International Journal of Computational Intelligence Systems (2024) 17:225
https://doi.org/10.1007/s44196-024-00640-3

RESEARCH ARTICLE

Enhancing Multi-step Brent Oil Price Forecasting with Ensemble Multi-scenario Bi-GRU Networks

Mohammed Alruqimi[1] · Luca Di Persio[1]

Received: 10 June 2024 / Accepted: 19 August 2024

Abstract

Accurate crude oil price forecasting is crucial for economic stability, investment planning, and strategic decision-making across various industries. Despite numerous research efforts in applying deep learning to time-series forecasting, achieving high accuracy in multi-step predictions for volatile time-series like crude oil prices remains a significant challenge. Moreover, most existing approaches primarily focus on one-step forecasting, and the performance often varies depending on the dataset and specific case study. This paper introduces ensemble-based deep-learning models to capture Brent oil price volatility and enhance the multi-step price prediction. Our methodology employs a two-pronged approach. First, we present an empirical comparison of deep-learning models and architectures, including RNNs, CNNs, and transformers, for forecasting Brent oil prices. We also examine the impact of various external factors on forecasting accuracy. Then, we introduce a novel approach that employs ensemble GRU-based models to enhance prediction accuracy across multiple forecasting scenarios. Extensive experiments were conducted using a dataset of historical Brent prices encompassing the COVID-19 pandemic, which significantly impacted energy markets. The results demonstrate that the proposed model outperforms benchmark and established models, achieving a 9.3% reduction in MSE compared to the closest benchmark model for a 3-day forecasting horizon.

Keywords Crude oil price forecasting · Brent oil analysis · Time-series forecasting · Ensemble learning

1 Introduction

Brent crude oil, one of the major global benchmarks for oil prices, plays a critical role in the energy markets and broader economic landscape. Forecasting its price is essential for various stakeholders, including policymakers, investors, and energy companies. However, crude oil price market prediction is known for its inherent complexity and obscurity. The high degree of volatility, unpredictable, irregular events, and complex interconnections among market factors make it extremely challenging to accurately forecast the fluctuations in crude oil prices. The dynamic interplay of supply and demand and changes in oil prices are influenced by external factors, such as economic growth, financial markets, geopolitical conflicts, warfare, and political considerations [1–3]. A variety of methodologies have been utilised to predict crude oil prices. The older approaches primarily rely on employing economic and statistical methods, such as VAR [4], ARIMA, GARCH [5], VMD [6], and Walvet decomposition [7]. These traditional methods, which typically rely on mathematical assumptions, struggle to accurately capture oil prices' complex, nonlinear dynamics, and the interplay of various influencing factors [8]. To overcome these limitations, recent studies increasingly leverage machine learning approaches [2, 9, 10]. In contrast to the traditional statistical and econometric models, AI provides a valuable alternative approach to capture complex nonlinear characteristics of the crude oil price movement. In recent years, there has been an effort to exploit the emergent advances in deep-learning machine models (such as Transformers and GANs) for time-series analysis [11, 12]. Despite the great success achieved by these models in processing natural languages and generating images and videos, Long short-term memory (LSTM) and gated recurrent unit (GRU) networks still maintain their pop-

✉ Mohammed Alruqimi
m.alruqimi@gmail.com; mohammed.alruqimi@univr.it

Luca Di Persio
luca.dipersio@univr.it

1 Department of Computer Science, University of Verona, Verona, Italy

ularity for time-series forecasting [13–15]. Their popularity could be attributed to their ability to effectively capture temporal dependencies, which is crucial for accurately predicting future values in a time-series. They are also significantly less complex to train than more modern models such as GANs and transformers. Also, a vast amount of data are required to train transformers, which makes them inefficient for tasks involving smaller datasets, such as daily-based oil price forecasting. Nevertheless, achieving accurate oil price forecasting remains a challenging task, particularly in terms of multi-step forecasting. Additionally, most current research focuses on one-day-ahead forecasting or utilises complex combinations of layers and models with high computational complexity.

This study proposes blind forecasts generated from three different scenarios across Bi-GRU networks. To this end, we start by assessing the forecast performance of various architectures of deep-learning models (popular deep-learning models used in the literature for price forecasting) to optimise our model selection. Then, we propose a novel effective model (abbreviated as ERS-Bi-GRU, denoting its main elements-Ensemble, Residual, Sentimental, and Bi-Directional Gated Recurrent Units) for Brent price forecasting. Our experiments additionally evaluate the influence of incorporating three external factors—the USD index (USDX), Saudi energy sector index (TENI), and sentiment score (SENT)—on enhancing prediction accuracy. The contributions of this paper are threefold:

- Introduction of an effective ensemble model for forecasting multi-step Brent crude oil prices.
- Evaluation of various established deep-learning architectures and combinations for oil price forecasting, discerning the optimal architecture within this domain. To the best of our knowledge, this is the first study to compare RNN-based models and more recent and complex models like transformers for oil price prediction.
- Evaluate the influence of external factors on Brent price movement.

The proposed model (ERS-Bi-GRU) has been compared against well-known benchmarks and established models in the literature to evaluate its forecasting accuracy. The obtained results indicate that the proposed model outperforms the benchmark models. The rest of the paper is organised as follows. Section 2 provides an overview of recent literature related to oil price forecasting. Section 4 describes the dataset used in this study. Section 5 outlines the methodology, Section 6 describes the experimental setup and results, and finally the conclusion.

2 Literature Review

Recent years have witnessed strong growth in adopting machine learning for time-series forecasting, driven by its remarkable ability to uncover intricate and nonlinear patterns within the data. Researchers have employed various Ml networks, such as long-short-term memory (LSTM) [3, 12, 16], gated recurrent unite (GRU) [17–19], convolutional neural network (CNN) [20], and transformer models [21] to predict crude oil prices based on historical price data and some relevant features. Machine learning approaches often require extensive feature engineering and can be sensitive to the quality and availability of data. Many researchers incorporated additional data sources, such as macroeconomic and technical indicators, social media sentiment, and news articles to enhance prediction accuracy. Additionally, advancements in natural language processing and sentiment analysis techniques have allowed for a more comprehensive understanding of the impact of geopolitical events and news on oil prices [3, 19, 22, 23]. A popular approach combines different neural networks with statistical and economic methods to improve crude oil forecasting. For instance, many researchers have merged RNN networks with CNN and self-attention mechanisms to capture temporal, local, and long-term dependencies in historical price data [24, 25]. Statistical and time-series analysis methods [such as variational mode decomposition (VMD), empirical mode decomposition (EMD), Granger causality, Gaussian process, and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)] are also prevalent alongside neural networks in recent approaches (i.e., [3, 19]).

