applied sciences

Article

Prediction of Complex Stock Market Data Using an Improved Hybrid EMD-LSTM Model

Muhammad Ali [1], Dost Muhammad Khan [1,*], Huda M. Alshanbari [2,*] and Abd Al-Aziz Hosni El-Bagoury [3]

1 Department of Statistics, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan
2 Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
3 Higher Institute of Engineering and Technology, El-Mahala El-Kobra 31951, Egypt
* Correspondence: dostmuhammad@awkum.edu.pk (D.M.K.); hmalshanbari@pnu.edu.sa (H.M.A.)

Abstract: Because of the complexity, nonlinearity, and volatility, stock market forecasting is either highly difficult or yields very unsatisfactory outcomes when utilizing traditional time series or machine learning techniques. To cope with this problem and improve the complex stock market's prediction accuracy, we propose a new hybrid novel method that is based on a new version of EMD and a deep learning technique known as long-short memory (LSTM) network. The forecasting precision of the proposed hybrid ensemble method is evaluated using the KSE-100 index of the Pakistan Stock Exchange. Using a new version of EMD that uses the Akima spline interpolation technique instead of cubic spline interpolation, the noisy stock data are first divided into multiple components technically known as intrinsic mode functions (IMFs) varying from high to low frequency and a single monotone residue. The highly correlated sub-components are then used to build the LSTM network. By comparing the proposed hybrid model with a single LSTM and other ensemble models such as the support vector machine (SVM), Random Forest, and Decision Tree, its prediction performance is thoroughly evaluated. Three alternative statistical metrics, namely root means square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), are used to compare the aforementioned techniques. The empirical results show that the suggested hybrid Akima-EMD-LSTM model beats all other models taken into consideration for this study and is therefore recommended as an effective model for the prediction of non-stationary and nonlinear complex financial time series data.

Keywords: ARIMA; nonlinear complex data; empirical mode decomposition; LSTM; support vector machine; decision tree; random forest; recurrent neural network; data processing

check for updates

Citation: Ali, M.; Khan, D.M.; Alshanbari, H.M.; El-Bagoury, A.A.-A.H. Prediction of Complex Stock Market Data Using an Improved Hybrid EMD-LSTM Model. *Appl. Sci.* **2023**, *13*, 1429. https://doi.org/10.3390/app13031429

Academic Editors: Guojun Wang, Muhammad Arif, Oana Geman and Muhammad Arfan Jaffar

Received: 1 January 2023
Revised: 15 January 2023
Accepted: 18 January 2023
Published: 21 January 2023

1. Introduction

The history of modern stocks dates back to 1602 when trade for the Dutch East India Company used to be made in Amsterdam, the Netherlands. In the early days, the buying and selling were only for this specific company, the first offshoots were exchanged in 1607, and the profit dispensation between its shareholders was made several years later [1].

A stock market is a place where shares are converted, merchandised, and distributed. It turns out to be a significant network for huge corporations to increase assets from stockholders. On the other hand, after issuing the stocks, a great number of wealth streams into the stock marketplace, which increases the structure of commercial assets by encouraging investment attentiveness, which significantly helps the growth of the product economy. Whereas through the flow of stocks, capital is shared, and the growth of investment is successfully promoted. Therefore, the stock market is well thought out to be an indicator of the financial and economic activities in a country or region. Specifically, the buying and selling prices of the stock market frequently serve as a barometer for both the value and quantity of stock since they can indicate a connection between supply and demand.

At the same time, several determinants, which include political, financial, and the market with the addition of technologies and shareholder sentiments, will altogether lead to variations in indices values. Consequently, stock prices are uninterruptedly changing, and this disparity offers breathing space for hypothetical engagements and upsurges the insecurity concerned with the indexes. This type of alarming situation not simply carries a financial deficit to the stockholders but additionally might cause definite restrictions on the financial development of corporations and states.

To be precise, examining as well as forecasting indexes in an appropriate way is crucial in respect of stakeholder selections inclusive of state financial constancy. In order to examine the financial market such as collecting, sorting, and combining numerous appropriate pieces of evidence, it must be assisted in comprehending as well as forecasting the tendency relevant to indexes together with conforming stockholder choices in such a manner to decrease chances of losses along with maximum profits [2]. As long as we can forecast the difficulties of the stock market precisely, it will offer a robust foundation for the justifiable growth of the financial sector of a country. Thus, the methodological examination and prediction of the stock market could be beneficial for the stockholders to gain profits, which will also be helpful for the growth of national financial resources.

