Article

Prophesying the Short-Term Dynamics of the Crude Oil Future Price by Adopting the Survival of the Fittest Principle of Improved Grey Optimization and Extreme Learning Machine

Asit Kumar Das [1], Debahuti Mishra [1], Kaberi Das [1], Pradeep Kumar Mallick [2], Sachin Kumar [3], Mikhail Zymbler [3] and Hesham El-Sayed [4,*]

[1] Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be) University, Bhuwaneshawar 751030, India; asitdasbbsr@gmail.com (A.K.D.); debahutimishra@soa.ac.in (D.M.); kaberidas@soa.ac.in (K.D.)

[2] School of Computer Engineering, KIIT Deemed to be University, Bhuwaneshawar 751024, India; pradeep.mallickfcs@kiit.ac.in

[3] Department of Computer Science, South Ural State University, 454080 Chelyabinsk, Russia; kumars@susu.ru (S.K.); mzym@susu.ru (M.Z.)

[4] College of Information Technology, United Arab Emirates University, Abu Dhabi 15551, United Arab Emirates

[*] Correspondence: helsayed@uaeu.ac.ae

Abstract: Crude oil market analysis has become one of the emerging financial markets and the volatility effect of the market is paramount and has been considered as an issue of utmost importance. This study examines the dynamics of this volatile market of crude oil by employing a hybrid approach based on an extreme learning machine (ELM) as a regressor and the improved grey wolf optimizer (IGWO) for prophesying the crude oil rate for West Texas Intermediate (WTI) and Brent crude oil datasets. The datasets are augmented using technical indicators (TIs) and statistical measures (SMs) to obtain better insight into the forecasting ability of this proposed model. The differential evolution (DE) strategy has been used for evolution and the survival of the fittest (SOF) principle has been used for elimination while implementing the GWO to achieve better convergence rate and accuracy. Whereas, the algorithmic simplicity, use of less parameters, and easy implementation of DE efficiently decide the evolutionary patterns of wolves in GWO and the SOF principle updates the wolf pack based on the fitness value of each wolf, thereby ensuring the algorithm does not fall into local optimum. Furthermore, the comparison and analysis of the proposed model with other models, such as ELM–DE, ELM–Particle Swarm Optimization (ELM–PSO), and ELM–GWO shows that the predictability evidence obtained substantially achieves better performance for ELM–IGWO with respect to faster error convergence rate and mean square error (MSE) during training and testing phases. The sensitivity study of the proposed ELM–IGWO provides better results in terms of the performance measures, such as Theil's U, mean absolute error (MAE), average relative variance (ARV), mean average percentage error (MAPE), and minimal computational time.

Keywords: crude oil forecasting; survival of the fittest (SOF); extreme learning machine (ELM); differential evolution (DE); particle swarm optimization (PSO); grey wolf optimizer (GWO); improved grey wolf optimizer (IGWO)

1. Introduction

In the current financial market, the crude oil trading and prophesying of prices provide excellent opportunities to the world's economy as the volatility has risen sharply in recent years and the traders have seen strong trends in producing consistent and good short-term swings in trades and long-term investment strategies. Crude oil is one of the most important energy sources used globally and due to its importance, a vast financial market based on physical trading, as well as derivatives trading, exists. This commodity is especially

important to the businesses that heavily depend on fuel, and has been one of the major imports and exports of numerous countries. The importance of crude oil posed a vast financial trading market and there is a need for some automated and computationally effective forecasting models to predict the future and options for investments in this commodity. Stakeholders, such as policymakers, trading companies, businesses, and investors [1–3], may use precise forecasting to reduce losses and increase revenues in their transactions. A commodity trading company and a company that utilizes crude oil as a raw material wishing to profit from commodity markets, require a robust forecasting model to predict the prices in advance. The outbreak of the COVID-19 pandemic and the recent crisis among the European nations and Russia casts doubts on the market for crude oil [4].

Crude oil prices suffer from high volatility and fluctuations, which is the key issue of the related financial market. To address such issues, the investors and traders are coming up with several new automated models to produce more accurate results with respect to the forecasting of crude oil prices. As a result, different policy makers, businesses, and institutions are suspicious about this type of trading, which makes it very hard to accurately predict the price in such volatile markets. This causes the market to fluctuate more than previously in the near term. Statistical models were among the first to be exploited to establish predictive models in crude oil price forecasting. The widely used traditional autoregressive moving average (ARMA) [5] and autoregressive integrated moving average (ARIMA) [6–9] models were adopted for forecasting the crude oil price, whereas some other classical forecasting models are also proposed to address the same problem using generalized autoregressive conditional heteroscedastic (GARCH) [9–13]. However, when the forecasting price under investigation is essentially linear or near-linear, the aforementioned models offer reasonable prediction results and only capture a limited amount of irregularities observed in crude oil price data. In [6–13], the authors utilized the ARIMA to forecast crude oil where they predicted the short-term price. GARCH, exponential generalized autoregressive conditional heteroscedastic (EGARCH), and Glosten–Jagannathan–Runkle GARCH (GJR–GARCH) forecasting models are also frequently used to tackle the non-linearity problem, and it was observed that these models suffer from few limitations, such as the transformation of non-linear data into a smooth linear form, the conversion of non-stationary data to a stationary form and non-credibility of the obtained results from a traditional model, as it forecasts the values of transformed data.

The wide use of artificial intelligence (AI) and machine learning (ML) strategies were also explored for the development of many such forecasting models to resolve the aforementioned irregularities for which the prices of crude oil vary in forecasting markets. The ML approaches have shown the ability to cope with noisy and irregular patterns of data problems on their own to address the drawbacks of these aforementioned models and also can solve the problem of handling the chaotic and nonlinear pattern of data over traditional approaches [14–20]. It was observed that the support vector machine (SVM) [17–20], back-propagation neural networks (BPNN) [14,15,21], and extreme learning machine (ELM) [22–24] along with a few other ML approaches such as random forest, fuzzy logic, and a few other models are commonly employed while developing the forecasting models [14–26] to achieve better predictive performance. Recent literature also demonstrates the use of hybridized models for the purpose wherein, the capability of two or more models are combined in order to develop hybridized predictive models. Some of these models utilize the capabilities of nature-inspired optimization techniques to effectively reduce the difference between real and predicted pricing. The widely used meta-heuristics techniques to develop the hybridized models are ant colony optimization (ACO) [27–29], particle swarm optimization (PSO) [30–35], artificial bee colony (ABC) [36], cuckoo search (CS) [37], differential evolution (DE) [38–41], grey wolf optimization (GWO) [42–46], etc.

It was seen that almost all the prediction forecasting models are complex ones and none are able to claim to be one hundred percent effective or accurate. The identification and specification of predictive models pose a big challenge to the research community because they need to be more precise and clear about the purpose of the model and need

a process to judge the capability of the designed predictive or forecasting model. The above-mentioned peculiarities of crude oil market analysis have motivated us to make an attempt at developing a computational effective experimental model harnessing the capabilities of ML and nature-inspired optimization techniques to forecast the future price of crude oil, which may be of some help to aid investors, increase profit, and enhance competitive advantage.

GWO is one of the meta-heuristic techniques proposed by Mirjalili [42] based on the inspiration of grey wolves, which mimics the hunting mechanism and leadership hierarchy of grey wolves. The ease of implementation, less storage, and computational requirements made the GWO more attractive for researchers. Additionally, GWO has faster convergence due to the continuous reduction of search spaces and thereby avoids local minima by controlling only two parameters, making GWO more stable and robust. Considering the above advantages of GWO [42–46], Jie-Sheng Wang et al. [47] proposed an improved version of GWO, (IGWO), by utilizing the evolution and elimination mechanisms for achieving better convergence and accuracy by adapting it to the survival of the fittest (SOF) principle of the biological updating of nature. In other words, a proper compromise had been achieved between the exploration and exploitation in IGWO over GWO. Authors have explored the advantages of DE such as algorithmic simplicity, use of less parameters, and ease of implementation to select the evolutionary pattern of wolves and they tried to update the wolf packs based on SOF to get rid of issues associated with getting trapped in local minima. The basic operations of DE, such as mutation, crossover, and selection, made IGWO have a strong exploration ability and, in the later stages of convergence, the difference between individuals and a population with a small value made the algorithm strong enough with respect to exploitation ability. Authors have explored this IGWO with twelve benchmark functions and also compared the performance of this IGWO with DE, PSO, ABC, CS, GWO, etc., and witnessed interesting results with respect to convergence and accuracy [47].

Due to the increase in complexity level and prevalence of non-linear patterns, short-term predictions are a more challenging task in the case of crude oil price forecasting. This is basically due to the different types and grades of crude oil and benchmarks that are being utilized as a pricing reference. West Texas Intermediate (WTI), Brent Blend, and Dubai Crude are the most often used benchmark datasets found in the literature [48–50]. In this work, ELM was explored as a predictive network considering its fast and efficient learning speed, fast convergence, good generalization ability, easy implementation mechanism, because it does not require adjusting the input weights and hidden layer biases during the implementation of the algorithm and it produces only one optimal solution. However, as ELM is memory-heavy and suffers from high space and time complexity, the traditional ELMs need to be optimized. The proposed hybrid forecasting model in this study was motivated by the improved version of GWO (IGWO) [47] to develop an empirical computational forecasting model. The IGWO [47] method is utilized to achieve the right balance between exploration and exploitation, as well as to speed up the convergence rate and enhance the accuracy of the ELM network. The main contributions of this paper can be stated as follows;

(a). A hybrid ELM [22–24] based on a short-term (1-day, 3-day, 5-day, 7-day, 15-day, and 30-day) forecasting model ELM–IGWO is proposed, which effectively combined ELM's potential with the grey wolf multi-population search strategy.

(b). The hybridizing optimization methods offer their own set of benefits and also need meticulous readjustment of specific parameters of the algorithm. In the ELM network, the global best is regarded as the least objective value (error), whereas, the global worst is considered the greatest objective value (error). To start with, the biases and weights are chosen randomly, and the weight is modified using optimization techniques in the following iteration, and so on until the process is complete.

(c).   The IGWO [47] method is utilized to achieve the right balance between exploration and exploitation, as well as to speed up the convergence rate and thereby enhance the accuracy of the ELM network.

(d).   In the proposed model of IGWO, where DE speeds up the convergence rate, the SOF principle in IGWO [47] is capable of handling the non-linearity nature of the crude oil price.

(e).   Two datasets, the WTI crude oil and the Brent future oil datasets [48,51], were experimented on and were used to expand the model's feasibility by augmenting the original crude oil datasets with a few technical indicators (Tis) and statistical measures (SMs) to increase the dimensionality of the original datasets [52,53]. These were finally given as inputs into the proposed ELM–IGWO crude oil forecasting model.

(f).   The proposed forecasting model is compared based on mean squared error (MSE) with other hybrid ELM-based forecasting models such as ELM–DE, ELM–PSO, and ELM–GWO to validate the superiority of the results obtained.

(g).   The forecasting ability of the proposed model is established based on actual price vs. predicted price for both datasets for the three combinations of the augmented form of datasets such as original dataset +Tis, original dataset +SMs, and original dataset + Tis + SMs.

(h).   Finally, the model's validation was performed based on MSE, Theil's U, mean absolute error (MAE), average relative variance (ARV) and mean absolute percentage error (MAPE), and comparison based on CPU time utilization (in seconds) [53].

The rest of the paper is organized as follows: brief discussions on the different crude oil prediction models are presented as a literature survey in Section 2. The methodologies adopted in this approach are discussed in Section 3, which includes the ELM, DE, PSO, GWO, and IGWO. The experimental scenario is discussed in Sections 4 and 5 discusses the results of this study. Finally, Section 6 concludes this work with the future scope.

2. The Literature Review

In this article, different econometric models are studied to forecast the price of crude oil. GARCH models are used to characterize crude oil price variations. Wang and Wu used GARCH family models to estimate the volatility of four distinct energy commodities in 2012 [9–13]. Many statistical forecasting models were proposed that show a better accuracy using ARIMA [5–9]. In another approach, the authors have used the Markov-switching AR–ARCH model to predict three distinct crude oil prices [54]. It was observed that better results can be obtained by using a self-exciting threshold auto regressive algorithm as suggested by the authors in [55] when the data are stationary and linear in nature. However, they cannot account for the nonlinearity and complexities of crude oil pricing. Due to those challenges, some researchers have used AI and ML in their research to achieve the same objective [14–20]. Furthermore, various artificial neural network-based models were also proposed as these technologies can handle intricacy and nonlinearity data and hence can provide a better accuracy over traditional approaches of forecasting crude oil price [56].

The authors in [57] have investigated and observed that the parameter sensitivity and overfitting of the data utilized can influence the use of a single technique for the prediction of oil prices. In order to improve the accuracy of forecasting models, researchers have created hybrid techniques based on optimization strategies to predict crude oil prices. To forecast the WTI crude oil spot prices, a hybrid model was proposed where the deconstruction of data takes place using the Harr-A-Trous wavelet transform before applying it to the BPNN algorithm and this hybrid technique appears to beat the benchmark models, according to researchers [14,15,21]. Some hybrid works were also carried out by merging the genetic algorithm (GA) with the SVM which was proven to give better results than the traditional approaches [58]. In another approach, the author utilized value at risk (VAR) for determining the determinants of the price, and the determinant value is fed to the GA for optimizing the parameters of the SVM to accurately predict crude oil price named as VARSVM [59]. The authors in [48] used a mix of variational mode decomposition (VMD)

and an ARIMA model to forecast crude oil prices. VMD is used to extract risk variables, which are then modeled with ARMA–GARCH [60].

The optimal parameters are obtained using optimization methods of the forecasting techniques such as ANN and SVM with the much-generalized optimizations techniques such as; GA, PSO, ABC, etc. to discover the model's best parameters. The PSO optimizer is one of the most prominent computational algorithms used in forecasting research. Finding the best settings via conventional PSO proved perplexing, and it also resulted in local minima [30–35,61]. However, an adaptive PSO technique was developed to address the flaw identified by the authors in [62] which aids in the discovery of the best system and control parameters. Another approach was also developed using poly-hybrid PSO for intelligent parameter modification. A hybrid model was proposed where a computational technique was dubbed the flower pollination algorithm (FPA), which can optimize parameters of the forecasting models very well. In terms of resolving optimization difficulties, this FPA technique outperforms GA and PSO [63,64]. In another approach, the author combines FPA and BPNN to create a hybrid model that forecasts OPEC nations' petroleum consumption. By providing fewer forecast errors, this model beats other hybrid models [54]. The trade patterns were changed due to the tight oil revolution along with worldwide demand and supply. To analyze movements of the Brent–WTI spread over 15 years (2005–2020), Isabella Ruble et al. [49] developed three scenarios using ARIMAX–GARCH and Markov-switching models, which estimate the impact of crude oil trader's decisions on the Brent–WTI spread. In order to identify the decoupling and recoupling between WTI and Brent crude oil prices, Loretta Mastroeni et al. [50] proposed a dynamic time warping (DTW) algorithm. Authors have presented DTW-based indexes considering relative alignment index and wrapping index, which show the greatest decoupling between WTI and Brent occurs because of WTI local market conditions. Overall, the existing literature shows that academicians and researchers have attempted to estimate crude oil prices using traditional models for a long time. Because of the shifting of trading conditions, oil market participants must be aware of new approaches for predicting oil prices. The investigation in this study looks at how automated learning approaches along with nature-inspired algorithms may be used to predict future crude oil prices.

