




Research Article

A New Approach for Forecasting Crude Oil Prices Using Median Ensemble Empirical Mode Decomposition and Group Method of Data Handling

Waqas Ahmad,¹ Muhammad Aamir ¹, Umair Khalil,¹ Muhammad Ishaq ²,
Nadeem Iqbal ³, and Mukhtaj Khan ³

¹Department of Statistics, Abdul Wali Khan University Mardan, Mardan, Khyber Pakhtunkhwa, Pakistan

²School of Natural Sciences, National University of Sciences and Technology, Islamabad, Pakistan

³Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, Khyber Pakhtunkhwa, Pakistan

Correspondence should be addressed to Mukhtaj Khan; mukhtaj.khan@awkum.edu.pk

Received 28 January 2021; Revised 6 May 2021; Accepted 19 May 2021; Published 31 May 2021

Academic Editor: Bin Liu

Copyright © 2021 Waqas Ahmad et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The accuracy of time series forecasting is more important and can assist organizations to take up-to-date decisions for better planning and management. Several classical econometrics and computational approaches show promising results for the ordinary time series forecasting tasks, but they are not satisfactory in crude oil price forecasting. Ensemble empirical mode decomposition (EEMD) not only resolves the problem of nonlinearity and nonstationarity of time series prediction but also creates some problems (i.e., mood mixing and splitting). In this study, we proposed a new hybrid method that combines the median ensemble empirical mode decomposition and group method of data handling (MEEMD-GMDH) to reduce mood splitting problems and forecast crude oil price. MEEMD is achieved by replacing the mean operator with the median operator during the EEMD process. For testing and validation purposes of the different models, the two-seat stamp benchmarked crude oil price data are used (i.e., Brent and West Texas Intermediate (WTI)). To check the proposed model performance, different evaluation measures are used including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Diebold-Mariano (DM) test. All the forecasting accuracy measures confirmed that our proposed model performs well in crude oil prices forecasting as compared to other hybrid models.

1. Introduction

The global market price fluctuates dramatically and increases in the long term. Commodity prices' fluctuation causes a massive impact on the global economy, for example, soaring the cost of imports, stimulating inflation, sluggish growth in the economy, and decreasing the efficacy of macroeconomic policy. Thus, analyzing the characteristics of international market variations in goods in order to forecast the price and pattern is critical for the world economy. Goods are vital for global growth from the perspective of nations and governments and have a significant strategic effect on national economic stability. If consumer prices can be more reliably estimated, goods can be imported at low prices to significantly

relieve imported inflation pressures. As import prices drop, government subsidies to businesses may be lowered and fiscal policy stability increases. Moreover, the currency reserves of a nation could be used to accommodate the flexibility of exchange rates to improve the resilience of monetary policies regardless of the decline of foreign exchange spending.

In the opinion of producers, goods are raw materials for the aviation, shipping, and food processing sectors. They are also items for the oil mining industry and nonferrous metallurgy businesses. Fluctuations of product markets influence the costs and earnings of companies. The effective fluctuation of the value of goods helps farmers to accurately schedule production, minimize the cost, and achieve greater profitability. As for exchange firms, sharp commodity price

swings are also the result of significant losses. The probability of market volatility is therefore underdetermined due to the lack of an analysis team, operational team, and the required decision-making process. In fact, the expected price contrasts considerably with the real price when dealing with product prospects. By creating a model of projection for consumer goods and a method of mitigation fluctuations in prices, trading firms are able to avoid risks and lower trade losses. In general, it is important and urgent to study commodity price predictions to provide rational support for the government and business decision-makers.

Crude oil is one of the core major natural products, with demand and supply exceeding 80 million barrels per day, because it covers two-thirds of the world's direct energy consumption [1]. Oil assumes an undeniably important role in the global economy since about 66% of the world's energy utilization comes from unrefined petroleum and gasoline. Sharp oil price value improvements are most likely going to shake aggregate economic activity, especially since Jan 2004, the world's oil cost has been rising rapidly and is creating striking fluctuations for the world economy. Consequently, unss oil prices are a source of major zeal for many analysts, research experts, and organizations. The price of crude oil is essentially dictated by its demand and supply but is more clearly affected by numerous unpredictable past/present/future occurrences, such as climate change, stock levels, GDP development, and political perspectives. These realities lead to a distinctly varying and nonlinear market and the basic component of maintaining the intricate dynamic is not understood.

In 2019, global oil consumption reached 1,0075 million barrels per day considering data from the International Energy Agency (IEA). Oil indeed plays the most important role in fulfilling global energy needs. Asian emerging-market countries have become the key contributors to the rising demand for crude oil. The fast economic development led them to dramatically raise the demand for crude oil. The crude oil demand has increased due to the expeditious economic growth. China's oil consumption, for instance, has risen from an average of 69,700 barrels per day in 2005 to 145,100 barrels per day in 2019. As a demand factor, rising crude oil prices would result in higher production costs for the nonoil companies and a shrink in profit [2]. As crude oil is a very critical commodity to the global economy, many leading governments, investors, and scholars have invested a lot of effort in building models to predict fluctuations in their prices and important properties. Given its complexity, price charts are susceptible to factors such as supply and demand, speculative activities, competition between suppliers, development of technology, and endless war [3, 4].

Due to the nonlinear and complex nature, it is difficult for humans to understand the high volatility in crude oil prices. In the past, crude oil prices for the West Texas Intermediate (WTI) peaked in July 2008 at USD 145.31 per barrel. But the price fell sharply to USD 30.28 per barrel due to the financial crisis, which was about 80 percent from the high at the end of 2008. Prices climbed to \$113 per barrel in April 2011, when the economy boomed, but in February

2016, it dropped again to \$27 per barrel, owing to certain political causes and demand and supply variations [5]. From different perspectives, the impact of crude oil price fluctuations on the national economy is reflected in two aspects. In the first aspect, soaring crude oil prices have seriously affected the economic empowerment of oil-importing economies. The second aspect is that the decline in crude oil prices (such as the decline in 1998) has caused serious budgetary deficit problems for oil-exporting countries [6]. Since crude oil price series are generally considered to be nonlinear and nonstationary time series, they can be accurately influenced by several factors; therefore, accurately predicting the price of oil can be quite challenging. Since the oil price pattern displayed nonlinear, nonstatic, or multiscaling elements, researchers started to analyze oil price volatility by using multiscale techniques, such as the wavelet analysis and the analytical decomposition mode (EMD). These techniques have a strong time and frequency resolution and can increase the regularity of the variations. There are several methods of analyzing and predicting oil prices developed by researchers that can be separated approximately into single models and mixed models. Single models include observational approaches, methods for causal inference, times, and math. Combined models are made according to such laws by integrating single models.

