accurate market forecasts to develop effective economic policies and mitigate the potential impact of market volatility on national economies. Fluctuations in crude oil prices, for instance, can have cascading effects on inflation, trade balances, and overall economic stability (Uzo-Peters et al. 2018; Xiuzhen et al. 2022; Periwal 2023). Similarly, precious metal prices often indicate broader economic sentiments and can influence monetary policies and international trade relationships.

In this context, the science of forecasting plays a pivotal role in providing foresight into future trends in crude oil and precious metal prices. Advanced analytical models (Kou et al. 2021, 2022; Li et al. 2022a, b; Lahmiri 2023a), statistical methods (Lahmiri et al. 2022; Lahmiri 2023b), machine learning (Lahmiri et al. 2023), and deep-learning algorithms (Amirifar et al. 2023; Amirshahi and Lahmiri 2023a, b; Lahmiri and Bekiros 2019, 2020, 2021) enable analysts to search through vast datasets, identify patterns, and make predictions that are invaluable for both short-term traders and long-term investors (Abdullah Ahmed and Bin Shabri 2014; Zhao et al. 2015; Das et al. 2022; Jiang et al. 2022; Liang et al. 2023). Driven by this motivation, this study investigates forecasting

¹ <https://www.gold.org/goldhub/data/gold-reserves-by-country>.

methodologies within the domains of crude oil and precious metals markets to enhance the precision of price predictions.

Recent innovations in deep learning models seem promising for time-series forecasting; however, the crude oil and precious metals forecasting literature struggles to use these models for price prediction. This study attempts to fill this gap in the forecasting literature by applying several deep- and machine-learning models to predict the daily closing prices of crude oil, gold, and silver. First, the time-series data of daily spot prices of two prominent crude oils, West Texas Intermediate (WTI) and Brent, and two precious metal markets, gold and silver, are gathered and normalized. Then, several input sequences are prepared using the sliding window method with four different window lengths. Next, the dataset is split into training, validation, and test sets using a time-based splitting approach. Finally, a comprehensive set of 16 forecasting models, consisting of 12 deep-learning models, 2 baseline-ensemble models, and 2 baseline machine-learning models, is implemented to predict the next-day market price. The deep learning models used in the current study include long short-term memory (LSTM), bidirectional LSTM (BiLSTM), gated recurrent units (GRU), bidirectional GRU (BiGRU), Time2Vector BiLSTM (T2V-BiLSTM), Time2Vector BiGRU (T2V-BiGRU), convolutional neural networks (CNN), hybrid CNN-BiLSTM, hybrid CNN-BiGRU, temporal convolutional networks (TCN), hybrid TCN-BiLSTM, and hybrid TCN-BiGRU models. Two baseline ensemble models are the random forest and LightGBM gradient-boosting models, and two baseline machine-learning models are the support vector regression (SVR) and k-nearest neighborhood (KNN) models.

Each of the employed models has its strengths and limitations. LSTM models are a type of recurrent neural networks (RNN) that are popular for their ability to capture long-term dependencies, overcome the gradient vanishing problem, and handle variable-length sequences; however, LSTMs can be computationally expensive and prone to overfitting, requiring regularization techniques (Yu et al. 2019). GRU models, another type of RNN, have a simpler architecture, resulting in faster training and inference times; however, they may have limitations in capturing complex patterns compared with LSTM models. Bidirectional models, such as BiLSTM or BiGRU, consider both forward and backward information, making them more robust to variations in the input sequence order; however, they are computationally complex and require more memory resources (Khan et al. 2021). CNNs are effective at capturing local patterns and features within time-series data. CNNs learn filters to detect specific temporal patterns and are translation invariant, meaning they can detect patterns regardless of their position in the input sequence; however, CNNs have limitations, such as the requirement for fixed-length inputs, limited consideration of temporal ordering, and the ability to capture long-term dependencies. Hybrid CNN–LSTM models combine the strengths of both CNNs and LSTMs, capturing spatial and temporal features. They are suitable for tasks that require capturing complex patterns in time-series data; however, they can be less interpretable than standalone models (Gharghory 2021). TCNs are designed to capture long-term dependencies efficiently. They use dilated convolutions to capture information from several past time steps. TCNs are adaptable to different time-series lengths without padding or truncation; however, they can be complex to design and tune and are sensitive to input scaling (Gopali et al. 2021). Ensemble machine-learning models such as random forest

and LightGBM are also used in time-series analysis. Random forest combines multiple decision trees and offers high prediction accuracy and robustness against outliers. LightGBM is an efficient gradient-boosting framework that effectively handles large datasets. Both models have their accuracy and generalization strengths but cannot explicitly capture temporal dependencies (Ke et al. 2017). SVR is a flexible model that can capture linear and nonlinear relationships; it focuses on support vectors, which greatly influence the model's decision boundary. SVR can handle high-dimensional datasets and complex relationships between variables; however, the performance of SVR depends on selecting appropriate hyperparameters, and it does not explicitly model temporal dependencies. KNN is an instance-based algorithm that makes predictions based on the similarity of training instances; it requires no training phase but suffers from the curse of dimensionality and cannot capture temporal dependencies.

Our paper compares the forecasting performance of these models by mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE) error functions. This paper primarily aims to answer the following questions through empirical experiments. (1) What is the best deep-learning model that can predict crude oil, gold, and silver spot prices reliably and precisely? (2) In response to the first question, does a particular model outperform other models for crude oil and precious metals prices? (3) Which input sequence length is more informative for each market's price prediction? (4) Are hybrid models effective in forecasting crude oil, gold, and silver spot prices? (5) What conclusions about the properties of each deep-learning model can be drawn in the context of crude oil and precious metals time-series forecasting?

The arrangement of the rest of this manuscript is as follows. “[Literature review](#)” section provides an overview of the relevant prior research and summarizes our contributions to the existing literature. “[Methodology](#)” section explains the methods and performance evaluation criteria used in this study. “[Empirical analysis and results](#)” section describes the datasets, demonstrates the results, and discusses our findings. Finally, “[Conclusion](#)” section summarizes the paper and presents some managerial implications and policy suggestions.

Literature review

Accurately forecasting financial markets is a critical guide for determining economic policies. Consequently, researchers have dedicated their efforts to developing and improving models that capture the intrinsic behavior and dynamics of financial market time series. The prediction methods used in these studies generally comprise statistical or econometrics, machine learning, and deep-learning methods. Several forecasting modeling approaches have recently been applied to crude oil and precious metals. For instance, Zhao et al. (2018) proposed a numerical vector trend forecasting method for predicting the daily spot price of Brent crude oil, outperforming traditional models such as autoregressive integrated moving average (ARIMA), SVR, and wavelet analysis models. Similarly, Szarek et al. (2020) proposed a new stochastic distribution, skewed Student's t-distribution, for silver, copper, and gold time-series estimation, which accounts for the time-dependent parameters and non-Gaussian behavior of time-series data.

Drachal (2022) employed the Bayesian symbolic regression method to address variable uncertainty in monthly crude oil price forecasting.

Due to the nonlinearity, nonstationarity, and heteroscedasticity of crude oil and precious metal markets, classical statistical forecasting models such as vector autoregressive (VAR), ARIMA, and autoregressive distributed lag (ARDL) struggle to perform well in forecasting tasks. These models make assumptions about the normality and stationarity of price data, which often do not hold for many time-series data for commodity markets. As a result, recent studies have used machine- and deep-learning models, which excel in handling nonlinear data and do not rely on the normality assumption for accurate price predictions. In the literature, three main types of deep neural networks are used for sequence modeling, and they can be applied for time-series forecasting (Lim and Zohren 2021). These networks include (i) RNNs and their variants, such as LSTM (Hochreiter and Schmidhuber 1997) and GRUs (Cho et al. 2014), (ii) CNNs (Lecun et al. 1998) and their recent variant, TCN (Lea et al. 2016), and (iii) transformer (Vaswani et al. 2017) and its variants (Devlin et al. 2018; He et al. 2020; Liu et al. 2019).

Several studies used statistical, machine learning, and deep-learning models to account for the importance of gold price forecasting. Alameer et al. (2019) used a multi-layer perceptron model with a whale optimization algorithm for gold next-month price forecasting. This model demonstrates a lower forecasting error than ARIMA model forecasts. Madziwa et al. (2022) employed an ARDL model to forecast annual gold prices using lagged gold prices, gold demand, and treasury bill rates as predictors. In another study, Zhang and Ci (2020) used the US Consumer Price Index, crude oil price, exchange rate, and Dow Jones Industrial Price Index in a deep belief network to predict monthly gold prices. Risse (2019) predicted gold excess returns to the risk-free rate of return using the ana SVR model. SVR finds the nonlinear relationship in the data by mapping a linear function into a high-dimensional feature space. Tree-based ensemble models have demonstrated promising performance in forecasting gold prices. Yuan (2023) leveraged the XGBoost (Chen and Guestrin 2016) and LightGBM (Ke et al. 2017) models for gold and bitcoin price forecasting. Furthermore, deep-learning methods have been increasingly used for gold price prediction. For instance, using association rules and the LSTM mode, Boongasame et al. (2022) predicted the price of gold. Vidal and Kristjanpoller (2020) developed a hybrid of convolutional neural networks and long- and short-term memory models (CNN–LSTM), which incorporate historical log-return series and time-series data in an image format to predict the volatility of gold spot prices. Likewise, various studies have used deep-learning models for crude oil price forecasting. Orojo et al. (2019) employed a multirecurrent network to forecast a one-month ahead WTI crude oil price. Lin et al. (2022) forecasted crude oil futures prices using a BiLSTM-Attention-CNN model with wavelet transform. Swamy and Lagesh (2023) explored the effectiveness of investor sentiments from Twitter in predicting the daily gold price by a wavelet analysis method and unveiled a strong correlation between Twitter sentiments and the gold price. Fang et al. (2023a, b) forecasted Brent crude oil prices using an improved slope-based method based on empirical mode decomposition (EMD) and feedforward neural network (FNN) methods.

Conversely, the literature on forecasting other precious metal markets is relatively limited. Sroka (2022) utilizes block bootstrap methods to forecast daily silver prices,

while Salisu et al. (2020) tested the impact of Google Trends on forecasting the prices of four precious metal markets using an ARDL model. Zhang et al. (2022a, b) introduced a new objective function to forecast commodity markets, including silver prices. To our knowledge, there is no precedent study to forecast the silver price using machine- and deep-learning models. We attempt to fill this void in the literature.

Given the ongoing improvements in natural language processing (NLP) tasks, recent studies have incorporated news text and Google Trends features into their forecasting models. These approaches leverage the valuable information in the textual data to enhance the accuracy of predictions. For example, Li et al. (2019) extracted text data from online news media and created sentiment features that were grouped by their topics using a latent Dirichlet allocation method. Their topic-sentiment forecasting model shows that text features complement financial features for crude oil price forecasting. Similarly, Bai et al. (2022) constructed features from news headlines for WTI crude oil forecasting. Fang et al. (2023a, b) employed a FineBERT approach to extract sentiment information from crude oil-related news, which was then integrated into a hybrid attention-based BiGRU model for WTI price forecasting. Kerttly de Medeiros et al. (2022) demonstrated performance enhancement using a mixed data sampling model incorporating mixed-frequency data and a textual sentiment indicator for oil price forecasting. Salisu et al. (2020) utilized an econometric ARDL model to show that search engine data from Google Trends significantly positively affect precious metal returns. Similarly, Tang et al. (2020) considered Google Trends a useful predictor in a multivariate empirical mode decomposition method for forecasting Brent crude oil spot prices. Other EMD methods have been used by Wang et al. (2018), Qin et al. (2019), Yang et al. (2020), G. Li et al. (2022a, b), and Guo et al. (2022) in their proposed crude oil forecasting models. Liang et al. (2023) also used historical crude oil prices in a deep reinforcement learning algorithm to forecast multistep ahead WTI, Brent, and Oman prices. A recent review paper (Mohamed and Messaadia 2023) highlights that artificial neural networks and support vector machines (SVMs) are the most popular artificial intelligence techniques used to forecast crude oil prices. Collectively, these studies showcase the growing significance of advanced forecasting methods to enhance the accuracy and reliability of predictions in the crude oil and precious metal markets.

Some studies have achieved improved forecasting performances by developing ensemble models. Zhao et al. (2017) combined the advantages of stacked denoising autoencoders (SDAE) and bootstrap aggregation (bagging) techniques to model the nonlinear and complex relationships of oil price factors and to generate multiple data sets for training a set of base learners. Wang et al. (2020) proposed an ensemble of five linear and nonlinear submodes to produce the prediction intervals of crude oil spot prices while optimizing the weights of submodes using the gray wolf optimizer. Zhang et al. (2021) developed an ensemble deep-learning model for electricity price series prediction. Jiang et al. (2022) combined a decomposition-ensemble approach optimized by the seagull algorithm with sentiment analysis to forecast future crude oil prices. Su et al. (2022) proposed a hybrid forecasting model using SVM, extreme learning machines, XGBoost, and LSTM models to predict crude oil futures series. Sun et al. (2022) proposed a secondary decomposition–reconstruction–ensemble approach for crude oil price forecasting.

The temporal convolutional networks (TCNs) (Lea et al. 2016) are variants of CNN models that employ casual convolutions and dilations to predict sequential data with temporality and large receptive fields. A simple convolution can only look back at a fixed timing window, whereas a TCN uses dilated convolutions to achieve a large receptive field with fewer convolutional layers. TCNs capture long-term patterns using a hierarchy of temporal convolutional filters, and in that manner, they tend to outperform bidirectional LSTM models and are a magnitude faster to train. A TCN was first developed for action detection in video data settings to account for spatial and temporal input features (Lea et al. 2016). However, recently, TCNs have drawn more attention from scholars and have been applied to various time-series data. For instance, Lara-Benítez et al. (2020) utilized a TCN model to forecast electricity demand and prices in Spain. In the environmental milieu, Yan et al. (2020) predicted the El Niño-Southern Oscillation, an index measuring the earth's climate variability, by applying an ensemble empirical mode decomposition–TCN model. This model shows improved prediction performance compared with the LSTM model.

Considering temporal patterns in predicting time-series data is a significant challenge for many models. Some recent studies have introduced learnable time representations to account for temporal patterns in sequential data (Xu et al. 2019, 2021; Li et al. 2017). Among these studies, Kazemi et al. (2019) introduced the Time2Vector method to represent sequential data as periodic and nonperiodic vectors that can capture complex temporal patterns in data. Yang et al. (2021) improved the performance of an attention neural network for nonintrusive load monitoring by applying the Time2Vector method. This current study applies Time2Vector embedding to input series and incorporates the resulting periodic and nonperiodic features into several deep-learning models to forecast crude oil, gold, and silver prices. Table 1 summarizes the literature on crude oil and precious metal forecasting.