More recently, Transformer-based approaches have been explored in the literature [11, 26], raising questions about their performance compared to sequential processing mechanisms like RNNs. This is the first paper to provide an empirical assessment and comparison between Transformer-based and RNN-based models for oil price forecasting. Furthermore, while many existing approaches focus primarily on single-step forecasting, robust frameworks are still needed to effectively handle multi-step forecasting scenarios.

3 Preliminaries

This section briefly describes the deep-learning network architectures examined in this work. These architectures also serve as benchmarks for evaluating our proposed model.

As listed in Table 1, the targeted networks selected for the experiments include LSTM, GRU, CNN-LSTM, CNN-LSTM-att, Transformer, Autoformer, Informer, and TimsNet. GRU and LSTM networks have been selected due to their established effectiveness in time-series fore-

Table 1 Experimental design

Models	LSTM, GRU, Bi-LSTM, Bi-GRU, CNN-Bi-LSTM, CNN-Bi-LSTM-att, Transformer, Autoformer, Informer, TimesNet
External variables	USDX, TENI, SENT
Window size	Tuning range 5–22
Prediction horizon	1,3

casting tasks [12, 17]. Combining CNN and LSTM networks and attention is another popular approach widely employed in the literature. CNNs excel at extracting local features from time-series data by applying learnable filters that capture spatial relationships within specific time windows. The CNN extracts relevant local features, while the LSTM and GRU model long-term dependencies. Additionally, the self-attention mechanisms effectively direct attention across the time-series data. On the other hand, transformers have demonstrated remarkable achievements across various domains beyond just NLP. Over the last three years, transformer-based architectures (including Informer [26], Autoformer [11]) have been adapted for time-series tasks. TimesNet [27] is another cutting-edge CNN-based-structure model introduced in April (2023).

3.1 Long Short-Term Memory (LSTM)

LSTM is a popular type of recurrent neural network (RNN) architecture widely used in deep learning [28]. It was introduced to address the vanishing and exploding gradient problems that hinder the training of traditional RNNs. LSTM models are particularly effective in capturing long-term dependencies in sequential data such as time-series.

3.2 Gated Recurrent Unit (GRU)

GRU is a Recurrent Neural Network (RNN) variant that offers certain advantages compared to the popular Long Short-Term Memory (LSTM) model. GRU is known for its efficiency and faster computation, requiring less memory than LSTM. However, LSTM tends to perform better in scenarios involving datasets with longer sequences, as it can effectively capture and retain long-term dependencies [29].

3.3 CNN-LSTM

Convolutional Neural Networks (CNNs) are deep-learning models designed explicitly for analysing visual data, such as images or videos. They employ a hierarchical structure of interconnected layers, including convolutional layers for feature extraction and pooling layers for dimensionality reduction. CNNs excel at extracting local features from time-series data by applying learnable filters that capture spatial relationships within specific time windows. This capability is precious for identifying the data's cyclical patterns or localised events. By combining these strengths, CNN-LSTM architectures synergistically address the limitations of individual models. The CNN extracts relevant local features, while the LSTM leverages these features to model long-term dependencies

3.4 CNN-LSTM-Attention

The emergence of self-attention mechanisms in transformer architectures has revolutionised natural language processing tasks. By allowing models to weigh the importance of different input tokens, self-attention enables capturing long-range dependencies efficiently. Building upon this concept, the CNN-LSTM-Self Attention approach combines convolutional neural networks (CNNs) for spatial feature extraction, long short-term memory (LSTM) networks for capturing temporal dependencies, and self-attention mechanisms for effectively directing attention across the time-series data. This fusion of techniques empowers the model to capture both local patterns and long-term dependencies. It is a promising approach for time-series forecasting tasks where understanding short-term fluctuations and overarching trends is crucial.

3.5 Transformer

A transformer model is a neural network that learns context and thus meaning by tracking relationships in sequential data like the words in this sentence, first described in 2017 by Ashish Vaswani [30]. Transformers excel in capturing complex relationships in data and have become a cornerstone in modern deep-learning architectures, epitomised by models, such as BERT, GPT, and others. Their success extends beyond NLP, finding applications in various domains, including computer vision and time-series analysis. Transformer-based models have been proposed to adopt various self-attention mechanisms to discover long-range dependencies and enhance long-term forecasting.

3.6 Informer

The Informer is a transformer-based model designed to handle long dependencies in Long-Sequence Time-Series Forecasting. This model, as described in [26], addresses some drawbacks of long-sequence time-series forecasting transformers. The encoder of the Informer processes long-sequence inputs and replaces canonical self-attention with ProbSparse attention. The decoder handles lengthy sequence inputs by zero-padding the target elements, calculating the weighted attention composition of the feature map, and promptly generating output elements generatively.

3.7 Autoformer

In exploring the ongoing research efforts to adapt the transformer architecture for time-series prediction, this paper introduces the Autoformer model [11]. The Autoformer includes several enhancements, such as series decomposition and an auto-correlation mechanism. The encoder uses series decomposition blocks (blue) to remove the long-term trend-cyclical aspect, focusing on modelling seasonal patterns. The decoder gradually accumulates the trend from hidden variables, with past seasonal data utilised by the encoder–decoder auto-correlation mechanism.

3.8 TimesNet

TimesNet (Temporal 2D-Variation Modelling for General Time-Series Analysis) is a (CNN)-based model proposed for general time-series analysis tasks, including forecasting, classification, imputation, and anomaly detection. This model was introduced in 2023 in this paper[27] and has demonstrated state-of-the-art performance on various benchmark datasets. TimesNet transforms the 1D time-series data into 2D tensors to simultaneously represent intra-period and inter-period variations. The fundamental structure of TimesNet involves stacking of TimesBlocks in a residual manner. These TimesBlocks are adept at capturing diverse temporal 2D-variations from k different reshaped tensors. The fusion process is carried out based on normalised amplitude values.

4 Dataset

4.1 Dataset Description

The initial dataset used in this paper encompasses eight variables: closing prices of Brent oil, the USD index, the Saudi Energy index, the Saudi Tadawul All Share index, the S&P 500 index, the Natural Gas index, the Gold index, and a sentimental score. These variables span 2,380 observations from January 2012, to April 2021, capturing the period impacted by the COVID-19 pandemic on energy and stock markets. This time range was specifically chosen due to the availability of sentiment scores for this period.