In the past decades, future observation and prediction based on time series data have attracted great attention in many research fields. To predict the future behavior of a particular phenomenon, many techniques have been developed to address this issue, such as cointegration analysis, vector error correction model (VECM), vector autoregression (VAR), linear-regression (linR), random walk model, GARCH, and ARIMA models. Other than that, computational approaches such as empirical mode decomposition (EMD), artificial neural network (ANN), and ensemble empirical mode decomposition (EEMD) have also been used. Gülen [7] used a cointegration methodology to predict the WTI crude oil price. Lanza et al. [8] utilized the error correction model (ECM) to predict crude oil prices. Another famous methodology is the GARCH model; likewise, [9] used the GARCH properties to predict Brent crude oil price. Mohammadi and Su [10] applied the ARIMA-GARCH model on weekly crude oil spot prices in eleven international markets, to forecast the conditional mean and volatility. ANN and ARIMA models were used to predict the future price of WTI crude oil [11]. They documented a comparative analysis between the ANN and ARIMA models to show the techniques with the best results based on the forecasting accuracy measures including Mean Absolute Error (MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE). The scholars concluded that the ANN had better prediction results than the ARIMA model. Mirmirani and Li [11] investigated US oil prices using vector autoregression (VAR) and ANN and concluded that BPN-GA attains the best results. Ahmad [12] predicted the Oman crude oil prices using the ARIMA model and proved that ARIMA (1, 1, 5)*(1, 1, 1) achieved the best results. Aamir et al. [13] used the complete ensemble

empirical mode decomposition with adaptive noise (CEEMDAN) hybrid model to forecast crude oil prices for both Brent and WTI markets. Three-layer feedforward neural network (FNN) was incorporated by [14] for forecasting short-term crude oil price. The authors [15, 16] forecast the crude oil price by using a support vector machine (SVM) and contrasted the results with back-propagation neural network (BPNN) and ARIMA models. Their findings indicated that SVM outperformed the BPNN and ARIMA models. The aim of the authors [17] was to achieve some improvements in the prediction of oil price volatility by using an Artificial Neural Network Generalized Autoregressive Conditional Heteroscedasticity ANN-GARCH hybrid model combined with financial variables. They concluded that the hybrid model improves the volatility prediction accuracy by more than 30% through the results measured by the heteroscedasticity-adjusted mean square error (HMSE) model. Lin and Sun [2] used CEEMDAN-MLGRU based decomposition method to forecast WTI crude oil price. Li et al. [18] used a new hybrid-based model, namely, EEMD-SBL-ADD, and concluded that the proposed model is promising for forecasting crude oil price. The authors [19] also worked on Box-Jenkins (ARIMA) and neural network (NN) models to forecast repairable system failure analysis. They concluded that both models performed better with short-term forecasting; however, NN as compared to the ARIMA model gave satisfactory performances. Moreover, the authors in [20] used machine-learning (decision tree) models to forecast crude oil prices. This study also concluded that the decision tree models achieved higher prediction accuracy than benchmark models such as multiple linR and ARIMA. Due to nonlinear features, the classical time series models were not capable of predicting crude oil prices accurately [21, 22].

Motivated by the potential of median ensemble empirical mode decomposition (MEEMD) in signal decomposition, we proposed a new method for the prediction of crude oil prices combining MEEMD, namely, the median ensemble empirical mode decomposition and group method of data handling (MEEMD-GMDH), stimulated by the capability of MEEMD in the signal breakdown in order to minimize mood splitting and mixing problems. In particular, there are three phases of MEEMD-GMDH. First, MEEMD is used to decompose the daily raw price range for crude oil into several relatively simple components. Secondly, we use GMDH inputs from the autoregressive term to separately forecast each component. Finally, the predicted results for each component are aggregated as a result of the final forecast.

This paper contains the following main contributions:

- (1) We are proposing a new MEEMD framework that uses the median operator rather than the mean operator during an ensemble noisy intrinsic mode function (IMF) trial.
- (2) We forecast crude oil prices by integrating MEEMD-GMDH following the “decomposition and ensemble” framework. To the best of our knowledge, this blend is used for forecasting purposes for the first time.

- (3) Experimental outcomes show that the approach proposed is beneficial for forecasting crude oil prices.

The remainder of the study is structured in the sequence as follows: In Section 2, we concisely explain MEEMD, GMDH, ANN, and ARIMA models. Section 3 formulates the proposed MEEMD-GMDH model in detail. To evaluate the proposed model, results and discussion are presented in Section 4 and finally, and Section 5 concludes this paper.

2. Methodology

2.1. MEEMD. EMD is one of the popular and widely used decomposition methods for nonlinear and nonstationary time series forecasting. The EMD decomposes the data into IMFs along with a residual. Intermittent recurrence of signals in EMD is usually due to mode mixing and mode splitting. Mode mixing is defined as one IMF containing different scales while mode splitting is defined as the spread of one scale over two or more IMFs. To remove the effect of mode mixing from EMD, an additional white noise term is added to the original signal before applying EEMD. The new added white noise solves the problem of mode mixing; however, it inevitably creates new mode splitting due to two main reasons: (i) Signals having a scale located in the overlapping region of the EMD equivalent filter would have a finite probability of mode splitting [23]; (ii) The added white noise cannot guarantee full uniformity across all scales, which may lead unexpected signal intermittency and irregularity [24]. To reduce the mode splitting problem, [25] proposed MEEMD. It is the variation of the EEMD method that uses the median operator instead of a mean operator during the ensemble noisy IMF trial. The steps of MEEMD are as follows:

- (i) Create the ensemble:

$$Z_n(t) = X(t) + \mu_n(t). \quad (1)$$

For $n = 1, \dots, N$, where $\mu_n(t) \sim N(0, \sigma^2)$.

- (ii) Perform and decompose every member of $Z_n(t)$ into a maximum number (M_n) of IMFs using standard (EMD) to obtain the IMFs $\{d_n^m(t)\}_{m=1}^{M_n}$ and one residue $r_n(t)$.

$$Z_n(t) = \sum_{m=1}^{M_n} d_n^m(t) + r_n(t). \quad (2)$$

- (iii) Use the median operator to obtain final IMFs within MEEMD. The IMFs are computed as

$$d_m(t) = \text{median}\{d_1^m(t) + d_2^m(t) + \dots + d_N^m(t)\}. \quad (3)$$

In normal distribution $N \sim (\mu, \sigma^2)$, the median is asymptotically normal [2]. That is,

$$\begin{aligned} \text{mean} &= m, \\ \sigma^2 &= \frac{\pi}{2N} \epsilon^2. \end{aligned} \quad (4)$$

The added noise for MEEMD is as follows:

$$\sigma N = \sqrt{\frac{\pi}{2N}} \epsilon, \quad (5)$$

where σN is the final standard deviation which is equal to the difference between input and sum of IMFs. The flowchart of the MEEMD is presented in Figure 1.

2.2. ARIMA Model. Box and Jenkins (1976) introduced the Box-Jenkins technique (ARIMA models) in the area of time series analysis [26]. An ARIMA model predicts a value in a response time series as a linear combination of its own p^{th} previous values, the previous error, and current and previous values of other time series. Box-Jenkins versions are highly versatile due to the use of both AR and MA concepts. The time series (Y_t) interdependencies are measured by AR terms, while the previous error conditions depend on the MA terms. The following form is given for an ARMA order model (p , d , and q) for a univariate series.

The ARIMA process of order (p , d , and q) is defined as follows:

$$Y_t = \gamma_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}. \quad (6)$$

Here, Y_t is the target value of time series t , Y_{t-j} is the lagged previous values, and ϕ_j is the coefficients of lagged previous time series values, whereas θ_j is the coefficients of the previous error term, ϵ_t is an error term with normally distributed $(0, \delta^2)$, and ϵ_{t-1} is the previous error terms. As most of the time series are usually nonstationary, the ARIMA model needs stationary data [27]. This can be achieved by differentiating the time series data. The ACF and PACF plots can be used to select the appropriate order of AR and MA terms.