Additionally, we extracted the statistics of research carried out in crude oil forecasting performed by academicians and researchers worldwide during 1975–2023 from the Scopus database. These data were extracted on 1 March 2022 using the keyword "crude oil forecasting", which is shown in Figure 1a–c and represents the research carried out in this particular application from the year 1975–2023, which also indicated that more research was carried out during the years 2018–2022 in a source of publications such as journal articles, conference proceedings, reviews, book chapters, books, etc., and in different subject areas that include crude oil forecasting. Additionally, here, we discussed a few recent ML-based hybrid crude oil forecasting models during 2019–2022 and these are summarized in Table 1. To be in line with studied and presented literature in this section, this current work aims at the development of an accurate and computational effective crude oil forecasting model which captures the irregularities of this volatile and risky financial market [1–4,65–67] by exploring ELM with an IGWO optimization algorithm for WTI and Brent crude oil.
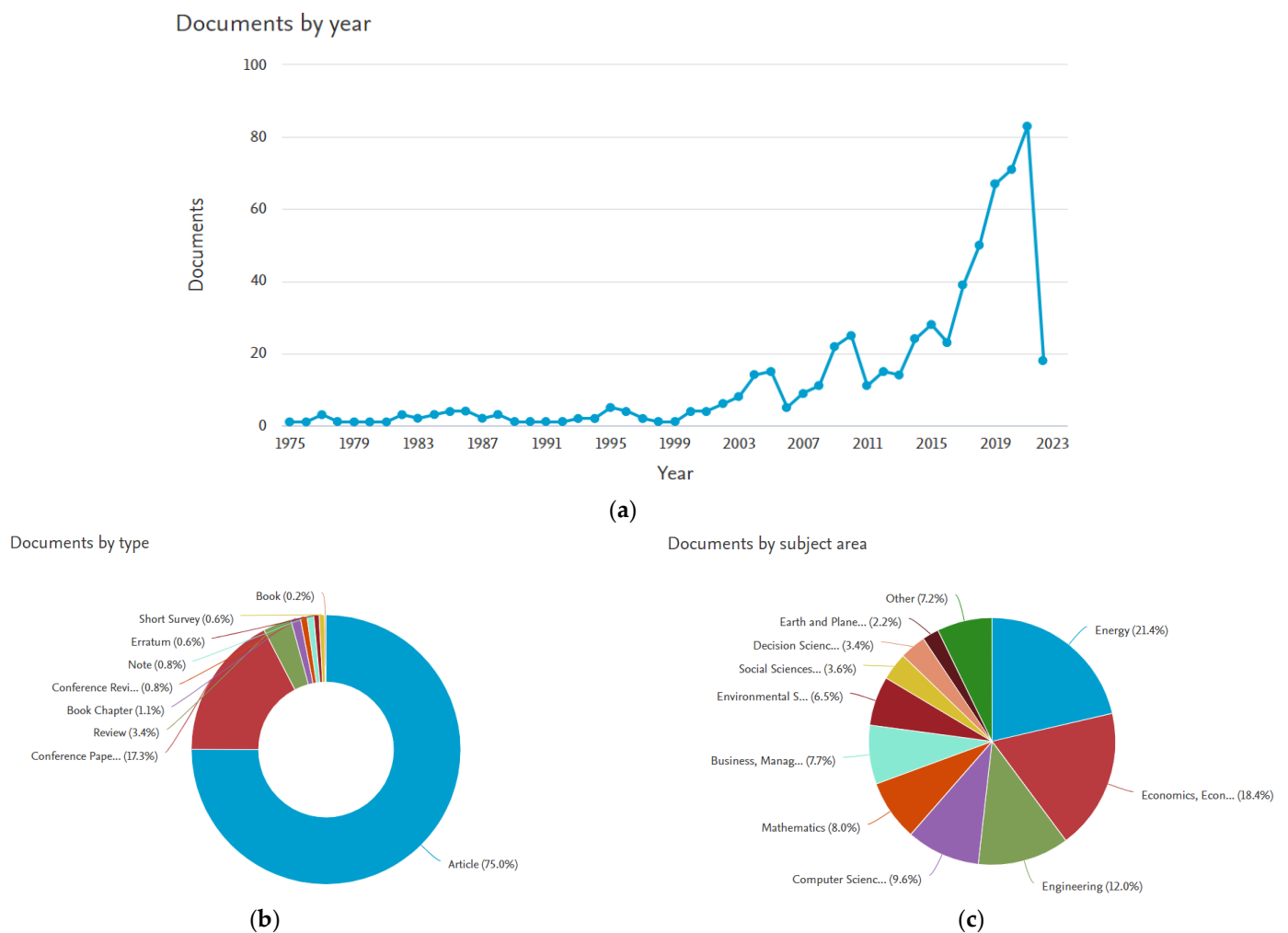
Documents by year

(a)

Documents by type

(b)

Documents by subject area

(c)

Figure 1. (**a**). Yearly representation of crude oil prediction; (**b**). Type of documents published for crude oil prediction; (**c**). Type of documents with respect to subject area published for crude oil prediction.

Table 1. Summary of the recent studies on crude oil price forecasting during 2019–2022.

Source	Year	Objective(s)	Model(s) Adopted	Finding(s)	Pros	Cons
Kaijian Hea et al. [68]	2022	To calculate the value at risk factor of crude oil	Md VaR based risk forecasting model using multiple mode decomposition model and Quantile regression neural network	The proposed model is used to calculate VaR by using a semi-parametric data-driven approach for both normal market and transient market conditions.	The VaR estimated by the proposed MD VaR model gives more reliability and accuracy	The complex risk structure has not been analyzed.
Zhonglu Chen et al. [69]	2022	To predict the realized volatility of China's crude oil futures	Mixed data sampling (MIDAS) modeling framework: MIDAS–RV Model	In this work, both the jump and leverage effects are used to predict the RV of Chinese crude oil futures, where jump is used for the short-term and leverage effects used for long-term prediction.	The model is robust and economic. During the COVID-19 epidemic, the model is used to guide the investors as well as market participants. It also reduces the investment risks	-

Table 1. *Cont.*

Source	Year	Objective(s)	Model(s) Adopted	Finding(s)	Pros	Cons
Yuan Zhao et al. [70]	2021	To improve the accuracy of online real-time price forecasting for crude oil.	PSO–VMD algorithm and the and PVMD–SVM–ARMA model	In this paper, an improved variational mode decomposition is used with PSO to optimize the parameters and then the proposed hybrid model established the characteristics of sub-sequences, and the interval prediction model is constructed with the combination of point prediction model and Bootstrap sampling and finally, these can be used as a predictor to predict the online real-time price of crude oil.	The model poses flexibility and accuracy, which can give more information for people working with oil.	The method used for prediction is only mint for small samples and short-term periods.
Binrong Wu et al. [71]	2021	Effective crude oil price predicting	Convolutional neural network (CNN), VMD	In order to predict the crude oil price, here, a combined approach of Google trends and news text information was proposed which applied the following methodologies: relationship investigation, deep learning techniques, and decomposition techniques.	This approach provides satisfactory accuracy for crude oil price prediction. Google trends and news text information can promote each other.	The Google trends used in this approach are a complex process. Except for VMD, no other decomposition techniques were used to improve the accuracy.
Taiyong Li et al. [72]	2021	To reduce the complexity rises in Crude oil price forecasting	VMD and random sparse Bayesian learning (RSBL) VMD–RSBL	Here the authors have introduced random samples and random lags (features) into SBL and built an individual forecasting model with the combination of VMD and RSBL.	The proposed VMD–RSBL model is effective and efficient	The method is not applied for multivariate price forecasting
Hooman Abdollahi [73]	2020	To improve the accuracy of crude oil price prediction by considering the characteristics that exist in the oil price time series	Model by hybridizing the methods such as complete ensemble empirical mode decomposition, support vector machine, particle swarm optimization, and Markov-switching generalized autoregressive conditional heteroskedasticity	The algorithm is used as an effective tool for predicting nonlinear components and both the forecasted nonlinear components and volatile components result in the reliable forecast of the oil price.	The model predicted the volatile components with appropriate accuracy and gives the best performance while forecasting the component with low volatility.	The model was not applied to forecast other energy commodities in order to prove the robustness and generalizability.
Ling Lin et al. [74]	2020	To forecast crude oil price by considering the long memory, asymmetric, heavy-tail distribution, nonlinear, and non-stationary characteristics of crude oil price.	The model combined the complex long memory GARCH-M with wavelet analysis	The model can forecast the crude oil price during periods of extreme incidents. MS–DR method and crude oil volatility are being used here to model large fluctuations within the forecasting interval.	The model proposed can provide beneficial information required for the process of forecasting and Helps the investors to determine overall trends in oil prices reduce the market risks.	-

Table 1. *Cont.*

Source	Year	Objective(s)	Model(s) Adopted	Finding(s)	Pros	Cons
Shangkun Deng et al. [75]	2019	To predict the changes of crude oil price, as well as execute simulated trading.	Hybrid model based on multiple timeframes dynamic time wrapping and GA is developed for direction forecasting and simulation trading	The proposed method includes four components such as; (1) Data Pre-processing, (2) Multiple timeframes DTW prediction, (3) GA parameters optimization, and (4) Prediction, trading, and evaluation	The model gives high performances in terms of hit ratio, accumulated return, and Sharpe ratio, and the results are significantly superior to that of benchmark methods. It can also provide beneficial information to investors, energy-related enterprises, and government officers engaged in policy decisions.	The model should work with short term data

3. Methodologies Adopted

The various methodologies adopted for experimentation of the proposed forecasting model such as ELM and a few nature-inspired optimization techniques used for experimentation such as DE, PSO, standard GWO, and IGWO are discussed in detail.

3.1. Extreme Learning Machine (ELM)

In this forecasting model, randomly selected input parameters and a hidden layer parameters-based ML algorithm are firstly presented by [22–24] named ELM. They proposed this algorithm based on the single hidden layer feed-forward network. The advantage of using this algorithm is that it does not require any fine-tuning of parameters at each iteration, which makes for faster convergence of error, and also the ELM has a faster learning rate and higher speculation execution than gradient-based learning techniques, such as back-propagation. In this algorithm, a generalized inverse operation is applied to the hidden layer to get the weights of the output. Customary gradient-based learning has various issues including neighborhood minima, over-fitting, and erroneous parameter setting, which can be avoided by utilizing the ELM. To overcome those, the parameters of ELM were optimized using IGWO, which had been motivated by the work of Jie-Sheng Wang et al. [47].

3.2. Differential Evolution (DE)

DE is one of the extensively used evolutionary algorithms in areas addressing optimization issues. Based on Darwin's theory of evolution, this algorithm finds scope in problems that are characterized by nonlinearity, discontinuity, multimodality, and non-differentiability. Being an evolutionary algorithm, DE is capable of generating new off-springs by perturbing the solutions with scaled difference vectors and new offsprings are generated by recombining solutions pertaining to different values using genetic operators such as mutation selection and crossover [39]. In every generation the current individual solution is replaced with a new offspring with a better solution and this optimization algorithm's implementation is considerably simpler and more straightforward than many of the other meta-heuristic search algorithms, which is perhaps the reason why many researchers have studied it extensively [40]. DE is considerably different from other evolutionary algorithms as it mutates the secondary parents with distinct members of the current population with ascended differences, which is a property called self-referential mutation [41]. The first step in this algorithm is to initialize a random population Np, which is characterized by d number of real-valued decision vectors. Here, each vector is a genome/chromosome and is a candidate solution to the d—dimensional optimization problem. After initialization, the mutation is performed. There are many mutation strategies available and a generic con-

vention of naming mutation strategies is $DE/x/y/z$, where DE represents DE, x represents the string with the vector that is perturbed, y indicates the number of vectors considered for perturbation and z represents the type of crossover operation that is to be used.

3.3. Particle Swarm Optimization (PSO)

The PSO is a meta-heuristic algorithm based on the concept of swarm intelligence. It was proposed by Kennedy and Eberhart in 1995 and solves complex engineering problems effectively. The principle of this algorithm is to find a place for a swarm of flying birds to land where the availability of food is maximized as well as the risk of the existence of predators is minimized. The PSO is a population-based distributed learning scheme. The key steps of this algorithm are presented in a concise manner [30–33]. The various application areas of PSO are forecasting, classification, clustering, and function approximation. The PSO is applied to the fields of sensory networks, security, smart grids, the financial sector, healthcare, and manufacturing. The PSO is a simple optimization algorithm but performance-wise is slower in learning and accuracy-wise is satisfactory. Many variants of PSO were reported to improve on these two shortcomings. Multi-objective PSO was also developed to solve multi-objectives, multi-variables, and multi-constraint optimization problems [34,35].

3.4. Standard GWO Algorithm

In [42], the author presented a populace knowledge framework known as the GWO, as given in Figure 2, motivated by grey wolf predation conduct for optimization of parameters. This algorithm is proposed by Mirjalili based on the inspiration of grey wolves, which mimics the hunting mechanism and leadership hierarchy of grey wolves. The leadership hierarchy is simulated by *alpha*, *beta*, *delta*, and *omega* wolves.
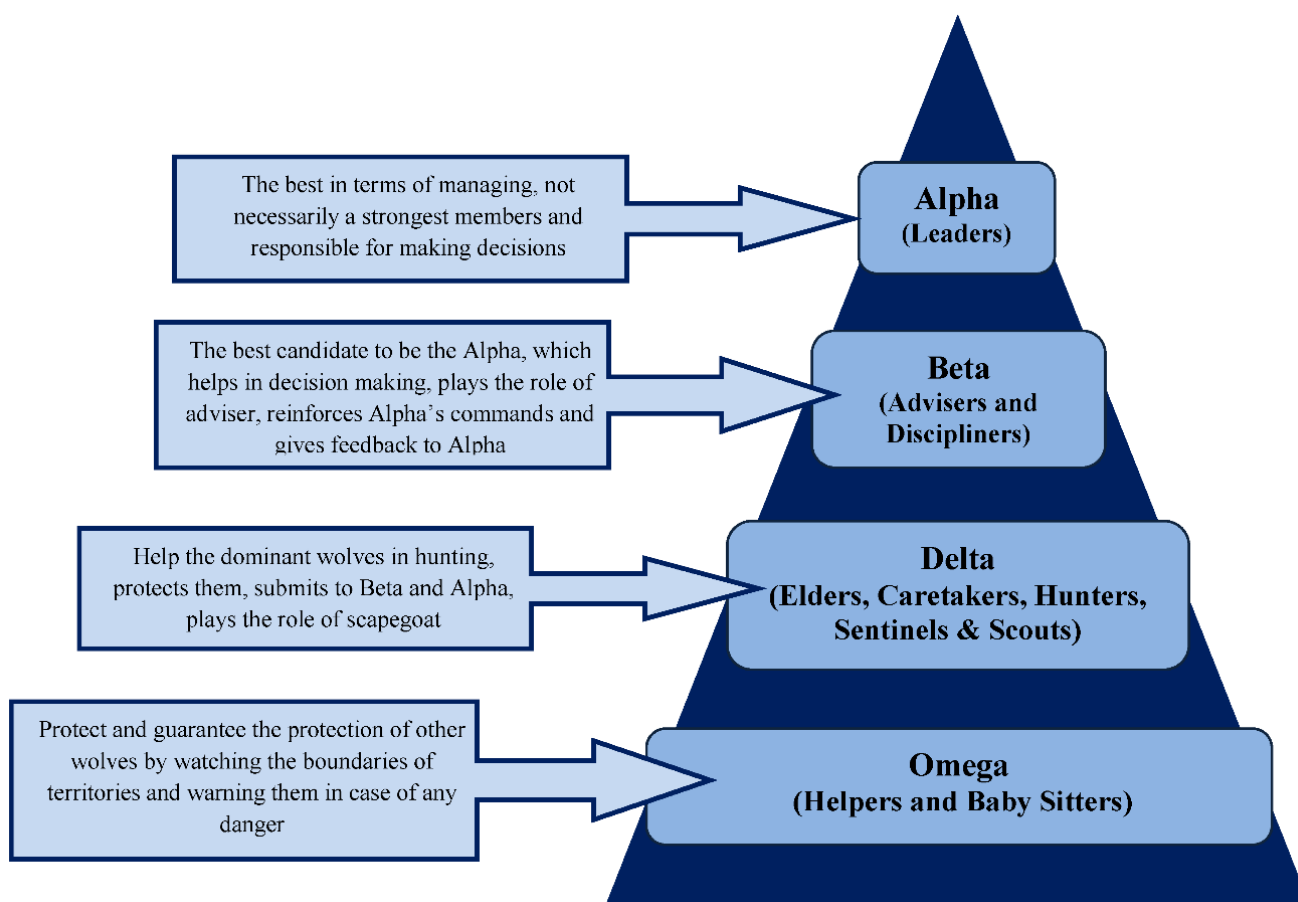

Figure 2. Pyramidal structure of working principle of GWO.