Figure 1 displays the time-series plot of Brent's daily closing price. The figure shows that sharp fluctuations in crude oil logistics and production are often linked to world events. Extensive research has explored the influence of industrial commodities, indices (such as USD, S&P 500, Natural Gaz, and Gold), and sentiment analysis extracted from news articles, financial reports, and even tweets on the energy market [22, 23, 31–33]. External factors that affect the movement of crude oil prices could be categorised into supply, demand, financial, political, and major events factors [31]. These factors are inter-correlated; for instance, supply and demand movement may be subject to political factors and world events.

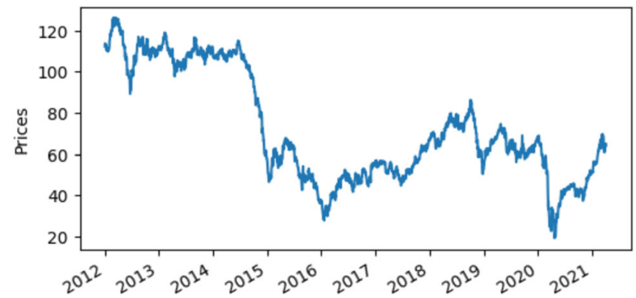

Fig. 1 Brent oil price trend from 2012 to 2021

4.2 Correlation-Based Feature Selection

Correlation-based Feature Selection is a method to identify and select the most relevant features from a dataset by analysing the relationship between each feature and the target variable. This technique helps select features that can play the most significant role in predicting the target variable. We created a filtered dataset that includes only three external variables from our initial dataset to be used as input features to our models alongside the Brent oil prices, based on their strong correlation with Brent prices. We used a Spearman correlation coefficient threshold of 0.6 to identify these variables, as shown in Fig. 2. Therefore, the three variables enumerated below have been designated for subsequent experimentation alongside the target variable (Brent crude oil prices). The impact of these variables on the prediction performance has been further evaluated experimentally during the model training, as detailed in Sect. 6

1. Sentiment score (SENT): We used an accumulative sentimental score provided by CrudeBERT_Plus model and presented in [22]. CrudeBERT is a variant of FinBERT that has been fine-tuned towards assessing the impact of market events on crude oil prices, focusing on frequently occurring market events and their effects on market prices according to Adam Smith's theory of supply and demand. Mainly, CrudeBERT dataset used headlines originating from 1034 unique news sources, of which the majority has been published on the Dow Jones newswires

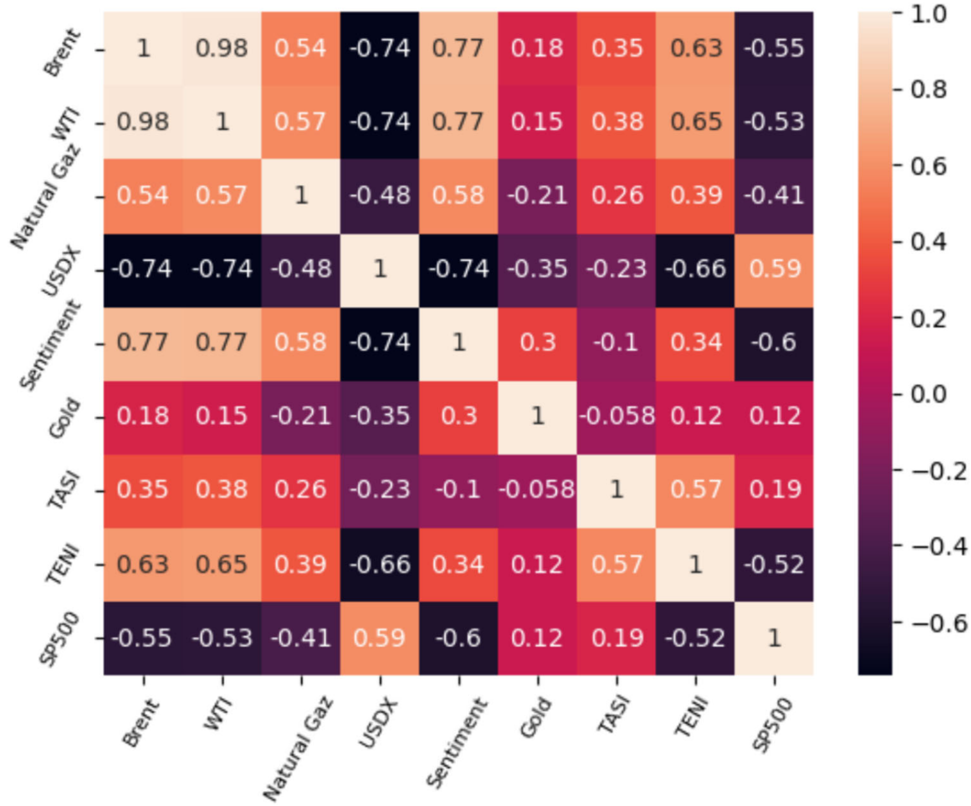

Fig. 2 Heatmap Spearman Correlation

(approx. 21,200), followed by Reuters (approx. 3,000), Bloomberg (approx. 1,100), and Platts (approx. 870). More details about generating this sentiment score are found in the paper [22]; the data are publicly available and can be reached following this link.

2. USDX Index: The U.S. dollar index (USDX) measures the value of the U.S. dollar relative to a basket of foreign currencies. There is a negative correlation between crude oil and the USDX index. USDX historical dataset has been obtained for the same period from Investing.
3. Saudi energy sector index (TENI) contains two companies working in the energy sector (Arabian Drilling Co and Rabigh Refining & Petrochemical Co). The historical dataset of this index has been obtained for the same period from https://uk.investing.com/indices/tpisi-historical-data.

4.3 Data Preparation

Our training data have been standardised by removing the mean and scaling to unit variance. To incorporate observations of all features, we run a left-join merge between the Brent crude oil price time-series and the other three input time-series. On the other hand, the missing values in the three external factors have been filled in by running a Linear Interpolation function: Let's X_t is a null value, $X_t = (X_{t-1}+X_{t+1})/2$. The dataset has been split into train/valid and tested as follows: Training set: from 2012-01-03 to 2019-10-10; Validation set: from 2019-10-11 to 2020-06-23; and Test set from 2020-06-24 to 2021-04-01.