2.3. Artificial Neural Network Model. ANN is a very successful model for time series forecasting [28]. One of the characteristics of ANNs is their universal approximation; that is, ANNs can estimate any nonlinear continuous function up to any desired degree of accuracy [29, 30]. The most commonly used for time series forecasting is single hidden layer feedforward neural network (SLFN). The relationship between output (y_t) and the inputs (y_1, \dots, y_{t-p}) has the following output:

$$y_t = \phi_0 + \sum_{j=1}^q \phi_j g \left(\theta_{0j} + \sum_{i=1}^p \theta_{ij} y_{t-i} \right) + \mu_t, \quad (7)$$

where θ_{ij} ($i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$) and ϕ_j ($j = 0, 1, 2, \dots, q$) are the weights, ϕ_0 and θ_{0j} are the bias term, Y_{t-i} ($i = 1, 2, \dots, p$) represents the input nodes, and $j = (1, 2, \dots, q)$ is hidden nodes. Using the logistic function as the hidden layer activation function g , the first layer is the input layer where the data are introduced to the network, the second layer is the hidden layer, and the last layer is the output layer where the result of a given input is produced. The ANN architecture is shown in Figure 2.

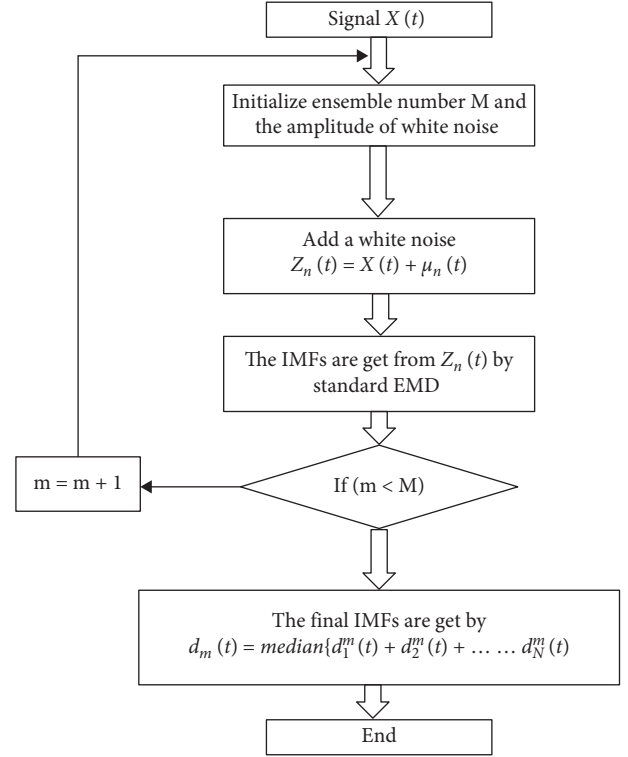


FIGURE 1: Flowchart of median ensemble empirical mode decomposition (MEEMD).

2.4. Group Method of Data Handling (GMDH). The idea of GMDH was first proposed by Ivakhnenko in 1966, as an inductive learning algorithm [31]. According to [32], GMDH methodology solves higher-order regression polynomials, that is, solving modeling and classification problems. In the time series forecasting, the GMDH algorithm identifies the relationship between the variables based on their lag values. The GMDH methodology automatically chooses the process to follow in the algorithm after training the relationship between variables. The authors in [33] analyzed that GMDH has the ability to generalize and can fit the complexity of nonlinear systems. The Ivakhnenko polynomial is defined as follows:

$$y = \alpha_0 + \sum_{i=1}^T \alpha_i x_i + \sum_{i=1}^T \sum_{j=1}^T \alpha_{ij} x_i x_j + \sum_{i=1}^T \sum_{j=1}^T \sum_{k=1}^T \alpha_{ijk} x_i x_j x_k + \dots \quad (8)$$

Here, y represents the response variable, α_0 represents weights (coefficients), and $x_i x_j$ represents lagged time series data. GMDH model consists of the following five steps:

- Step 1. *Step 1.* Divide the data into training and testing sets. Choose input $x_1, x_2, x_3, \dots, x_T$ (variables). T is equal to the number of inputs. Construct the GMDH model using train data, while evaluating the estimated model using the testing dataset.
- Step 2. For partial description of GMDH, choose the new k variables ($Z = z_1, z_2, \dots, z_k$) where k is the number of combinations i.e.,

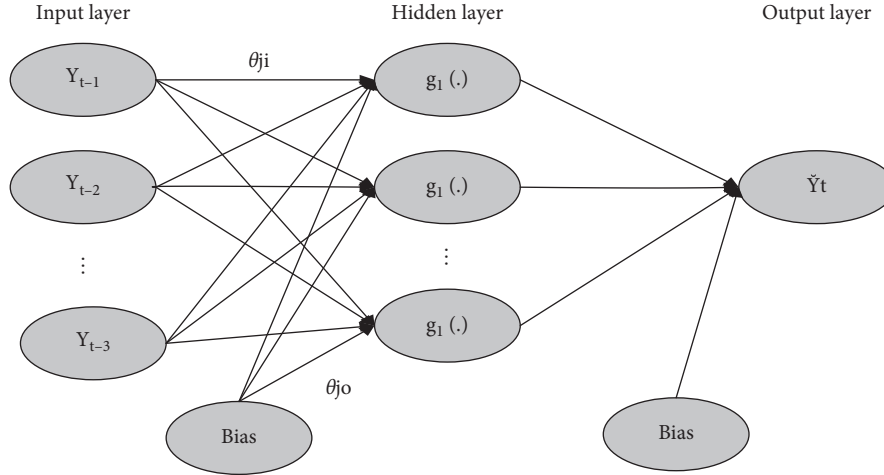


FIGURE 2: The architecture of the ANN model.

$k = (T^* (T - 1)/2)$. The partial description is the form of

$$f = (x_1, x_2, \dots, x_n) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_j + \alpha_3 x_{ij} + \alpha_4 x_i^2 + \alpha_5 x_j^2 + \dots \quad (9)$$

Many researchers consider that the partial description is a transfer function. There are many types of transfer functions. In this study, we use the radial basis function (RBF). The radial basis function is the form of

$$z_k = e^{-g_k^2}. \quad (10)$$

Step 3. Estimate the vector of coefficients of partial description using the least square method.

$$A_i = (X_i^T X_i)^{-1} X_i Y. \quad (11)$$

Here, $A_i = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n)$ is a polynomial coefficient. $Y = (y_1, y_2, \dots, y_n)$ is the observed values, and

$$X = \begin{bmatrix} x_{i1} & x_{j1} & x_{i1}x_{j1} & x_{i1}^2 & x_{j1}^2 \\ x_{i2} & x_{j2} & x_{i2}x_{j2} & x_{i2}^2 & x_{j2}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{in} & x_{jn} & x_{in}x_{jn} & x_{in}^2 & x_{jn}^2 \end{bmatrix}. \quad (12)$$

Step 4. In this step, identify new (inputs) for the second layer. Based on some criteria to choose the input (variables) for the second layer, choosing the best variables is based on some performance index, MSE, and Relative Mean Square Error (RMSE). The best neuron out of k neurons identified the value of MSE of checking dataset such that

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - z_{i,h})^2, \quad \text{for } h = 1, 2, \dots, k. \quad (13)$$

Those values of Z whose MSE is less than the threshold; then stop the process. Otherwise, new input is $[x_1, x_2, \dots, x_k]$.