Forecasting indexes is still considered a challenging problem in financial time series because of their structure of unpredictability. Numerous parallel research articles show the significance of research in financial economics amidst the alarming situation for the stockholders to gain maximum profits. Time series analysis is mainly dependent on the orthodox approach of index projection [3].

De Gooijer et al. [4] examined the research articles available in journals administrated by the International Institute of Forecasters and concluded that the majority of the researchers utilized time-series methods for prediction. The most well-known and widely used classical time-series approaches are autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA) [5–8]. The main idea of all such techniques relies upon the time series only whilst omitting additional key determinants such as the contextual findings. Explicitly, considering the current values as the dependent variable and their lagged values as an independent variable to establish a meaningful relationship between them. Additionally, such techniques frequently demand numerous propositions and earlier understanding, such as what is the statistical distribution of the data, the justifiable range for numerous parameters, and their connections. Although, as a complicated structure with a good deal of effective attributes and an unpredictable situation, the stock market inclines to show robust attributes, which forms the classical systematic approaches unsuccessful.

Furthermore, the volume of the data used by building the model and then for the prediction of the stocks are frequently enormous, increasing countless difficulties in training the model. With such characteristics, it becomes inefficient to forecast the stock market by implementing the well-known classical approaches. In the recent past, financial experts have been extensively applying machine learning techniques for prediction problems, especially support vector regression (SVR) [9], as well as artificial neural networks (ANNs) [10]. Moreover, the development of deep learning techniques became an emerging field for machine learning because of its outstanding capability to handle nonlinear time series without any input features to build its architecture. Such techniques have a strong data handling capacity that became helpful in solving the complexities triggered by such nonlinear financial time series. For this reason, the integration of deep learning and financial time series has a remarkably outstretched ability, but limited studies are available regarding the prediction of the stock market with the help of deep learning techniques [11,12].

In this study, we forecast the stock market by integrating our proposed new version of EMD based on the Akima spline interpolation technique and LSTM network. The main strength of this research work for the prediction of financial time series is summarized in the following way. The decomposition method of EMD and its different versions are ordinary techniques that are extensively implemented in the field of engineering to process signals as well as financial time-series investigation and prediction [13], therefore, to extract the trend component from the stock price time-series data, it was suggested to implement the new version of EMD. The primary objective is to steadily break down the variations of the diverse composite time-series data to achieve a sequence of IMFs, and a unique monotone residue component [14].

Hence, the complex financial time series is decomposed, and then the prediction is made from the decomposed simple systems to obtain enhanced efficacy. Secondly, we suggest implementing a hybrid LSTM model that is developed by using all the subcomponents (residue) that are correlated with the actual stock closing prices, and finally, the prediction is made by aggregating individual forecasts obtained from each hybrid LSTM model. Though several researchers have practiced deep learning for financial time series prediction, many of them utilized it for classification problems, and considerably few scholars have used it for regression problems. Additionally, amongst various frequently used deep learning techniques, such as convolutional neural networks (CNN), deep belief networks, and LSTM, we selected the improved hybrid LSTM model with multiple hidden layers having the outstanding ability to examine the dependencies among time-series data at various time points through its memory function. Thus, an effort has been made to forecast the stock market with the help of implementing the aforementioned hybrid new version of EMD based on the Akima spline interpolation technique and LSTM network. Specifically, we integrated LSTM with Akima-EMD and suggested a stock market forecasting model. The characteristics of our proposed model are twofold, i.e., to improve the forecast accuracy as well as to minimize time delay.

To sum up, a financial time series forecasting model that combines Akima-EMD and hybrid LSTM is suggested. The remaining sections of this article are summarized in the following order: related work in Section 2, methodology and proposed ensemble hybrid model in Section 3, investigational results in Section 4, model comparison in Section 5, and conclusion of the paper in Section 6.