Gradient-boosting methods are powerful predictive models for many tasks. Borisov et al. (2021) compared the performance of tree-based ensembles, such as XGBoost, LightGBM, and CatBoost (Prokhorenkova et al. 2018), with some deep-learning models, including but not limited to multilayer perceptron, regularization learning networks, neural oblivious decision ensembles, and transformers. They assert that machine learning tree-based models outperform deep-learning models in several prediction tasks with tabular data; however, their study does not include deep-learning models for sequential data and is silent about forecasting financial market prices. To address this shortfall, in the current study, we will use tree-based ensemble models such as random forest and LightGBM compared with 12 deep-learning models and two other machine-learning models (KNN and SVR) to forecast daily crude oil and precious metals market prices.

This study makes significant contributions to the literature on forecasting commodity market prices.

- Considering that there is limited literature on using deep-learning models to forecast the price of commodity markets, this study implements and compares various types of state-of-the-art deep-learning models for crude oil and precious metal spot price forecasting. Hence, our study encompasses several forecasting results that provide comprehensive insights for crude oil, gold, and silver market players and investors.

Table 1 Literature review of crude oil and precious metal forecasting

Method type	Author(s)	Method(s)	Features		Data			
			Financial market	Google trend or sentiment	WTI	Brent	Gold	Silver
Statistical and Econometric methods	Zhao et al. (2018)	VTFM	✓			✓		
	Szarek et al. (2020)	SGT	✓				✓	✓
	Drachal (2022)	BSR	✓		✓	✓		
	Kertily de Medeiros et al. (2022)	MIDAS	✓	✓		✓		
	Salisu et al. (2020)	ARDL	✓	✓			✓	✓
	Tang et al. (2020)	MEMD	✓	✓		✓		
	Madziwa et al. (2022)	ARDL	✓				✓	
	Swamy and Lagesh (2023)	wavelet analysis method	✓	✓			✓	
Machine Learning methods	Sroka (2022)	block bootstrap-ARIMA	✓					✓
	Zhao et al. (2017)	Ensemble (SDAE-bagging)	✓		✓			
	Bai et al. (2022)	AdaBoost.RT	✓	✓	✓			
	Risse (2019)	SVR	✓				✓	
	Yuan (2023)	XGBoost & LightGBM	✓				✓	
	Wang et al. (2020)	EPPA	✓			✓		
	Su et al. (2022)	ensemble	✓		OPEC crude oil			
	Sun et al. (2022)	ensemble	✓		✓	✓		
Deep Learning methods	Li et al. (2019)	CNN	✓	✓	✓			
	Fang et al. (2023a, b)	FineBERT-VMD-Att-BiGRU	✓	✓	✓			
	Fang et al. (2023a, b)	ISBN-EMD-FNN	✓			✓		
	Liang et al. (2023)	Deepreinforcement learning	✓		✓	✓		
	Alameer et al. (2019)	MLP	✓				✓	
	Zhang and Ci (2020)	Deep belief network	✓				✓	
	Boongasame et al. (2022)	LSTM	✓				✓	

Table 1 (continued)

Method type	Author(s)	Method(s)	Features		Data			
			Financial market	Google trend or sentiment	WTI	Brent	Gold	Silver
	Vidal and Kristjanpoller (2020)	CNN-LSTM	✓				✓	
	Orojo et al. (2019)	MRN	✓		✓			
	Lin et al. (2022)	BiLSTM-Attention-CNN	✓		✓	✓		

- Most studies on precious metals focus only on gold price predictions; however, this study forecasts the price of both gold and silver to maintain a more general understanding of the precious metal markets.
- To the best of our knowledge, this study is the first in forecasting literature that applies the TCN model, Time2Vector embedding module, and hybrid TCN-BiLSTM and TCN-BiGRU models to forecast the spot price of WTI, Brent, Gold, and Silver time series.
- The forecasting period in the test dataset of this study, from 2020-01-03 to 2022-03-25, covers two critical global events that significantly affected financial markets. First, the financial crisis during the COVID-19 pandemic significantly impacted all financial markets; in particular, crude oil prices plunged in April 2020. Second, the Russia–Ukraine conflict in February 2022 was associated with a sharp rise in crude oil, gold, and silver prices. Therefore, the results of this study and the proposed models can be used during financial crises and extreme global situations. Figure 5 shows the line chart of the WTI, Brent, gold, and silver prices for reference.

Methodology

LSTM and BiLSTM

LSTM and BiLSTM are structural variants of RNN models that can remember important information from time-series sequences (Lin et al. 2022). In particular, BiLSTM concatenates two LSTM layers in opposite directions. The interior structure of a common LSTM cell is shown in Fig. 1a. An LSTM unit consists of an input gate, a forget gate, and an output gate. These gates facilitate information flow and help the cell forget unnecessary information. First, the forgetting gate decides what information from the inputs and previous hidden states to discard. Second, the input gate decides what information from the inputs and previous cell states to keep and updates the cell state. Finally, the output gate obtains the output h_t by multiplying the o_t of the input information processed by the sigmoid activation function and the cell state vector transformed by the tanh activation function. The equations of a forward pass in an LSTM unit are as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{1}$$

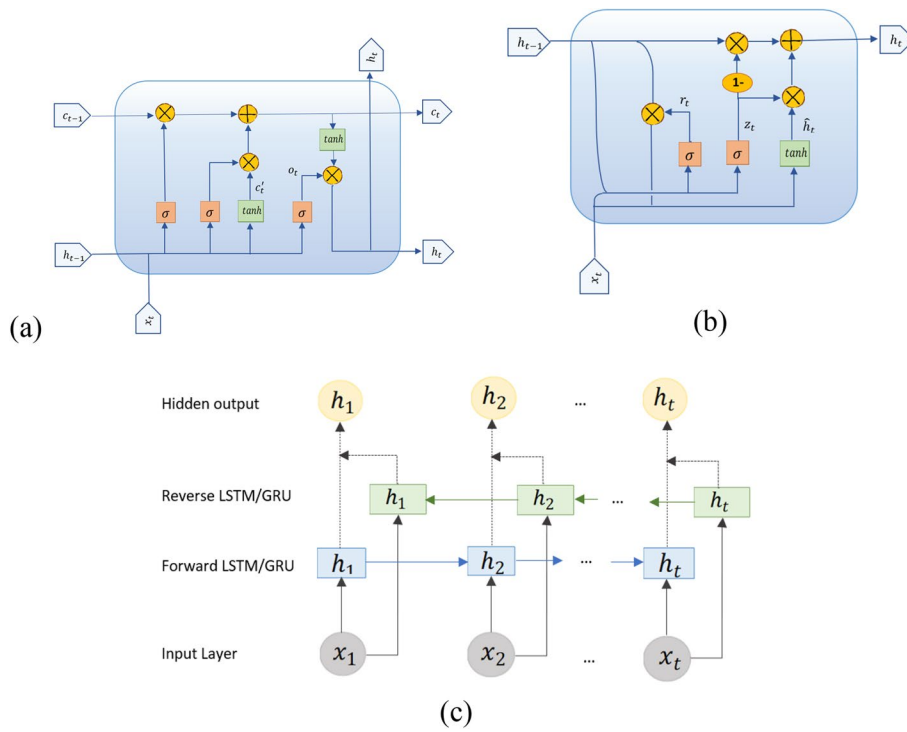


Fig. 1 **a** LSTM internal cell structure, **b** GRU internal cell structure, **c** A single layer BiLSTM or BiGRU model

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{2}$$

$$c'_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot c'_t \tag{4}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{5}$$

$$h_t = o_t \odot \tanh(c_t), \tag{6}$$

where $x_t \in \mathbb{R}^d$ is the input vector, and $h_t \in \mathbb{R}^h$ is the hidden state vector. Furthermore, f_t is the forget gate vector, i_t is the input gate vector, o_t is the output gate vector, c'_t is the temporary cell state vector, $c_t \in \mathbb{R}^h$ cell state vector, and $W \in \mathbb{R}^{h \times d}$, $U \in \mathbb{R}^{h \times h}$, and $b \in \mathbb{R}^h$ represent the parameter matrices and vectors.

In a BiLSTM model, from opposite directions, h_t is concatenated to construct the bidirectional hidden state. The formulas of bidirectional h_t are as follows:

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \tag{7}$$

$$\overleftarrow{h}_t = LSTM\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{8}$$

$$h_t = \left[\overrightarrow{h}_t, \overleftarrow{h}_t\right] \tag{9}$$

GRU and BiGRU

Like the LSTM, the GRU is a variant of RNN cells that can forget insignificant information and help the model use longer data sequences. GRU has fewer parameters than LSTM because it eliminates the output gate.

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{10}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{11}$$

$$\hat{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \tag{12}$$

$$h_t = z_t \odot \hat{h}_t + (1 - z_t) \odot h_{t-1} \tag{13}$$

where $x_t \in \mathbb{R}^d$ is the input vector, and $h_t \in \mathbb{R}^h$ hidden state vector. Additionally, z_t is the forget gate vector, r_t is the reset gate vector, \hat{h}_t is the candidate activation vector, $W \in \mathbb{R}^{h \times d}$, $U \in \mathbb{R}^{h \times h}$, and $b \in \mathbb{R}^h$ represent the parameter matrices and vectors, and σ is the sigmoid activation function. For certain sequential datasets, GRUs outperform LSTM models (Chung et al. 2014; Gruber and Jockisch 2020). The internal structure of the GRU cell is depicted in Fig. 1b.

For a bidirectional GRU model, hidden state vectors from two opposite directions are concatenated as follows:

$$\vec{h}_t = GRU\left(x_t, \vec{h}_{t-1}\right) \tag{14}$$

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{15}$$

$$h_t = \left[\vec{h}_t, \overleftarrow{h}_t\right] \tag{16}$$

Figure 1c shows the architecture of a single-layer bidirectional LSTM (BiLSTM) or bidirectional GRU (BiGRU) model.

CNN

A CNN is a FNN model proposed by Lecun et al. (1998). CNNs are very popular in computer vision applications, such as facial recognition systems, object localization, object detection, and semantic segmentation. CNNs are effective at capturing local patterns and features within a time series. The convolutional layers learn filters to detect specific

temporal patterns, making CNNs well suited for capturing local dependencies and short-term patterns in time-series data. CNNs are inherently translation invariant, meaning they can detect patterns regardless of their position in the input sequence. This property is helpful for time-series analysis because the same patterns may occur at different time steps. The local perception and weight sharing of CNN can significantly reduce the number of parameters, thus improving the efficiency of model learning (Lu et al. 2020); however, they suffer from limitations such as the requirement for fixed-length inputs, lack of consideration of temporal ordering, and limited ability to detect long-term temporal dependencies.

The architecture of this model is generally constructed from two layers: the convolution layer and the pooling layer. The convolution layer extracts useful features from the input series by applying several convolution kernels to the inputs, as indicated in Eq. 17, which downsamples the input for final forecasting. Then, a pooling layer is applied to the output of the convolution layer to reduce the dimensionality of the model.

$$l_t = \sigma(x_t * k_t + b_t) \quad (17)$$

where l_t is the output of the convolution layer, σ is the activation function, $x_t \in \mathbb{R}^d$ is the input vector, $k_t \in \mathbb{R}^d$ is the parameter vector of the convolution kernel, and b_t is the bias term.

TCN

The intrinsic weaknesses of CNN, including fixed-size inputs and mismatched input and output dimensions, restrict its application in time-series forecasting. The TCN (Lea et al. 2016) is a variant of the CNN that employs causal and dilated convolutions appropriate for sequential data with temporality and large receptive fields. Causal means no information leakage from the future to the past, and the receptive field means the set of sample elements of the original input that affect a specific element of the output. A TCN model can show full coverage of the input history by setting a proper dilated factor and kernel size. Furthermore, the TCN has a simple network structure and outperforms standard recurrent networks, such as the RNN and LSTM networks, regarding the effectiveness and efficiency of time-series predictions (Yan et al. 2020). Figure 2 shows a general representation of our TCN model with dilated causal convolutions. This model's architecture consists of the following.

Dilated convolution layer: The dilated convolution architecture modifies Kronecker-factored convolutional filters, enabling a larger receptive field with fewer parameters and layers (Zhou et al. 2015). For a sequence of $x_t \in \mathbb{R}^d$ and a filter $f : \{0, \dots, k-1\} \rightarrow \mathbb{R}$, the dilated convolution operation $*_D$ on entries s of the sequence is defined as follows:

$$F(s) = (x_t *_D f)(s) = \sum_{i=1}^{k-1} f(i) \cdot x_{s-D \cdot i} \quad (18)$$

where D is the dilation factor, k is the filter size, and $s - D \cdot i$ assures that only past data are convoluted. A tanh function transforms the output of the dilated causal convolution layer.

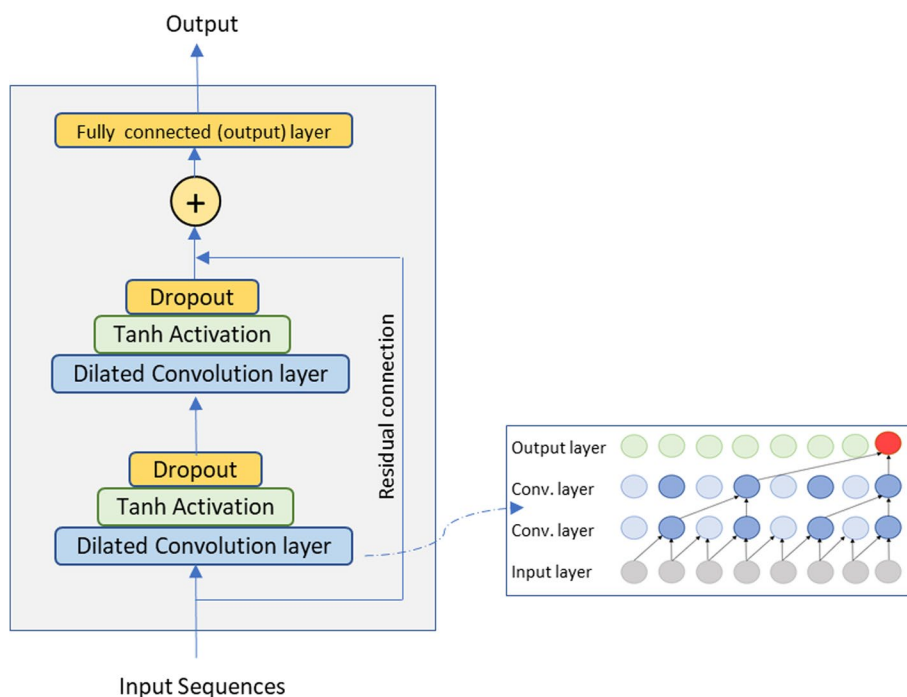


Fig. 2 (left) The architecture of a TCN model with a stack of two dilated causal convolutional layers and a residual connection. (right) a dilated causal convolution layer with dilated factors $D = \{1, 2, 4\}$ and kernel size $k = 2$

Dropout layer: A dropout layer with a probability of 0.2 is applied after each dilated convolution layer to regularize the model and eliminate the overfitting problem.