5 Methodology

This work presents a hybrid approach, based on deep-learning models and times-series volatility analysis, for multi-step forecasting of Brent crude oil prices. Mathematically, given a set of time-series inputs $X_t = \{x_{t-2}, x_{t-1}, \ldots, x_t\}$, where each vector x_t contains multiple features (e.g., historical oil prices, sentimental indicators, and market data), the deep-learning model learns a complex function f_θ parameterised by θ. The objective is to forecast the prices over multiple horizons h: $x_{t+h} = f_\theta(X_t) + \epsilon_t$ where ϵ_t is the error term. Additionally, we investigate how external factors affect prediction accuracy and determine the most effective deep-learning architecture and configuration to enhance forecasting performance. To this end, we followed the steps outlined below, illustrated by Fig. 3.

First, we collected eight historical data (eight time-series), performed a correlation-based test to select only the most correlated variables for the next phase, and prepared our filtered dataset, as described in Sect. 4. The filtered dataset comprises daily historical observations of Brent crude oil prices, along with USDX, SENT, and TENI. Second (in the first

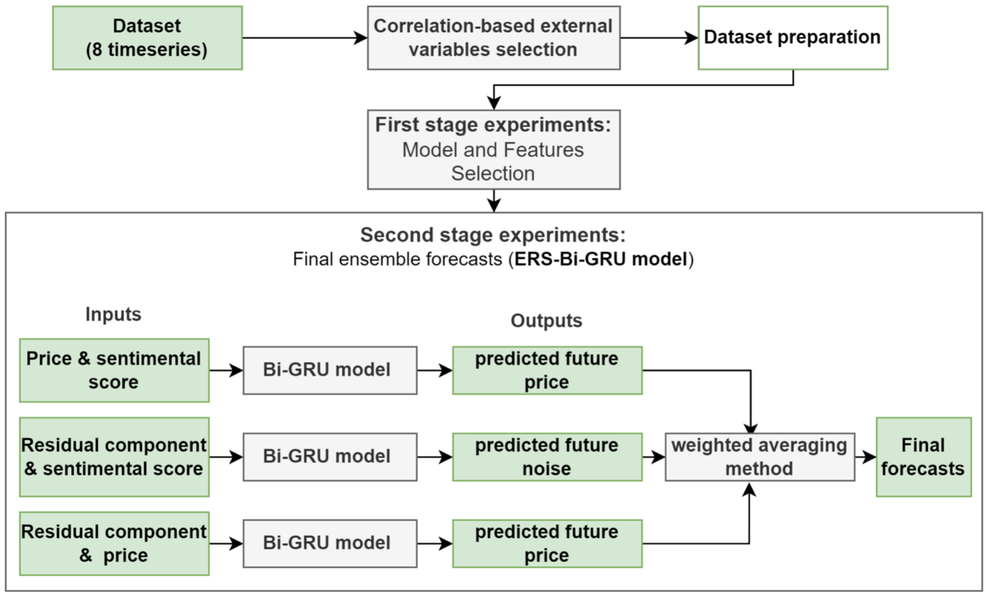

Fig. 3 Methodology flowchart and ERS-BI-GRU model

stage experiments), we conducted experiments using different architectures of GRU, LSTM, CNN, and transformer-based models adapted for time-series forecasting. A wide range of experiments was carried out, considering various model architectures, time steps, and forecasting horizons. We trained each model with univariate and multivariate inputs, incorporating the three external factors (USDX, NETI, and SENT) individually and collectively. We compared the performance of each model in one-step and multi-step horizon forecasting, as discussed in Sect. 6. We evaluated the results obtained by each model using MAE, MSE, and RMSE metrics. This phase's objective is to identify the most effective model experimentally. It features for our case study, which will be used in the following (final) phase to produce the ensemble forecast. Finally, based on the results and observations from the previous phase, we built the ERS-Bi-GRU model, which merges forecasts of three Bi-GRU networks performing three forecasting scenarios, as detailed in Sect. 6. The ERS-GRU model demonstrated superior performance compared to other targeted and created benchmark models.

6 Experiments

This section highlights a representative sample of the experiments carried out, the evaluation metrics used, and the best results achieved by each model. However, the models' training experiments were conducted on a machine running Windows 11 with 16GB of RAM. The models were built and trained using the PyTorch deep-learning framework.

6.1 Error Evaluation Metrics

To evaluate the performance of each model, the following commonly used evaluation metrics were employed:

Mean Absolute Error (MAE) measures the average absolute difference between the predicted and actual values. It is calculated as:

$$\mathrm{MAE}(y, \hat{y}) = \frac{\sum_i^n |y_i - \hat{y}_i|}{n}$$

Mean Squared Error (MSE) calculates the average squared difference between the predicted and actual values. It is computed as

$$\mathrm{MSE}(y, \hat{y}) = \frac{\sum_i^n (y_i - \hat{y}_i)^2}{N}$$

Root-Mean-Squared Error (RMSE) is the square root of the MSE, providing a measure of the average magnitude of the error. It is given by:

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}}$$

Mean Squared Prediction Error (MSPE) is a statistical measure used to evaluate the accuracy of a predictive model. It quantifies the average squared difference between the predicted values and the actual observed values.

$$MSPE(y, \hat{y}) = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2.$$

Coefficient of Determination (R^2 or R-squared) is a statistical measure of how well the regression predictions approximate the actual data points. The following equation expresses the latter:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

These evaluation metrics allow for quantifying the accuracy and precision of the model's predictions. Lower MSE, MAE, MSPE, and RMSE values indicate better performance, indicating smaller discrepancies between the predicted and actual values, while a high R^2 value implies better performance. For a fair comparison, the models have been trained under the same procedures but with a customised hyperparameter configuration for each model. MAE, MSE, MSPE, and RMSE are calculated using actual, not scaled, prices.

6.2 Experimental Design

Table 1 shows the experiment parameters. To evaluate the performance of each one of the targeted models [GRU, LSTM, Bi-GRU, Bi-LSTM, LSTM-CNN, LSTM-CNN-attention, Autoformer, Informer, Transformer, TimesNet], a systematic approach was employed, considering various window sizes (tuning range from 5 to 22) with one-step (one-ahead) and multi-step (three-ahead days) forecasts, with more focus on the multi-step forecasting. For each model, two training scenarios were considered: (1) utilising only Brent crude price lags (univariate) and (2) incorporating different combinations of the three external factors, namely USDX, SENT, and TENI (multivariate). This approach enabled a comprehensive assessment of the models from multiple perspectives. To ensure a unified, systematic comparison, the models were applied to the same dataset as described in Section (4). A unified dataset splitting and evaluation methodology was adopted to maintain consistency throughout the experiments. Hyperparameters were tuned separately for each model based on the MSE during the training phase, considering various factors, such as window size, prediction horizon, input types (univariate/multivariate), and the inclusion of external factors.