2. Literature Review

As a type of recurrent neural network (RNN), LSTMs are capable of learning long-term interdependencies in data. This makes them an appropriate technique for processing and forecasting significant scenarios with comparatively extended intervals and delays in time series. In this section of the study, the research work related to the advancement of LSTM and its implementation for stock market prediction is discussed. It is witnessed from the previous studies that RNN is a widely used technique to forecast the indexes [15–17].

However, this technique suffers from the vanishing gradient problem, which makes the prediction results misleading. An LSTM network can be used to solve the problem of vanishing gradients since it replaces the hidden layer units of RNNs with memory cells, making it more suitable for stock price prediction [18].

Several studies have demonstrated that the LSTM network produces better predictions than any other neural network. Based on this characteristic, LSTM turns out to be a significant tool in natural language recognition [19–21], time series prediction [22–24], specifically stock price prediction [25–28], as well as other areas such as water desalination [29–31], material sciences [32], welding of materials [33], laser technology [34,35], metal cutting [36], and material processing [37].

To forecast the KOSPI 200 index, a novel hybrid model that integrates LSTM and a type of GARCH model was proposed by [38]. According to this study, the proposed hybrid model outperforms the other models in terms of minimum values of several statistical metrics, including MAE, mean squared error (MSE), heteroscedasticity-adjusted MAE (HMAE), and heteroscedasticity-adjusted MSE (HMSE).

The decomposition-and-ensemble framework is a well-known hybrid method that works on the principle of "divide-and-conquer" [39,40]. The core idea of such hybrid models for the purpose of forecasting complex financial time series is to divide the actual data into small components, technically known as IMFs, and a monotone residue. Now it is quite simple for the model to predict this subseries correspondingly. Furthermore, the final prediction results of the actual series can be obtained by combining the forecasted values obtained from ensemble modeling of the different independent IMFs.

In particular, a hybrid attention-based EMD-LSTM model to forecast the indexes was proposed by [41]. The LSTM forecasting model that works on the mechanism of attention is capable for obtaining the association between the forecasted values and input attributes, to reduce the prediction error. With the help of regression analysis, the value of the coefficient of correlation is nearly equal to one, suggesting a better prediction performance from the suggested EMD-LSTM-ATTE technique. The EMD-LSTM-ATTE model prediction performance is better than the competing forecasting models, according to numerical values of statistical metrics such as MAE, RMSE, and MAPE.

The model proposed by [42] for forecasting the stock market utilizes stockholders' emotional tendencies to build a deep learning algorithm. Firstly, the investor's sentiment is taken into consideration, which can successfully enhance the prediction performance of the model. Secondly, since the stock price data is very complex and nonlinear, therefore, its accurate prediction is often not only difficult but also challenging. Thirdly, the stacked LSTM model is implemented because of its use of examining associations amongst time-series data through its memory function. Based on the investigational results, it is evident that the revised stacked LSTM model provides the best prediction performance and reduces the time delay at the same time.

To forecast stock values, the authors of [43] presented a hybrid deep learning neural network built on the CEEMD, transformer, LSTM, GRU, and high-frequency adaptive structure. The efficiency of the suggested model is checked on 100 stocks selected from CSI-300. Various methods, including SVM, LSTM, and CNN, are compared with the proposed model. In terms of minimum values for MSE and MAE, the proposed model, which is called FDG-trans, outperforms all other models.

To detect the distributed denial of service attack (DDoS), Ref. [44] proposed an effective and adaptive intrusion detection system by using LSTM and convolution neural network (CNN). Based on the CIC-DDoS2019 dataset, a proposal has been developed for detecting different types of DDoS attacks. A CICFlowMeter-V3 network was used to develop the dataset. Several performance measures were used to evaluate the model, including precision, recall, F1-score, and accuracy. It was found that the proposed CNN-LSTM model was capable of reaching a high degree of precision (100%) concerning all the evaluation metrics.