The fittest solution is *alpha*, the second- and third-best solutions are *beta* and *delta*. The rest of each member of the population is represented by *omega*. The GWO involves three main steps of hunting such as prey searching, prey encircling, and prey attacking [42–46]. During the searching and hunting phases, all *omega* wolves are guided by the *alpha*, *beta*, and *delta* wolves. When prey is found, the process begins. Then, the *alpha*, *beta*, and *delta* wolves lead and guide the *omega* wolves so that the prey is encircled. The GWO involves a smaller number of search parameters but provides competitive performance compared to other meta-heuristic methods.

Assuming that in *d*-dimensional space, the grey wolf pack $\{Q_i.\ i = 1, 2, 3, \ldots, n\}$ consists of *n* grey wolves. The GWO algorithm is described as follows:

(a). Enclosure stage: In this stage, they initially circle the prey after the wolf determines the situation of their prey. Numerically it can be presented using Equations (1) and (2). The distance between the prey and the wolf is represented by D_{gp}, whereas the position of the prey and the wolf after the *t*th iteration is represented by $Q_p(t)$ and $Q(t)$, respectively. The coefficient factors *M.N* are illustrated in Equations (2) and (3), respectively.

$$D_{gp} = \left| N.Q_p(t) - Q(t) \right| \tag{1}$$

$$Q(t + 1) = Q_p(t) - M.D_{gp} \tag{2}$$

$$M = 2e \times R_{n1} - e \tag{3}$$

$$N = 2 \times R_{n2} \tag{4}$$

In Equations (3) and (4), the values of random numbers are kept between [0, 1] and are represented as R_{n1} and R_{n2}, respectively. In this algorithm, *t* is the iteration number, *max* represents the maximum number of iterations. Here, with an increase in iterations, the *e* value decreases from 2 to 0 as given in Equation (5).

$$e = 2 - 2\left(\frac{t}{max}\right) \tag{5}$$

(b). Hunting stage: During this stage, the *alpha* wolf will rapidly discover the situation of the prey and quest for the prey. At the point when the *alpha* wolf discovers the situation of the prey, the wolves *alpha*, *beta*, and *delta* have a specific comprehension of the situation of the prey and expect *w* to move toward the situation of the prey. This is the prey cycle of wolves. After the attack stage is finished, the *alpha* wolf leads wolves *beta* and *delta* to chase down their prey. During the time spent hunting, the situation of individual wolves will move with the getaway of the prey. Where Q_{aplha}, Q_{beta} and Q_{delta} represent the current positions of wolves *alpha*, *beta*, and *delta* as mentioned in Equations (6)–(8), respectively, and $Q(t)$ indicates the current grey wolf position. N_1, N_2 and N_3 are *random vectors* and the location of the wolf was updated using Equation (9).

$$Q_1 = Q_{alpha} - A_1 \left| N_1 Q_{alpha}(t) - Q(t) \right| \tag{6}$$

$$Q_2 = Q_{beta} - A_2 \left| N_2 Q_{beta}(t) - Q(t) \right| \tag{7}$$

$$Q_3 = Q_{delta} - A_3 \left| N_3 Q_{delta}(t) - Q(t) \right| \tag{8}$$

$$Q(t + 1) = \frac{Q_1 + Q_2 + Q_3}{3} \tag{9}$$

In GWO, exploration means that the wolf leaves the original search path and searches for a new search space and exploitation refers to searching for a direction in an unknown region. In other words, wolves try to continue for a search by exploiting the original search space to a certain extent with a detailed search in a region that is explored. Therefore, the GWO is good at obtaining a compromise between exploration and exploitation [65].

3.5. Improved Grey Wolf Optimizer (IGWO)

The SOF principle is the key factor in the proposed IGWO algorithm, which had been proposed by Jie-Sheng Wang et al. [47]. The detailed idea of this approach is that the IGWO is proposed to expand the GWO's exploitation with the SOF feature and is explained in this section, this motivated us to hybridize this IGWO with ELM for developing a forecasting model to predict the future price of crude oil datasets. The SOF idea of the organic refreshing of nature and organic advancement is added to the traditional GWO in the IGWO. The differential advancement strategy was picked as the developmental example of wolves in this work since it enjoys the benefits of a straightforward idea, has less calculation boundaries, and is simple to execute. Therefore, the algorithm does not fall into the local optimal solution, and the wolf pack is refreshed by the SOF standard. That is, sorting the wellness esteems that relate to each wolf that rises after every cycle of the algorithm. The evolution operation with GWO was stated clearly in [47]. The SOF development law causes wolves to become stronger over time. In addition, the development activity is added to fundamental GWO to work on the algorithm's looking through speed. DE was picked up as an advanced method to adopt the difference among the individuals to recombine the population and obtain intermediate individuals to obtain the next generation's population through a competition between parent and offspring based on three basic operations: mutation, crossover, and selection and finally, the wolf's position is updated after this evolution operation.

(a) *Mutation Operation*

In differential development, the most conspicuous aspect is the mutation operation as the offsprings are dependent on this. When a person is chosen, two weight disparities are added to the individual to achieve variety. The difference vector of the parents is the core variation element of differential development, and each vector comprises two distinct individuals (Q_{r1}^t, Q_{r2}^t) of the parent (the tth generation). The following is the definition of the difference vector. Mutation activity is the most prominent part of differential development. At the point when an individual is picked, two weight disparities are added to the person to accomplish the assortment. The distinction vector of the parents is the center variety component of differential development, and every vector contains two unmistakable people (Q_{r1}^t, Q_{r2}^t) of the parent (the tth generation). Coming up next is the meaning of the distinction vector.

$$Dd_{12} = Q_{r1}^t - Q_{r2}^t \qquad (10)$$

where r_1 and r_2 are the index numbers of two distinct populations. As a result, the mutation operation may be summarized in Equation (11). To ensure that wolves can advance in the way that is best for their development, an ideal variety factor should be made. Therefore, this paper chooses magnificent wolves as guardians. *Beta* and *delta* are picked up as two parents after countless reproduction preliminaries and afterward converged with the *alpha* wolf to create a variety factor. Accordingly, Equation (12) is utilized to develop the variety factor. A unique scaling factor is utilized to give the calculation a solid investigation capacity in the beginning phases to fall into local optima and a high double-dealing capacity in the last stages to accelerate the combination. Thus, the scaling factor F changes from large to small contingent upon the number of cycles in Equation (13). The scaling factors are represented by f_{min} and f_{max}.

$$V_i^{t+1} = Q_{r1}^t - Q_{r2}^t \qquad (11)$$

$$V_i^{t+1} = Q_a^t + F \times (Q_{r1}^t - Q_{r2}^t) \qquad (12)$$

$$F = f_{min} + (f_{max} - f_{min}) \times \frac{Max\_iteration - (iteration - 1)}{Max\_iteration} \qquad (13)$$

(b) *Crossover Operation*

This operation makes the wolves' objective vector individual Q_i^t go through a hybrid activity with the variety vector V_i^{t+1}, bringing about a test individual U_i^{t+1}. To guarantee that every individual Q_i^t advances, an irregular picking approach is utilized to guarantee that somewhere around the slightest bit of U_i^{t+1} is provided by U_i^{t+1}. The hybrid likelihood factor, CR, is used to figure out what piece of U_i^{t+1} is provided by V_i^{t+1} and what touch is contributed by Q_i^t for the excess pieces of U_i^{t+1}. The mathematical formulation of the crossover operation is given in Equation (14).

$$U_{ij}^{t+1} = \begin{cases} V_i^{t+1}, & rand(j) \leq CR \ or \ j = randn(i) \\ Q_i^t, & rand(j) \geq CR \ or \ j \neq randn(i) \end{cases} \qquad j = 1, 2, 3, \ldots\ldots D \qquad (14)$$

The random uniform distribution is obeyed by $rand(j) \in [0, 1]$, j is the jth variable, CR is the crossing probability and $rand(i) \in [1, 2, 3, \ldots D]$. When the CR is more than one, V_i^{t+1} is able to contribute more to U_i^{t+1}, as shown in Equation (14). When the CR is equal to one, $U_i^{t+1} = V_i^{t+1}$. If the CR is lower, Q_i^t will be able to contribute more to U_i^{t+1}.

(c)   *Selection Operation*

The selection procedure employs the greedy choice approach and it creates the experimental individual U_i^{t+1} after the mutation and crossover operations, and then completes it with Q_i^t and is expressed in Equation (15). In this equation, the fitness function is represented by f and Q_i^{t+1} is the tth generation individual. The person with the best fitness is chosen as an individual of $(t+1)$ generation from U_i^{t+1} and Q_i^t, and the individual of the tth generation is replaced.

$$Q_{ij}^{t+1} = \begin{cases} U_i^{t+1}, & f(U_i^{t+1}) < f(Q_i^t) \\ Q_i^t, & f(U_i^{t+1}) < f(Q_i^t) \end{cases} \qquad i = 1, 2, 3, \ldots\ldots n \qquad (15)$$

The basic operations of DE, such as mutation, crossover, and selection, made this IGWO have a strong exploration ability and in the later stages of convergence, the difference between individuals and a population with a small value made the algorithm strong enough with respect to exploitation ability. The updating mechanism based on the SOF principle for IGWO suggested by authors in [47] can be stated as follows. "As it is known that, in reality some of the vulnerable wolves need to be eliminated or discarded to overcome the challenges associated with uneven distribution of prey, hunger, disease and few other specific reasons and new wolves must undergo the SOF principle and join the wolf community or wolf pack". This SOF principle updates the wolf pack and makes the algorithm stronger so as not to fall into the local optima problem. The key principles of this IGWO [47] can be stated as:

(a).   The strength of the wolves with respect to fitness value is measured by keeping the number of wolves fixed in a wolf pack and the concept says the wolf with the largest fitness value is chosen as the better one.

(b).   In this way, at each iteration, the fitness values of the wolves are sorted in ascending order thereby eliminating the wolves with a larger fitness value, and random new wolves are generated matching the number of wolves eliminated.

(c).   It can be observed that, when the numbers of wolves with a high fitness value are large, that leads to generating the same large number of new wolves and this type of case leads to slow convergence speed because of large search space. Similarly, if the fitness value is chosen to be too small, the diversity of the population is not guaranteed, which results in the inability of exploring new solution spaces. Therefore, the authors have proposed to have a random fitness value.

(d). In this work, the fitness value was chosen to be between the range $[n/\varepsilon, \; n/0.75 \times \varepsilon]$ using Equation (16), where the total number of wolves is represented as n, and the wolf updating scaling parameter is termed as ε.

$$R = [n/\varepsilon, \; n/0.75 \times \varepsilon] \qquad (16)$$

The following are the major steps in the procedure for the IGWO algorithm:

(1) Creating the population of grey wolves by defining the factors such as e, M, and N for a randomly generated wolf location $Qi(I = 1, 2, \ldots n)$.

(2) Determining each *wolf's fitness* and, based on the fitness value, defining the best three wolves as *alpha*, *beta*, and *delta*, respectively.

(3) Making necessary adjustments and updating the positions of the other wolves, i.e., the w wolf, according to Equations (5)–(11).

(4) The evolution of the proposed algorithm is carried out based on constructing a variation factor, using *alpha*, *beta*, and *delta* as described in Equation (12). After the crossover and selection operations, choosing a fit animal to be the next generation's wolf, selecting the top three of them, and defining them as *alpha*, *beta*, and *delta*, in that order.

(5) Sorting the wolves' fitness values by eliminating the wolves with the highest fitness value and creating random R wolves using Equation (16).

(6) Making necessary changes to factors e, M, and N using Equations (3)–(5).

(7) Determining the *alpha*'s position and fitness value if the termination condition is satisfied. If optimal position is not obtained, go back to step one (2).

4. Experimentation and Result Analysis

This section outlines the schematic representation of the proposed ELM–IGWO forecasting model (shown in Figure 3), the datasets description, and its' preliminary statistics. In this approach, the comparison of the proposed ELM–IGWO was carried out with ELM–DE, ELM–PSO, and ELM–GWO. The datasets WTI future crude oil and Brent future crude oil dataset were utilized for forecasting [51]. Here, the short-term dynamics of the crude oil future price were predicted for 1 day, 3 days, 5 days, 7 days, 15 days, and one month ahead and the simulation of the proposed work is carried out on MatLab 2017a with an i5 11th generation processor.
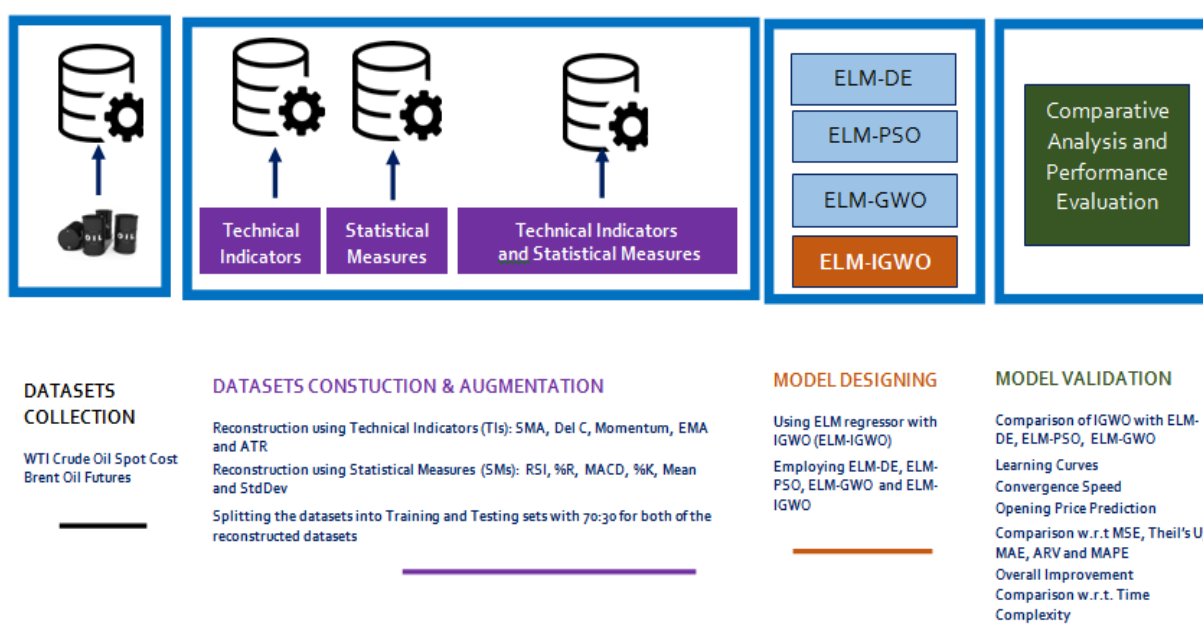

Figure 3. Schematic layout of the proposed crude oil forecasting model.

4.1. Model Description

A crude oil forecasting model based on ELM–IGWO was proposed in this study. The TIs and SMs [52,53] were used to increase the dimensionality of the original datasets and to get an augmented form of the datasets. These augmented datasets were given as the input to the proposed forecasting model. Figure 3 depicts an exchange rate price prediction model utilizing IGWO for crude oil datasets. The model was trained as a prediction model using ELM, with a time frame of a few days ahead as the output. The result for a one-day forecast contained the value for the following day, the output for a three-day prediction contained the value for the next three days, and so on. The result of the ELM's calculation is analyzed with the expected value to reflect the errors. Randomly produced weights and bias have a significant impact on the predictive model, as well as resulting in non-optimal solutions. As a result, the IGWO [47] was used to optimize the weights and bias.

This proposed ELM–IGWO forecasting model has four stages of operation. In the first phase, two crude oil datasets such as WTI crude oil spot cost and Brent oil futures [51] were collected for the last five years from 2016 to 2021. Three variations datasets were reconstructed by augmenting TIs (original dataset + TIs), SMs (original datasets + SMs), and TIs and SMs (original dataset + TIs + SMs), and then those augmented datasets were divided into training and testing sets with a 70:30 ratio in the second phase.

In the third phase, the ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO models were experimented with and finally, in the fourth phase, a well-defined comparison was made among all the experimented forecasting models and learning curves, convergence speed was recorded along with the predicted value of the opening price. The proposed ELM–IGWO was validated based on Theil's U, MAE, ARV, and MAPE. Finally, the overall improvement of the proposed forecasting model and time complexity was measured over ELM–DE, ELM–PSO, and ELM–GWO.

4.2. Parameters Considered

The various parameters chosen for ELM, DE, PSO, GWO, and IGWO during experimentation are shown in Table 2.

Table 2. Various parameters and their experimental settings.