7 Findings and Discussion

7.1 LSTM and GRU Models

Despite the emergence of newer advanced models in deep learning, LSTM and GRU models have maintained their significance and remain competitive options for time-series prediction tasks. These models excel in preserving the temporal order. The simplicity of model tuning and their lower computation complexity compared to Transformers or GANs are further advantages. Results obtained from GRU and LSTM models are shown in Tables 2 and 3. The best results achieved were by the SENT-GRU-1 model (MAE 0.0647, MSE 0.0069, and RMSE 0.0833) in terms of 1-day ahead and (MAE 1.0411, MSE 2.0097, and RMSE 0.4176) in term of a 3-days ahead forecasting with the SENT-Bi-GRU model. It is worth mentioning that the evaluation metrics (MAE, MSE, and RMSE) have been calculated using the original values (re-normalised values).

Table 2 GRU-based models

Model	Horizon	MAE	MSE	RMSE
GRU	3	1.2119	2.4764	1.5736
Bi-GRU	3	1.1049	2.1991	1.4829
SENT-Bi-GRU	3	**1.0411**	**2.0097**	**1.4176**
USD-Bi-GRU	3	1.0874	2.1378	1.4621
TENI-Bi-GRU	3	1.0960	2.1537	1.4675
SENT-GRU-1	1	**0.0647**	**0.0069**	**0.0833**
TENI-GRU-1	1	0.3104	0.1530	0.3912

7.2 Transformers and TimesNet Experiments

We examined a base Transformer model along with two well-known transformer-based architectures specifically designed for time-series tasks, namely Autoformer [11] and Informer [26]. Additionally, we examined another advanced time-series model named TimesNet, which is based on a CNN architecture using a temporal 2D-variation modelling approach [27]. Similar to the experiments presented in the previous section, we conduct experiments incorporating the three external variables in various combinations. Among the models mentioned above, the best results were obtained regarding 3-day ahead forecasting by incorporating the sentiment score with Autoformer (MAE 2.8073, MSE 5.7048, and RMSE 2.3884), as shown in Table 4. These transformer-based models exhibit inferior performance compared to sequence models. This disparity can be attributed to the self-attention mechanism applied to transformer-based models, which is somewhat "anti-order" and can lead to temporal information loss. This loss is usually not a significant concern in semantic-rich applications like natural language processing (NLP), where the meaning of a sentence remains largely preserved even if the word order is altered. However, when dealing with time-series data, the primary focus is modelling the temporal relation within a continuous sequence of points. Additionally, it is essential to acknowledge the relatively higher training time and complexities associated with Transformers regarding the difficulty in fine-tuning the hyperparameters compared to more traditional models like GRU or LSTM.

Table 3 LSTM-based models

Model	Horizon	MAE	MSE	RMSE
SENT-Bi-LSTM	3	1.0455	2.0231	1.4223
USD-Bi-LSTM	3	1.1690	2.3917	1.5465
TENI-Bi-LSTM	3	1.1570	2.3038	1.5178
SENT-Bi-CNN-LSTM	3	1.3955	3.3865	1.8402
SENT-Bi-CNN-LSTM-att	3	1.4524	3.2913	1.8141

Table 4 Transformer-based models

Model	Horizon	MAE	MSE	RMSE
SENT-Autoformer	1	1.3854	3.2847	1.8123
SENT-Autoformer	3	2.8073	5.7148	2.3884
SENT-Informer	1	3.7948	18.6935	4.3236
SENT-Informer	3	5.0463	31.8270	5.6415
SENT-TimesNet	1	1.4099	3.3947	1.8424
SENT-TimesNet	3	3.4176	7.9913	2.8269
SENT-Transformer	1	3.4512	15.251	3.9052
SENT-Transformer	3	6.5671	30.4323	5.5165
Autoformer	1	1.5357	3.8618	1.9651
Autoformer	3	5.383	36.9292	6.0769

7.3 External Variables

The impact of three external time-series variables on prediction accuracy has been investigated: USDX, TENI, and the cumulative sentimental score (SENT). The initial part of the models' names indicates the integrated variables; for example, "SENT-Bi-GRU" signifies a Bi-GRU model where the sentiment index serves as an input. The results shown in Tables 2 and 3 suggest that the cumulative sentimental score (SENT) is the most effective external time-series variable for improving the prediction accuracy of Brent crude oil price. SENT is more effective, because it captures the overall sentiment of the market, which can be a valuable predictor of future prices. USDX and TENI, on the other hand, may not be as effective predictors, because they only capture specific aspects of the market, such as the strength of the US dollar or the price of oil.

7.4 SENT-Bi-GRU

Among all the previous experiments conducted, the SENT-BI-GRU network, which is a Bi-GRU network integrated with sentiment index, demonstrated the highest accuracy performance. SENT-Bi-GRU is a simple and effective bidirectional GRU architecture incorporating a cumulative sentiment index. The architecture of this network consists of a bidirectional GRU network followed by two fully connected layers, all connected through ReLU activation functions. The best performance has been achieved with a window size of five time steps, a batch size of 16, and the AdamW optimiser.

7.5 ERS-Bi-GRU

We incorporated a residual analysis into the previous model (SENT-Bi-GRU) to refine the predictive capabilities further. The main idea is to ensemble the forecasts of two Bi-GRU networks. One network incorporates a sentimental index as an input feature, while the other includes a residual component as an input feature (Fig. 4 depicts the residual component). The forecasts from these two networks are then fused using a weighted averaging method, resulting in a more robust and accurate prediction. The objective is to enhance capturing fluctuations and irregularities in the oil market. We assume that we can model the price fluctuations with the help of the sentiment index and the residual component.