To predict noisy intraday stock prices, Ref. [45] proposed a hybrid model combining CEEMD, entropy, GRU, and history attention (CEGH). Four steps make up the implementation of the suggested model. In the first step, the complex stock price data is decomposed into different components, technically known as IMFs, with the help of complete ensemble empirical mode decomposition (CEEMD). Then, for each IMF, the sample entropy (SampEn) values and approximation entropy (ApEn) values are used to remove noise. The remaining IMFs were then aggregated into four groups, and a feedforward neural network (FNN) or recurrent gate unit with historical attention (GRU-HA) was used to forecast the comprehensive signals. Integrating the outcomes of each group's predictions yields the final forecast. For checking the prediction performance of the suggested model, two stock markets such as China and U.S., are used as a real-world scenario. It is evident from the empirical results that the proposed CEGH model outperforms the other models in terms of prediction accuracy.

A hybrid model that combines the features of a Graph Convolutional Network (GCN) and a Bidirectional Long Short-Term Memory (BiLSTM) network was suggested by [46] to predict crude oil prices. It provides new possibilities for the analysis of time series and improves existing results used in previous studies. Based on the minimum values of RMSE, MSE, MAPE, and the R-squared (R^2), the prediction accuracy of the proposed BiLSTM-GCN is better than the BiLSTM, GCN, and the orthodox models.

A slope-based method (ISBM) is based on empirical mode decomposition (EMD) and feed-forward neural networks (FNN), namely, the EMD-ISBM-FNN method proposed by [47] to decompose and forecast crude oil prices. The complex nonlinear and nonstationary Brent crude oil data are divided into several IMFs and single montone residues in the first step of the EMD approach. In order to train the architect of the FNN model, these subcomponents are then used as input features. Compared to the single FNN and EMD-FNN, the performance of the suggested model, known as the EMD-ISBM-FNN, is higher.

The available literature is the motivation factor behind implementing a hybrid Akima-EMD-LSTM model to forecast the stock market index. Since it is difficult to predict the nonlinear and nonstationary patterns of the stock market with the help of the classical time series and other models used in the previous literature. Hence, consideration is given to EMD and its different types, including the proposed new version of breaking down the complex stock prices into simple subcomponents that can be modeled easily with the help of the LSTM network. In this research article, we propose a new version of EMD that is based on the Akima spline interpolation technique and then integrate this new method with the LSTM network that uses the correlated subcomponents (residue) with the actual data. Detail description of our proposed model is given in Section 3.5.

3. Materials and Methods

The method of EMD, EEMD, SEMD, the deep learning algorithm known as the LSTM network, and the proposed new version that is based on the Akima spline interpolation technique is comprehensively described in the following subsection.

3.1. Empirical Mode Decomposition (EMD)

The method of EMD was introduced by [48] in 1998. The complex signal is divided into distinct oscillatory components that range in frequency from low to high using the well-known Hilbert-Huang transform (HHT) approach, leaving just a single monotone residue. There are two essential requirements for any IMF: (i) the upper and lower envelopes must have zero means, and (ii) the number of zero-crossings and the number of extrema must differ by one. With a signal $y(t)$, the EMD method can successfully separate signals into their many components. This approach is often used for prediction problems simply because it is reliable, straightforward, and effective and doesn't depend on any major model assumptions [49–51]. The following is a description of the EMD process:

Step 1: Identify every local extremum (local maximum and minimum) in the signal $\{y_i\ (t)\}$.

Step 2: Determine the upper signal envelope $\{U\ (t)\}$, and lower envelope $\{L\ (t)\}$.

Step 3: To obtain the meaning of the upper and lower envelopes, i.e., $M\ (t)$, join all the minima and maxima using the cubic spline interpolation approach.

$$Mean(t) = \frac{U(t) + L(t)}{2} \tag{1}$$

Step 4: In order to acquire the first component, the mean envelope determined in step 3 is subtracted from the actual signal, i.e.

$$k_1(t) = y(t) - Mean(t) \tag{2}$$

If $k_1\ (t)$ satisfies the two requirements for the IMF as described above, it should be treated as the first IMF; otherwise, steps 1 through 4 are repeated by treating $k_1\ (t)$ as a novel signal.