Forecasting Models	Parameters	Values
ELM	Hidden Neurons	10
	Weight Range	[0, 1]
DE	Crossover Probability	0.25
	Scaling factor range [Max, Min]	[0.2, 0.9]
PSO	Search coefficient C1	2.5
	Search coefficient C2	1.3
	Inertia Weight	0.8
GWO	e (Decreased linearly)	2.0
IGWO	e (Decreased linearly)	2.0
	Crossover probability	0.2
	Scaling factor range [Max, Min]	[0.2, 0.9]

The population size and maximum iteration were both considered as 100 for all optimization algorithms.

4.3. Data Description and Data Augmentation

The detailed descriptions of the used TIs and SMs are mentioned in Table 3. The literature suggests that short-term forecasting mostly uses 10–30 days of crossover and for long-term forecasting, it is advisable to consider the window of 50–200 days, though in reality there is no set combination for choosing window sizes. In this study, the data of the corresponding analytical measures are determined using historical exchange rate data

and a window size of 12 days [20]. The idea behind choosing this 12-day window size is that it reacts more quickly to the changes in price trends than the previous pair rather than choosing a longer-term of 50–200 days, which may be more susceptible to whipsaws.

Table 3. Formula for TIs and SMs.

Technical Indicators and Statistical Measures	Formula
Simple Moving Average (*SMA*)	$SMA = \frac{Summation\ of\ n\ days\ open\ price}{n}$
Del C	$Del\ C = (i^{th} + 1)open\ price - i^{th}\ open\ price$
Momentum	$Momentum = \frac{Open\ Price\ (p)}{Open\ Price\ (p-n) \times 100}$
Exponential Moving Average (*EMA*)	$EMA = (Price_{Today} \times SF) + (EMA_{Yeasterday} \times (1 - SF)$ $SF(Smoothing\ factor) = \frac{2}{k+1},\ k = the\ length\ of\ EMA$
Average True Range (*ATR*)	$ATR = \frac{ATR_{t-1}X(n-1)+TR}{n}$ $TR = Max((Today's High - Today's Low), (Today's High - Today's Open), (Today's Open - Today's Low))$
Relative Strength Index (*RSI*)	$RSI = 100 - \frac{100}{(1+RS)}\ where\ RS = \frac{AVG\ n'Days\ upOpen}{AVG\ n'days\ down\ Open}$
William's %R	$\%R = \frac{HIghest\ High - Open}{Highest\ High - Lowest\ low} \times 100$
Moving average convergence divergence (*MACD*)	$MACD = 12\ days\ EMA - 26\ days\ EMA$
Stochastic Oscillator (%*K*)	$\%K = 100 \times \frac{(O - L_{12})}{(H_{12} - L_{12})}$ where O is most recent opening price, L_{12} is the lowest price of the t previous trading session and H_{12} is the highest price of the t previous trading session
Mean	$Mean = \frac{1}{N} \sum\limits_{i=1}^{N} a_i$ where a_1, a_2, a_3 are the values that contained in dataset.
Standard Deviation (StdDev)	$Standard\ Deviation = \sqrt{\frac{1}{N} \sum\limits_{i=1}^{N} (x_i - \mu)^2}$

The WTI crude petroleum spot cost and Brent crude oil spot cost (in US dollars per barrel) series from [48–50] is utilized as the trial test in this experiment to obtain the information for the exchange rate projection. A total of 70% of the dataset was utilized for training and 30% was utilized for testing in this experiment. Table 4 depicts the datasets along with their data range. The dimensionality of the dataset was augmented based on TIs and SMs as discussed above, such as *SMA*, *Del C*, *Momentum*, *EMA*, *ATR*, *RSI*, Williams %*R*, *MACD*, stochastic oscillator %*K*, and statistical measures such as StdDev and mean were computed on the datasets, as mentioned in Table 3.

Table 4. Description of data samples and data range.

Datasets	Total Samples	Data Range	Training Sample	Test Sample
WTI Crude Oil Spot Cost	1341	25 July 2016 to 23 August 2021	940	401
Brent Oil Futures	1311	25 July 2016 to 23 August 2021	917	394

In this proposed approach, all the common parameters were kept constant for all experimental findings. Here, the common parameters were population size and the number of iterations and both were set to 100 in this experiment. The algorithm-specific parameters

for ELM have a single hidden layer, with fifteen nodes taken into account to deliver a reasonable output of the required data.

5. Result Analysis

This section focuses on the results obtained for the ELM–IGWO forecasting model. The comparative error rate of ELM–DE, ELM–PSO, EM–GWO with ELPM–IGWO, and also the convergence rate with respect to short-term predictive days of different intervals for both the datasets of ELM–IGWO are discussed. The actual vs. predicted results obtained for those two datasets for three different combinations of datasets are also presented for short-term forecasting. The validation of the proposed model based on MSE, Theil's U, MAE, ARV, MAPE, and also the time taken for all the experimented forecasting models are recorded to validate and get better insight into the proposed model.

5.1. Description and Analysis of Error Convergence Graphs

The hybrid optimization methods offer their own set of benefits and also need meticulous readjustment for specific parameters of the algorithm. Algorithm-specific parameters, each of which is unique to the algorithm, have a significant impact on its performance. A local best solution is the outcome of improper tuning specific parameters used in the algorithm. In the proposed approach, due to the parametric non-tuning capability of ELM, this is used as an objective function. The global best is regarded as the least objective value (error), whereas, the global worst is considered the greatest objective value (error). The network architecture used in this study has fifteen nodes in the ELM hidden layer. For the first iteration, the biases and weights are chosen randomly, and the weight is modified using optimization techniques in the following iteration, and so on until the process is complete.

The convergence of error for each of the discussed optimization algorithms along with some previous approaches are presented in this study for both of the datasets. The error convergence graph of ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO is given in Figure 4a,b for WTI and Brent crude oil datasets, respectively. The convergence speed during the training processes for the WTI crude oil dataset, shown in Figure 4a, can be observed as the proposed ELM–IGWO is converging at approximately the 45th iteration whereas, ELM–PSO and ELM–GWO are converging at approximately the 87th and 50th iteration. Though ELM–DE is converging at the 25th iteration the error rate is high in comparison to the rest of other models and the proposed ELM–IGWO. The convergence error observation for the Brent crude oil dataset during the training phase from Figure 4b can be stated as: the proposed ELM–IGWO is converging at the 24th iteration, whereas, ELM–DE, ELM–PSO, and ELM–GWO are converging at the 78th, 80th, and 80th iterations, respectively. It can be summarized that the proposed ELM–IGWO is really outperforming with respect to MSE in both the training and testing phases of experimentation in comparison to the other three compared forecasting models.

Figure 5a–d illustrate the error convergence graphs for the WTI crude oil rate dataset using the ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO models, respectively, for different intervals of days. From those figures, it can be inferred that the proposed ELM–IGWO is converged at the 60th iteration with <0.047 MSE, and ELM–GWO is somehow being converged at the same iteration with an MSE of <0.49 for 30 days ahead of prediction. Similarly, it can be observed that the proposed ELM–IGWO is showing good performance over the rest of the compared models for all the days of predictions made and is converging at approximately the 60th iteration with an MSE < 0.05 for all six days of predictions ahead. The ELM–DE and EML–PSO are converging approximately at the 65th iteration with an MSE < 0.18 and an MSE < 0.14 for all six days of prediction timeframes. The ELM–GWO is converging at approximately the 60th iteration for all six days of predictions made with an approximate MSE < 0.14. From this figure, it can be seen that in all cases, the proposed ELM–IGWO has a good convergence rate for all six days of predictions made with less error during the training phase for the WTI crude oil rate dataset.
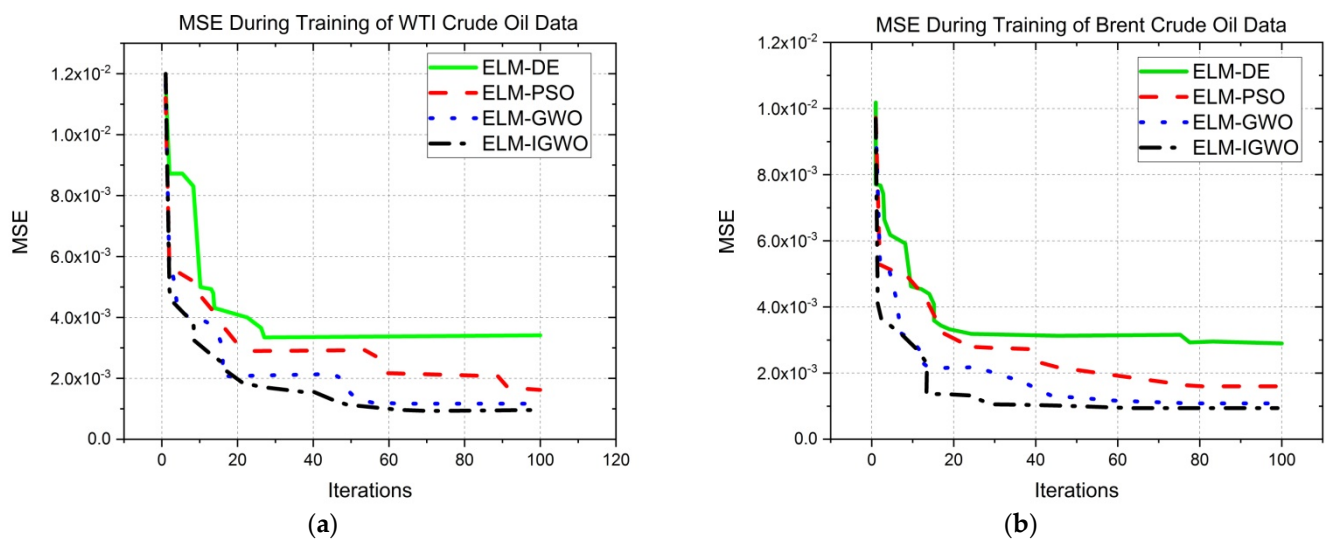
Figure 4. Comparative error convergence rate of ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO using: (**a**) Training data using WTI crude oil dataset and; (**b**) Training data using Brent crude oil dataset.
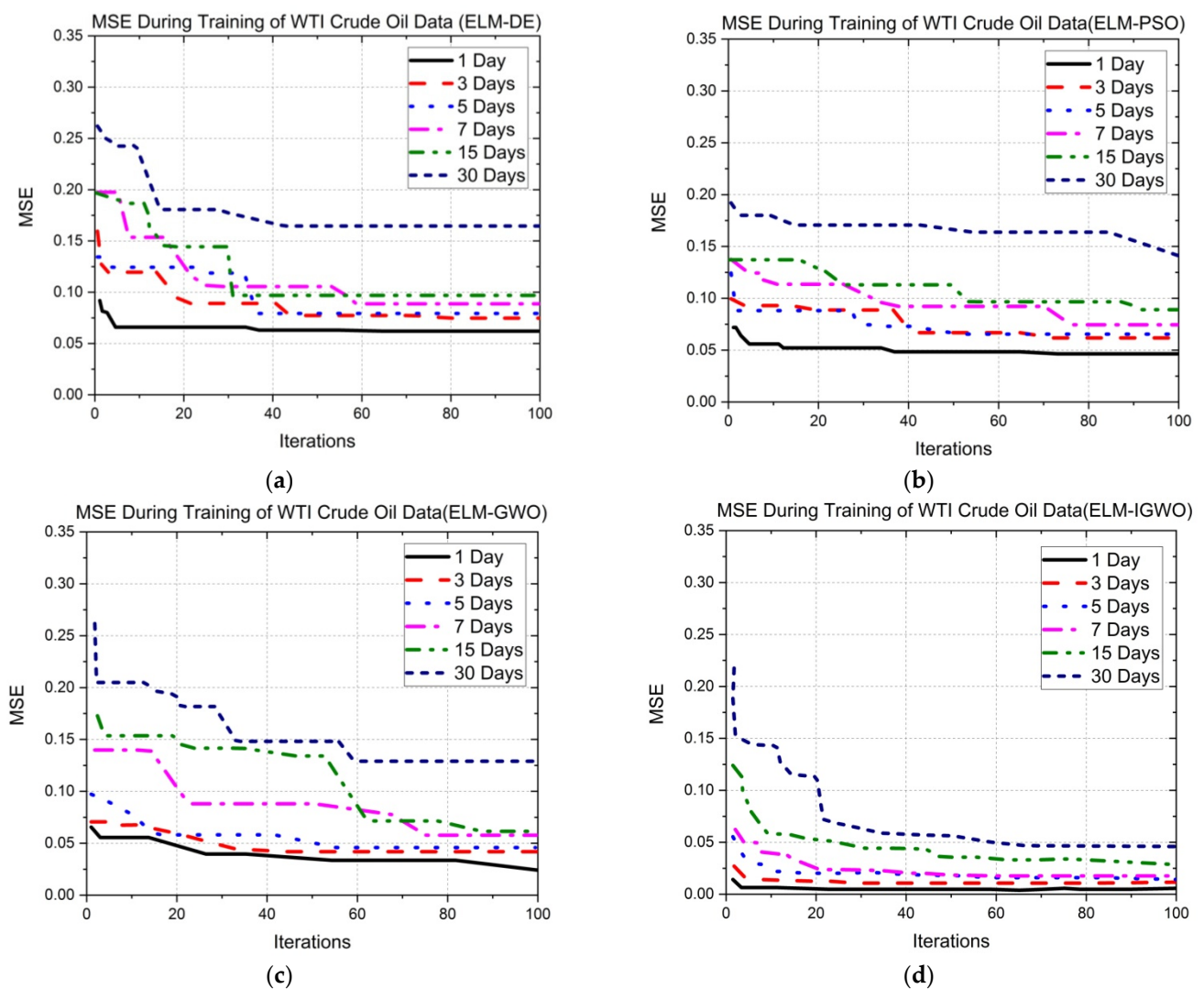

Figure 5. Convergence speed of WTI crude oil dataset during training using: (**a**) ELM–DE; (**b**) ELM–PSO; (**c**) ELM–GWO; and (**d**) ELM–IGWO for different intervals of days.

Figure 6a,d represent the error convergence graphs using different models, i.e., ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO, respectively, for different days of interval. The suggested ELM forecasting model utilizing IGWO converges better than ELM–PSO in the case of Brent crude oil, which can be understood from those figures. The proposed ELM–IGWO is outperforming the rest of all the other models with respect to the rate of convergence and error rate for all six timeframes of forecasting made, and all the curves are converging very well from the 5th iterations (only 30-day curves are from the 55th iteration) with an MSE < 0.05. However, the rest of the models were converging from the 20th, 10th, and 6th iterations with an MSE < 0.15, MSE < 0.13, and MSE < 0.10 for ELM–DE, ELM–PSO, and ELM–GWO, respectively, for six days of forecasting made for the Brent crude oil dataset during the training phase.