In criteria 2, ignore the weakest variables, and replace x_1, x_2, \dots, x_k by those columns of z_1, z_2, \dots, z_k that best estimate the response variable in the checking set.

$$x_1 = z_1, x_2 = z_2, \dots, x_k = z_k. \quad (14)$$

Step 5. In this step, check the stopping criterion. Whether a set of polynomials of the model is further improved, the lowest value of MSE obtained in the current layer is compared with the smallest value of MSE obtained in the previous layer. If an improvement is achieved, one goes back and repeat steps 1 to 5; otherwise, the process is stopped, and the algorithm has been completed. And finally, the GMDH model is shown in Figure 3.

3. The Proposed MEEMD-GMDH Model

Due to nonlinear features, the classical time series models are not capable of predicting crude oil prices accurately [21]. Therefore, inspired by the advantage of MEEMD in this study, a novel approach that integrates MEEMD-GMDH is used for forecasting crude oil price. The decomposition and ensemble framework of MEEMD-GMDH consists of the following steps and is shown in Figure 4.

Step 1. Decomposition of data: MEEMD is applied to decompose the $\{X(t)\}$ crude oil prices series into two parts: (i) IMFs components and (ii) one residue component $r_n(t)$.

Step 2. Individual prediction: divide the data into training and testing sets. Construct the GMDH model using train data, while evaluating the estimated model using the testing dataset.

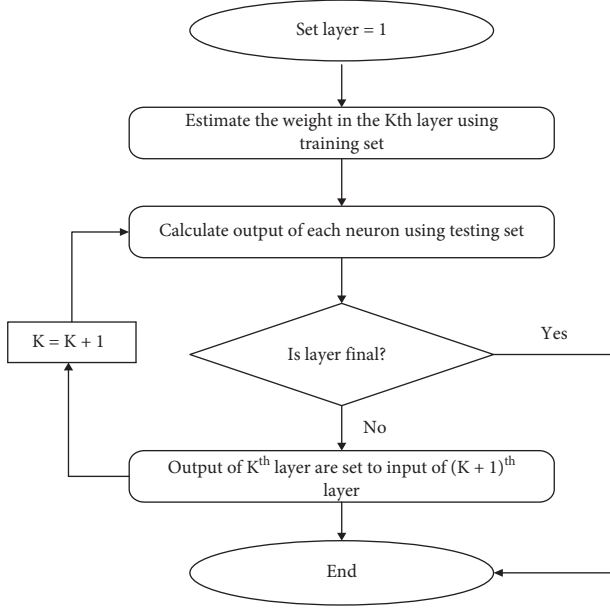


FIGURE 3: GMDH network flowchart.

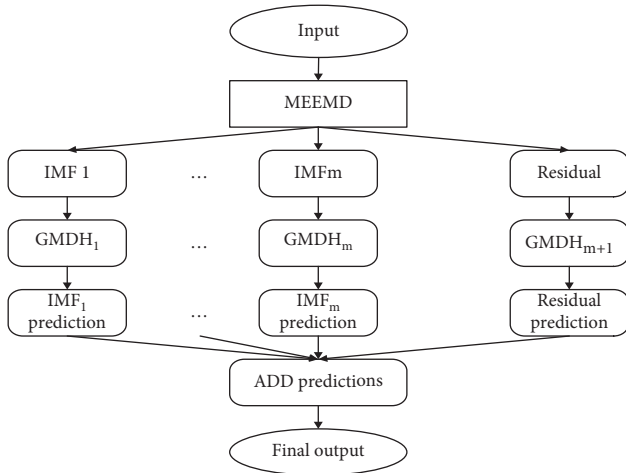


FIGURE 4: Flowchart of the proposed MEEMD-GMDH model.

Step 3. Ensemble prediction: the test results for all IMFs from Step 2 are composed by adding as the final prediction results.

4. Experimental Results

4.1. Data Description. In this study, daily crude oil prices of time series data are utilized, that is, WTI and Brent. The WTI series consists of 8000 observations from Feb 10, 1989, to Oct 10, 2019; 80 percent (6400 observations) are used as a training set while 20 percent (1600 observations) are used as a testing set. The Brent dataset consists of 12000 observations from Dec 10, 1973, to Oct 10, 2019; 80 percent (9600 observations) are used as a training set whereas 20 percent (2400 observations) are used as an assessment set to check the model performances. The distribution of training and

checking data series is 80 percent and 20 percent, respectively [34].

4.2. Evaluation Criteria. The forecasting accuracy measures are the most important criteria when competing models occur. In this study, the forecasting capacity of the models is measured using four criteria, as presented in Table 1, where n is the number of data points, Y_t represents observed values, and \hat{Y}_t represents the predicted values.

The Diebold-Mariano (DM) test statistic compares the prediction error of the two models, where $\bar{d} = (1/n) \sum_{t=1}^n d_t$, $d_t = (y_t - f_{1,t})^2 - (y_t - f_{2,t})^2$, $f_{1,t}$ is the forecasts achieved from the 1st model, and $f_{2,t}$ is the forecasts from the second model. The original WTI crude oil series is decomposed into 10 IMFs and one residue component, while the Brent series is decomposed into 11 IMFs and one residue component using MEEMD. We add the white noise with a standard deviation of 0.02 and the number of ensemble sizes is equal to the size of data. Both of them are illustrated in Figures 5 and 6.

4.3. Fitting Proposed Model to the Data. In designing the GMDH model, one must determine the number of input variables. The selection of input corresponding to the number of variables plays an important role in many successful applications of the GMDH model. According to [35], no theory can be used to guide the selection of the number of inputs. To make the MEEMD-GMDH model, we choose the best order (p , d , and q) of the ARIMA model for every k th IMFs based on AIC and BIC that are used to determine the input variables for the proposed model.

4.4. Predictive Performance of Single Models. We proposed a hybrid model that includes two components: a decomposition by MEEMD and forecasting by the GMDH model. The compared single models include GMDH, ANN, and classical time series ARIMA. In this study, the comparison of the three competing single models concerning forecasting evaluation (accuracy) for testing datasets is presented in Table 2.

From Table 2, among all these models, GMDH model attains the smallest value (better performance) on the metrics (RMSE, MAE, and MAPE). As shown in Table 2, the single ANN model performed better than the classical ARIMA model. Table 2 also indicated that the ARIMA model performed worst, because the classical econometric and time series method does not perform well for nonlinear time series.

The forecasting evaluation measures for single models are shown in Table 2; some interesting conclusions can be drawn:

- (1) The forecasting evaluation performance (RMSE) for both crude oil markets is presented in Table 2. The decision is made from the RMSE value that GMDH (single model) got the lowest value and outperformed the other single models (ANN; ARIMA) for both markets.

TABLE 1: Criteria to assess the competing models.