The proposed models, denoted as Ensemble Sentiment-Residual BiGRU (ESR-BiGRU), involve the following steps:

1. Train a Bi-GRU model to future crude oil prices. The model takes the historical crude oil prices and sentiment index as inputs. Let P_t represent the crude oil price at time t and S_t denote the sentiment index simultaneously. The forecast from Bi-GRU can be denoted as $forecast_1 = F_{\text{Bi-GRU}}(P_t, S_t)$.
2. Train a Bi-GRU model: This model takes sentiment index S and the residual component R (calculated by removing the trend from the oil prices) as inputs and learns to predict the unpredictable variations in oil prices. Let R_t represent the residual at time t and S_t denote the sentiment index simultaneously. The forecast from Bi-GRU can be denoted as $forecast_2 = F_{\text{Bi-GRU}}(R_t, S_t)$.
3. Train a Bi-GRU model: This model takes historical crude oil prices P and the residual component R as inputs and learns to predict future crude oil prices. The forecast from this Bi-GRU is denoted as $forecast_3 = F_{\text{Bi-GRU}}(P_t, R_t)$.
4. Calculate the final forecast: The last forecast is obtained by combining the three forecasts as follows: $Forcat_{final} = ((forecast_3 - forecast_2 * W_1) * W_2) + (forecast_1 * W_3)$.

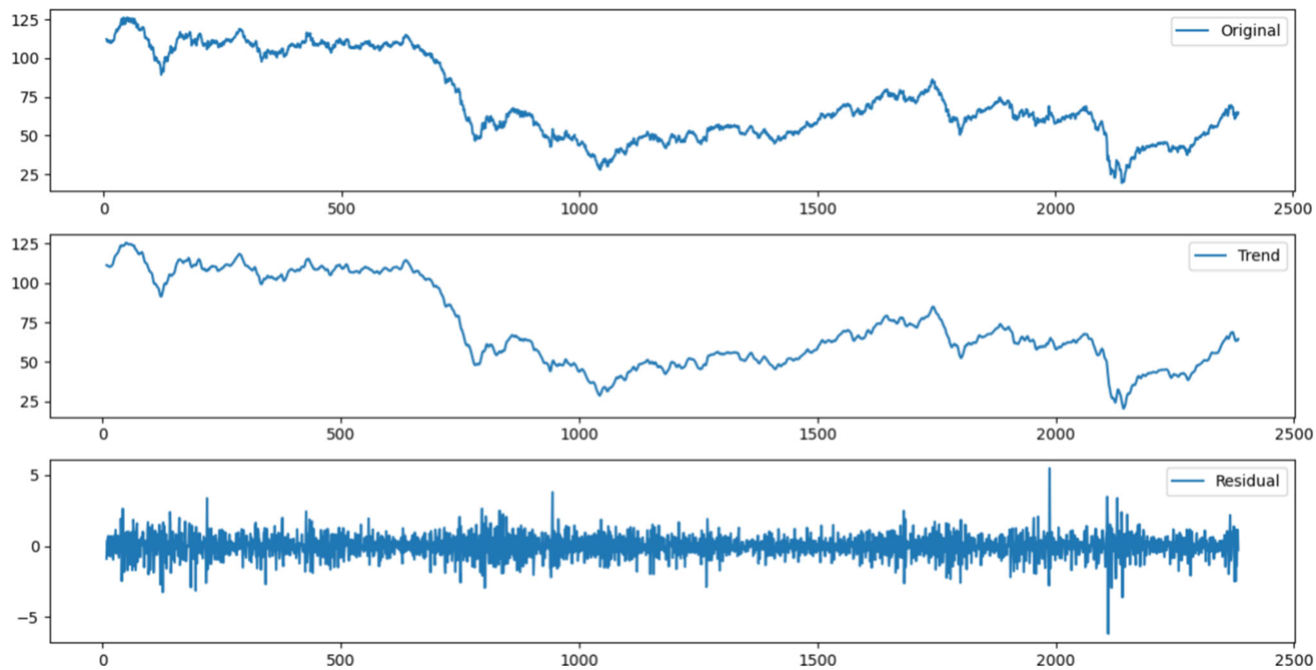

Fig. 4 Crude Brent oil price components

Table 5 ERS-Bi-GRU model results versus other created benchmark models. The values are calculated using the actual, not scaled prices

	Horizon	MAE	MSE	RMSE	MSPE	R^2
ERS-Bi-GRU	3	**1.04475**	**1.9946**	**1.4123**	**0.00079**	**0.9750**
Bi-GRU	3	1.1049	2.1991	1.4829	0.00087	0.97239
SENT-Bi-GRU	3	1.0411	2.0097	1.4176	0.00081	0.97476
SENT-Bi-LSTM	3	1.0455	2.0231	1.4223	0.00082	0.97460
SENT-Bi-CNN-LSTM	3	1.3955	3.3865	1.8402	0.00140	0.95748
SENT-Bi-CNN-LSTM-att	3	1.4524	3.2913	1.8141	0.06280	0.95867
SENT-Autoformer	3	2.8073	5.7048	2.3884	0.10241	0.92862
SENT-TimesNet	3	3.4176	7.9913	2.8269	0.27340	0.90014