(**a**)

(**b**)

(**c**)

(**d**)
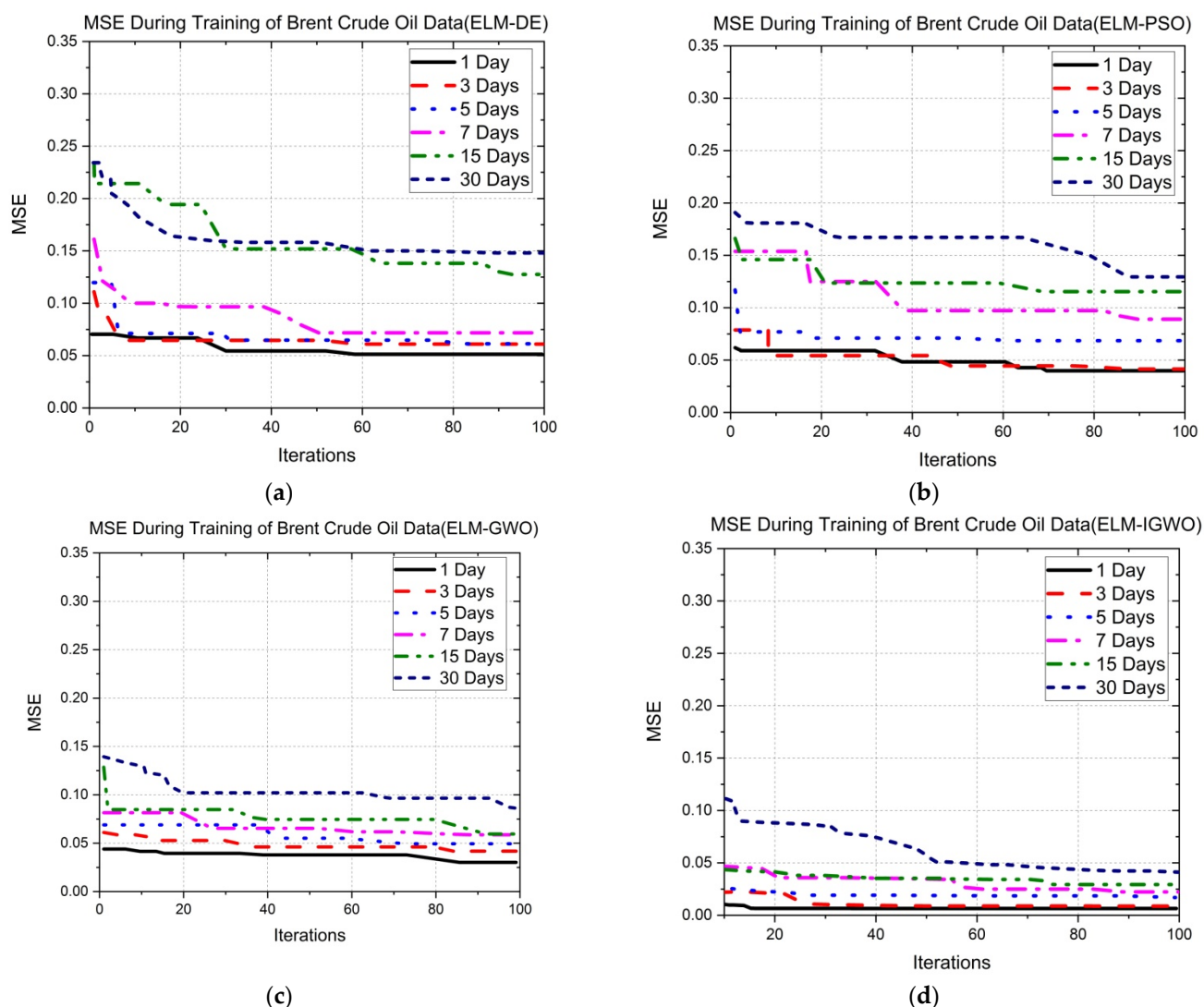
Figure 6. Convergence speed of Brent crude oil dataset during training using: (**a**) ELM–DE; (**b**) ELM–PSO; (**c**) ELM–GWO; and (**d**) ELM–IGWO for different intervals of days.

5.2. Forecasting Results and Discussions

In this section, the results obtained from different TIs and the SMs are presented and discussed and compared with each other and the proposed algorithm. A technical analysis graph of the ELM–IGWO algorithm is presented in Figure 7 for different intervals of days based on TIs (original datasets + TIs). Furthermore, the price forecasting graphs are also presented in Figure 8 for different intervals of days, which show the actual values along with the projected values of the WTI crude oil based on SMs (original dataset + TSMs). To

have a better picture of the result obtained from Figures 7 and 8, a comparative analysis is presented in Figure 9 including the TIs as well as SMs (original dataset + TIs + SMs). Figure 10 shows the predicted and actual opening price of crude oil utilizing the Brent future oil dataset for different day frames based on TIs (original dataset + TIs). Whereas, Figure 11 presents the forecasting curves of the predicted and actual crude oil price for different day frames utilizing the Brent future oil dataset based on SMs (original dataset + SMs). Furthermore, the combined impact of the TIs and SMs original dataset + TIs+ SMs) are shown in Figure 12, utilizing the Brent future oil dataset.
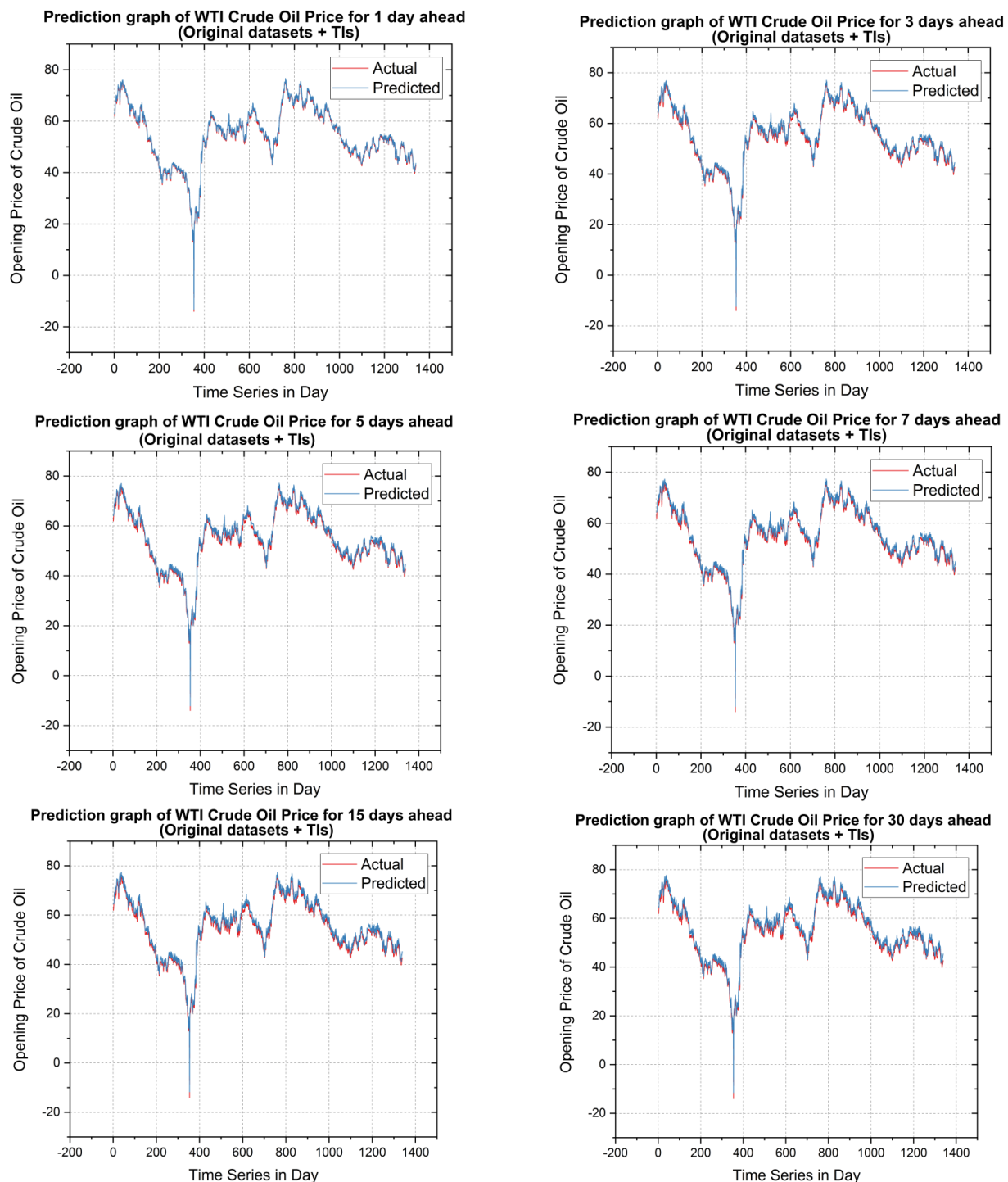
Figure 7. Open price prediction of WTI crude oil dataset of ELM–IGWO for 1 day, 3 days, 5 days, 7 days, 15 days and 30 days using TIs (Original datasets + TIs).
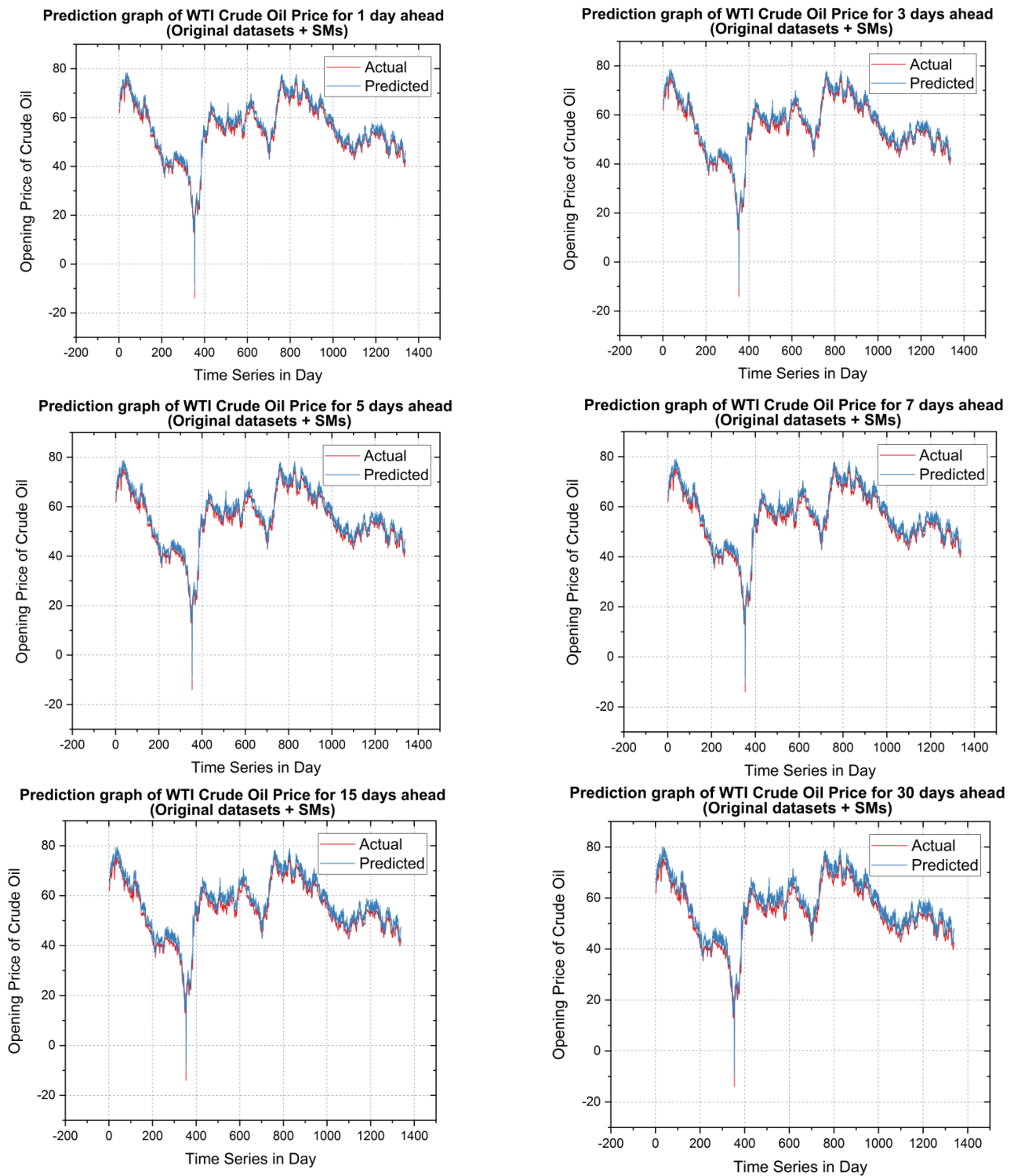
Figure 8. Open price prediction of WTI crude oil dataset of ELM–IGWO for 1 day, 3 days, 5 days, 7 days, 15 days, and 30 days using SMs (Original dataset + SMs).
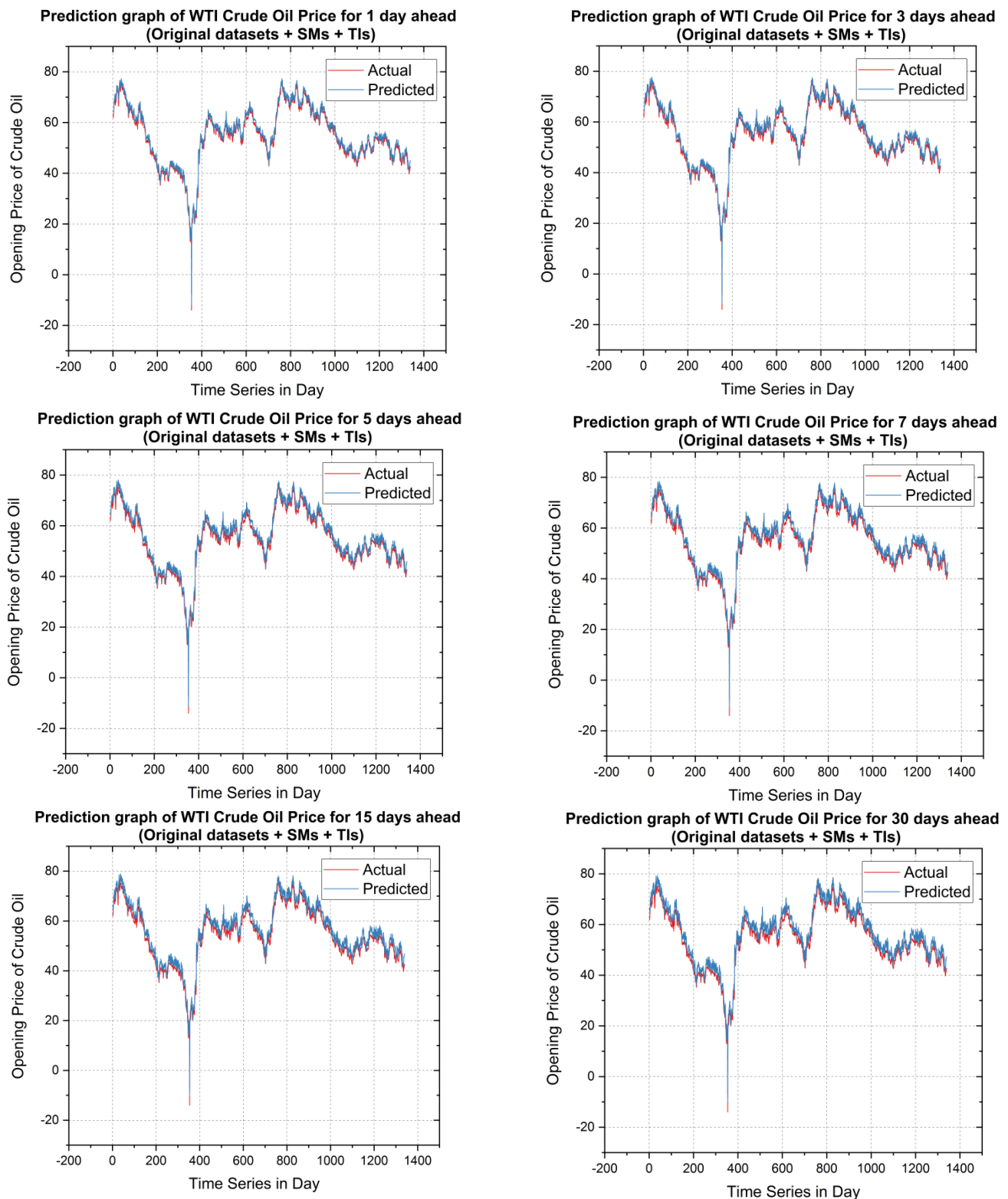
Figure 9. Open price prediction of WTI crude oil dataset of ELM–IGWO for 1 day, 3 days, 5 days, 7 days, 15 days, and 30 days using combined performance indicator, i.e., SMs and TIs (Original dataset + TIs + SMs).
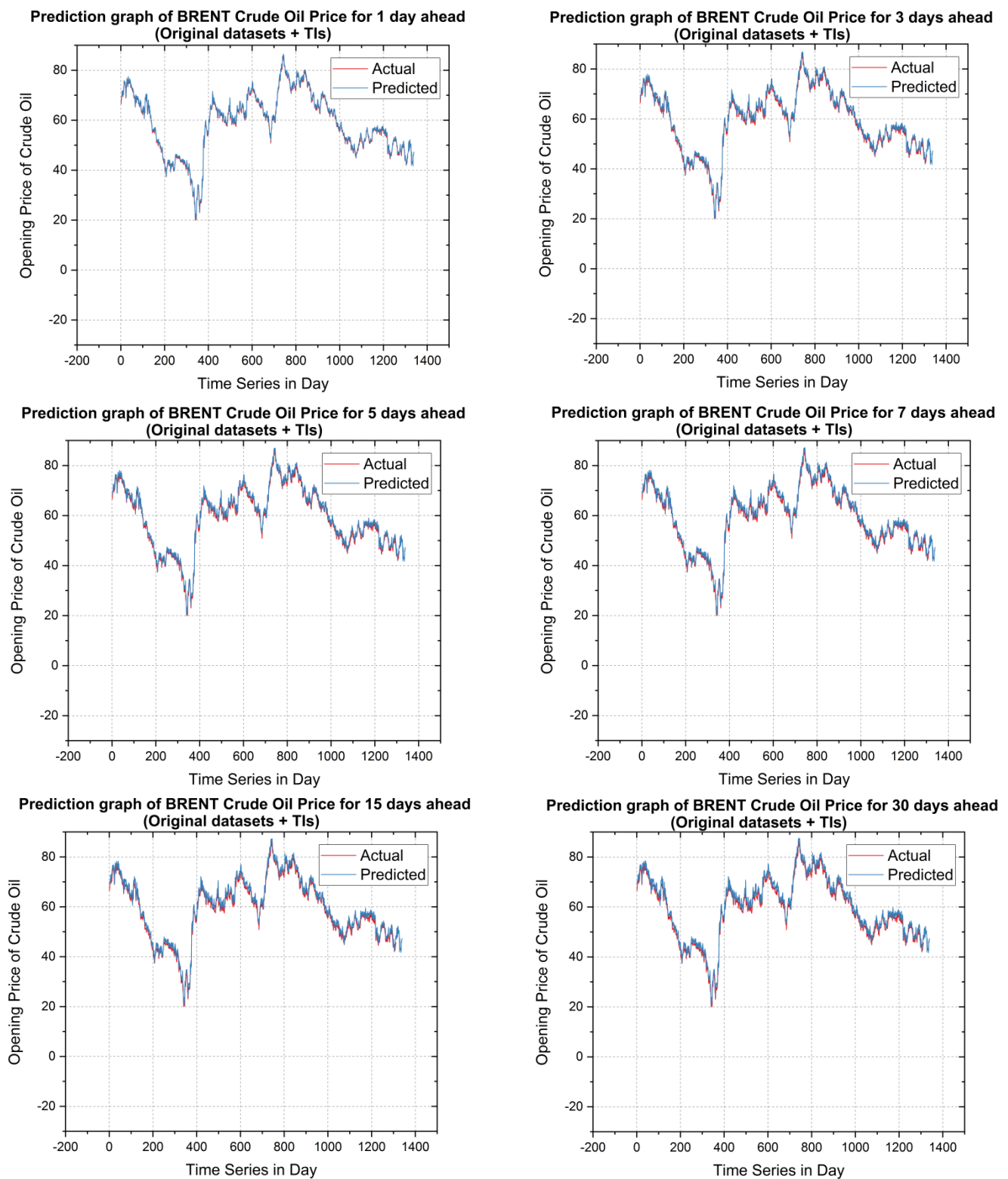
Figure 10. Open price prediction of Brent crude oil dataset of ELM–IGWO for 1 day, 3 days, 5 days, 7 days, 15 days, and 30 days usingTIs (Original dataset + TIs).
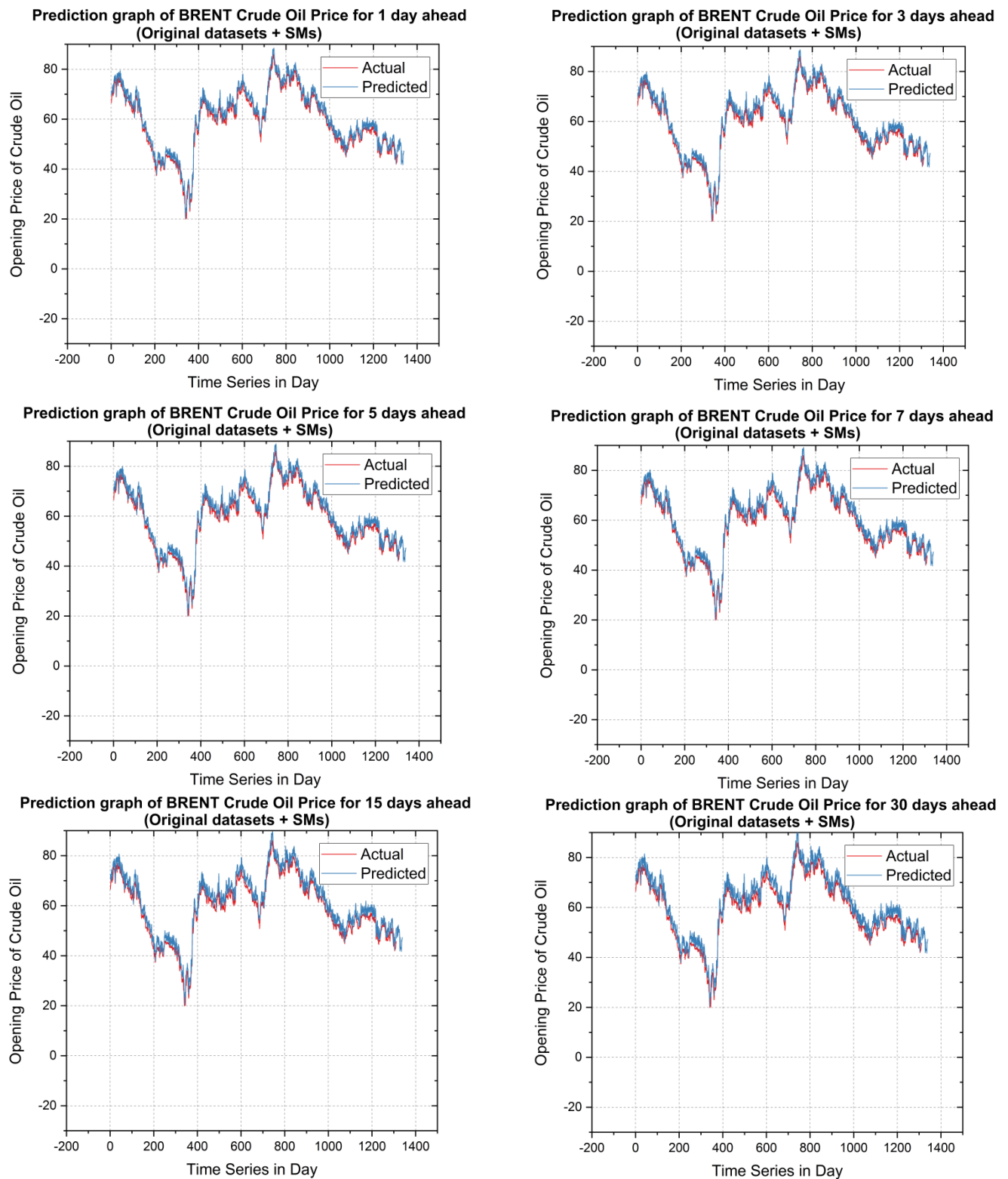
Figure 11. Open price prediction of Brent crude oil dataset of ELM–IGWO for 1 day, 3 days, 5 days, 7 days, 15 days, and 30 days using SMs (Original dataset + SMs).
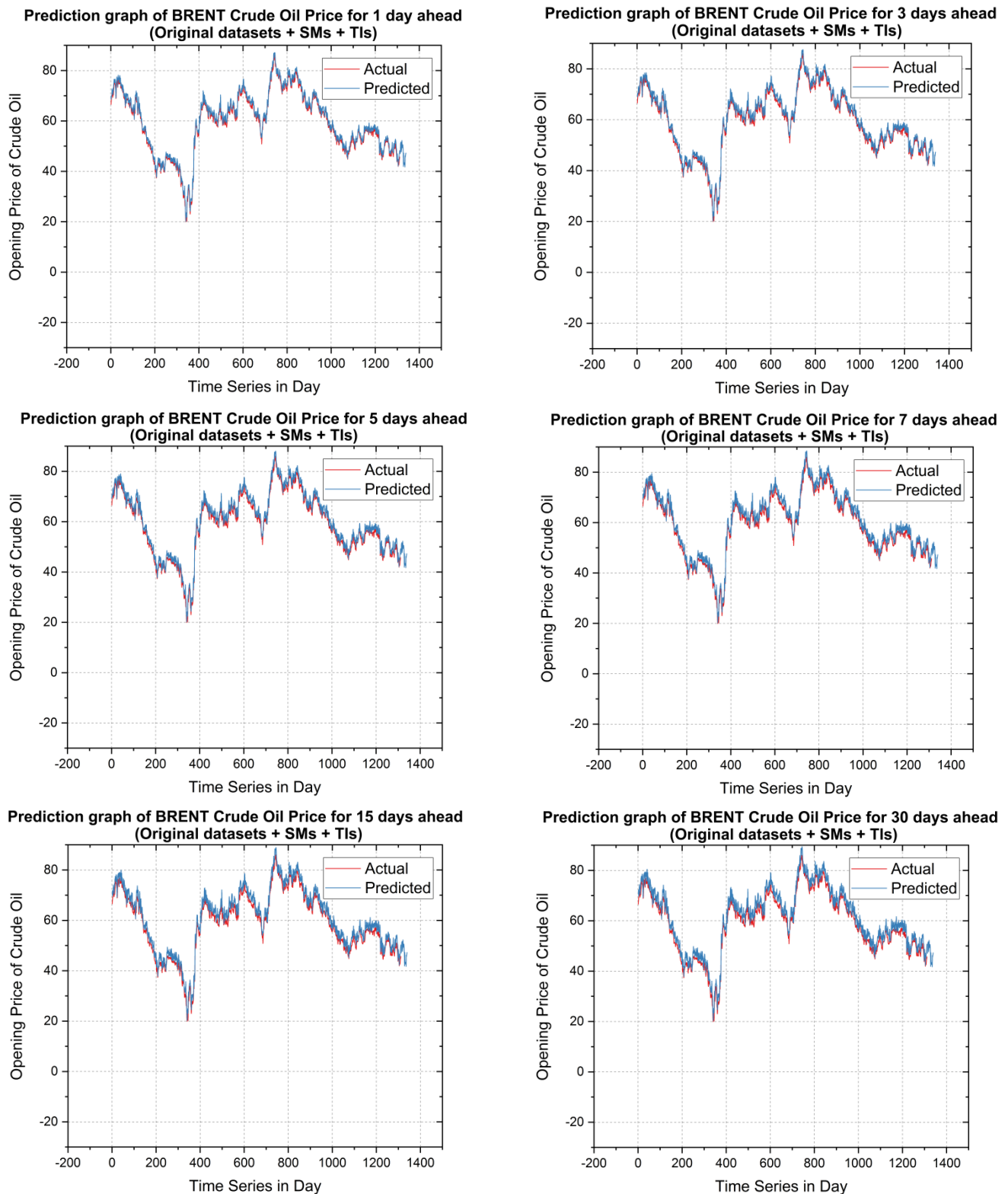
Figure 12. Open price prediction of Brent crude oil dataset of ELM–IGWO for 1 day, 3 days, 5 days, 7 days, 15 days, and 30 days using combined TIs and SMs (Original dataset + TIs + SMs).