Criterion	Formula
Root Mean Square Error	$RMSE = \sqrt{(1/n) \sum_{t=1}^n (\hat{Y}_t - Y_t)^2}$
Mean Absolute Error	$MAE = (1/n) \sum_{t=1}^n \hat{Y}_t - Y_t $
Mean Absolute Percentage Error	$MAPE = (1/n) \sum_{t=1}^n (\hat{Y}_t - Y_t)/Y_t * 100$
Diebold-Mariano (DM)	$DM = (\bar{d})/\sqrt{\text{var}(\bar{d})}$

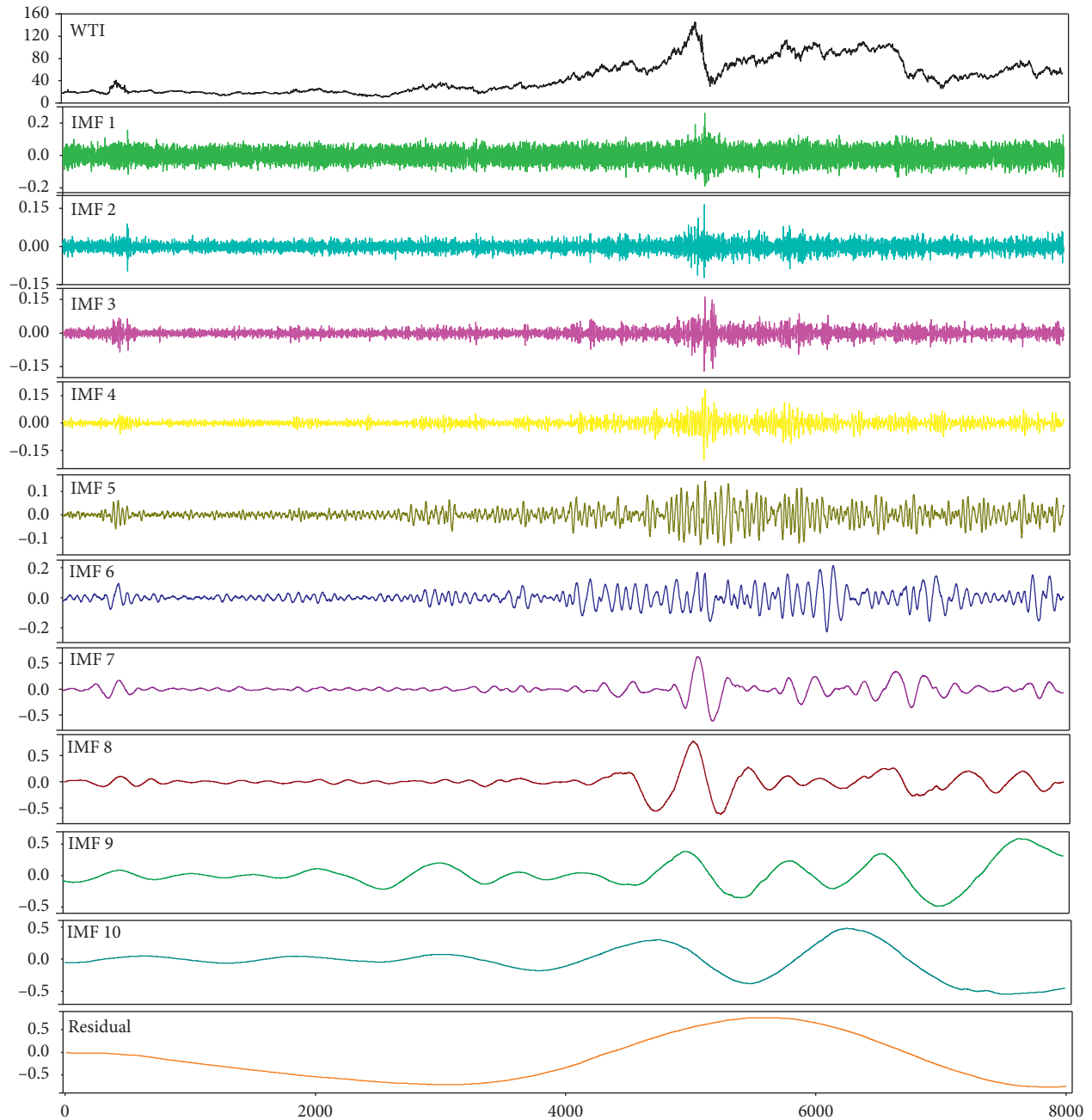


FIGURE 5: The decomposition results of WTI series using MEEMD.

- (2) The MAPE value of GMDH (single model) 0.2994% and 0.7584% for both markets (Brent and WTI) lies in the classification of perfect forecasts.
- (3) The ANN model attained second (lowest values) and the ARIMA model has achieved the third rank in the forecasting performance.

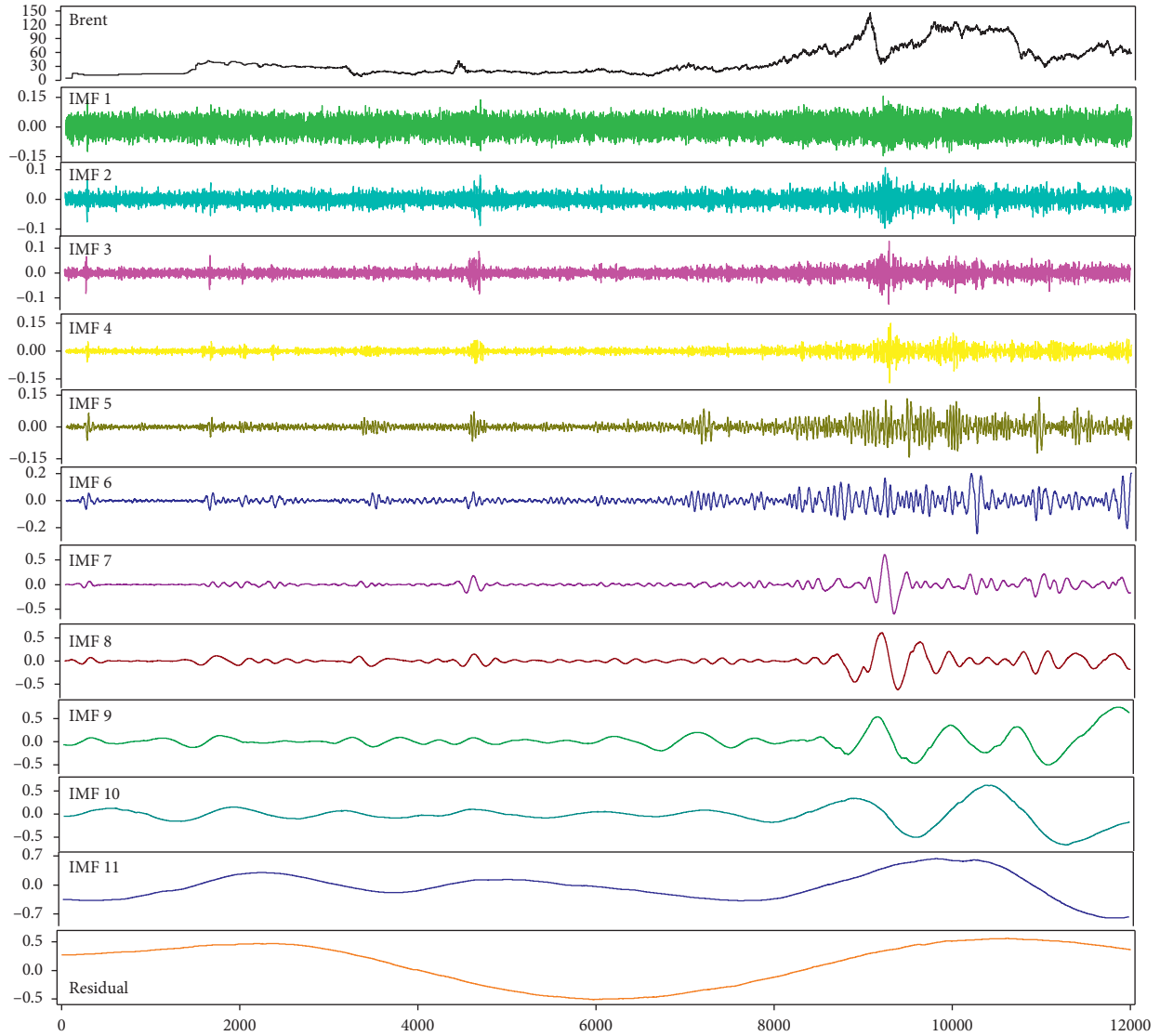


FIGURE 6: The decomposition results of the Brent series using MEEMD.