Residual block: We used a stack of two dilated causal convolution layers together, and the results from the final convolution were added back to the inputs to obtain the outputs of the block. The residual connection avoids the vanishing and/or exploding gradient problem in deep-learning models.

Fully connected layer: The output of the residual block is then inputted into a fully connected layer to predict the next-day price.

In Fig. 2, the TCN model has a stack of two layers, a residual connection, and a fully connected layer. Each layer in the stack has a dilated causal convolution, a tanh activation function, and a dropout for regularization. The dilation factors for the dilated convolution layer are $D = 1, 2, 4$ and a filter size of $k = 2$. When $D = 1$, the dilated convolution becomes a basic convolution.

In recurrent-type neural networks, operations apply sequentially. In contrast, in a TCN model, all sequences are convolved simultaneously in each dilated convolutional layer; hence, the training of TCN is much faster than in STM or GRU models (Lea et al. 2016).

Time2Vector (T2V-BiLSTM and T2V-BiGRU)

Time-series input can be considered a sequence in which a dependency across time exists among the sample data rather than being identically and independently

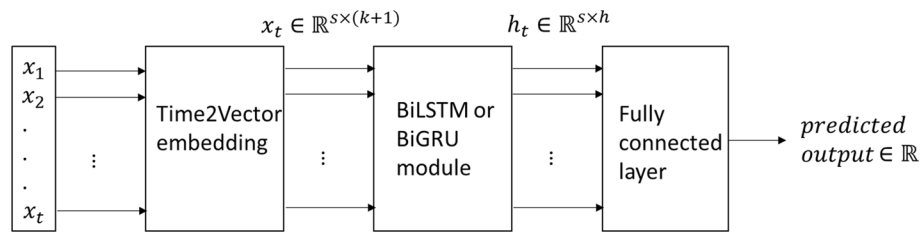


Fig. 3 T2V-BiLSTM or T2V-BiGRU models. s is the input sequence length, k is the T2V output size, h is the recurrent hidden size

distributed (i.i.d); therefore, it is essential to account for time features while developing a time-series forecasting model. Vector embedding has been successfully used in many NLP tasks (Pennington et al. 2014; Mikolov et al. 2013; Almeida and Xexéo 2019). Similarly, Time2Vector (Kazemi et al. 2019) is a learnable vector embedding for time that can be easily combined with many deep-learning models. Time2Vector is a decomposition technique that encodes a temporal signal into periodic and nonperiodic patterns, allowing the model to understand and learn from the time-dependent patterns. It eliminates the need for explicit feature engineering when dealing with time-related features. By incorporating temporal information meaningfully, Time2Vector can improve the performance of time-series models.

For a given scalar notion of time τ , Time2Vec of τ is a vector of size $k + 1$ defined as follows:

$$T2V(\tau)[i] = \begin{cases} w_i\tau + b_i, & \text{if } i = 0. \\ \mathcal{F}(w_i\tau + b_i), & \text{if } 1 \leq i \leq k. \end{cases} \tag{19}$$

where $T2V(\tau)[i]$ is the i^{th} element of $T2V(\tau)$. \mathcal{F} is a periodic activation function, and w and b are learnable weight and bias parameters, respectively. Following the indicated activation function in the original T2V paper (Kazemi et al. 2019), we use a sine function as \mathcal{F} . Time2Vector (T2V) assures that the time scale will not affect the learned periodic and nonperiodic time features (Yang et al. 2021).

To construct the T2V-BiLSTM and T2V-BiGRU models, first, the input sequences are transformed by Time2Vector embeddings, then the embedded input vectors are entered into a single-layer BiLSTM or BiGRU model, and finally, the output is predicted through a fully connected layer. Figure 3 presents a schematic of the T2V-BiLSTM or T2V-BiGRU model. Figure 4 summarizes the complete data preprocessing, model training, and prediction process for this study’s test set.

Hybrid models

To verify the applicability of hybrid models in forecasting daily crude oil, gold, and silver prices, we used CNN-BiLSTM, CNN-BiGRU, TCN-BiLSTM, and TCN-BiGRU models. CNNs in the initial layers of the hybrid model can learn low-level spatial features, such as local patterns, while the BiLSTM layers can learn high-level temporal dependencies. This hierarchical representation learning allows the model to capture local and global dependencies in the time-series data. CNNs and TCNs are well suited for feature

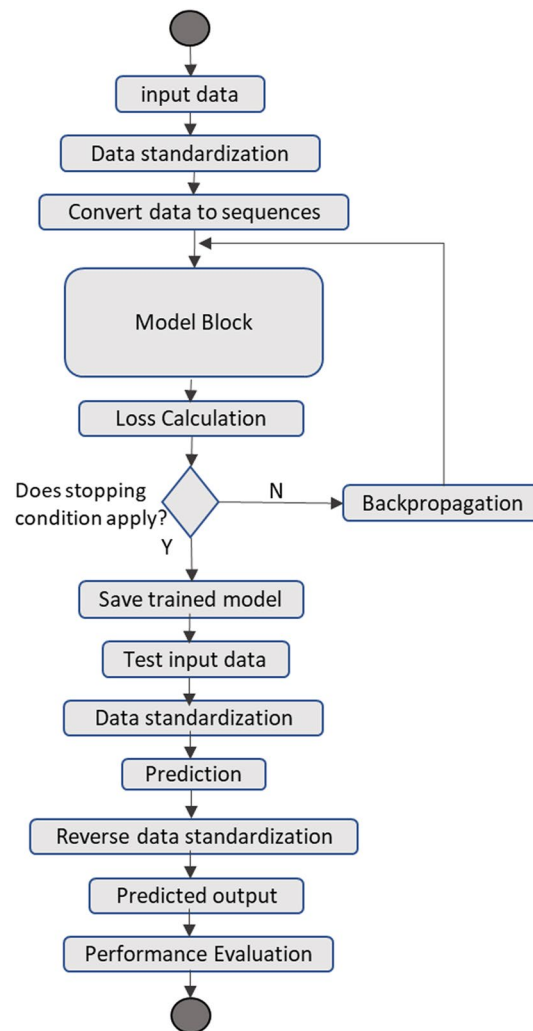


Fig. 4 The price time-series forecasting flow chart

extraction from raw data, including time-series data. They can automatically learn relevant features and reduce the dimensionality of the input, which can be beneficial for downstream BiLSTM or BiGRU layers to learn more meaningful representations. The explanation of each model structure is as follows.

CNN-BiLSTM and CNN-BiGRU models: First, a one-dimensional convolution layer is applied to input sequences in the CNN module. Then, a max pooling layer is applied to the output of the convolution layer to extract the essential features. Next, the output of the pooling layer is entered into a single-layer BiLSTM or BiGRU module, and the final output is predicted through a fully connected layer.

TCN-BiLSTM and TCN-BiGRU models: First, a TCN module receives the input sequences. Next, the output of the TCN is introduced into a single-layer BiLSTM or BiGRU module, and the final output is predicted through a fully connected layer.

Ensemble and machine-learning models

This study uses random forest and LightGBM, a gradient-boosting technique among the ensemble machine-learning models. Random forest generally provides high prediction accuracy because of the aggregation of multiple decision trees. It is less prone to overfitting than individual decision trees. By combining multiple trees and using techniques such as bagging and random feature selection, random forest reduces variance and improves the model's generalization ability. It is also robust to outliers and missing values; however, it lacks autocorrelation modeling because random forest treats each data point independently and does not explicitly consider the temporal dependencies between consecutive observations in the time series. Random forest is not well suited for extrapolation, especially for long-term forecasts; thus, it may be difficult to capture and project future trends extending beyond the observed data range. While random forest is generally robust to overfitting, it can still be sensitive to noisy data; it may overfit the noise if the dataset contains a substantial amount of noise or irrelevant features, leading to degraded performance.

LightGBM is a powerful and efficient gradient-boosting framework that performs excellently in various machine-learning tasks. LightGBM is highly efficient and can handle large datasets with millions of instances and features. It uses a histogram-based algorithm to achieve faster training and prediction times than traditional gradient-boosting implementations. The main advantage of LightGBM is low memory usage due to the use of a compact data structure for representing the dataset during training. Like other gradient-boosting algorithms, LightGBM can be prone to overfitting if not properly regularized or tuned. LightGBM may struggle to capture complex feature interactions compared with deep-learning models.

SVR is a machine-learning model that captures linear and nonlinear relationships between variables. It can handle high-dimensional datasets and capture complex relationships between variables. The algorithm focuses on the support vectors, the data points that influence the model's decision boundary most. Outliers have less impact on this model because of the use of a margin. SVR allows using different kernel functions, such as linear, polynomial, radial basis function, and sigmoid. This flexibility enables the modeling of various relationships between the input and target variables; however, SVR performance highly depends on selecting appropriate hyperparameters, such as kernel type, regularization parameter, and kernel-specific parameters. Training an SVR model can be computationally expensive, especially when dealing with large datasets or complex kernel functions. SVR does not account for the temporal dependencies among observations for time-series datasets.

KNN is an instance-based, nonparametric algorithm that uses different distance metrics, such as Euclidean distance, Manhattan distance, or cosine similarity, to make predictions. The KNN does not explicitly learn a model from the training data. Instead, it stores the entire training dataset and uses it during prediction, eliminating the need for a time-consuming training phase. As the number of training instances increases, the algorithm's prediction time can be significant because it requires calculating distances to all training samples. Some limitations of KNN models are the curse of dimensionality, sensitivity to the scale of features, intensive memory requirement, time-consuming predictions with large datasets, and lack of capturing temporal dependencies.

Evaluation criteria

This study adopts the following three metrics to calculate the forecasting error and evaluate the prediction performance: MAE, MAPE, and RMSE. MAE measures the difference between two continuous variables and calculates the mean value of all absolute errors. MAPE is a scaleless error value that measures the relative forecasting error. RMSE represents the standard deviation of the residual error between the predicted and observed values. The models' prediction performance increases with decreasing error measures. The formula for the above evaluation criteria is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (20)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (21)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (22)$$

where n is the sample size, and y_i and \hat{y}_i are the true and predicted values for sample i , respectively.

Empirical analysis and results

Data description and preprocessing

The daily closing prices of WTI and Brent crude oil, gold, and silver were collected from 2000-01-04 to 2022-03-25 (Fig. 5). The original spot price data for WTI and Brent crude oil are derived from the US Energy Information Administration (<https://www.eia.gov>), while the spot prices of gold and silver are from KITCO (<https://www.kitco.com>). We used data from the same trading days across all four markets to obtain an identical sample size for all time series.

To find the best hyperparameters and evaluate the models' real-world performances, evaluating them on a separate validation set and a test set representing future unseen data is essential. Splitting the time-series datasets is challenging because of temporal dependencies, seasonality, and trends. If we split the data randomly, it breaks the temporal order, and the model may be trained on future data, leading to data leakage and overfitting. Moreover, if the training set does not capture the full range of seasonality or fails to include representative trend patterns, the model's ability to generalize to unseen data may be compromised. Ensuring the training set contains consecutive past observations to predict future observations, includes multiple seasonal cycles, and adequately captures the underlying trends is crucial. Time-based splitting and rolling window approaches can address these challenges in time-series analysis. In time-based splitting, we split the data based on a specific date or time, ensuring that the training set only contains past observations and the test set contains future observations. In the rolling window approach, a sliding window is used to create samples in the training, validation, and test sets, where each sample includes past observations and the corresponding future

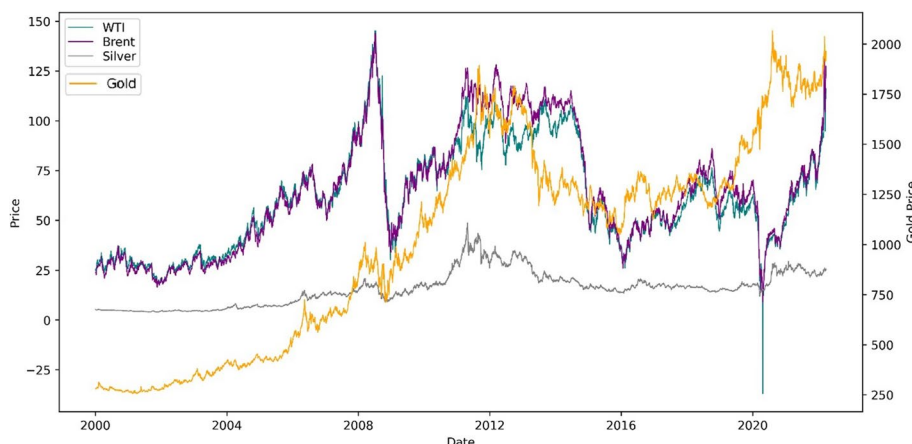


Fig. 5 WTI, and Brent crude oil, Gold and Silver price movements from 2000-01-04 to 2022-03-25

target observation. Thus, for each market, the entire dataset is split into three parts: 65% training data (from 2000-01-04 to 2014-06-15), 25% validation data (from 2014-06-16 to 2020-01-02), and 10% test data (from 2020-01-03 to 2022-03-25). The test data period includes the financial crisis due to the COVID-19 pandemic and the sharp decline in crude oil prices in April 2020. Therefore, test data include highly volatile price data, making forecasting even more challenging.

Since deep-learning models are sensitive to the scale of data, we normalized each dataset into [0,1] intervals to limit the effect of noise, speed up the updating of neural network parameters, and enhance the training performance of the model. The formula to standardize the data is as follows:

$$x'_t = \frac{x_t - \min(x_t)}{\max(x_t) - \min(x_t)} \tag{23}$$

where x_t and x'_t denote the data before and after standardization, respectively. Table 2 summarizes the sample’s descriptive statistics and statistical tests for WTI and Brent crude oil, gold, and silver. The total sample size for all markets is 5426. All four market spot prices show significant characteristics of skewness, while WTI, Brent, and gold also represent significant leptokurtic properties at a 5% significance level. Furthermore, the significant Jarque–Bera test statistics at a 1% significance level show that the WTI, Brent, gold, and silver price time series do not comply with the normal distribution; hence, these markets can be treated as nonstationary signals.