In the classical ELM approach, a randomly generated bias, weights for the input, and the hidden layer are selected. Generally, the output weights are selected iteratively but, in this case, analytically, the output weights are calculated. This approach works on the principle of the pseudo-inverse approach. In this approach, the *b* value is calculated for a

single term and hence this will reduce the computational timing and extra tuning effort of the parameters in the hidden layer. The b value is calculated after collecting all the training data and hence the testing is carried out upon that b value and, in the testing phase, the MSE value is obtained. In a conventional ELM approach, the memory demand is more as the model has to train with all the training data being taken into consideration. Furthermore, it also takes more time for the computation. In this approach, the node in the hidden layer is more and another drawback is that a non-optimal solution may be incorporated, reducing the model's performance as the output weights are more reliant on the input weights and the hidden bias, which is chosen randomly.

Thus, to have a better result, the optimal biasing and the weights have to be selected. To achieve the best results, this study employs the ELM with the proposed improved version of GWO, i.e., IGWO [47]. In this approach, the convergence curve is simulated and it shows that the convergence of the IGWO in comparison to the other three algorithms, i.e., DE, PSO, and GWO have a quicker response. The outputs of these two algorithms can fall into the local minima within the current search space as, due to the multimodal functions, it has a number of local minima. Thus, to overcome these drawbacks, an integrated mechanism of IGWO discussed earlier, was used which helped to overcome it falling into the lower function value. This helps to obtain a quicker convergence rate from the earlier result.

The results of the conventional ELM optimized with DE, PSO, GWO, and IGWO are presented in this study. Another comparison is made between the simulation results of TIs and SMs. Furthermore, to extend the search result, this paper also carried out a simulation of the above using a combination of TIs and SMs. The WTI crude oil rate datasets are used in the simulations. In this experiment, a total of 1341 samples were used for simulation. The samples were obtained between 25 July 2016 and 23 August 2021. The window size of 12 is chosen for different intervals of days within a month when windowing the process. To determine the algorithm's generalization potential, the WTI crude oil rate datasets and Brent future oil dataset were each categorized into two halves, i.e., a training dataset and testing dataset, which was carried out at a ratio of 7:3. Initially, the model is trained using the training dataset, which comprised 70 percent of the total data samples, i.e., 940, and the remaining portion of the data samples, i.e., 401 were used for testing the model for predicting the crude oil price on the WTI dataset. Likewise, out of 1311 samples, 917 samples were utilized for training and the rest were utilized for testing purposes.

5.3. Validation and Discussion

In this work, the goal function for ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO, and the activation functions for ELM are rectified as a linear unit and the MSE fitness function was used to train all of the models. Furthermore, the assessment of the performance of the proposed algorithm is carried out using some of the measurement tools such as MSE, Theil's U, MAE, ARV, MAPE, etc. For ease of visualization and simplicity of understanding, the MSE values of each model are grouped by SMs, TIs, and a mix of TIs and SMs, and are presented in Tables 5 and 6 for the WTI and Brent datasets, respectively. From those tables, it can be observed that the performance of TI is obtaining a better result in the case of the ELM compared to the performance of SM and the combined TIs and SMs. In this proposed model, the MSE is calculated for the entire dataset, i.e., for the training and testing datasets, and for the testing datasets, the remaining performance matrices are used.

An empirical comparison of the performance measures obtained is reflected in Table 7. The use of ELM is readily indicated in this table of performance measurements. This table clearly shows that the performance of ELM with IGWO has an advantage over the other three ELM-based hybrid approaches. Furthermore, a comparative statement of time consumed during the execution of the previous state-of-the-art methods, i.e., ELM–DE, ELM–PSO, ELM–GWO with ELM–IGWO is presented in Figures 13 and 14 and clearly depicts that the proposed hybrid method outperforms the other hybrid models in terms of execution time measured in seconds for the WTI crude oil and Brent crude oil datasets, respectively.