TABLE 2: Forecasting accuracy comparison of single models.

Models	Brent			WTI		
	GMDH	ANN	ARIMA	GMDH	ANN	ARIMA
RMSE	0.1772	1.1372	2.2543	0.5173	1.2397	2.0031
MAE	0.3126	1.0292	1.3469	0.5763	0.9235	2.8719
MAPE	0.2994	1.2885	1.5164	0.7584	1.5234	1.5035

TABLE 3: Forecasting accuracy of hybrid models of WTI series.

Models	EEMD			MEEMD		
	GMDH	ANN	ARIMA	GMDH	ANN	ARIMA
RMSE	0.3872	0.6580	0.6680	0.0426	0.1043	0.4297
MAE	0.5061	0.6244	0.6061	0.0351	0.1025	0.3915
MAPE	0.2864	0.4943	1.2306	0.0087	0.3770	0.5850

4.5. Predictive Performance of Hybrid Models. The most important comparison of hybrid models based on decomposition methods for both crude oil series is shown in Tables 3 and 4.

Regarding the hybrid models (i.e., EEMD-GMDH, EEMD-ANN, EEMD-ARIMA, MEEMD-GMDH, MEEMD-ANN, and MEEMD-ARIMA), Tables 3 and 4 show the experimental results of RMSE, MAE, and MAPE on both markets, Brent and WTI, respectively.

The forecasting evaluation criterion RMSE results are shown in Tables 3 and 4. The conclusion is drawn from RMSE that the proposed MEEMD-GMDH model significantly outperformed the other models for both Brent and WTI crude oil markets, while on other hand, MEEMD-based models (MEEMD-ANN and MEEMD-ARIMA) also attain the lowest value (better performance) than the corresponding EEMD-based (EEMD-ANN and EEMD-ARIMA) models.

TABLE 4: Forecasting accuracy of hybrid models of Brent series.

Models	EEMD			MEEMD		
	GMDH	ANN	ARIMA	GMDH	ANN	ARIMA
RMSE	0.6278	0.7142	0.7926	0.0240	0.0643	0.6780
MAE	0.6135	0.6742	0.8045	0.0360	0.1325	0.6820
MAPE	0.4543	0.4783	0.9762	0.0051	0.5370	0.9227

The forecasting evaluation performance (MAE) for both crude oil markets is presented in Tables 3 and 4. The decision is made from the MAE value that the proposed model got the lowest value and outperformed the other models for both markets. From Tables 3 and 4, we observe that the MEEMD-based model also performs better than the corresponding EEMD-based models.

The forecasting evaluation criterion MAPE on both markets for all selected models is presented in Tables 3 and 4. The decision made from MAPE value that the proposed model significantly outperformed the other models for both markets. The models MEEMD-ANN and MEEMD-ARIMA also attain the lowest values and perform well; then, the other benchmarked models and ranked second and third, respectively, in terms of MAPE. The MAPE values of the proposed model are 0.0087 and 0.0051 for both Brent and WTI markets, respectively, which lies in the classification of highly accurate forecasts.

From the forecasting evaluation measures shown in Tables 3 and 4, some interesting conclusions can be drawn:

- (1) Among all these models, the proposed method performed well based on RMSE, MAE, and MAPE presented in Tables 3 and 4.
- (2) The hybrid models based on MEEMD have better performance than those based on EEMD.
- (3) The proposed model performs better for long-term dependence than other classical and machine-learning methods for predicting crude oil prices.
- (4) Both MEEMD and EEMD hybrid models performed better than single time series and machine-learning models.
- (5) The suggested MEEMD-GMDH framework is way above all other comparable models in terms of MAPE, RMSE, and DM test, by utilizing the benefits of MEEMD and GMDH. All of these mean that the MEEMD-GMDH can effectively forecast crude oil prices.

Next, to confirm the superiority of the proposed model, we apply the DM test. For WTI dataset, the DM test statistic and their p -values are shown in Table 5, while for the Brent series, the DM test statistic and their corresponding p -values are presented in Table 6.

The assumptions of the DM test are the two methods that have the same number of predictions. The DM test confirmed the above conclusion. The MEEMD-based model statistically outperformed ANN and ARIMA models, and their p -values are less than <0.01 for both markets which shows the superiority of the MEEMD-GMDH model, while on the other hand, EEMD hybrid models and their p -values

TABLE 5: DM test results for WTI series.

Tested models	EEMD		MEEMD	
	ANN	ARIMA	ANN	ARIMA
GMDH	-10.674 (0.000)*	-5.421 (0.000)*	-6.783 (0.000)*	-18.346 (0.000)*
ANN		-2.456 (0.000)*		8.342 (0.000)*

*Statistically significant at 1%.

TABLE 6: DM test results for Brent series.

Tested models	EEMD		MEEMD	
	ANN	ARIMA	ANN	ARIMA
GMDH	-34.993 (0.000)*	-25.266 (0.000)*	-23.261 (0.000)*	-5.456 (0.000)*
ANN		7.456 (0.000)*		4.321 (0.000)*

*Statistically significant at 1%.

are also less than 0.01. Finally, the proposed model performed better than other models in this study.

4.6. Monte Carlo Simulations. In this section, simulation is performed to check the robustness and generalizability of the proposed MEEMD-GMDH model [36]. As we know, the nature of the crude oil prices data is the combination of the stochastic and deterministic components. The MEEMD and EEMD procedure divided the original time series into IMFs in such a way that the first IMF is more stochastic as compared to the second IMF and the second IMF is more stochastic than the third IMF and so on, whereas the last IMF is completely deterministic. Synthetic time series datasets which are composed of additive white noise and sine function are described in two different scenarios as follows [37–39]:

- (1) The first synthetic time series consisting of a sine function represents the deterministic component, whereas the normal distribution represents the stochastic component. That is,

$$Y_t = \sin(2\pi t) + \varepsilon(0, 1). \quad (15)$$

- (2) The second synthetic time series consisting of the sine function represents the deterministic component, whereas the ARMA model represents the stochastic component with an error of 0.25.

$$Y_t = \sin(2\pi t) + \text{ARMA}(1, 1). \quad (16)$$

- (3) Different time series are generated using equations (15) and (16) with a different number of observations, that is, 500, 1000, 2000, 5000, and 10000, and decompose all the series using MEEMD and EEMD. The distribution of training and testing data series is 80 and 20 percent, respectively, of every series. The forecast accuracy measures RMSE, MAE, and MAPE for testing datasets are presented in Tables 7 and 8,

TABLE 7: Forecasting accuracy of all models for first synthetic data.