For these forecasting tasks, $x_t = \{x_1, x_2, \dots, x_s\}$ is the input vector, where x_i is the price data at day i and s is the sequence length (sliding window length), and $y_t = \{x_{s+1}\}$ is the target. We created inputs for different sequences before sending a series into the model. In this study, we train 16 deep- and machine-learning models with four different sliding window lengths of 5, 30, 60, and 90 days to predict the next-day WTI, Brent, gold, and silver prices. We have considered 5 as a relatively short sliding window length and 30, 60, and 90 as relatively long to capture any seasonality or trend in the data. We will compare deep- and machine-learning models to determine how they forecast commodity price time series with longer input sequences.

Table 2 Descriptive statistics

Index	WTI	Brent	Gold	Silver
Count	5426	5426	5426	5426
Mean	61.47	64.27	1020.98	15.73
Standard deviation	25.79	29.33	510.85	8.51
Min	− 36.98	9.12	255.95	4.07
Max	145.31	143.95	2067.15	48.7
Skewness ^a	0.39*	0.41*	− 0.07**	0.61*
Kurtosis ^b	− 0.598*	− 0.817*	− 1.249*	0.084
Jarque-Bera ^c	218.903*	301.375*	357.452*	342.924*

^a ¹Null hypothesis is that the series are not skewed

^b Null hypothesis is that the series show normal kurtosis

^c Null hypothesis is that the series are not normally distributed

*, ** denote the rejection of the null hypothesis at the 1% and 5% significance level, respectively

Empirical results

Crude oil and precious metals are essential commodities in financial markets. This study aims to forecast the daily price of WTI and Brent crude oil, gold, and silver through deep-learning models and compare the prediction performance of deep-learning models with random forest, LightGBM, SVR, and KNN models as baseline machine-learning models, hence, our results indicate the best deep-learning model for forecasting crude oil, gold, and silver daily prices. We will experiment with the performance of all models across four sliding window lengths of 5, 30, 60, and 90 days to indicate the suitable input length for superior performance with each model. The deep-learning models used in this study are LSTM, BiLSTM, GRU, BiGRU, T2V-BiLSTM, T2V-BiGRU, CNN, CNN-BiLSTM, CNN-BiGRU, TCN, TCN-BiLSTM, and TCN-BiGRU models.

We used grid search on the validation dataset to tune and select the optimal hyperparameters of each model. The common hyperparameters among all models are the number of epochs, batch size, dropout rate, and learning rate, equal to 50, 32, 0.2, and 0.001, respectively. Table 3 presents the selected hyperparameters of four best-performing models in this study. Due to the large scale of the study and space limitations, we only presented the selected hyperparameters of BiGRU, T2-BiGRU, TCN, and TCN-BiGRU models for each market. The hyperparameters of the other models are available upon request from the corresponding author.

After each training step, the weights of the models are updated by the Adam optimizer with a scheduled learning rate (*lr*) as follows:

$$lr = \begin{cases} lr_0 & \text{if } epochs < 5 \\ lr * e^{(-0.1)} & \text{otherwise} \end{cases} \tag{24}$$

The initial learning rate (*lr*₀) is 0.001, applied from epoch one through epoch five, and then exponentially decreases for each epoch after epoch five. In this study, the models were trained to minimize the MSE loss function. The objective function of the training process is as follows:

Table 3 Selected hyperparameters of models

Models	Markets	Hidden size			T2V size			Num. filters			Kernel size						
		5	30	60	90	5	30	60	90	5	30	60	90				
BiGRU	WTI	64	128	128	128												
	Brent	64	128	128	128												
	Gold	128	128	128	128												
	Silver	64	128	128	128												
T2V-BiGRU	WTI	64	128	128	128	32	90	32	128								
	Brent	64	128	150	128	64	128	128	128								
	Gold	128	150	128	128	90	64	128	128								
	Silver	256	256	256	128	128	128	128	32								
TCN	WTI									32	16	16	16	2	4	2	4
	Brent									32	16	16	16	2	4	2	4
	Gold									64	64	64	32	2	16	4	4
	Silver									32	64	64	64	2	4	2	4
TCN-BiGRU	WTI	64	64	64	64					32	16	16	16	2	4	2	4
	Brent	64	64	64	64					32	16	16	16	2	4	2	4
	Gold	128	128	128	128					64	64	64	32	2	16	16	4
	Silver	64	128	64	128					32	64	64	32	2	16	4	4

Other common hyperparameters among all models and all four markets are: epochs = 50, batch size = 32, dropout rate = 0.2, and initial learning rate = 0.001

$$\text{Objective function} = \text{Minimize } MSE = \text{Minimize } \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (25)$$

where \hat{y}_i is the predicted price, and y_i is the true target price for sample i .

Overfitting in financial market price forecasting experiments can lead to misleading and unreliable results. Overfitting occurs when a model is too complex and can capture the noise in the data rather than the underlying patterns. The consequences of overfitting in financial market price forecasting can be severe. Traders reliant on the overfitted model may make poor investment decisions, leading to significant losses. Furthermore, the overfitted model may be susceptible to market changes, making it difficult to use in real-world situations. Techniques such as cross-validation, dropout, early stopping, and pruning (for random forest and LightGBM) are employed to mitigate the risk of overfitting in crude oil and precious metals market price forecasting. Cross-validation involves partitioning the data into training and validation sets and evaluating the model on the validation set to assess its generalization performance. Model regularization in this study is achieved through a dropout layer in the models' architectures and early stopping after 10 epochs during training. Early stopping will end the training process if the validation error does not improve. To further assure the robustness of the forecasting results, all reported errors and predicted values are the average outputs from 10 runs of each model.

All deep-learning models are implemented using Tensorflow Keras, and machine-learning models are created using Sklearn. The experiments were conducted using Python 3.8 and run on a computing system with a 70 W Tesla T4 NVIDIA-SMI GPU, CUDA version 11.2, and 16 GB RAM.

WTI price forecasting

To show the computational performance of our deep-learning models for WTI next-day spot price forecasting, we draw the forecasting performance of LSTM, BiLSTM, GRU, BiGRU, T2V-BiLSTM, T2V-BiGRU, CNN, CNN-BiLSTM, CNN-BiGRU, TCN, TCN-BiLSTM, and TCN-BiGRU models, which we compare with the baseline models, i.e., random forest, LightGBM, KNN, and SVR models. Each model was executed 10 times to reduce randomness and improve the robustness of the results. Table 4 presents the MAE, MAPE, and RMSE values for the forecasted next-day WTI prices in the test dataset across all models. Among the evaluated models and considering two out of three performance criteria, the TCN model consistently achieves the lowest MAE and MAPE for WTI price forecasting across all input sliding window sizes. However, when considering the RMSE metric, the BiGRU model outperforms the other models for input sequences of lengths 5 and 30. Conversely, for input sequences of lengths 60 and 90, the TCN-BiGRU and T2V-BiGRU models demonstrate superior performance, respectively. In addition to the superior prediction performance, the forecasting error of the TCN model is not significantly affected by the input sequence length, as we obtain MAE values of 1.510, 1.455, 1.444, and 1.472 with sequence lengths of 5, 30, 60, and 90, respectively. Comparing this with other models, we can see that most models' performance is more sensitive to the input sequence length. Using bidirectional models has proved effective

Table 4 WTI price forecasting performance

Models	MAE				MAPE				RMSE			
	5	30	60	90	5	30	60	90	5	30	60	90
LSTM	1.950	1.872	1.875	1.873	7.190	8.005	8.605	8.284	3.787	3.719	3.720	3.720
GRU	1.893	1.787	1.765	1.631	8.300	6.092	5.672	4.965	3.682	3.584	3.580	3.539
BILSTM	1.821	1.654	1.679	1.904	7.532	5.966	5.720	5.286	3.588	3.511	3.496	3.747
BIGRU	1.570	1.559	1.575	1.699	6.868	5.094	5.230	6.725	3.415	3.432	3.439	3.488
T2V-BILSTM	1.985	2.196	1.668	1.670	5.594	5.033	6.274	6.204	3.770	3.986	3.461	3.458
T2V-BIGRU	1.889	1.606	1.511	1.523	5.107	5.987	3.853	5.060	3.691	3.451	3.392	3.405
CNN	1.887	2.068	3.201	3.201	7.277	8.213	8.957	9.357	4.000	6.064	5.085	5.071
CNN-BILSTM	1.972	2.581	2.563	3.650	6.733	6.158	6.566	8.207	4.496	4.845	4.444	5.695
CNN-BIGRU	1.851	2.406	2.244	2.885	8.133	5.850	7.793	7.196	4.243	4.650	4.161	4.815
TCN	1.510	1.455	1.444	1.472	3.829	3.663	3.530	3.882	3.550	3.495	3.559	3.552
TCN-BILSTM	2.063	2.354	1.704	2.494	5.374	5.566	4.301	5.377	3.857	4.267	3.419	4.456
TCN-BIGRU	1.787	2.015	1.672	1.911	5.437	5.471	7.023	5.033	3.520	3.872	3.372	3.720
Random Forest	3.209	3.151	1.887	1.868	7.209	7.131	4.307	4.286	4.869	4.801	3.574	3.569
LightGBM	1.487	1.551	1.554	1.558	3.490	3.622	3.624	3.643	3.213	3.338	3.392	3.386
SVR	2.231	4.199	3.861	5.586	5.591	9.224	8.560	16.124	4.947	5.733	5.687	7.821
KNN	1.772	3.358	5.049	6.559	4.075	7.599	10.681	12.830	3.440	5.586	7.792	10.028

To assure the robustness of models' performances, the average of errors in ten runs of the models are reported here. Bold indicates the lowest value

in NLP tasks (Arbane et al. 2023; Huang et al. 2023; G. Liu and Guo 2019; Raza and Schwartz 2023); however, little attention has been paid to using these models for price time-series forecasting. In this study, all three performance criteria from Table 4 show that bidirectional recurrent models, such as BiLSTM and BiGRU, perform better than unidirectional models, such as LSTM and GRU, for all sequence lengths. Bidirectional RNNs exploit the network memory to process information from backward and forward directions. Therefore, interdependency among data samples is learned better compared to unidirectional models that only use forward-direction information processing. Our findings comply with Yang and Wang (2022) and Siami-Namini et al. (2019), who found that the BiLSTM model outperformed the LSTM model for time-series prediction. Furthermore, it is evident from Table 4 that GRU-type models such as GRU, BiGRU, T2V-BiGRU, CNN-BiGRU, and TCN-BiGRU perform better than LSTM-type models such as LSTM, Bi LSTM, T2V-Bi LSTM, CNN-Bi LSTM, and TCN-Bi LSTM in WTI price forecasting.

To evaluate the effectiveness of Time2Vector embedding in WTI price forecasting, we compare the MAE, MAPE, and RMSE of the BiLSTM and BiGRU models with those of the T2V-BiLSTM and T2V-BiGRU models, respectively. Using the T2V input embedding, the MAE of the BiLSTM and BiGRU models with input sequence 5 increases from 1.821 and 1.570 to 1.985 and 1.889, respectively. In contrast, the MAE of the BiLSTM and BiGRU models with input sequence 90 decreases from 1.904 and 1.699 to 1.670 and 1.523, respectively. Arguably, Time2Vector embedding does not improve forecasting with smaller input sequences, 5 and 30, while it improves the WTI price forecasting performance for longer sequences of 60 and 90. To study the impact of hybrid models, such as CNN-BiLSTM and CNN-BiGRU, we compared their performance with single BiLSTM and BiGRU models. Combining the CNN model with recurrent-type models has a detrimental effect on the forecasting performance of WTI prices, as evidenced by an increase in MAE across all sequence lengths. This outcome occurs because the CNN module downsamples the input sequence, and some information that might be useful for BiLSTM or BiGRU models will be lost, resulting in higher forecasting errors. Similarly, a single TCN model outperforms the hybrid TCN-BiLSTM and TCN-BiGRU models. The TCN model can see the entire sequence in its receptive field and use the best temporal features to forecast the WTI price; therefore, combining it with a recurrent-type model will only increase the complexity of the model and cause an overfitting problem without significant improvements in forecasting performance.

Upon examining the forecasting errors of ensemble tree-based models, i.e., random forest and LightGBM, it becomes clear that random forest performs poorly in predicting WTI prices, whereas LightGBM demonstrates exceptional forecasting capabilities. The MAPE and RMSE values of LightGBM across sequence lengths of 5, 30, and 90 days are consistently the lowest among all 16 forecasting models. Consequently, LightGBM can be considered an approximate match to the TCN model as the top-performing method for WTI price forecasting. Moreover, the performance of LightGBM exhibits a slight decline as the input sequence lengths increase; however, this decrease in performance is not significant, indicating that LightGBM is relatively insensitive to variations in the input sequence length. Conversely, using the SVR and KNN models, it becomes clear that the performance of conventional machine-learning models tends to deteriorate as

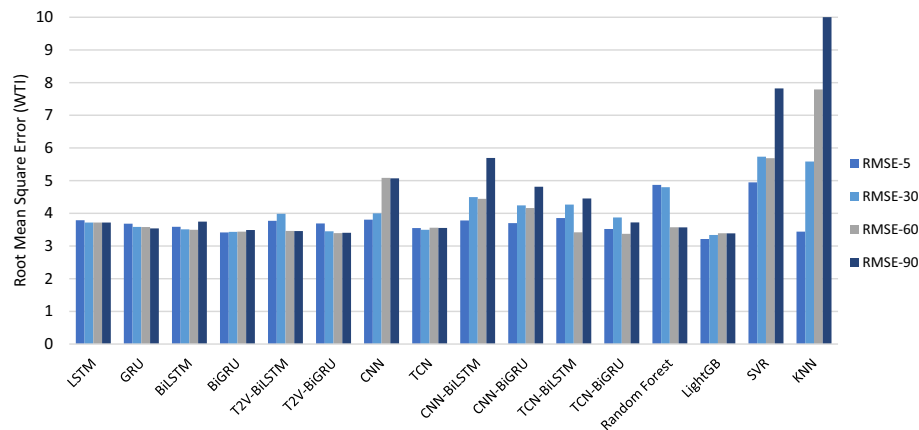


Fig. 6 RMSE of WTI crude oil next-day price forecasting models

the input sequences grow. In contrast, deep-learning models are less affected by larger input sequences, demonstrating their robustness. All deep-learning models outperform the SVR and KNN models for larger input sequences; however, for smaller sequences, such as those with a length of 5, the KNN model performs better than the deep-learning models, except for the BiGRU and TCN models. This discrepancy can be attributed to the data within each sequence serving as input features for the KNN model. As the sequence length increases, the KNN model faces greater challenges in identifying the nearest neighbors required for accurately predicting the target price.