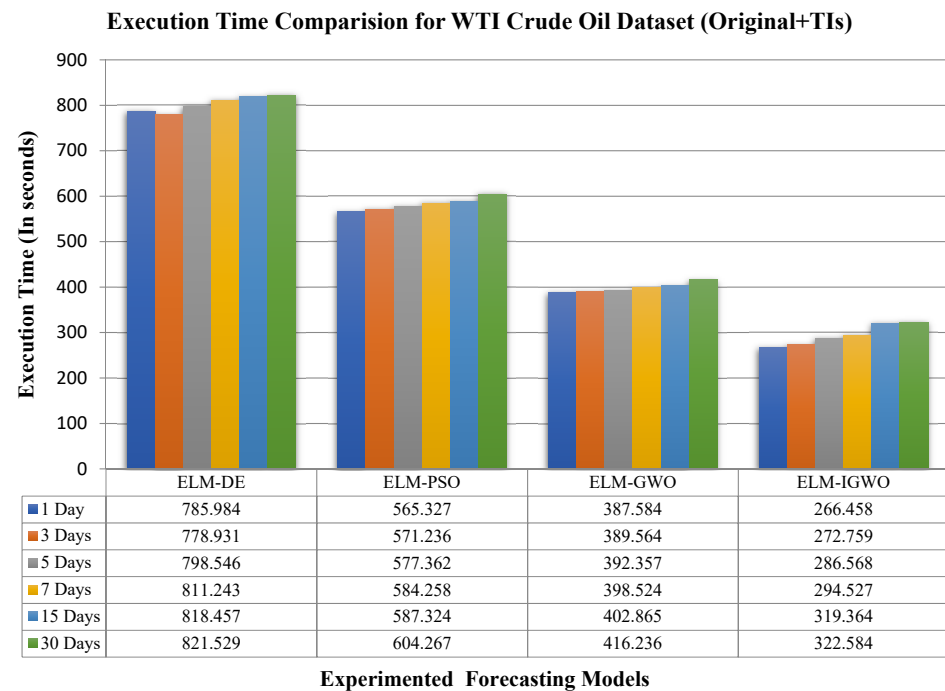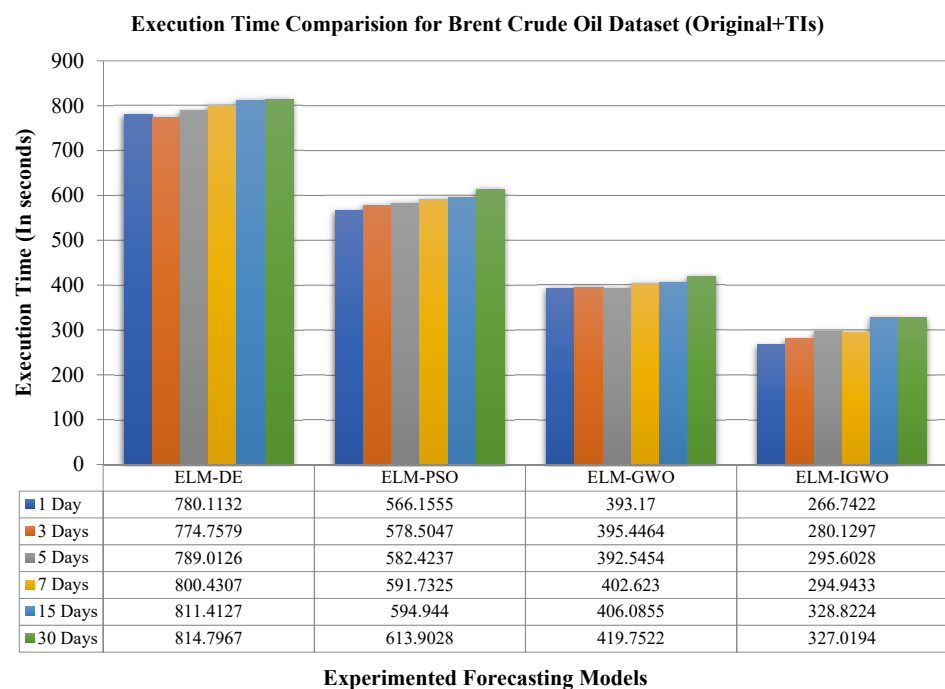**Table 5.** Comparison of MSE for WTI crude oil rate based on ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO.

No. of Days	ELM–DE			ELM–PSO			ELM–GWO			ELM–IGWO		
	TIs	SMs	TIs and SMs	TIs	SMs	TIs and SMs	TIs	SMs	TIs and SMs	TIs	SMs	TIs and SMs
1 Day	3.262×10^{-4}	3.136×10^{-4}	6.076×10^{-4}	3.494×10^{-4}	4.136×10^{-4}	5.076×10^{-4}	1.368×10^{-4}	1.572×10^{-4}	5.248×10^{-4}	1.053×10^{-4}	1.08×10^{-4}	5.042×10^{-4}
3 Days	3.591×10^{-4}	3.438×10^{-4}	7.664×10^{-4}	3.671×10^{-4}	3.988×10^{-4}	7.664×10^{-4}	2.986×10^{-4}	1.575×10^{-4}	9.347×10^{-4}	2.872×10^{-4}	1.48×10^{-4}	9.52×10^{-4}
5 Days	2.357×10^{-4}	2.433×10^{-4}	2.182×10^{-4}	3.883×10^{-4}	3.937×10^{-4}	4.178×10^{-4}	1.714×10^{-4}	1.643×10^{-4}	2.178×10^{-4}	1.744×10^{-4}	1.4×10^{-4}	2.1×10^{-4}
7 Days	3.493×10^{-4}	3.769×10^{-4}	3.452×10^{-4}	3.992×10^{-4}	4.169×10^{-4}	4.478×10^{-4}	1.421×10^{-4}	4.232×10^{-4}	1.82×10^{-4}	1.038×10^{-4}	4.16×10^{-4}	1.64×10^{-4}
15 Days	3.263×10^{-4}	4.832×10^{-4}	7.097×10^{-4}	4.247×10^{-4}	4.832×10^{-4}	6.563×10^{-4}	1.691×10^{-4}	4.48×10^{-4}	1.164×10^{-4}	1.26×10^{-4}	4.32×10^{-4}	1.07×10^{-4}
30 Days	11.44×10^{-4}	11.64×10^{-4}	14.58×10^{-4}	9.42×10^{-4}	9.67×10^{-4}	12.49×10^{-4}	9.374×10^{-4}	8.771×10^{-4}	11.61×10^{-4}	8.52×10^{-4}	8.47×10^{-4}	11.57×10^{-4}

Table 6. Comparison of MSE for Brent crude oil rate based on ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO.

No. of Days	ELM–DE			ELM–PSO			ELM–GWO			ELM–IGWO		
	TIs	SMs	TIs and SMs	TIs	SMs	TIs and SMs	TIs	SMs	TIs and SMs	TIs	SMs	TIs and SMs
1 Day	2.962×10^{-4}	3.467×10^{-4}	5.189×10^{-4}	2.653×10^{-4}	3.457×10^{-4}	5.136×10^{-4}	1.276×10^{-4}	1.837×10^{-4}	4.567×10^{-4}	1.103×10^{-4}	1.238×10^{-4}	4.308×10^{-4}
3 Days	3.191×10^{-4}	3.892×10^{-4}	6.738×10^{-4}	2.897×10^{-4}	3.324×10^{-4}	6.438×10^{-4}	2.583×10^{-4}	1.983×10^{-4}	5.462×10^{-4}	2.387×10^{-4}	1.948×10^{-4}	4.987×10^{-4}
5 Days	4.556×10^{-4}	4.896×10^{-4}	5.344×10^{-4}	3.988×10^{-4}	4.231×10^{-4}	4.874×10^{-4}	1.897×10^{-4}	2.278×10^{-4}	3.773×10^{-4}	1.664×10^{-4}	2.086×10^{-4}	3.479×10^{-4}
7 Days	6.278×10^{-4}	6.395×10^{-4}	7.378×10^{-4}	4.994×10^{-4}	5.362×10^{-4}	5.436×10^{-4}	1.354×10^{-4}	3.256×10^{-4}	2.321×10^{-4}	1.203×10^{-4}	4.325×10^{-4}	2.089×10^{-4}
15 Days	5.839×10^{-4}	5.852×10^{-4}	6.997×10^{-4}	4.667×10^{-4}	4.875×10^{-4}	5.573×10^{-4}	2.479×10^{-4}	3.988×10^{-4}	2.658×10^{-4}	2.263×10^{-4}	3.132×10^{-4}	2.019×10^{-4}
30 Days	9.563×10^{-4}	10.187×10^{-4}	11.158×10^{-4}	8.347×10^{-4}	9.246×10^{-4}	11.473×10^{-4}	8.894×10^{-4}	9.093×10^{-4}	10.568×10^{-4}	8.224×10^{-4}	8.787×10^{-4}	9.588×10^{-4}

Table 7. Comparison of different performance evaluation measures.

Methods	Theil's U	MAE	ARV	MAPE
ELM–DE	0.0014587	0.1356	0.004985	0.42578
ELM–PSO	0.0002784	0.0687	0.001256	0.12586
ELM–GWO	0.0001658	0.0027	0.000653	0.07842
ELM–IGWO	0.0001013	0.0009	0.000854	0.02967

Execution Time Comparision for WTI Crude Oil Dataset (Original+TIs)

	ELM-DE	ELM-PSO	ELM-GWO	ELM-IGWO
1 Day	785.984	565.327	387.584	266.458
3 Days	778.931	571.236	389.564	272.759
5 Days	798.546	577.362	392.357	286.568
7 Days	811.243	584.258	398.524	294.527
15 Days	818.457	587.324	402.865	319.364
30 Days	821.529	604.267	416.236	322.584

Experimented Forecasting Models

Figure 13. Comparison of execution time (in seconds) for different days for ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO for WTI crude oil dataset.

Execution Time Comparision for Brent Crude Oil Dataset (Original+TIs)

	ELM-DE	ELM-PSO	ELM-GWO	ELM-IGWO
1 Day	780.1132	566.1555	393.17	266.7422
3 Days	774.7579	578.5047	395.4464	280.1297
5 Days	789.0126	582.4237	392.5454	295.6028
7 Days	800.4307	591.7325	402.623	294.9433
15 Days	811.4127	594.944	406.0855	328.8224
30 Days	814.7967	613.9028	419.7522	327.0194

Experimented Forecasting Models

Figure 14. Comparison of execution time (in seconds) for different days for ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO for Brent crude oil dataset.

6. Conclusions and Future Scope

ELM was explored as a predictive network considering its fast and efficient learning speed, fast convergence, good generalization ability, and easy implementation mechanism. Additionally, it does not require any adjustment to the input weights and hidden layer biases during the implementation of the algorithm and it produces only one optimal solution. However, as the ELM is memory-heavy and suffers from high space and time complexity, the traditional ELMs need to be optimized. Therefore, in this study, an integrated ELM model based on IGWO (ELM–IGWO) was presented for prophesying the crude oil future price for short-term intervals such as for 1 day, 3 days, 5 days, 7 days, 15 days, and 30 days for the WTI crude oil and Brent crude oil datasets. The datasets are augmented using TIs and SMs to get better insight into the forecasting ability. Those augmented forms of the datasets, such as original dataset + Tis, original dataset + SMs, and original dataset + Tis+ SMs, are given as the input to the proposed ELM–IGWO crude oil forecasting model using a window size of 12 days. The idea behind choosing this 12-day window size is that it reacts more quickly to the changes in price trends than the previous pair.

The IGWO [47] method is utilized to achieve the right balance between exploration and exploitation, as well as to speed up the convergence rate and enhance the accuracy of the ELM network. The key advantages of the traditional GWO are its simple structure, ease of implementation, less memory and computational requirements, faster convergence due to continuous reduction of search space and fewer decision variables (*alpha*, *beta*, and *delta* wolves), ability to avoid local minima, and having only two control parameters to tune the algorithmic performance has made us realize how to improve the GWO and to make use of IGWO for optimizing the ELM and developing a forecasting model. The IGWO undergoes two phases of improvement from GWO to acquire a better searching performance by influencing the exploration and exploitation ability of the DE and the SOF principle. The weakness of falling into local optimum was greatly improved by adding the DE and SOF mechanisms to the standard GWO.

The DE was used to decide the evolutionary pattern of wolves by adding the evolution operation to the standard GWO. The basic operators of DE, such as mutation, crossover, and selection, are explored and added to obtain the difference among individuals to recombine the population and obtain intermediate individuals to get the population of the next generation, which evolves through a competition between the parent and offspring individuals, leading to the selection of outstanding individuals of wolves as parents through a form of variation factor. This variation factor ensures that the wolves can evolve towards a generation of good wolves, which is performed by selecting the *beta* and *delta* wolves as parents and then combining them with the *alpha* wolf. Additionally, a dynamic scaling factor is used to make the algorithm have high exploration and exploitation abilities and increase computational speed. After the mutation operation, the crossover operator with variation vector produces test individuals and the selection operation chooses the best individual with the best fitness value.

The SOF principle originated from Darwinian evolutionary theory to describe the mechanism of natural selection. The biological concept of the fitness (reproductive success) mechanism was added to the standard GWO to update the wolf pack by measuring the fitness value. In order to update the wolf pack, the fitness values are sorted in ascending order at each iteration of the algorithm and the wolves with the high fitness values are kept. The wolves with lower fitness values are eliminated and new wolves are randomly generated. The controlling factor for the generation of new wolves is randomly chosen between the range $[n/\varepsilon, \ n/0.75 \times \varepsilon]$ because, if the numbers of wolves with a high fitness value are large, that leads to generating the same large number of new wolves and this type of case leads to slow convergence speed because of the large search space and similarly, if the fitness value is chosen to be too small, the diversity of the population is not guaranteed, which results in an inability to explore new solution spaces.

The proposed model was compared with ELM–DE, ELM–PSO, and ELM–GWO based on the error convergence rate for both datasets. Figure 4a,b show the error convergence graph of ELM–DE, ELM–PSO, ELM–GWO, and ELM–IGWO for both the WTI and Brent crude oil datasets, respectively, during the training phase. Similarly, the convergence speed is measured and recorded, which are depicted in Figure 5a,d for WTI crude oil and Figure 6a,d for Brent crude oil datasets, respectively. From those two sets of figures, it can be observed that, in all the cases, the proposed ELM–IGWO has a good convergence rate for all short-term forecasting made, with less error during the training phase for the WTI crude oil rate dataset and for the Brent crude oil ELM–IGWO and is outperforming the rest of all the other models with respect to the rate of convergence and error rate for all six timeframes of forecasts made. Moreover, all the curves are converging very well from the 5th iteration (only 30-days curves are from the 55th iteration) with an MSE < 0.05. During the validation phase, the MSE, Theil's U, MAE, ARV, and MAPE are used as performance measures and the measured values are given in Tables 5–7 for the WTI and Brent datasets, respectively. From those tables it can be observed that the performance of the TIs (original dataset + TIs) is obtaining a better result in the case of ELM over the performance of the SMs (original dataset + SMs) and the combined TIs and SMs (original dataset +TIs + SMs). Finally, a comparison based on CPU time utilization (in seconds) was recorded for both the datasets based on the augmented form of original datasets + Tis, which are shown in Figures 13 and 14. The proposed model effectively combined ELM's potential with the grey wolf multi-population search strategy. The simulation results show that the suggested model beats the other three predictive models when it comes to projecting the WTI crude oil rate value and the Brent crude oil rate value. A comparison of TIs, SMs, and a combination of TIs and SMs was constructed. The performance of the TIs is excellent when compared to the SMs and with the combined measures, as shown in the results obtained from figures and error comparison tables.

As an enhancement of this study, the proposed model's ability could be checked by capturing the time shift, i.e., whether the knowledge of the WTI crude oil prices would improve the forecasting of the Brent crude oil prices or not. Moreover, while augmenting the datasets by TIs and SMs, a window size of 12 days was used and this can further be explored with varying window sizes. The above may be considered as a pathway for further research in this domain.

Author Contributions: Conceptualization, A.K.D., D.M., K.D. and S.K.; methodology, D.M., P.K.M. and M.Z.; validation, A.K.D. and D.M.; formal analysis, A.K.D.; writing—original draft preparation, A.K.D. and D.M.; writing—review and editing, A.K.D., D.M. and S.K.; supervision, D.M., S.K., M.Z. and H.E.-S.; funding acquisition, H.E.-S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on request to the first author.

Conflicts of Interest: The authors declare no conflict of interest in the publication of this article.