Number of observations	Models	EEMD			MEEMD		
		GMDH	ANN	ARIMA	GMDH	ANN	ARIMA
500	RMSE	0.7543	1.0234	2.4567	0.3199	0.7089	1.4563
	MAE	0.7432	1.0032	2.4321	0.2582	0.8694	1.0234
	MAPE	0.9183	1.2313	2.0235	0.7253	1.0342	0.8456
1000	RMSE	1.2345	2.0934	3.5672	0.6462	0.7843	1.5321
	MAE	1.0326	2.0742	2.9531	0.3719	0.7653	1.3267
	MAPE	1.9482	2.6148	3.0219	0.6037	0.8723	1.7364
2000	RMSE	1.0234	2.3456	2.5342	0.3621	0.4932	1.3942
	MAE	0.9934	1.4532	2.4329	0.2899	0.3444	1.2345
	MAPE	0.8894	2.8743	1.6893	0.3523	0.6953	0.9654
5000	RMSE	0.8743	1.0043	2.0001	0.3945	0.5673	1.8721
	MAE	0.8362	0.8732	1.8863	0.3173	0.5542	1.8290
	MAPE	0.9795	1.3452	2.8743	0.8358	0.9731	1.2236
10000	RMSE	0.8456	1.4567	2.5693	0.2867	1.9567	3.1043
	MAE	0.8123	1.4221	2.7123	0.2224	1.9256	2.9876
	MAPE	0.7632	1.1345	1.4576	0.5211	2.1345	2.1376

TABLE 8: Forecasting accuracy of all models for second synthetic data.

Number of observations	Models	EEMD			MEEMD		
		GMDH	ANN	ARIMA	GMDH	ANN	ARIMA
500	RMSE	0.7943	1.6321	2.6523	0.1592	0.6321	1.7643
	MAE	0.6793	1.6432	2.3987	0.1151	0.6124	1.7432
	MAPE	1.0032	2.1953	2.4404	0.7757	0.8943	2.0032
1000	RMSE	1.9123	2.8621	2.8732	0.5319	1.9456	2.3452
	MAE	1.8943	2.5001	2.7632	0.4983	1.4387	1.0643
	MAPE	2.0232	2.7348	2.1198	0.6183	2.4532	1.8476
2000	RMSE	0.4323	1.7634	3.2123	0.1493	0.3456	2.2345
	MAE	0.2341	1.4532	2.2134	0.1245	0.3324	2.4321
	MAPE	0.4567	1.5567	2.6785	0.2346	0.6783	3.4321
5000	RMSE	1.5432	2.9783	4.1283	0.1734	2.3457	2.2198
	MAE	1.0003	2.4431	3.9991	0.1368	1.2963	1.8645
	MAPE	1.7634	3.1863	4.9921	0.9703	2.7234	2.0036
10000	RMSE	0.7431	2.5673	2.9528	0.1419	1.3176	3.2134
	MAE	0.7219	1.9090	2.1947	0.1126	1.0954	2.5674
	MAPE	0.8732	1.3421	2.4788	0.6234	0.8743	3.4532

respectively, for scenarios 1 and 2 for all models, that is, EEMD-GMDH, EEMD-ANN, EEMD-ARIMA, MEEMD-ARIMA, MEEMD-ANN, and the proposed MEEMD-GMDH model.

From Tables 7 and 8, it is observed that the MEEMD improved the performance of the ARIMA, ANN, and GMDH models as compared to EEMD. Thus, for forecasting the crude oil prices, the MEEMD is recommended for data decomposition. The model MEEMD-GMDH outperforms all of the models for a different number of observations, that is, 500, 1000, 2000, 5000, and 10000. The MAPE values of the model MEEMD-GMDH are less than 1 for all sets of observations which demonstrated the highly accurate forecasts [37, 38]. The experimental findings of both scenarios demonstrated that all ensemble methodologies were effective but MEEMD was more effective. Moreover, the forecasting accuracy measures in terms of MAE, RMSE, and MAPE

highlighted that the model MEEMD-GMDH is the most efficient method for forecasting daily crude oil prices.

5. Conclusion

One of the most important quantitative models with significant interest in the literature is time series forecasts. Oil prices are a crucial factor influencing the economic agenda and policies of government and trading enterprises, because of the importance of the role of crude oil in the world economy. Proactive experience of their potential movements will also contribute to improved decision-making at all levels of government and management. The forecasts for oil prices are very complicated since the financial time series is extremely unpredictable, nonlinear, and erratic.

Despite the attempts to fix the issue with new mathematical approaches and because of its inherent complexity,

apparently volatile existence, and different variables influencing the fluctuation of the demand in crude oil, oil prices are still difficult to tackle. Incorporate strategies have been more indispensable than ever to apply computational approaches for predicting and encouraging investment decisions.

Achieving an accurate prediction of a time series is a very important but difficult task because of its attributes of nonlinearity and nonstationarity. In this paper, we proposed a hybrid model called MEEMD-GMDH for crude oil price forecasting. The MEEMD method uses a median operator instead of a mean operator during the ensemble noisy IMF trial during the standard EEMD process. The advantage of MEEMD is to reduce the mood splitting problem of IMFs. This is the first time that MEEMD-GMDH has been applied to predict crude oil prices. The experimental results show that our new proposed methodology goes beyond other decomposition hybrid models (EMD, EEMD, and CEEMD). This shows that MEEMD-GMDH is a superior and promising alternative to the autoregressive integrated moving average model, ANN, and other machine-learning approaches studied by other researchers. In addition to crude oil prices, for its robustness and routine testing, the MEEMD-GMDH methodology can be implemented with more complex tasks. Both theoretical and observational literature evidence indicate that the hybrid model is less generic or error dependent on the use of dissimilar models or models that vary strongly. Moreover, the hybrid procedure can reduce the model instability, usually present in the statistical inference and time series prevision, due to potential unreliable or evolving data trends. A literature analysis of the crude oil forecast reveals that there have been limited studies on AI and complex methods. The main objective of this approach is to help decision-makers to reduce the risks of crude oil and improve the accuracy of crude oil price forecasts. Moreover, the research results of this study are crucial to national economic growth and sustainable development.

The future work could be extended in two aspects: (1) to predict other time series, such as gold price series, electricity, and wind speed, one can apply MEEMD-GMDH; (2) to attain more accurate and special decomposition of time series data, one can apply more advanced average operators such as weighted mean, quartiles, or geometric mean.