Figure 6 presents the RMSE for the WTI next-day spot price forecasting models to find the best sliding window length for each forecasting model. Our experiments with WTI price forecasting show that using only recurrent-type models such as LSTM, GRU, BiLSTM, BiGRU, T2V-BiLSTM, and T2V-BiGRU, we obtain better prediction performance compared with using only CNN or a hybrid of CNN with Recurrent-type models such as CNN-BiLSTM and CNN-BiGRU. Recurrent-type models are not very sensitive to the input sequence length, and they even perform slightly better with relatively longer input sequences because longer sequences enable the model to learn more upward, downward, and complex patterns and generalize better in predicting unseen data. Nonetheless, since the CNN models cannot memorize important information from past data points, the forecasting error of CNN-type models, such as a single CNN, CNN-BiLSTM, and CNN-BiGRU, increases with the input sequence length. The RMSE of TCN-BiLSTM and TCN-BiGRU is generally smaller than the RMSE of CNN-BiLSTM and CNN-BiGRU models; therefore, we can conclude that among the hybrid models, the TCN module performs better than the CNN module in extracting the essential temporal features. Figure 6 shows that the input sequence of 60 days of lagged data points is generally better than other sliding window lengths such as 5, 30, or 90 days for WTI daily price forecasting; however, the CNN, CNN-BiLSTM, and CNN-BiGRU models perform better with an input sequence of 5 days than the other sequence lengths for WTI price prediction. Among the machine-learning models, Ensemble tree-based models emerge as the leading models for forecasting WTI prices. Notably, the random forest model exhibits subpar performance with shorter input sequences. LightGBM consistently performs well across all input sequences, demonstrating its robust forecasting capabilities.

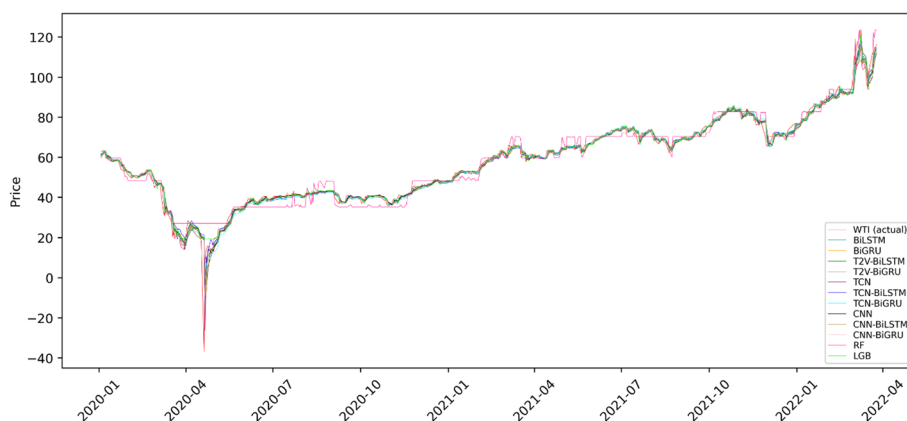


Fig. 7 Comparison of WTI crude oil price forecasting models on the test dataset

In contrast, the forecasting performance of the SVR and KNN models deteriorates as the input sequence length increases, suggesting that these models struggle to capture complex patterns and relationships effectively within longer data sequences.

Our observations regarding WTI forecasting align with Qin et al. (2023), where the GRU model demonstrated superior performance compared with random forest, SVR, and LSTM models, achieving a lower MAPE value. Similarly, our results corroborate with J. Yuan et al. (2023), highlighting that LightGBM exhibited significantly better performance than the LSTM and SVR models.

Figure 7 compares the line chart of predicted WTI prices in the test dataset with the actual WTI price value from 2020-01-03 to 2022-03-25. The predicted values at the end of April 2020 indicate that the TCN model surpasses the LightGBM model in accurately capturing sharp changes in the WTI price. The TCN model demonstrates superior performance in detecting and predicting abrupt fluctuations in price, showcasing its ability to capture and respond to sudden market dynamics with greater precision than the LightGBM model.

Brent price forecasting

Table 5 shows the errors, MAE, MAPE, and RMSE, of our forecasting models for Brent next-day spot price forecasting. We compared the forecasting performance of the LSTM, BiLSTM, GRU, BiGRU, T2V-BiLSTM, T2V-BiGRU, CNN, CNN-BiLSTM, CNN-BiGRU, TCN, TCN-BiLSTM, and TCN-BiGRU models with the baseline models, random forest, LightGBM, KNN, and SVR models. According to the lowest values of the MAE and RMSE measures for all input sequence lengths, 5, 30, 60, and 90, the TCN is the best-performing model in predicting the Brent crude oil price in the test dataset. Considering the MAPE for input sequences with 5 lagged data points, the TCN model has the best Brent price prediction performance; for input sequences of lengths 30, 60, and 90, the T2V-BiGRU model outperforms other models. Furthermore, the TCN model is not particularly sensitive to the input sequence length. The TCN achieves a robust and stable forecasting performance for all input sequence lengths as the MAE with 5, 30, 60, and 90 sequences are 1.295, 1.353, 1.315, and 1.301, respectively. The performance of most other models exhibits higher sensitivity to changes in the input sequence length

Table 5 Brent price forecasting performance

Models	MAE				MAPE				RMSE			
	5	30	60	90	5	30	60	90	5	30	60	90
LSTM	1.856	1.806	1.789	1.821	3.689	3.563	3.509	3.630	2.860	2.826	2.800	2.830
GRU	1.760	1.721	1.436	1.412	3.343	3.277	2.752	2.715	2.760	2.670	2.289	2.257
BiLSTM	1.703	1.492	1.556	1.562	3.328	2.895	2.974	3.013	2.643	2.364	2.446	2.449
BiGRU	1.421	1.428	1.439	1.425	2.888	2.861	2.895	2.833	2.222	2.217	2.237	2.228
T2V-BiLSTM	1.893	1.503	1.417	1.403	3.738	2.922	2.788	2.683	2.922	2.371	2.211	2.200
T2V-BiGRU	1.721	1.386	1.372	1.403	3.345	2.669	2.641	2.671	2.671	2.189	2.171	2.204
CNN	1.542	1.879	2.818	5.194	2.653	3.679	5.773	5.150	2.459	2.934	4.335	4.032
CNN-BiLSTM	1.857	2.456	2.680	2.641	3.618	4.887	5.241	5.121	2.844	3.807	4.126	4.096
CNN-BiGRU	1.769	2.414	2.265	2.279	2.136	5.026	4.684	4.233	2.747	3.694	3.478	3.248
TCN	1.295	1.353	1.315	1.301	2.637	2.905	2.658	2.697	2.035	2.059	2.052	2.045
TCN-BiLSTM	2.063	2.320	1.653	1.879	4.062	4.620	3.215	3.592	3.171	3.658	2.467	2.923
TCN-BiGRU	1.768	2.035	1.363	2.061	3.440	4.061	2.639	4.121	2.746	3.236	2.129	3.217
Random Forest	1.989	2.681	2.717	2.667	3.811	5.425	5.475	5.275	2.909	3.585	3.565	3.501
LightGBM	1.375	1.426	1.4525	1.478	2.719	2.910	2.963	3.063	2.130	2.174	2.190	2.210
SVR	2.130	2.787	4.480	4.530	4.685	5.192	9.343	8.417	3.284	4.102	6.284	6.448
KNN	1.594	3.224	5.440	6.421	3.197	6.737	10.72	12.94	2.483	4.984	8.379	10.14

To assure the robustness of models' performances, the average of errors in ten runs of the models are reported here. Bold indicates the lowest value

for Brent crude oil. For instance, the MAEs of the CNN model grow with increasing sequence length as it obtains MAEs of 1.542, 1.879, 2.818, and 5.194 with sequence lengths of 5, 30, 60, and 90, respectively. Similar to our findings for WTI crude oil price forecasting, we found that BiLSTM and BiGRU models generally outperform unidirectional LSTM and GRU models in forecasting Brent crude oil prices. By juxtaposing the MAE, MAPE, and RMSE of the GRU-type models (such as GRU, BiGRU, T2V-BiGRU, CNN-BiGRU, and TCN-BiGRU) with those of the LSTM-type models (such as LSTM, BiLSTM, T2V-BiLSTM, CNN-BiLSTM, and TCN-BiLSTM) we found that a GRU unit is a more appropriate recurrent unit for Brent crude oil price forecasting.

The impact of Time2Vector embedding in Brent crude oil price forecasting is assessed by comparing the MAE, MAPE, and RMSE of the T2V-BiLSTM and T2V-BiGRU models with the BiLSTM and BiGRU models, respectively. Table 5 shows that T2V embedding improves the forecasting performance of the BiLSTM model for input sequences of 60 and 90 while it stimulates the performance of the BiGRU model for input sequences of 30, 60, and 90. The results of Brent crude oil price forecasting confirm that T2V embedding favorably influences forecasting with longer input sequences. For the hybrid models, our results indicate that combining the CNN model with recurrent-type models adversely affects the performance of the BiLSTM and BiGRU models for Brent crude oil price forecasting. The same pattern appears when comparing the forecasting performance of a single TCN model with the TCN-BiLSTM and TCN-BiGRU hybrid models in predicting Brent daily prices. The TCN model outperforms the hybrid models.

Comparing the forecasting errors of the random forest, LightGBN, SVR, and KNN models with our deep-learning models indicates that the forecasting performance of deep-learning models is superior to that of machine-learning models. However, the ensemble LightGBM model stands as an exception, demonstrating remarkable

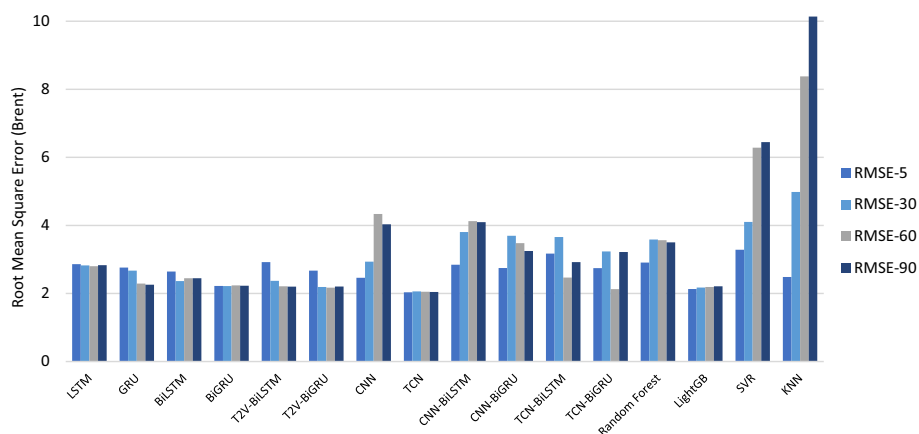


Fig. 8 RMSE of Brent next-day price forecasting models

performance as the second-best model among all 16 models for forecasting Brent crude oil prices across all input sequence lengths. This exceptional performance sets LightGBM apart from the other models, emphasizing its robustness and effectiveness in accurately predicting Brent crude oil prices, regardless of the input sequence length; however, for the short sequence length of 5, the KNN performs better than the deep-learning models, except for the BiGRU, CNN, and TCN models.

Figure 8 represents the RMSE of the forecasting models implemented in this study to predict the next-day Brent crude oil price in the test dataset. Our results denote that the recurrent-type models such as LSTM, GRU, BiLSTM, BiGRU, T2V-BiLSTM, and T2V-BiGRU outperform the CNN and hybrid models such as CNN-BiLSTM, CNN-BiGRU, TCN-BiLSTM, and TCN-BiGRU in terms of Brent price forecasting. Figure 8 shows that, in general, the efficacy of recurrent-type models in predicting the Brent price is enhanced with relatively longer input sequences; however, the CNN and hybrid models do not perform well with longer input sequences. The RMSE of TCN-BiLSTM and TCN-BiGRU are mainly lower than the RMSE of CNN-BiLSTM and CNN-BiGRU models; therefore, we can infer that the TCN module performs better than the CNN module in extracting the critical temporal features of Brent crude oil price. Examining the ensemble and conventional machine-learning models, namely random forest, LightGBM, SVR, and KNN, indicates that the optimal forecasting input sequence for Brent price prediction is five days. The LightGBM model achieves superior forecasting across all input sequences and, thus, is not significantly affected by changes in the input sequence length. As a general observation, the forecasting performance of these baseline models declines as the input sequence length increases, which indicates that shorter input sequences provide more accurate and reliable predictions than longer sequences when using these models for forecasting Brent prices. Regardless of the machine learning-type models, CNN, CNN-BiLSTM, and CNN-BiGRU models that perform better with shorter input sequences, our experiments indicate that the best input sequence length for Brent crude oil forecasting is 60 days of past data. Hence, the lowest RMSE values across most of the deep-learning models in this study are achieved for an input sequence length of 60 for Brent crude oil price forecasting.

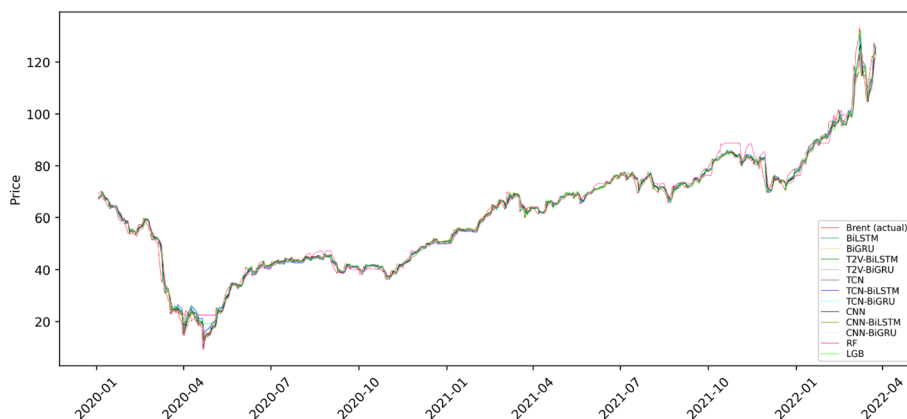


Fig. 9 Comparison of Brent crude oil price forecasting models on the test dataset

Our results validate the conclusions drawn by Zhao et al. (2017), indicating that deep-learning models outperform machine-learning models, such as SVR, in forecasting crude oil prices. Figure 9 compares the line chart of predicted Brent crude oil prices in the test dataset with the actual Brent price values from 2020-01-03 to 2022-03-25. Analyzing the predicted value during the abrupt Brent price change periods shows that the TCN model outperforms the LightGBM model in accurately capturing sharp changes in Brent price. Thus, TCN is a more reliable model for predicting the sudden changes in Brent price.