Step 5: In order to produce $r_1\ (t)$, the first IMF identified in step 4 will be subtracted from the signal, i.e.,

$$r_1(t) = y(t) - k_1(t) \tag{3}$$

Step 6: The filtering procedure from step 1 is used once more in this step, where $r_1\ (t)$ is treated as a new signal. The aforementioned procedure will keep going till the final IMF is extracted from the signal. Finally, the real signal $y\ (t)$ can be decomposed in such a way that the overall trend of the signal will be a smooth monotonic residue established in the final stage of EMD, mathematically:

$$y(t) = \sum_{i=1}^{n} k_i(t) + r_n \tag{4}$$

where $k_1(t)$, $k_2(t) \ldots k_n(t)$ are several IMFs with varying frequencies that range from high to low, and r_n is the residual. Figure 1 depicts the flowchart for this decomposition.

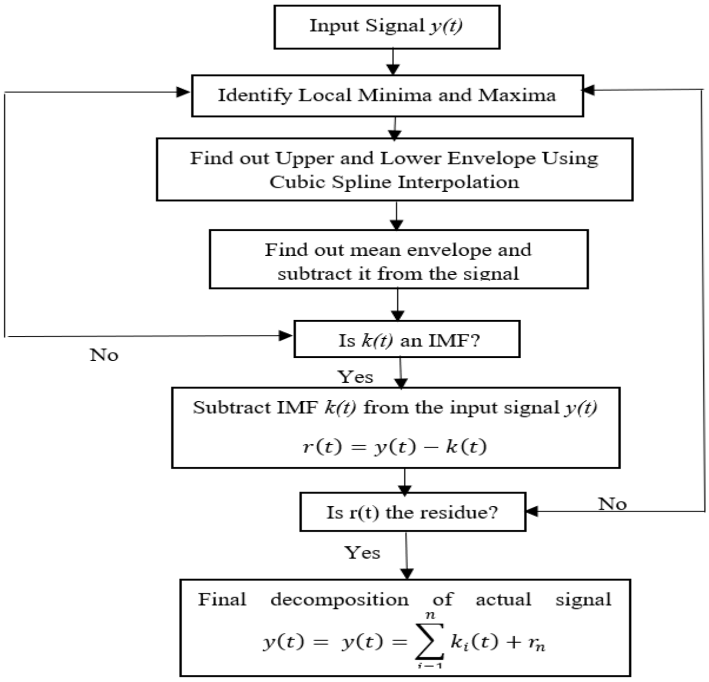

Figure 1. Flowchart of the EMD algorithm.

3.2. Ensemble Empirical Mode Decomposition (EEMD)

Mode mixing is the main disadvantage of the classical EMD while decomposing complex signals into different subgroups. To remove this drawback, Wu and Huang [52] suggested the ensemble empirical mode decomposition (EEMD) technique in 2009. The procedure of this technique is outlined below:

Step 1: Select the ensemble number, *m*, and the white noise amplitude, *n*.

Step 2: To create a new series, $y_i^{'}(t)$, add a white noise series, $w_i(t)$ to the actual signal, $y(t)$.

$$y_i^{'}(t) = y(t) + w_i(t) \tag{5}$$

Step 3: Divide the signal $y_i^{'}$ using EMD into several IMFs and a monotone residue.

Step 4: Repeat the above steps 2 and 3 by adding diverse white noise series, respectively, and

Step 5: Attain the (ensemble) means of the appropriate IMFs of the decompositions as the concluding outcome.

3.3. Statistical Empirical Mode Decomposition (SEMD)

The method of SEMD introduced by [53] uses smoothing instead of cubic spline interpolation for the purpose of extracting the first mode. The smoothing technique is somehow more useful than cubic spline interpolation, specifically when the signal is corrupted with noise. The step-by-step procedure of this method is outlined in the following lines.

Step 1: Make *K* test datasets $T_1, \ldots, T_k, \ldots, T_K$ from a signal $x(t)$.

Step 2: To generate $\widetilde{T}_k$ impute the k^{th} test dataset using the local average of two nearby points.

Step 3: Apply the SEMD technique to break down the composite signal $T_1, \ldots, T_{k-1}$, $\widetilde{T}_k$, $T_{k+1}, \ldots, T_k$ into an $h_{1,\lambda}$, and the remaining signal r_λ with the specified smoothing parameter.

Step 4: Determine the forecasted values for the remaining signal that was assessed at the k^{th} segment, namely $r_\lambda^k(t)$.