Data Availability

The data used to support the findings of this study are included within the supplementary information files.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Supplementary Materials

The crude oil prices data used in this paper consist of WTI and Brent. (*Supplementary Materials*)

References

- [1] P. K. Verleger, "Adjusting to volatile energy prices," *Peterson Institute*, vol. 39, 1994.
- [2] H. Lin and Q. Sun, "Crude oil prices forecasting: an approach of using CEEMDAN-based multi-layer gated recurrent unit networks," *Energies*, vol. 13, no. 7, p. 1543, 2020.
- [3] Y. Wang, Y. Wei, and C. Wu, "Detrended fluctuation analysis on spot and futures markets of West Texas Intermediate crude oil," *Physica A: Statistical Mechanics and Its Applications*, vol. 390, no. 5, pp. 864–875, 2011.
- [4] L. Yu, W. Dai, and L. Tang, "A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting," *Engineering Applications of Artificial Intelligence*, vol. 47, pp. 110–121, 2016.
- [5] T. Li, M. Zhou, C. Guo et al., "Forecasting crude oil price using EEMD and RVM with adaptive PSO-based kernels," *Energies*, vol. 9, no. 12, p. 1014, 2016.
- [6] S. Abosedra and H. Baghestani, "On the predictive accuracy of crude oil futures prices," *Energy Policy*, vol. 32, no. 12, pp. 1389–1393, 2004.
- [7] S. G. Gülen, "Efficiency in the crude oil futures market," *Journal of Energy Finance & Development*, vol. 3, no. 1, pp. 13–21, 1998.
- [8] A. Lanza, M. Manera, and M. Giovannini, "Modeling and forecasting cointegrated relationships among heavy oil and product prices," *Energy Economics*, vol. 27, no. 6, pp. 831–848, 2005.
- [9] C. Morana, "A semiparametric approach to short-term oil price forecasting," *Energy Economics*, vol. 23, no. 3, pp. 325–338, 2001.
- [10] H. Mohammadi and L. Su, "International evidence on crude oil price dynamics: applications of ARIMA-GARCH models," *Energy Economics*, vol. 32, no. 5, pp. 1001–1008, 2010.
- [11] S. Mirmirani and H. C. Li, "A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil," *Advances in Econometrics*, vol. 19, pp. 203–223, 2004.
- [12] M. I. Ahmad, "Modelling and forecasting Oman crude oil prices using Box-Jenkins techniques," *International Journal of Trade and Global Markets*, vol. 5, no. 1, pp. 24–30, 2012.
- [13] M. Aamir, A. Shabri, and M. Ishaq, "Crude oil price forecasting by CEEMDAN based hybrid model of ARIMA and kalman filter," *Jurnal Teknologi*, vol. 80, 2018.
- [14] I. Haidar, S. Kulkarni, and H. Pan, "Forecasting model for crude oil prices based on artificial neural networks," in *Proceedings of the 2008 International Conference on Intelligent Sensors*, pp. 103–108, Sydney, Australia, December 2008.
- [15] A. Khashman and N. I. Nwulu, "Intelligent prediction of crude oil price using Support Vector Machines," in *Proceedings of the 2011 IEEE 9th International Symposium on Applied Machine Intelligence and Informatics (SAMII)*, pp. 165–169, Smolenice, Slovakia, January 2011.
- [16] W. Xie, L. Yu, S. Xu, and S. Wang, "A new method for crude oil price forecasting based on support vector machines," in *Proceedings of the International Conference on Computational Science*, pp. 444–451, Faro, Portugal, June 2006.
- [17] W. Kristjanpoller and M. C. Minutolo, "Forecasting volatility of oil price using an artificial neural network-GARCH model," *Expert Systems with Applications*, vol. 65, pp. 233–241, 2016.
- [18] T. Li, Z. Hu, Y. Jia, J. Wu, and Y. Zhou, "Forecasting crude oil prices using ensemble empirical mode decomposition and sparse Bayesian learning," *Energies*, vol. 11, no. 7, p. 1882, 2018.

- [19] S. L. Ho, M. Xie, and T. N. Goh, "A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction," *Computers & Industrial Engineering*, vol. 42, no. 2-4, pp. 371-375, 2002.
- [20] E. Chen and X. J. He, "Crude oil price prediction with decision tree based regression approach," *Journal of International Technology and Information Management*, vol. 27, pp. 2-16, 2019.
- [21] R. M. N. Y. Sarpong-Streeter, R. A. L. Sokkalingam, M. B. Othman, D. L. C. Ching, and H. b. Sakidin, "A hybrid autoregressive integrated moving average-phGMDH model to forecast crude oil price," *International Journal of Energy Economics and Policy*, vol. 9, no. 5, pp. 135-141, 2019.
- [22] N. Bokde, A. Feijóo, D. Villanueva, and K. Kulat, "A review on hybrid empirical mode decomposition models for wind speed and wind power prediction," *Energies*, vol. 12, no. 2, p. 254, 2019.
- [23] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 01, no. 01, pp. 1-41, 2009.
- [24] L. Xie, X. Lang, A. Horsch, and Y. Yang, "Online oscillation detection in the presence of signal intermittency," *Control Engineering Practice*, vol. 55, pp. 91-100, 2016.
- [25] X. Lang, N. ur Rehman, Y. Zhang, L. Xie, and H. Su, "Median ensemble empirical mode decomposition," *Signal Processing*, vol. 176, Article ID 107686, 2020.
- [26] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, John Wiley & Sons, Hoboken, NJ, USA, 2015.
- [27] M. Aamir and A. Shabri, "Modelling and forecasting monthly crude oil price of Pakistan: a comparative study of ARIMA, GARCH and ARIMA Kalman model," *AIP Conference Proceedings*, vol. 2341, Article ID 060015, 2016.
- [28] R. Sharda, "Neural networks for the MS/OR analyst: an application bibliography," *Interfaces*, vol. 24, no. 2, pp. 116-130, 1994.
- [29] G. Zhang, B. Eddy Patuwo, and M. Hu, "Forecasting with artificial neural networks," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35-62, 1998.
- [30] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.
- [31] D. Srinivasan, "Energy demand prediction using GMDH networks," *Neurocomputing*, vol. 72, no. 1-3, pp. 625-629, 2008.
- [32] A. Shabri and R. Samsudin, "A hybrid GMDH and box-jenkins models in time series forecasting," *Applied Mathematical Sciences*, vol. 8, pp. 3051-3062, 2014.
- [33] T. Kondo, "The learning algorithms of the GMDH neural network and their application to the medical image recognition," in *Proceedings of the 37th SICE Annual Conference*, pp. 1109-1114, Tokushima, Japan, July 1998.
- [34] M. Aamir and A. Shabri, "Improving crude oil price forecasting accuracy via decomposition and ensemble model by reconstructing the stochastic and deterministic influences," *Advanced Science Letters*, vol. 24, no. 6, pp. 4337-4342, 2018.
- [35] R. Samsudin and P. Saad, "A Time Series Forecasting Model Using Group Method of Data Handling (GMDH)." *Mathematical Problems in Engineering*, vol. 2020, Article ID 1325071, 23 pages, 2020.
- [36] N. D. Bokde, Z. M. Yaseen, and G. B. Andersen, "ForecastTB-an R package as a test-bench for time series forecasting-application of wind speed and solar radiation modeling," *Energies*, vol. 13, no. 10, p. 2578, 2020.
- [37] W. Gao, M. Aamir, A. B. Shabri, R. Dewan, and A. Aslam, "Forecasting crude oil price using kalman filter based on the reconstruction of modes of decomposition ensemble model," *IEEE Access*, vol. 7, pp. 149908-149925, 2019.
- [38] P. Xu, M. Aamir, A. Shabri, M. Ishaq, A. Aslam, and L. Li, "A new approach for reconstruction of IMFs of decomposition and ensemble model for forecasting crude oil prices," *Mathematical Problems in Engineering*, vol. 2020, Article ID 1325071, 23 pages, 2020.
- [39] M. Aamir, A. Shabri, and M. Ishaq, "Improving forecasting accuracy of crude oil prices using decomposition ensemble model with reconstruction of IMFs based on ARIMA model," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 14, no. 4, pp. 471-483, 2018.