Gold price forecasting

Table 6 presents the forecasting errors of gold price prediction with 16 deep- and machine-learning models. Considering the models’ resulting MAE, MAPE, and RMSE, the TCN model has the best gold price prediction performance for input sequences of 5 and 90 days. Moreover, for gold price predictions with input sequences of 30 and 60, the BiGRU and GRU models show superior performance. Our results show that in most cases, the deep-learning models performed remarkably better than the baseline random forest, LightGBM, SVR, and KNN models in predicting the price of gold. Compared with CNN-BiLSTM, TCN-BiLSTM, and TCN-BiGRU, the SVR model achieved lower MAE, MAPE, and RMSE values. The prediction with gold price data shows that bidirectional LSTM models perform better than unidirectional LSTM models for all input sequences. Meanwhile, the BiGRU model outperformed the GRU model exclusively for input sequences of 5 and 60 days. Comparing the gold price forecasting errors of the GRU-type models, such as GRU, BiGRU, T2V-BiGRU, CNN-BiGRU, and TCN-BiGRU, with those of the LSTM-type models, such as LSTM, Bi LSTM, T2V-Bi LSTM, CNN-Bi LSTM, and TCN-Bi LSTM, we found that the GRU-type models are more appropriate than the LSTM-type models for gold price forecasting.

Figure 5 shows that the dynamics of gold price movement from 2000-01-04 to 2022-03-25 differs from the WTI and Brent crude oil markets, and an upward trend is visible in Gold price movements throughout the time. Nevertheless, our deep-learning models could predict the gold price for the test data relatively well. In contrast to its performance in WTI and Brent price forecasting, the LightGBM model surprisingly did not

Table 6 Gold price forecasting performance

Models	MAE			MAPE			RMSE					
	5	30	60	90	5	30	60	90	5	30	60	90
	LSTM	26.922	29.530	33.946	35.389	1.516	1.667	1.921	2.003	33.259	35.913	40.370
GRU	18.212	17.169	21.505	24.501	1.007	0.950	1.204	1.347	25.262	23.364	28.569	32.884
BiLSTM	22.332	26.169	32.760	36.389	1.252	1.477	1.849	2.052	29.257	33.659	42.267	42.411
BiGRU	19.067	15.188	27.635	22.148	1.051	0.847	1.559	1.217	26.132	20.846	35.118	28.840
T2V-BiLSTM	25.081	33.579	36.574	32.406	1.411	1.895	2.054	1.810	31.366	42.016	45.554	69.549
T2V-BiGRU	19.931	24.353	24.487	35.058	1.108	1.362	1.385	1.998	28.047	30.963	31.426	60.803
CNN	23.284	34.055	53.341	64.353	1.304	1.911	2.986	3.698	31.049	45.085	65.778	81.940
CNN-BiLSTM	30.395	68.590	68.621	66.562	1.711	3.980	4.014	3.836	36.972	79.667	85.067	77.196
CNN-BiGRU	24.445	36.158	46.481	42.364	1.372	2.051	2.674	2.314	31.479	47.926	58.035	53.647
TCN	15.557	17.209	27.856	19.894	0.87	0.965	1.57	1.116	20.831	22.58	33.726	24.817
TCN-BiLSTM	35.674	96.560	81.277	59.714	2.008	5.657	4.692	3.398	43.678	112.973	94.097	39.860
TCN-BiGRU	42.830	78.336	66.335	55.971	2.420	4.549	3.833	3.198	51.533	93.987	81.011	42.960
Random Forest	96.265	73.143	156.52	73.434	5.613	4.1920	9.4799	4.2087	121.328	97.421	177.28	97.525
LightGBM	43.407	69.40	70.274	79.67	2.432	3.968	4.016	4.584	64.392	94.058	95.052	104.554
SVR	87.021	49.64	69.895	48.114	5.073	2.828	4.011	2.759	92.059	60.121	80.39	63.037
KNIN	45.454	96.477	109.76	135.61	2.553	5.621	6.450	8.099	65.257	116.41	131.78	156.72

To assure the robustness of models' performances, the average of errors in ten runs of the models are reported here. Bold indicates the lowest value

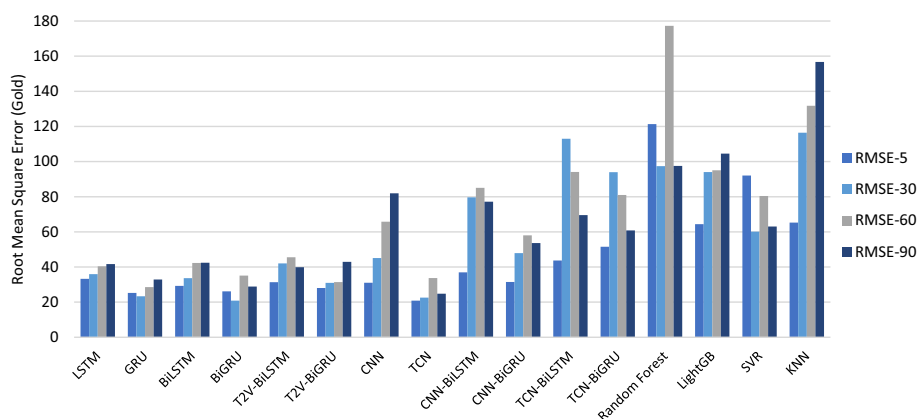


Fig. 10 RMSE of Gold next-day price forecasting models

exhibit strong generalization capabilities when predicting the gold price during the test data period. Despite its success in other forecasting tasks, the LightGBM model failed to provide accurate and reliable predictions for gold prices, indicating that the underlying dynamics and patterns of gold price data might differ significantly from those of WTI and Brent. Table 8 shows the coefficient of variation for the resulting MAEs of all forecasting models. The coefficient of variation is a scaleless value calculated by dividing the SD of the model MAEs through various input sequence lengths by the mean of those MAEs. The forecasting results of the gold market with the results of the WTI and Brent crude oil markets from Table 8 show that the models are more sensitive to the input sequence lengths of the gold market as the MAE forecasting error of each model varies markedly across the sequence lengths.

Figure 10 depicts the RMSE of our forecasting models to predict the next-day gold price in the test dataset. The recurrent-type models, such as LSTM, GRU, BiLSTM, BiGRU, T2V-BiLSTM, and T2V-BiGRU, generally have lower RMSE values compared to the CNN and hybrid models, such as CNN-BiLSTM, CNN-BiGRU, TCN-BiLSTM, and TCN-BiGRU. This result aligns with the research conducted by He et al. (2019) on gold price prediction, which demonstrated that a hybrid CNN–LSTM model did not exhibit superior performance compared with individual CNN or LSTM models.

A shorter input sequence of 5-day price data is more useful in gold price predictions with deep- and machine-learning models. The gold price forecasting performance generally deteriorates by increasing the input sequence length. The best prediction performance across all models and sequences was achieved through the BiGRU model using 30 days of gold price data. Based on the findings presented in Table 8, it is evident that LightGBM exhibits a higher coefficient of variation for MAE in Gold price forecasting than WTI and Brent crude oil. This outcome indicates that LightGBM is considerably sensitive to changes in the input sequence length when predicting the gold price. The higher coefficient of variation indicates that the performance of LightGBM may vary significantly when the input sequence length changes, underscoring the need for careful consideration and optimization of the input sequence length specifically for gold price forecasting with LightGBM. Figure 11 compares the line chart of predicted gold prices in the test dataset with the

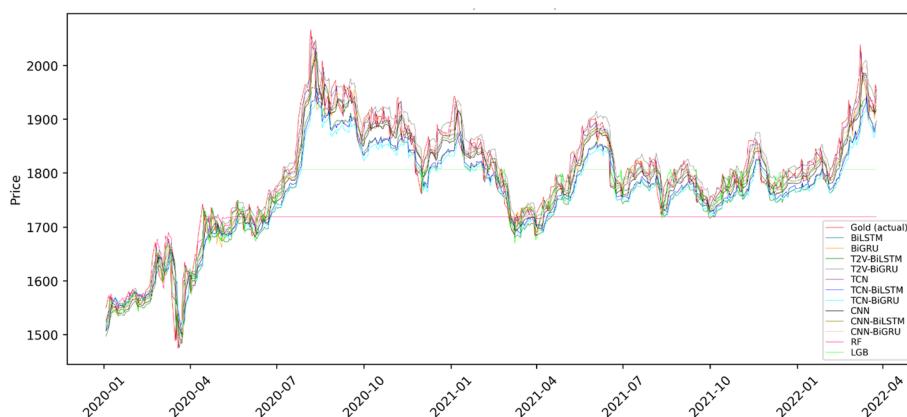


Fig. 11 Comparison of Gold price forecasting models on the test dataset

actual gold price values from 2020-01-03 to 2022-03-25. These results indicate that the random forest and LightGBM models do not generalize well in gold price forecasting. Comparing the performance of LightGBM and KNN models in predicting gold prices, our results demonstrate the superiority of LightGBM, which supports the study by Yuan (2023).

Silver price forecasting

As a precious metal, the daily spot price of silver is forecasted through the deep-learning models in this study and compared with the random forest, LightGBM, SVR, and KNN forecasts. Table 7 shows the MAE, MAPE, and RMSE of silver price predictions. The TCN model is the best-performing model across all input sequence lengths to forecast the daily silver price, as it scores the lowest MAE, MAPE, and RMSE among all models. Besides the TCN’s superior ability to forecast the silver price, this model is the least susceptible to the input sequence length, as shown by the MAE coefficient of variation in Table 8. The coefficient of MAE variation across all sequence lengths is 0.015 for the TCN model, the lowest among all models. The results of this study indicate that, except for the TCN-BiLSTM and TCN-BiGRU models with an input sequence of five days, our deep-learning models are superior to the SVR and KNN models in predicting the price of silver. For silver price forecasting, providing bidirectional information seems promising with the BiLSTM model as it reached lower MAE, MAPE, and RMSE values than the unidirectional LSTM; however, bidirectional information did not improve the forecasting performance of the GRU model for silver price prediction. Furthermore, the results from Table 7 indicate that GRU-type models have a relatively better forecasting performance than LSTM-type models for silver price prediction.

Using the ensemble (random forest and LightGBM) or conventional (SVR and KNN) machine-learning models, only LightGBM outperformed some of the deep-learning models, namely CNN, CNN-BiLSTM, CNN-BiGRU, TCN-BiLSTM, and TCN-BiGRU, in silver price forecasting. LightGBM was the best machine-learning model for silver price forecasting across all sequence lengths.

Comparing the MAE coefficient of variations between the silver and gold markets in Table 8 shows that the performance of our forecasting models is relatively less affected

Table 7 Silver price forecasting performance

Models	MAE				MAPE				RMSE			
	5	30	60	90	5	30	60	90	5	30	60	90
LSTM	0.490	0.463	0.497	0.491	2.186	2.086	2.239	2.221	0.721	0.694	0.741	0.710
GRU	0.440	0.394	0.381	0.385	1.967	1.753	1.667	1.684	0.658	0.590	0.549	0.556
BiLSTM	0.471	0.387	0.499	0.476	2.095	1.716	2.237	2.137	0.702	0.575	0.733	0.710
BiGRU	0.425	0.442	0.388	0.364	1.897	1.973	1.721	1.609	0.639	0.649	0.571	0.543
T2V-BiLSTM	0.478	0.589	0.491	0.434	2.138	2.666	2.206	1.933	0.717	0.850	0.730	0.653
T2V-BiGRU	0.472	0.548	0.462	0.370	2.125	2.483	2.065	1.641	0.703	0.794	0.679	0.553
CNN	0.551	0.730	0.752	0.768	2.454	3.273	3.400	3.468	0.786	1.030	1.090	1.140
CNN-BiLSTM	0.571	0.747	0.670	0.775	2.538	3.307	3.022	3.478	0.800	1.103	0.980	1.108
CNN-BiGRU	0.528	0.611	0.656	0.635	2.358	2.737	2.952	2.843	0.779	0.926	0.953	0.893
TCN	0.355	0.349	0.346	0.357	1.573	1.547	1.525	1.575	0.529	0.520	0.513	0.525
TCN-BiLSTM	0.647	0.942	0.746	0.603	2.881	4.269	3.345	2.668	0.960	1.327	1.013	0.867
TCN-BiGRU	0.714	0.928	0.553	0.467	3.186	4.155	2.476	2.086	1.058	1.307	0.780	0.651
Random Forest	1.831	0.831	0.848	2.476	8.076	3.562	3.646	11.30	2.149	1.059	1.075	2.764
LightGBM	0.509	0.475	0.486	0.475	2.154	2.047	2.070	2.061	0.710	0.665	0.684	0.646
SVR	0.602	1.164	1.077	1.130	2.656	5.006	4.609	5.034	0.775	1.321	1.359	1.513
KNN	0.667	1.170	1.847	2.686	2.805	5.078	8.022	11.47	0.913	1.501	2.295	3.300

To assure the robustness of models' performances, the average of errors in ten runs of the models are reported here. Bold indicates the lowest value

Table 8 coefficient of variation (CoV) for the MAE of forecasting models

	WTI	Brent	Gold	Silver
LSTM	0.020	0.016	0.124	0.031
GRU	0.061	0.116	0.164	0.068
BiLSTM	0.067	0.056	0.216	0.107
BiGRU	0.041	0.005	0.250	0.087
T2V-BiLSTM	0.137	0.148	0.153	0.131
T2V-BiGRU	0.108	0.114	0.248	0.158
CNN	0.274	0.577	0.423	0.144
CNN-BiLSTM	0.260	0.158	0.321	0.132
CNN-BiGRU	0.182	0.130	0.257	0.092
TCN	0.020	0.020	0.271	0.015
TCN-BiLSTM	0.162	0.143	0.388	0.205
TCN-BiGRU	0.081	0.179	0.248	0.305
Random Forest	0.258	0.121	0.341	0.465
LightGBM	0.019	0.027	0.201	0.028
SVR	0.301	0.301	0.251	0.229
KNN	0.429	0.452	0.339	0.476

H0: the mean CoV of Gold and Silver markets are equal ($p\text{-value} = 0.0074$)

H0: the mean CoV of WTI and Brent markets are equal ($p\text{-value} = 0.7142$)

The coefficient of variation is calculated by $\frac{\text{standard deviation}(MAE_i)}{\text{mean}(MAE_i)}$, $i = S_5, S_{30}, S_{60}, S_{90}$. Models with values in bold are least sensitive to the input sequence lengths for each market's price predictions. Bold indicates the lowest value

by changes in the input sequence length when predicting the silver market. This finding indicates that the forecasting models exhibit greater stability and consistency in their predictions for the silver market, regardless of variations in the input sequence length.