Step 5: Steps (ii) through (iv) should be repeated for $k = 1, \ldots, K$, and the prediction error is computed as

$$PE(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left\{x(t) - r_\lambda^k(t)\right\}^2 \tag{6}$$

3.4. Long Short-Term Memory (LSTM) Network

The extended recurrent neural network (RNN) is known as the long short-term memory (LSTM) network. There are several internal connections between nodes in the hidden layer of the RNN, which allows information to pass backward or forwards without restrictions [54–56]. The output of the RNN may be seen as a response to both the input layer and the previous state of each hidden unit, and the state of the preceding node can be considered as the input. The following two reasons explain why its effectiveness in the investigation is far less efficient than expected. At first, it could be difficult for an RNN to determine the right window size for historical observations, which can lead to insufficient retrieval of the data's variational properties. Secondly, the gradient of an RNN could potentially exhibit an ascending expansion or decline whilst utilizing the technique of gradient descent to handle historical data, which specifically results in gradient explosion. The extended connection structure of an RNN retains a lot of unimportant information without filtering, which is the motivation for the aforementioned approach. The conventional RNN is, therefore, a poor choice for handling long-term data. The hidden layer nodes of the RNN are replaced by unique memory cells (blocks) that employ filtering as well as the conversion of previous states and information in LSTM's expressive "gate" structure, which optimizes the RNN. Figure 2 shows the basic structure of memory cells. Input gate i_t, forget gate f_t, and output gate o_t make up each of the three gates that make up a memory cell. Based on the most recent timestamp, the input gate determines the amount of current information in the cell. The forget gate's contribution is to determine how much information from the previous cell state should be maintained and how much should be discarded, which prevents the internal cell values from increasing without bounds. The structure of the output gate allows it to filter the new state and output the filtered data. Following are the steps that describe the procedure of the LSTM network.

Firstly, the initial gate i_t filters and draws additional information belonging to the input x_t appearing in the current state (time t), as well as computes the value of the memory cell $\widetilde{c}_t$ for updating the state.

$$i_t = \delta(Z_i \cdot [h_{t-1}, x_t] + a_i) \tag{7}$$

$$\widetilde{c}_t = tanh(Z_c \cdot [h_{t-1}, x_t] + a_c \tag{8}$$

Determine the value of the forget gate f_t. The forget gate refines and stores the prior information to such an extent that they can represent long-term tendencies and eliminates irrelevant information.

$$f_t = \delta\left(Z_f \cdot [h_{t-1}, x_t] + a_F\right) \tag{9}$$

By removing a piece of the information from the old cell, moreover adding together the filtered candidate value, the old cell state c_{t-1} is restructured to the new cell state c_t.

$$c = f_t * c_{t-1} + i_t * \widetilde{c}_t \tag{10}$$

The * symbol in the last equation defined above is known as the dot product between matrices. Conclusively, the previously mentioned output gate o_t filters the updated state c_t, and the resulting output is determined based on the updated state and the output gate state.

$$o_t = \delta(Z_o \cdot [h_{t-1}, x_t] + a_0) \tag{11}$$

$$h_t = o_t * tanhtanh\ (c_t) \tag{12}$$

In the above Equations, (7)–(12) h_t is the hidden layer technically known as the activation of the memory cell, Z_c, Z_f, and Z_o, are suitable weight matrices, a_i, a_c, a_f, and a_0 represent the respective bias vectors, and $\sigma\ (\cdot)$ and *tanh* $(\cdot)$ are the sigmoid functions and hyperbolic tangent function, correspondingly. The schematic view of the LSTM network is presented in Figure 2.