Nomenclatures

ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
GARCH	Generalized Autoregressive Conditional Heteroscedastic
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedastic
GJR–GARCH	Glosten–Jagannathan–Runkle GARCH
AI	Artificial Intelligence
ML	Machine Learning

SVM	Support Vector Machine
BPNN	Back-Propagation Neural Networks
ELM	Extreme Learning Machine
ACO	Ant Colony Optimization
PSO	Particle Swarm Optimization
ABC	Artificial Bee Colony
CS	Cuckoo Search
DE	Differential Evolution
GWO	Grey Wolf Optimization
IGWO	Improved Grey Wolf Optimization
SOF	Survival Of Fittest
WTI	West Texas Intermediate
TIs	Technical Indicators
SMs	Statistical Measures
MSE	Mean Square Error
MAE	Mean Absolute Error
ARV	Average Relative Variance
MAPE	Mean Absolute Percentage Error
GA	Genetic Algorithm
VAR	Value At Risk
VMD	Variational Mode Decomposition
FPA	Flower Pollination Algorithm
DTW	Dynamic Time Warping
MIDAS	Mixed Data Sampling
CNN	Convolutional Neural Network
RSBL	Random Sparse Bayesian Learning
GARCH-M	GARCH with Long Memory
SMA	Simple Moving Average
EMA	Exponential Moving Average
ATR	Average True Range
RSI	Relative Strength Index
MACD	Moving Average Convergence Divergence
%K	Stochastic Oscillator
StdDev	Standard Deviation

References

1. Ahmad, M.I. Modelling and forecasting Oman Crude Oil Prices using Box-Jenkins techniques. *Int. J. Trade Glob. Mark.* **2012**, *5*, 24–30. [CrossRef]
2. Morana, C. A semiparametric approach to short-term oil price forecasting. *Energy Econ.* **2001**, *23*, 325–338. [CrossRef]
3. Sadorsky, P. Modeling and forecasting petroleum futures volatility. *Energy Econ.* **2001**, *23*, 325–338. [CrossRef]
4. Weng, F.; Zhang, H.; Yang, C. Volatility forecasting of crude oil futures based on a genetic algorithm regularization online extreme learning machine with a forgetting factor: The role of news during the COVID-19 pandemic. *Resour. Policy* **2021**, *73*, 102148. [CrossRef] [PubMed]
5. Monsef, A.; Hortmani, A.; Hamzeh, K. Prediction of Oil Price using ARMA Method for Years 2003 to 2011. *Int. J. Acad. Res. Account. Financ. Manag. Sci.* **2013**, *3*, 235–247.
6. Xiang, Y.; Zhuang, X.H. Application of ARIMA model in short-term prediction of international crude oil price. *Adv. Mater. Res.* **2013**, *798*, 979–982. [CrossRef]
7. Zhao, C.L.; Wang, B. Forecasting crude oil price with an autoregressive integrated moving average (ARIMA) model. In *Fuzzy Information & Engineering and Operations Research & Management*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 275–286.
8. Wu, F.; Cattani, C.; Song, W.; Zio, E. Fractional ARIMA with an improved cuckoo search optimization for the efficient Short-term power load forecasting. *Alex. Eng. J.* **2020**, *59*, 3111–3118. [CrossRef]
9. Mohammadi, H.; Su, L. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Econ.* **2010**, *32*, 1001–1008. [CrossRef]
10. Agnolucci, P. Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. *Energy Econ.* **2009**, *31*, 316–321. [CrossRef]
11. Maitra, S. GARCH Processes & Monte-Carlo Simulations for Crude-Oil Prediction. Available online: https://www.researchgate.net/publication/335977950_GARCH_Processes_Monte-Carlo_Simulations_for_Crude-Oil_Prediction (accessed on 16 March 2022).
12. Hou, A.; Suardi, S. A nonparametric GARCH model of crude oil price return volatility. *Energy Econ.* **2012**, *34*, 618–626. [CrossRef]

13. Wang, Y.; Wu, C. Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models? *Energy Econ.* **2012**, *34*, 2167–2181. [CrossRef]

14. Wang, S.Y.; Yu, L.; Lai, K.K. A novel hybrid AI system framework for crude oil price forecasting. *Lect. Notes Comput. Sci.* **2004**, *3327*, 233–242.

15. Ding, Y. A novel decompose-ensemble methodology with AIC-ANN approach for crude oil forecasting. *Energy* **2018**, *154*, 328–336. [CrossRef]

16. Chai, J.; Lu, Q.; Hu, Y.; Wang, S.; Lai, K.K.; Liu, H. Analysis and Bayes statistical probability inference of crude oil price change point. *Technol. Forecast. Soc. Chang.* **2018**, *126*, 271–283. [CrossRef]

17. Fan, L.; Pan, S.; Li, Z.; Li, H. An ICA-based support vector regression scheme for forecasting crude oil prices. Technol. Forecast. *Soc. Chang.* **2016**, *112*, 245–253.

18. Xie, W.; Yu, L.; Xu, S.; Wang, S. A new method for crude oil price forecasting based on support vector machines. In Proceedings of the International Conference on Computational Science, Reading, UK, 28–31 May 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 444–451.

19. Ahmed, R.A.; Shabri, A.B. Daily crude oil price forecasting model using AIMA, generalized autoregressive conditional heteroscedastic and support vector machines. *Am. J. Appl. Sci.* **2014**, *11*, 425–432. [CrossRef]

20. Nayak, R.K.; Mishra, D.; Rath, A.K. A Naı̈ve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. *Appl. Soft Comput.* **2015**, *35*, 670–680. [CrossRef]

21. Gupta, N.; Nigam, S. Crude Oil Price Prediction using Artificial Neural Network. *Procedia Comput. Sci.* **2020**, *170*, 642–647. [CrossRef]

22. Li, M.B.; Huang, G.B.; Saratchandran, P.; Sundararajan, N. Fully complex extreme learning machine. *Neurocomputing* **2005**, *68*, 306–314. [CrossRef]

23. Yu, L.; Dai, W.; Tang, L. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. *Eng. Appl. Artif. Intell.* **2016**, *47*, 110–121. [CrossRef]

24. Wang, J.; Athanasopoulos, G.; Hyndman, R.; Wang, S. Crude oil price forecasting based on internet concern using an extreme learning machine. *Int. J. Forecast.* **2018**, *34*, 665–677. [CrossRef]

25. Azadeh, A.; Asadzadeh, S.M.; Mirseraji, G.H.; Saberi, M. An emotional learningneuro-fuzzy inference approach for optimum training and forecasting of gas consumption estimation models with cognitive data. *Technol. Forecast. Soc. Chang.* **2015**, *91*, 47–63. [CrossRef]

26. Escribano, Á.; Wang, D. Mixed random forest, cointegration, and forecasting gasoline prices. *Int. J. Forecast.* **2021**, *37*, 1442–1462. [CrossRef]

27. Uthayakumar, J.; Metawa, N.; Shankar, K.; Lakshmanaprabu, S.K. Financial crisis prediction model using ant colony optimization. *Int. J. Inf. Manag.* **2020**, *50*, 538–556.

28. Pan, M.; Li, C.; Gao, R.; Huang, Y.; You, H.; Gu, T.; Qin, F. Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization. *J. Clean. Prod.* **2020**, *277*, 123948. [CrossRef]

29. Ghanbari, A.; Kazemi, S.; Mehmanpazir, F.; Nakhostin, M.M. A Cooperative Ant Colony Optimization-Genetic Algorithm approach for construction of energy demand forecasting knowledge-based expert systems. *Knowl.-Based Syst.* **2013**, *39*, 194–206. [CrossRef]

30. Wang, P.C.; Shoup, T.E. A poly-hybrid PSO optimization method with intelligent parameter adjustment. *Adv. Eng. Softw.* **2011**, *42*, 555–565. [CrossRef]

31. Xu, X.; Ren, W. A hybrid model of stacked autoencoder and modified particle swarm optimization for multivariate chaotic time series forecasting. *Appl. Soft Comput.* **2021**, *116*, 108321. [CrossRef]

32. Zhang, T.; Tang, Z.; Wu, J.; Du, X.; Chen, K. Multi-step-ahead crude oil price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm. *Energy* **2021**, *229*, 120797. [CrossRef]

33. Houssein, E.H.; Gad, A.G.; Hussain, K.; Suganthan, P.N. Major Advances in Particle Swarm Optimization: Theory, Analysis, and Application. *Swarm Evol. Comput.* **2021**, *63*, 100868. [CrossRef]

34. Larrea, M.; Porto, A.; Irigoyen, E.; Barragán, A.J.; Andújar, J.M. Extreme learning machine ensemble model for time series forecasting boosted by PSO: Application to an electric consumption problem. *Neurocomputing* **2021**, *452*, 465–472. [CrossRef]

35. Yang, X.; Yuan, J.; Yuan, J.; Mao, H. An improved WM method based on PSO for electric load forecasting. *Expert Syst. Appl.* **2010**, *37*, 8036–8041. [CrossRef]

36. Wang, J.; Wang, Z.; Li, X.; Zhou, H. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *Int. J. Forecast.* **2019**, *38*, 21–34. [CrossRef]

37. Cuong-Le, T.; Minh, H.-L.; Khatir, S.; Wahab, M.A.; Mirjalili, S.; Tran, M.T. A novel version of Cuckoo search algorithm for solving optimization problems. *Expert Syst. Appl.* **2021**, *186*, 115669. [CrossRef]

38. Wang, L.; Hu, H.; Ai, X.-Y.; Liu, H. Effective electricity energy consumption forecasting using echo state network improved by differential evolution algorithm. *Energy* **2018**, *153*, 801–815. [CrossRef]

39. Opara, K.R.; Arabas, J. Differential Evolution: A survey of theoretical analyses. *Swarm Evol. Comput.* **2019**, *44*, 546–558. [CrossRef]

40. Neri, F.; Tirronen, V. Recent Advances in Differential Evolution: A survey and Experimental Analysis. *Artif. Intell. Rev.* **2010**, *33*, 61–106. [CrossRef]

41. Das, S.; Mullick, S.S.; Suganthan, P.N. Recent advances in differential evolution—An updated survey. *Swarm Evol. Comput.* **2016**, *27*, 1–30. [CrossRef]

42. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]

43. Rajakumar, R.; Sekaran, K.; Hsu, C.-H.; Kadry, S. Accelerated grey wolf optimization for global optimization problems. *Technol. Forecast. Soc. Chang.* **2021**, *169*, 120824. [CrossRef]

44. Altan, A.; Karasu, S.; Zio, E. A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. *Appl. Soft Comput.* **2021**, *100*, 106996. [CrossRef]

45. Tikhamarine, Y.; Souag-Gamane, D.; Ahmed, A.N.; Kisi, O.; El-Shafie, A. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *J. Hydrol.* **2020**, *582*, 124435. [CrossRef]

46. Ma, X.; Mei, X.; Wu, W.; Wu, X.; Zeng, B. A novel fractional time delayed grey model with Grey Wolf Optimizer and its applications in forecasting the natural gas and coal consumption in Chongqing China. *Energy* **2019**, *178*, 487–507. [CrossRef]

47. Wang, J.-S.; Li, S.-X. An Improved Grey Wolf Optimizer Based on Differential Evolution and Elimination Mechanism. *Sci. Rep.* **2019**, *9*, 7181. [CrossRef]

48. Chai, J.; Xing, L.-M.; Zhou, X.-Y.; Zhang, Z.G.; Li, J.-X. Forecasting the WTI crude oil price by a hybrid-refined method. *Energy Econ.* **2018**, *71*, 114–127. [CrossRef]

49. Ruble, I.; Powell, J. The Brent-WTI spread revisited: A novel approach. *J. Econ. Asymmetries* **2021**, *23*, e00196. [CrossRef]

50. Mastroeni, L.; Mazzoccoli, A.; Quaresima, G.; Vellucci, P. Decoupling and recoupling in the crude oil price benchmarks: An investigation of similarity patterns. *Energy Econ.* **2021**, *94*, 105036. [CrossRef]

51. Available online: https://in.investing.com/commodities/crude-oil-historical-statistics (accessed on 30 December 2021).

52. Available online: https://www.ig.com/en/trading-strategies/10-trading-indicators-every-trader-should-know-190604 (accessed on 2 January 2022).

53. Lai, T.L.; Xing, H. *Statistical Models and Methods for Financial Markets*; Springer: New York, NY, USA, 2008.

54. Nademi, A.; Nademi, Y. Forecasting crude oil prices by a semiparametric Markov switching model: OPEC, WTI, and Brent cases. *Energy Econ.* **2018**, *74*, 757–766. [CrossRef]

55. de Albuquerquemello, V.P.; de Medeiros, R.K.; da Nobrega Besarria, C.; Maia, S.F. Forecasting crude oil price: Does exist an optimal econometric model? *Energy* **2018**, *155*, 578–591. [CrossRef]

56. Debnath, K.B.; Mourshed, M. Forecasting methods in energy planning models. Renew. Sustain. *Energy Rev.* **2018**, *88*, 297–325.

57. Tang, L.; Yu, L.; Wang, S.; Li, J.P.; Wang, S.Y. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting. *Appl. Energy* **2012**, *93*, 432–443. [CrossRef]

58. Guo, X.; Li, D.; Zhang, A. Improved Support Vector Machine Oil Price Forecast Model Based on Genetic Algorithm Optimization Parameters. *AASRI Procedia* **2012**, *1*, 525–530. [CrossRef]

59. Zhao, L.; Cheng, L.; Wan, Y.; Zhang, H.; Zhang, Z. A VAR-SVM model for crude oil price forecasting. *Int. J. Glob. Energy Issues* **2015**, *38*, 126–144. [CrossRef]

60. Jianwei, E.; Bao, Y.; Ye, J. Crude oil price analysis and forecasting based on variational mode decomposition and independent component analysis. *Phys. A Stat. Mech. Appl.* **2017**, *484*, 412–427.

61. Xiao, Y.; Xiao, J.; Lu, F.; Wang, S. Ensemble ANNs-PSO-GA Approach for Day-ahead Stock E-exchange Prices Forecasting. *Int. J. Comput. Intell. Syst.* **2014**, *7*, 272–290. [CrossRef]

62. Alfi, A.; Modares, H. System identification and control using adaptive particle swarm optimization. *Appl. Math. Model.* **2011**, *35*, 1210–1221. [CrossRef]

63. Yang, X.S. Flower pollination algorithm for global optimization. In Proceedings of the International Conference on Unconventional Computing and Natural Computation, Orléans, France, 3–7 September 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 240–249.

64. Chiroma, H.; Khan, A.; Abubakar, A.I.; Saadi, Y.; Hamza, M.F.; Shuib, L.; Gital, A.Y.; Herawan, T. A new approach for forecasting OPEC petroleum consumption based on neural network train by using flower pollination algorithm. *Appl. Soft Comput.* **2016**, *48*, 50–58. [CrossRef]

65. He, K.; Tso, G.K.; Zou, Y.; Liu, J. Crude oil risk forecasting: New evidence from multiscale analysis approach. *Energy Econ.* **2018**, *76*, 574–583. [CrossRef]

66. Czudaj, R.L. Heterogeneity of beliefs and information rigidity in the crude oil market: Evidence from survey data. *Eur. Econ. Rev.* **2022**, *143*, 104041. [CrossRef]

67. Dutta, A.; Bouri, E.; Saeed, T. News-based equity market uncertainty and crude oil volatility. *Energy* **2021**, *222*, 119930. [CrossRef]

68. He, K.; Zou, Y. Crude oil risk forecasting using mode decomposition based model. *Procedia Comput. Sci.* **2022**, *199*, 309–3142. [CrossRef]

69. Chen, Z.; Ye, Y.; Li, X. Forecasting China's crude oil futures volatility: New evidence from the MIDAS-RV model and COVID-19 pandemic. *Resour. Policy* **2021**, *75*, 102453. [CrossRef] [PubMed]

70. Zhao, Y.; Zhang, W.; Gong, X.; Wang, C. A novel method for online real-time forecasting of crude oil price. *Appl. Energy* **2021**, *303*, 117588. [CrossRef]

71. Wu, B.; Wang, L.; Lv, S.-X.; Zeng, Y.-R. Effective crude oil price forecasting using new text-based and big-data-driven model. *Measurement* **2021**, *168*, 108468. [CrossRef]

72. Li, T.; Qian, Z.; Deng, W.; Zhang, D.; Lu, H.; Wang, S. Forecasting crude oil prices based on variational mode decomposition and random sparse Bayesian learning. *Appl. Soft Comput.* **2021**, *113*, 108032. [CrossRef]

73. Abdollahi, H. A novel hybrid model for forecasting crude oil price based on time series decomposition. *Appl. Energy* **2020**, *267*, 115035. [CrossRef]

74. Lin, L.; Jiang, Y.; Xiao, H.; Zhou, Z. Crude oil price forecasting based on a novel hybrid long memory GARCH-M and wavelet analysis model. *Phys. AStat. Mech. Appl.* **2020**, *543*, 123532. [CrossRef]

75. Deng, S.; Xiang, Y.; Fu, Z.; Wang, M.; Wang, Y. A hybrid method for crude oil price direction forecasting using multiple timeframes dynamic time wrapping and genetic algorithm. *Appl. Soft Comput.* **2019**, *82*, 105566. [CrossRef]