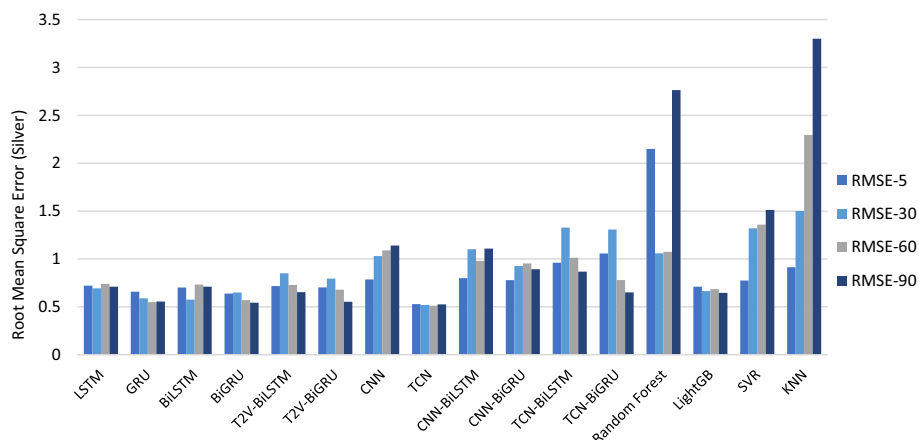


Fig. 12 RMSE of Silver next-day price forecasting models

Unlike the gold market, where the models show higher sensitivity to changes in the input sequence length, the silver market demonstrates a more robust and reliable forecasting performance across different input sequence lengths.

Figure 12 presents the RMSE of our deep-learning models to forecast the silver next-day price in the test dataset. Similar to the results of the WTI, Brent, and gold markets, the silver price forecasting error of the recurrent-type models such as LSTM, GRU, BiLSTM, BiGRU, T2V-BiLSTM, and T2V-BiGRU are generally lower than the forecasting error of the CNN and hybrid models such as CNN-BiLSTM, CNN-BiGRU, TCN-BiLSTM, and TCN-BiGRU. The best-performing model for predicting the silver price is the TCN model, which demonstrates robust forecasting performance across all input sequence lengths. Our results show that the recurrent-type models generally perform better with a longer input sequence of 90 days to predict the next-day silver price. The best prediction performance across all models and sequences is achieved through the TCN model using 60 days of past silver price data. Moreover, in hybrid models such as CNN-BiLSTM, CNN-BiGRU, TCN-BiLSTM, and TCN-BiGRU, the TCN module performs better than the CNN module in extracting the temporal features of the silver market price.

Figure 13 illustrates the line chart of the best-predicted silver prices in the test dataset with the actual silver price values from 2020-01-03 to 2022-03-25, showing that the TCN and random forest models are the best and least generalizing models in silver price forecasting.

Using MAPE as the metric, our silver price prediction results surpass those of a Gono et al. (2023), which employed random forest and XGBoost methods. Our best MAPE for silver price prediction, 1.52%, significantly outperforms the best MAPE of 5.98% achieved by Gono et al. (2023).

Our significant empirical findings can be summarized as follows.

1. TCN is the best model for generalizing and forecasting commodity market prices.
2. LightGBM is the best machine-learning model for forecasting commodity market prices; however, compared with the TCN model, it performs poorly in capturing and responding to sharp market dynamics.

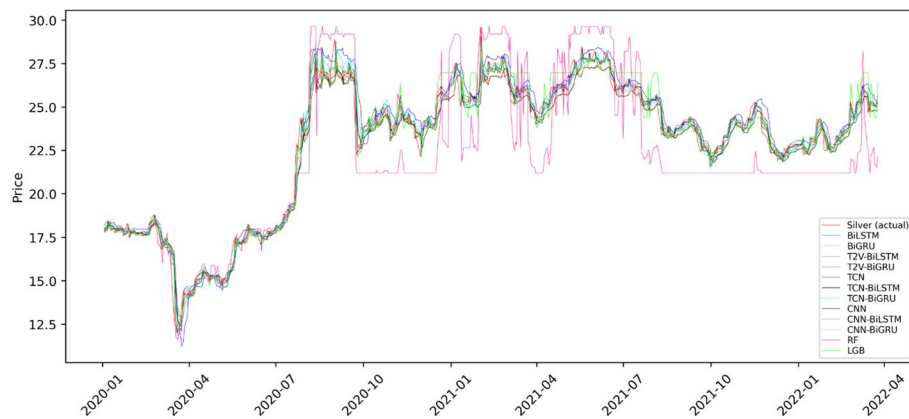


Fig. 13 Comparison of Silver price forecasting models on the test dataset

3. GRU-type models are the best recurrent-type deep-learning models in commodity price forecasting.
4. CNN-type models perform poorly in forecasting commodity market prices.
5. The TCN and LightGBM models are the most robust to input sequence lengths in predicting commodity market prices.
6. Using bidirectional models improves commodity price forecasting compared to only information from the forward price direction. This finding is also supported by Siami-Namini et al. (2019), indicating that BiLSTM-based modeling yields better predictions than regular LSTM-based models.
7. To achieve superior forecasting performance, it is essential to consider the proper input sequence length for each deep- or machine-learning model.
8. Among WTI, Brent, gold, and silver, gold is the most sensible market for the input sequence length in price forecasting.
9. Time2Vector embedding improves forecasting performance only when using longer input sequences.

Our findings provide valuable insights for analysts seeking to improve the accuracy of commodity market price forecasts. By examining the performance of various forecasting models and considering the impact of input sequence length on their predictive capabilities, our study offers guidance for selecting the most suitable models and input parameters for forecasting commodity market prices. With this knowledge, governments, energy sector managers, crude oil and precious metals investors can make sensible decisions. In a governmental context, crude oil and precious metal price forecasting helps governments in fiscal planning, economic policy decisions, resource allocation, revenue management, international trade negotiations, socioeconomic development, environmental policies, and geopolitical considerations. Accurate forecasts enable governments to make informed decisions that impact the national economy, public finances, and sustainable development. Accurate crude oil price forecasting gives managers valuable insights to optimize operations, manage risks, allocate budgets and resources efficiently, and make strategic decisions in the dynamic energy market. They can use forecasted prices to hedge against potential price fluctuations, secure favorable contracts, and

manage exposure to market volatility. Accurate crude oil price forecasting can provide a competitive advantage by enabling managers to make timely and informed decisions. They can anticipate market trends, respond quickly to price fluctuations, and stay ahead of competitors regarding pricing, supply chain management, and customer satisfaction.

Conclusion

Crude oil, particularly WTI and Brent, is crucial in global financial markets and economics. In recent years, crude oil prices have become more vulnerable to geopolitical and macroeconomic factors. Thus, understanding the dynamics of crude oil markets is inevitable. Furthermore, precious metals such as gold and silver are key commodities mined in particular countries, which makes the economies of these countries highly reliant on precious metal markets. Moreover, gold is a substitute asset for stock markets and is indispensable in financial investment portfolios. Therefore, developing an accurate forecasting model for crude oil, gold, and silver price movements is vital for policymakers, business owners, investors, and other stakeholders to mobilize timely political movements, foresee market trends, and properly design investment strategies to mitigate investment risks. In this study, we implement 12 deep-learning models, namely, LSTM, BiLSTM, GRU, BiGRU, T2V-BiLSTM, T2V-BiGRU, CNN, CNN-BiLSTM, CNN-BiGRU, TCN, TCN-BiLSTM, and TCN-BiGRU, to forecast the WTI, Brent, gold, and silver market prices and compare their forecasting performance with four baseline models, namely, random forest, LightGBM, SVR, and KNN models. We use each market's historical price information for this and apply four different sliding window lengths of 5, 30, 60, and 90 days. MAE, MAPE, and RMSE evaluation metrics are employed to assess the forecasting power of each model. We compared the forecasting performance of these models across various input sequence lengths and found that the TCN model is the best-performing model for forecasting the prices of WTI, Brent, gold, and silver. LightGBM exhibits comparable forecasting performance to the TCN model in accurately predicting WTI and Brent crude oil prices. Our results also indicate that the BiGRU and GRU models are the best for predicting gold spot prices with input sequences of 30 and 60, respectively. The best forecasting performance for each market is WTI through a TCN model with input sequence 60, MAPE 3.53%, Brent through a TCN model with input sequence 5, MAPE 2.64%, gold through a BiGRU model with input sequence 30, MAPE 0.85%, and silver through a TCN model with input sequence 60, MAPE 1.53%. Eventually, our study indicates using the TCN model for superior financial time-series price predictions. From the empirical results, we determine that the bidirectional LSTM and GRU models outperform the unidirectional LSTM and GRU models, respectively. Moreover, GRU-type models such as GRU, BiGRU, T2V-BiGRU, CNN-BiGRU, and TCN-BiGRU outperformed their LSTM-type peers in predicting WTI, Brent, gold, and silver prices.

Our study has several implications for policymakers and investors. First, the results of this study can assist investors and decision makers in promptly anticipating crude oil, gold, and silver market prices and adjusting their investment portfolios. Additionally, stakeholders can execute risk-hedging methods and lower their losses with timely predictions. In particular, gold is considered a suitable safe-haven asset for the stock and cryptocurrency markets (Junttila et al. 2018). Therefore, timely prediction of the gold market price will help stock market investors hedge their portfolios. Regarding

organizational-level and country-level relationships, organizations such as the Organization of the Petroleum Exporting Countries, World Petroleum Council, and International Energy Agency and government agencies can further apply the indicated method, for example, the TCN model, to devise profitable policies related to global crude oil prices. Finally, our study would be particularly valuable for forecasting crude oil, gold, and silver prices in case of extreme events such as the COVID-19 pandemic and the recent conflict between Russia and Ukraine, which were covered in the period considered in this study.

Several limitations must be acknowledged in our research on forecasting crude oil and precious metal prices. First, these markets' volatile and nonlinear nature poses difficulties in capturing all the intricate patterns and sudden price changes. Additionally, external factors such as natural disasters, geopolitical events, and supply–demand dynamics can significantly influence commodity prices and accurately incorporating these factors into forecasting models remains a complex task. Finally, it is essential to acknowledge the inherent uncertainty in forecasting and implement appropriate risk management strategies. Addressing these limitations will enhance the robustness and reliability of our research findings.

Some possible directions for improving crude oil and precious metals price forecasting exist. First, rather than using only historical price data, other features such as technical indicators, macroeconomic features, supply and demand data, production rate, and interconnections with other financial markets can be used to predict crude oil and precious metal prices. Second, incorporating the stakeholders' sentiments, which can be derived from news articles and social media platforms, might improve the forecasting performance of our proposed method. Finally, an alternative to using sequential data, other data structures, and learning methods, such as temporal graph neural networks, can be implemented to forecast price time-series data.

Abbreviations

ARDL	Autoregressive distributed lag
ARIMA	Autoregressive integrated moving average
BiGRU	Bidirectional gated recurrent units
BiLSTM	Bidirectional long short-term memory
CNN	Convolutional neural networks
CNN-BiGRU	Convolutional neural networks-bidirectional gated recurrent units
CNN-BiLSTM	Convolutional neural networks-bidirectional long short-term memory
CPI	Consumer price index
DBN	Deep belief network
EMD	Empirical mode decomposition
EEMD-TCN	Empirical mode decomposition-temporal convolutional network
ELM	Extreme learning machines
ENSO	El Niño-Southern oscillation
FNN	Feed-forward neural network
GRU	Gated recurrent units
IEA	International energy agency
ISBM	Improved slope-based method
KNN	k -Nearest neighbors
LDA	Latent Dirichlet allocation
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MEMD	Multivariate empirical mode decomposition
MIDAS	Mixed data sampling
MRN	Multi-recurrent network
NLP	Natural language processing
OPEC	Organization of the Petroleum Exporting Countries
RMSE	Root mean squared error

RNN	Recurrent neural networks
SDAE	Stacked denoising autoencoders
SVM	Support vector machines
SVR	Support vector regression
T2V	Time2Vector
T2V-BiGRU	Time2Vector bidirectional gated recurrent units
T2V-BiLSTM	Time2Vector bidirectional long short-term memory
TCN	Temporal convolutional networks
TCN-BiGRU	Temporal convolutional networks-bidirectional gated recurrent units
TCN-BiLSTM	Temporal convolutional networks-bidirectional long short-term memory
US	United States
VAR	Vector autoregressive
VTFM	Vector trend forecasting method
WOA	Whale optimization algorithm
WPC	World petroleum council
WTI	West Texas intermediate

Acknowledgements

Not applicable.

Author contributions

PF: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing original draft, writing review and editing. SL: Conceptualization, methodology, and review. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Declarations

Competing interests

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.