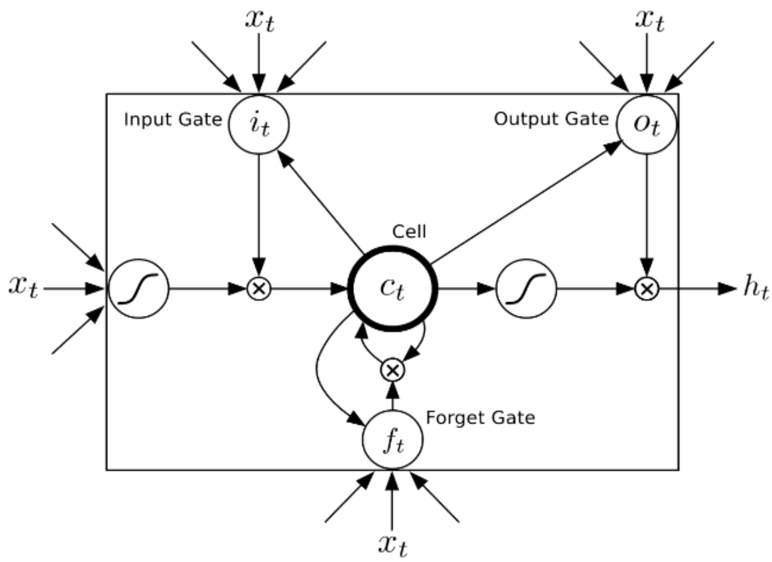

Figure 2. Flowchart of the basic LSTM network.

3.5. Proposed New Version of EMD Based on Akima Spline Interpolation Technique

The proposed new EMD algorithm is based on the Akima spline interpolation technique for clean as well as noisy signal $y(t)$ is summarized as follows.

- Extract the first oscillatory component $k^*(t)$ from the noisy signal $y(t)$ by the Akima spline interpolation technique.
- Determine the upper signal envelope $\{U\ (t)\}$ and lower envelope $\{L\ (t)\}$.
- Join all the minima, and maxima through the Akima spline interpolation technique concerning to find out the mean of both the upper and lower envelope, i.e., $M\ (t)$:

$$M(t) = \frac{U(t) + L(t)}{2} \tag{13}$$

- In order to acquire the first component, the mean envelope determined in step 3 will be subtracted from the actual signal, i.e.,

$$k(t) = y(t) - M(t) \tag{14}$$

- Repeat steps 1–4 for the component $k(t)$ until a stopping criterion is satisfied, and take the resulting $k(t)$ as $k^*(t)$.

If the remaining signal, i.e., $r(t) = y(t) - k^*(t)$ has still some oscillation components, then it can be further decomposed with the help of a new EMD. The addition of $|\delta_{i+1} + \delta_i|/2$, and$|\delta_{i-1} + \delta_{i-2}|/2$ terms forces $d_i = 0$ when $\delta_i = \delta_{i+1} = 0$, i.e., $d_i = 0$ when $v_i = v_{i+1} = v_{i+2}$, and hence it eliminates the overshoot problem when the data are constant for more than two consecutive nodes. After developing the proposed method, we will integrate it with the LSTM network to develop our new hybrid model for stock price prediction. The proposed method can be thought of as a two-stage process in which the EMD, EEMD, SEMD, and the new novel method are used in the first stage to implement the decomposition of the nonlinear and nonstationary stock price time series data into various subgroups known as IMFs, and the hybrid LSTM model has been built for prediction in the second stage. Figure 3 depicts the entire process graphically before going over each step of how to do it.