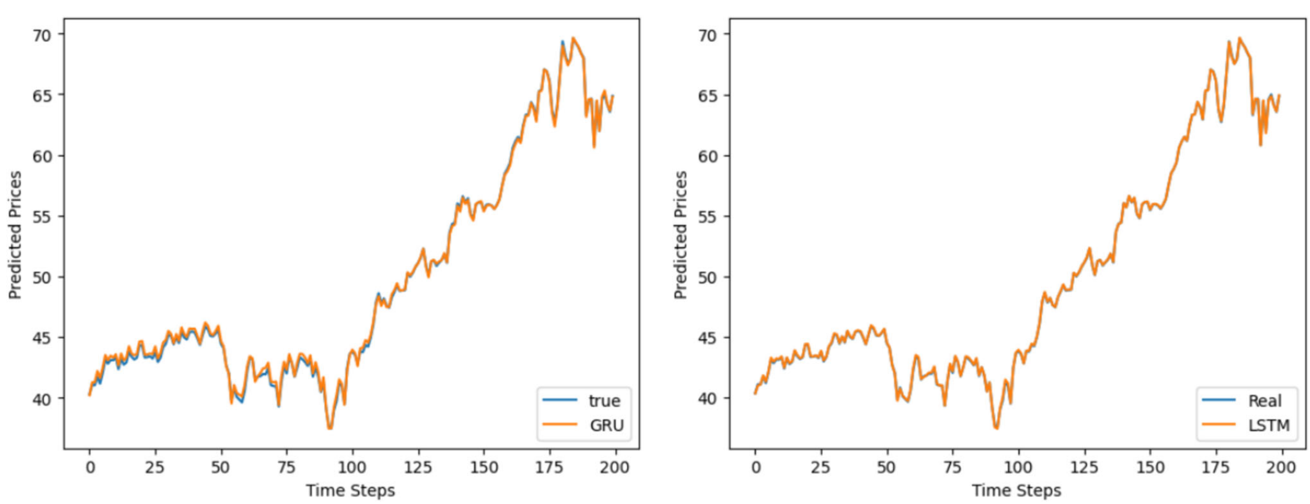

Fig. 5 One-ahead day forecasts using SENT-GRU and SENT-LSTM models

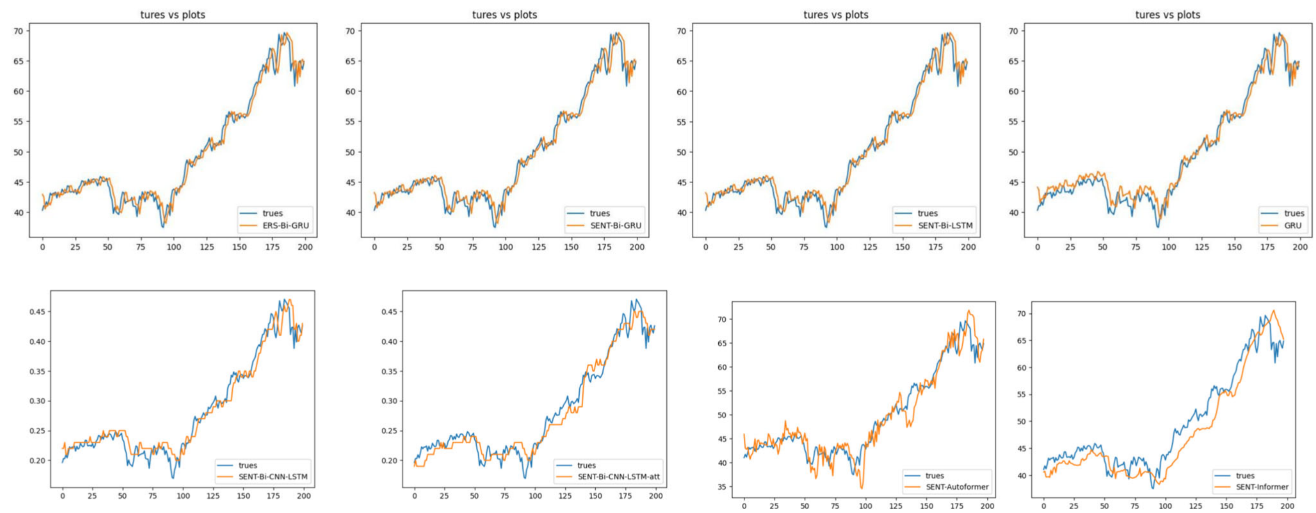

Fig. 6 3-ahead day forecasts, ERS-Bi-GRU model versus other benchmark models

This model outperforms all the other models examined in this paper and benchmark models, achieving MAE 1.04475, MSE 1.9946 and RMSE 1.4123.

7.6 Discussion and Benchmark Comparison

In general, the significance of the experiment results is as follows: First, simple models (GRU and LSTM) outperform transformer and TimesNet models for Brent crude oil price forecasting; second, incorporating crude oil-relevant sentimental index is demonstrated to be effective and promising to help in estimating the fluctuation level of the price. However, among the various models considered, our findings reveal that bidirectional (LSTM/GRU) models yield the best performance. While Bi-GRU outperforms slightly the Bi-LSTM. We obtained the best results of forecasting with the ESR-Bi-GRU model (MAE 1.04475, MSE 1.9946, and RMSE 1.4123) for a 3-day ahead forecasting, as shown in Table 5. It is worth mentioning that MAE, MSE, and RMSE have been calculated using the re-scaled predicted values. Optimal time-step values for different models are among (5, 10, 17, and 22), with 16 of batch size.

Figures 5, 6 depict samples of the forecasting results of the proposed model and the other created benchmark models. The proposed model (ERS-Bi-GRU) exhibits a high level of accuracy in forecasting the closing prices of Brent crude oil, demonstrating a solid alignment between its predictions and the observed values. This highlights the model's ability to provide precise forecasts for crude oil prices. This study focuses on forecasting Brent oil prices using a daily observation-based dataset, which limits the overall dataset size. Furthermore, our analysis is constrained to the period from 2012 to 2021 due to the unavailability of the sentiment index outside this timeframe.

8 Conclusion

This work follows a comprehensive methodology for accurate multi-step forecasting for crude Brent prices. We started by conducting a comparative analysis of the performance of various architectures of six deep-learning networks (GRU, LSTM, Autoformer, Informer, Transformer, and TimesNet) for Brent crude price forecasting. In addition, it introduces a sample and efficient model, ERS-Bi-GRU, for multi-step forecasting of Brent crude price. Our experiments go beyond the models by examining the significance of external factors that commonly influence the crude oil market. By incorporating these additional factors, the evaluation aims to provide a holistic understanding of the models' forecasting capabilities in real-world scenarios. The results demonstrated that sequence-based models (GRU and LSTM) outperform transformer-based models for forecasting crude oil prices. The results showed that the proposed ERS-Bi-GRU model outperforms benchmark models in the field and state-of-the-art models (Autoformer and TimesNet) regarding one/multi-step forecasting. The results also demonstrated the importance of extraneous factors to improve the forecasting accuracy. Finally, it is worth mentioning that the nature of the data significantly influences training models for forecasting tasks and is extremely sensitive to hyperparameter configurations. Therefore, efficient algorithms for optimising hyperparameter configuration are recommended. In future work, we seek to enhance this model by integrating optimisation algorithms (such as Particle, Swarm Optimization, Gravity Search, algorithm, and Gray Wolf Optimizer) for hyperparameter tuning. Additionally, we plan to incorporate the proposed model within a GAN architecture and add SDE solvers for modelling volatility and noise. Furthermore, we intend to apply this approach to financial data.

Author Contributions Mohammed Alruqimi (A): Methodology, Experiments implementation, writting the manuscript. Luca Di Persio (B): Methodology, reviewed the manuscript.

Funding No external funding was received for this research.

Data Availability The dataset used in this paper can be accessed through this GitHub repos.

Code Availability It will be available upon request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval and consent to participate Not applicable.