Received: 17 January 2023 Accepted: 7 February 2024

Published online: 16 July 2024

References

- Abdullah Ahmed R, Bin Shabri A (2014) Daily crude oil price forecasting model using Arima, generalized autoregressive conditional heteroscedastic and support vector machines. *Am J Appl Sci* 11(3):425–432
- Adekoya OB, Akinseye AB, Antonakakis N, Chatziantoniou I, Gabauer D, Oliyide J (2022) Crude oil and Islamic sectoral stocks: Asymmetric TVP-VAR connectedness and investment strategies. *Resour Policy* 78:102877
- Akbar M, Iqbal F, Noor F (2019) Bayesian analysis of dynamic linkages among gold price, stock prices, exchange rate and interest rate in Pakistan. *Resour Policy* 62:154–164
- Alameer Z, Elaziz MA, Ewees AA, Ye H, Jianhua Z (2019) Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. *Resour Policy* 61:250–260
- Almeida F, Xexéo G (2019) Word embeddings: a survey
- Amirifar T, Lahmiri S, Zanjani MK (2023) An NLP-deep learning approach for product rating prediction based on online reviews and product features. *IEEE Trans Comput Soc Syst*. <https://doi.org/10.1109/TCSS.2023.3290558>
- Amirshahi B, Lahmiri S (2023a) Hybrid deep learning and GARCH-family models for forecasting volatility of cryptocurrencies. *Mach Learn Appl* 12:100465
- Amirshahi B, Lahmiri S (2023b) Investigating the effectiveness of Twitter sentiment in cryptocurrency close price prediction by using deep learning. *Expert Syst*. <https://doi.org/10.1111/exsy.13428>
- Arbane M, Benlamri R, Brik Y, Alahmar AD (2023) Social media-based COVID-19 sentiment classification model using Bi-LSTM. *Expert Syst Appl* 212:118710
- Baek C (2019) How are gold returns related to stock or bond returns in the U.S. market? Evidence from the past 10-year gold market. *Appl Econ* 51(50):5490–5497
- Bai Y, Li X, Yu H, Jia S (2022) Crude oil price forecasting incorporating news text. *Int J Forecast* 38(1):367–383
- Balcilar M, Gabauer D, Umar Z (2021) Crude Oil futures contracts and commodity markets: new evidence from a TVP-VAR extended joint connectedness approach. *Resour Policy* 73:102219
- ben Khelifa S, Guesmi K, Urom C (2021) Exploring the relationship between cryptocurrencies and hedge funds during COVID-19 crisis. *Int Rev Financ Anal* 76:101777
- Bhowmik R, Wang S (2020) Stock market volatility and return analysis: a systematic literature review. *Entropy* 22(5):522

- Boongasame L, Viriyaphol P, Tassanavipas K, Temdee P (2022) Gold-price forecasting method using long short-term memory and the association rule. *J Mob Multimedia* 19(1):165–186
- Borisov V, Leemann T, Seßler K, Haug J, Pawelczyk M, Kasneci G (2021) Deep neural networks and tabular data: a survey. *IEEE Trans Neural Netw Learn Syst*:1–21
- 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, pp 785–794
- Cho K, van Merriënboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: encoder-decoder approaches
- Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling
- Das S, Nayak J, Kamesh Rao B, Vakula K, Ranjan Routray A (2022) Gold price forecasting using machine learning techniques: review of a decade. *Adv Intell Syst Comput Book Ser (AISC)* 1349:679–695
- Devlin J, Chang MW, Lee K, Google KT, Language, AI (2018) BERT: pre-training of deep bidirectional transformers for language understanding
- Drachal K (2022) Forecasting the crude oil spot price with bayesian symbolic regression. *Energies* 16(1):4
- Enwereuzoh PA, Odei-Mensah J, Owusu Junior P (2021) Crude oil shocks and African stock markets. *Res Int Bus Financ* 55:101346
- Fang T, Zheng C, Wang D (2023a) Forecasting the crude oil prices with an EMD-ISBM-FNN model. *Energy* 263:125407
- Fang Y, Wang W, Wu P, Zhao Y (2023b) A sentiment-enhanced hybrid model for crude oil price forecasting. *Expert Syst Appl* 215:119329
- Gharghory SM (2021) A hybrid model of bidirectional long-short term memory and CNN for multivariate time series classification of remote sensing data. *J Comput Sci* 17(9):789–802
- Gono DN, Napitupulu H (2023) Silver price forecasting using extreme gradient boosting (XGBoost) method. *Mathematics* 11(18):3813
- Gopali S, Abri F, Siami-Namini S, Namin AS (2021) A comparison of TCN and LSTM models in detecting anomalies in time series data. *IEEE Int Conf Big Data* 2021:2415–2420
- Gruber N, Jockisch A (2020) Are GRU Cells more specific and LSTM Cells more sensitive in motive classification of text? *Front Artif Intell* 3
- Guo J, Zhao Z, Sun J, Sun S (2022) Multi-perspective crude oil price forecasting with a new decomposition-ensemble framework. *Resour Policy* 77:102737
- He P, Liu X, Gao J, Chen W (2020) DeBERTa: decoding-enhanced BERT with Disentangled Attention. *International Conference on Learning Representations*
- He Z, Zhou J, Dai HN, Wang H (2019) Gold price forecast based on LSTM-CNN model. In: 2019 IEEE international conference on dependable, autonomic and secure computing, pp 1046–1053
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Huang Y, Liu Q, Peng H, Wang J, Yang Q, Orellana-Martín D (2023) Sentiment classification using bidirectional LSTM-SNP model and attention mechanism. *Expert Syst Appl* 221:119730
- Hussain Shahzad SJ, Raza N, Shahbaz M, Ali A (2017) Dependence of stock markets with gold and bonds under bullish and bearish market states. *Resour Policy* 52:308–319
- Jiang H, Hu W, Xiao L, Dong Y (2022) A decomposition ensemble based deep learning approach for crude oil price forecasting. *Resour Policy* 78:102855
- Junttila J, Pesonen J, Raatikainen J (2018) Commodity market based hedging against stock market risk in times of financial crisis: the case of crude oil and gold. *J Int Finan Markets Inst Money* 56:255–280
- Kazemi SM, Goel R, Eghbali S, Ramanan J, Sahota J, Thakur S, Wu S, Smyth C, Poupart P, Brubaker M (2019) Time2Vec: learning a vector representation of time
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 30:3146–3154
- Kertly de Medeiros R, da Nóbrega BC, Pitta de Jesus D, Philippe de Albuquerque V (2022) Forecasting oil prices: new approaches. *Energy* 238:121968
- Khan M, Wang H, Riaz A, Elfatyany A, Karim S (2021) Bidirectional LSTM-RNN-based hybrid deep learning frameworks for univariate time series classification. *J Supercomput* 77(7):7021–7045
- Kou G, Olgu Akdeniz Ö, Dinçer H, Yüksel S (2021) Fintech investments in European banks: a hybrid IT2 fuzzy multidimensional decision-making approach. *Financ Innov* 7:39
- Kou G, Yüksel S, Dinçer H (2022) Inventive problem-solving map of innovative carbon emission strategies for solar energy-based transportation investment projects. *Appl Energy* 311:118680
- Lahmiri S (2023a) Multifractals and multiscale entropy patterns in energy markets under the effect of the COVID-19 pandemic. *Decis Anal J* 7:100247
- Lahmiri S (2023b) A comparative study of statistical machine learning methods for condition monitoring of electric drive trains in supply chains. *Supply Chain Anal* 2:100011
- Lahmiri S, Bekiros S (2019) Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons Fractals* 118:35–40
- Lahmiri S, Bekiros S (2020) Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market. *Chaos, Solitons Fractals* 133:109641
- Lahmiri S, Bekiros S (2021) Deep learning forecasting in cryptocurrency high-frequency trading. *Cogn Comput* 13:485–487
- Lahmiri S, Bekiros S, Avdoulas C (2023) A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. *Decis Anal J* 6:100166
- Lahmiri S, Bekiros S, Bezzina B (2022) Complexity analysis and forecasting of variations in cryptocurrency trading volume with support vector regression tuned by Bayesian optimization under different kernels: an empirical comparison from a large dataset. *Expert Syst Appl* 209:118349

- Lara-Benítez P, Carranza-García M, Luna-Romera JM, Riquelme JC (2020) Temporal convolutional networks applied to energy-related time series forecasting. *Appl Sci* 10(7):2322
- Lea C, Flynn MD, Vidal R, Reiter A, Hager G.D (2016) Temporal convolutional networks for action segmentation and detection
- Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Li G, Yin S, Yang H (2022a) A novel crude oil prices forecasting model based on secondary decomposition. *Energy* 257:124684
- Li T, Kou G, Peng Y, Yu PS (2022b) An integrated cluster detection, optimization, and interpretation approach for financial data. *IEEE Trans Cybern* 52(12):13848–13861
- Li X, Shang W, Wang S (2019) Text-based crude oil price forecasting: a deep learning approach. *Int J Forecast* 35(4):1548–1560
- Li Y, Du N, Bengio S (2017) Time-dependent representation for neural event sequence prediction
- Liang X, Luo P, Li X, Wang X, Shu L (2023) Crude oil price prediction using deep reinforcement learning. *Resour Policy* 81:103363
- Lim B, Zohren S (2021) Time-series forecasting with deep learning: a survey. *Philos Trans R Soc Math Phys Eng Sci* 379(2194):20200209
- Lin Y, Chen K, Zhang X, Tan B, Lu Q (2022) Forecasting crude oil futures prices using BiLSTM-Attention-CNN model with Wavelet transform. *Appl Soft Comput* 130:109723
- Liu G, Guo J (2019) Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neuro-computing* 337:325–338
- Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) RoBERTa: a robustly optimized BERT pretraining approach
- Lu W, Li J, Li Y, Sun A, Wang J (2020) A CNN-LSTM-based model to forecast stock prices. *Complexity* 2020:1–10
- Madziwa L, Pillalamarri M, Chatterjee S (2022) Gold price forecasting using multivariate stochastic model. *Resour Policy* 76:102544
- Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality
- Mohamed NA, Messaadia M (2023) Artificial intelligence techniques for the forecasting of crude oil price: a literature review. In: *International conference on cyber management and engineering (CyMaEn)*, pp 340–343
- Murshed M, Tanha MM (2021) Oil price shocks and renewable energy transition: empirical evidence from net oil-importing South Asian economies. *Energy Ecol Environ* 6(3):183–203
- Orojo O, Tepper J, McGinnity TM, Mahmud M (2019) A multi-recurrent network for crude oil price prediction. In: *IEEE symposium series on computational intelligence (SSCI)*
- Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp 1532–1543
- Perival A (2023) The impact of crude oil price fluctuations on Indian economy. *Int J Res Appl Sci Eng Technol* 11(4):3173–3202
- Phan DHB, Sharma SS, Narayan PK (2016) Intraday volatility interaction between the crude oil and equity markets. *J Int Finan Markets Inst Money* 40:1–13
- Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst* 31
- Pullen T, Benson K, Faff R (2014) A Comparative analysis of the investment characteristics of alternative gold assets. *Abacus* 50(1):76–92
- Qin Q, Huang Z, Zhou Z, Chen C, Liu R (2023) Crude oil price forecasting with machine learning and Google search data: an accuracy comparison of single-model versus multiple-model. *Eng Appl Artif Intell* 123:106266
- Qin Q, Xie K, He H, Li L, Chu X, Wei YM, Wu T (2019) An effective and robust decomposition-ensemble energy price forecasting paradigm with local linear prediction. *Energy Econ* 83:402–414
- Raza S, Schwartz B (2023) Entity and relation extraction from clinical case reports of COVID-19: a natural language processing approach. *BMC Med Inform Decis Mak* 23(1):20
- Reboredo JC (2013) Is gold a safe haven or a hedge for the US dollar? Implications for risk management. *J Bank Finance* 37(8):2665–2676
- Risse M (2019) Combining wavelet decomposition with machine learning to forecast gold returns. *Int J Forecast* 35(2):601–615
- Salisu AA, Ogbonna AE, Adewuyi A (2020) Google trends and the predictability of precious metals. *Resour Policy* 65
- Sarwar S, Shahbaz M, Anwar A, Tiwari AK (2019) The importance of oil assets for portfolio optimization: the analysis of firm level stocks. *Energy Econ* 78:217–234
- Siarni-Namini S, Tavakoli N, Namin AS (2019) The Performance of LSTM and BiLSTM in forecasting time series. *IEEE Int Conf Big Data* 2019:3285–3292
- Sroka Ł (2022) Applying block bootstrap methods in silver prices forecasting. *Econometrics* 26(2):15–29
- Su M, Liu H, Yu C, Duan Z (2022) A new crude oil futures forecasting method based on fusing quadratic forecasting with residual forecasting. *Digital Signal Process* 130:103691
- Sun J, Zhao P, Sun S (2022) A new secondary decomposition-reconstruction-ensemble approach for crude oil price forecasting. *Resour Policy* 77:102762
- Swamy V, Lagesh MA (2023) Does happy Twitter forecast gold price? *Resour Policy* 81:103299
- Szarek D, Bielak Ł, Wyłomańska A (2020) Long-term prediction of the metals' prices using non-Gaussian time-inhomogeneous stochastic process. *Phys A Stat Mech Appl* 555
- Tang L, Zhang C, Li L, Wang S (2020) A multi-scale method for forecasting oil price with multi-factor search engine data. *Appl Energy* 257:114033
- Uzo-Peters A, Laniran T, Adenikinju A (2018) Brent prices and oil stock behaviors: evidence from Nigerian listed oil stocks. *Financ Innov* 4(1):8

- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *Vidal A, Kristjanpoller W* (2020) Gold volatility prediction using a CNN-LSTM approach. *Expert Syst Appl* 157(1348):1
- Wang J, Athanasopoulos G, Hyndman RJ, Wang S (2018) Crude oil price forecasting based on internet concern using an extreme learning machine. *Int J Forecast* 34(4):665–677
- Wang J, Niu T, Du P, Yang W (2020) Ensemble probabilistic prediction approach for modeling uncertainty in crude oil price. *Appl Soft Comput J* 95:106509
- Wang L, Ma F, Niu T, Liang C (2021) The importance of extreme shock: examining the effect of investor sentiment on the crude oil futures market. *Energy Econ* 99:105319
- Xiuzhen X, Zheng W, Umair M (2022) Testing the fluctuations of oil resource price volatility: a hurdle for economic recovery. *SSRN Electron J*
- Xu D, Ruan C, Korpeoglu E, Kumar S, Achan K (2021) A Temporal kernel approach for deep learning with continuous-time information
- Xu D, Ruan C, Kumar S, Korpeoglu E, Achan K (2019) Self-attention with functional time representation learning
- Yan J, Mu L, Wang L, Ranjan R, Zomaya AY (2020) Temporal convolutional networks for the advance prediction of ENSO. *Sci Rep* 10(1):8055
- Yang M, Li X, Liu Y (2021) Sequence to point learning based on an attention neural network for nonintrusive load decomposition. *Electronics* 10(14):1657
- Yang M, Wang J (2022) Adaptability of financial time series prediction based on BiLSTM. *Procedia Comput Sci* 199:18–25
- Yang S, Chen D, Li S, Wang W (2020) Carbon price forecasting based on modified ensemble empirical mode decomposition and long short-term memory optimized by improved whale optimization algorithm. *Sci Total Environ* 716:137117
- Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* 31(7):1235–1270
- Yuan Z (2023) Gold and bitcoin price prediction based on KNN, XGBoost and LightGBM model. *Highlights Sci Eng Technol* 39:720–725
- Zhang P, Ci B (2020) Deep belief network for gold price forecasting. *Resour Policy* 69:101806
- Zhang S, Chen Y, Zhang W, Feng R (2021) A novel ensemble deep learning model with dynamic error correction and multi-objective ensemble pruning for time series forecasting. *Inf Sci* 544:427–445
- Zhang Y, Wang J, Yu L, Wang S (2022a) An extreme bias-penalized forecast combination approach to commodity price forecasting. *Inf Sci* 615:774–793
- Zhang Z, He M, Zhang Y, Wang Y (2022b) Geopolitical risk trends and crude oil price predictability. *Energy* 258:124824
- Zhao L, Cheng L, Wan Y, Zhang H, Zhang Z (2015) A VAR-SVM model for crude oil price forecasting. *Int J Glob Energy Issues* 38(1/2/3):126
- Zhao LT, Wang Y, Guo SQ, Zeng GR (2018) A novel method based on numerical fitting for oil price trend forecasting. *Appl Energy* 220:154–163
- Zhao Y, Li J, Yu L (2017) A deep learning ensemble approach for crude oil price forecasting. *Energy Econ* 66:9–16
- Zhou S, Wu JN, Wu Y, Zhou X (2015) Exploiting local structures with the kronecker layer in convolutional networks

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.