References

1. Miao, H., Ramchander, S., Wang, T., Yang, D.: Influential factors in crude oil price forecasting. Energy Econ. **68**, 77–88 (2017). https://doi.org/10.1016/j.eneco.2017.09.010
2. Lu, Q., Sun, S., Duan, H., Wang, S.: Analysis and forecasting of crude oil price based on the variable selection-lstm integrated model. Energy Inform. **4**(2), 47 (2021). https://doi.org/10.1186/s42162-021-00166-4
3. Jiang, Z., Zhang, L., Zhang, L., Wen, B.: Investor sentiment and machine learning: predicting the price of China's crude oil futures market. Energy **247**, 123471 (2022). https://doi.org/10.1016/j.energy.2022.123471
4. Drachal, K.: Forecasting crude oil real prices with averaging time-varying var models. Resour. Policy **74**, 102244 (2021). https://doi.org/10.1016/j.resourpol.2021.102244
5. Mohammadi, H., Su, L.: International evidence on crude oil price dynamics: applications of arima-garch models. Energy Econ. **32**(5), 1001–1008 (2010). https://doi.org/10.1016/j.eneco.2010.04.009
6. Huang, Y., Deng, Y.: A new crude oil price forecasting model based on variational mode decomposition. Knowl. Based Syst. **213**, 106669 (2021). https://doi.org/10.1016/j.knosys.2020.106669
7. He, K., Yu, L., Lai, K.K.: Crude oil price analysis and forecasting using wavelet decomposed ensemble model. Energy **46**(1), 564–574 (2012). https://doi.org/10.1016/j.energy.2012.07.055. (**Energy and Exergy Modelling of Advance Energy Systems**)
8. Ftiti, Z., Jawadi, F.: On the oil price uncertainty. Energy J. **0**(Special I) (2019)
9. Cen, Z., Wang, J.: Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer. Energy **169** (2018). https://doi.org/10.1016/j.energy.2018.12.016
10. Lindemann, B., Müller, T., Vietz, H., Jazdi, N., Weyrich, M.: A survey on long short-term memory networks for time series prediction. Proc. CIRP **99**, 650–655 (2021) https://doi.org/10.1016/j.procir.2021.03.088. (14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 15-17 July 2020)
11. Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. CoRR (2021). arXiv:2106.13008
12. Deng, C., Ma, L., Zeng, T.: Crude oil price forecast based on deep transfer learning: Shanghai crude oil as an example (2021). https://doi.org/10.3390/su132413770
13. Eren, Y.: Küçükdemiral: a comprehensive review on deep learning approaches for short-term load forecasting. Renew. Sustain. Energy Rev. **189**, 114031 (2024). https://doi.org/10.1016/j.rser.2023.114031
14. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? (2022)
15. Shen, L., Wei, Y., Wang, Y.: Gbt: two-stage transformer framework for non-stationary time series forecasting. Neural Netw. **165**, 953–970 (2023). https://doi.org/10.1016/j.neunet.2023.06.044
16. Jovanovic, L., Jovanovic, D., Bacanin, N., Jovancai Stakic, A., Antonijevic, M., Magd, H., Thirumalaisamy, R., Zivkovic, M.: Multi-step crude oil price prediction based on lstm approach tuned by salp swarm algorithm with disputation operator. Sustainability **14**(21) (2022). https://doi.org/10.3390/su142114616
17. Wang, B., Wang, J.: Energy futures and spots prices forecasting by hybrid sw-gru with emd and error evaluation. Energy Econo. **90**, 104827 (2020). https://doi.org/10.1016/j.eneco.2020.104827
18. Zhang, S., Luo, J., Wang, S., Liu, F.: Oil price forecasting: a hybrid gru neural network based on decomposition-reconstruction methods. Expert Syst. Appl. **218**, 119617 (2023). https://doi.org/10.1016/j.eswa.2023.119617
19. Fang, Y., Wang, W., Wu, P., Zhao, Y.: A sentiment-enhanced hybrid model for crude oil price forecasting. Expert Syst. Appl. **215**, 119329 (2023). https://doi.org/10.1016/j.eswa.2022.119329
20. He, K., Yang, Q., Zou, Y.: Crude oil price prediction using embedding convolutional neural network model. Proc. Comput. Sci. **214**, 959–964 (2022). https://doi.org/10.1016/j.procs.2022.11.265. (**9th International Conference on Information Technology and Quantitative Management**)
21. Liu, C., Ruan, K., Ma, X.: Dmeformer: a newly designed dynamic model ensemble transformer for crude oil futures prediction. Heliyon **9**(6), 16715 (2023). https://doi.org/10.1016/j.heliyon.2023.e16715
22. Kaplan, H., Weichselbraun, A., Braşoveanu, A.M.P.: Integrating economic theory, domain knowledge, and social knowledge into hybrid sentiment models for predicting crude oil markets. Cogn. Comput. (2023). https://doi.org/10.1007/s12559-023-10129-4
23. Zhao, L.-T., Xing, Y.-Y., Zhao, Q.-R., Chen, X.-H.: Dynamic impacts of online investor sentiment on international crude oil prices. Resour. Policy **82**, 103506 (2023). https://doi.org/10.1016/j.resourpol.2023.103506
24. Hu, Zhenda: Crude oil price prediction using ceemdan and lstm-attention with news sentiment index. Oil Gas Sci. Technol. Rev. IFP Energies Nouvelles **76**, 28 (2021). https://doi.org/10.2516/ogst/2021010
25. Shan, L., Liu, Y., Tang, M., Yang, M., Bai, X.: Cnn-bilstm hybrid neural networks with attention mechanism for well log predic-

tion. J. Pet. Sci. Eng. **205**, 108838 (2021). https://doi.org/10.1016/j.petrol.2021.108838
26. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: beyond efficient transformer for long sequence time-series forecasting. CoRR (2020). arXiv:2012.07436
27. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: TimesNet: temporal 2D-variation modeling for general time series analysis (2023)
28. Sherstinsky, A.: Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. Physica D: Nonlinear Phenomena **404**, 132306 (2020). https://doi.org/10.1016/j.physd.2019.132306
29. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling (2014)
30. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need (2023)
31. Coleman, L.: Explaining crude oil prices fotios petropoulosing fundamental measures. Energy Policy **40**, 318–324 (2012). https://doi.org/10.1016/j.enpol.2011.10.012. (**Strategic Choices for Renewable Energy Investment**)
32. Miao, H., Ramchander, S., Wang, T., Yang, D.: Influential factors in crude oil price forecasting. Energy Econ. **68**, 77–88 (2017). https://doi.org/10.1016/j.eneco.2017.09.010
33. Li, J., Xu, Z., Xu, H., Tang, L., Yu, L.: Forecasting oil price trends with sentiment of online news articles. Asia-Pacific J. Oper. Res. **34**(02), 1740019 (2017). https://doi.org/10.1142/S021759591740019X