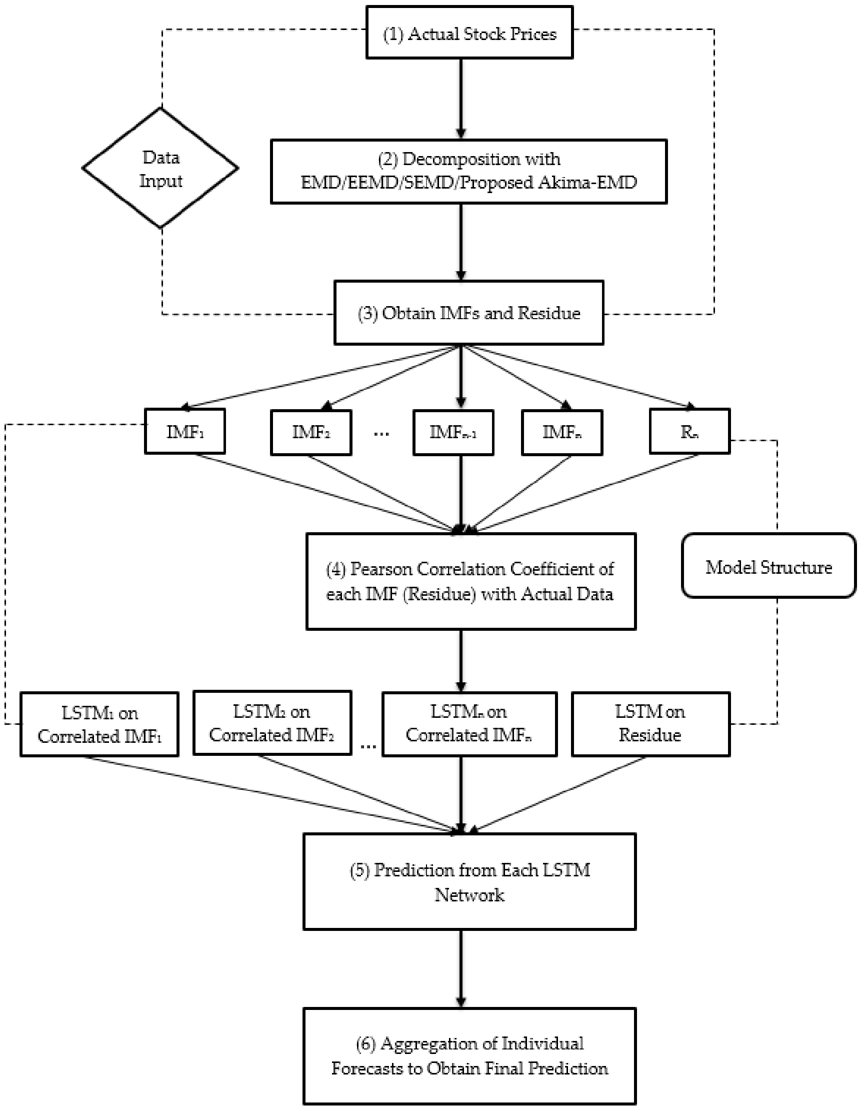

Figure 3. Flowchart of our improved hybrid model.

Step 1: Real-world stock index time series data for this study were gathered from the yahoo finance website. The original stock index time series data were then preprocessed to ensure that they met the requirements for the breakdown of the EMD and its various versions, including the proposed new one.

Step 2: The actual stock price time series data are divided into various IMFs using the EMD, EEMD, SEMD, and new Akima-EMD methods, having left only one residual.

Step 3: In this step, the Pearson coefficient of correlation between each IMF (residue) and actual stock prices are calculated. Those IMFs having approximately zero correlation with the original data are separated, and the remaining strongly correlated components with the actual data are used to build the hybrid Akima-EMD-LSTM model. The results of the Pearson coefficient of correlation and other statistical measures are presented in Section 4.2.

Step 4: In this step, we used different ensemble models such as EMD-LSTM, EEMD-LSTM, SEMD-LSTM, and the proposed Akima-EMD-LSTM for prediction. It is evident from the empirical results presented that our proposed hybrid ensemble model (Akima-EMD-LSTM) outperforms the other hybrid ensemble models in terms of minimum values of statistical metrics such as RMSE, MAE, and MAPE.

Step 6: Once the hybrid Akima-EMD-LSTM model is developed for the correlated subcomponents with the actual data, our next step is to predict the future daily closing stock prices from our proposed model.

Step 7: Finally, the final predicted values are obtained, and a comparison is made between the predicted and hold-out datasets. Furthermore, the suggested model is compared with other machine learning methods based on different performance metrics such as RMSE, MAE, and MAPE.

4. Results and Discussion

4.1. Preprocessing of the Data

In order to check the efficiency in respect of the suggested hybrid Akima-EMD-LSTM model, the daily closing price of the KSE-100 index of Pakistan Stock Exchange Limited is used. The historical data has been taken from the Yahoo Finance website (https://finance.yahoo.com) accessed on 31 August 2022. The length of the data is between 1 January 2015 and 25 August 2022. A visual representation of the data presented in Figure 4 confirms our claim that the stock prices are nonstationary, nonlinear, and volatile. The descriptive statistics of the data are also presented in Tables 1 and 2. Furthermore, the data are sliced into training and testing in such a way that 90% of the data are used for training, whereas the remaining 10% is used for testing the model. Because of numerous factors, there exist non-trading days on the stock market, data on these days are not included, and only the data for the trading days are considered for analysis.

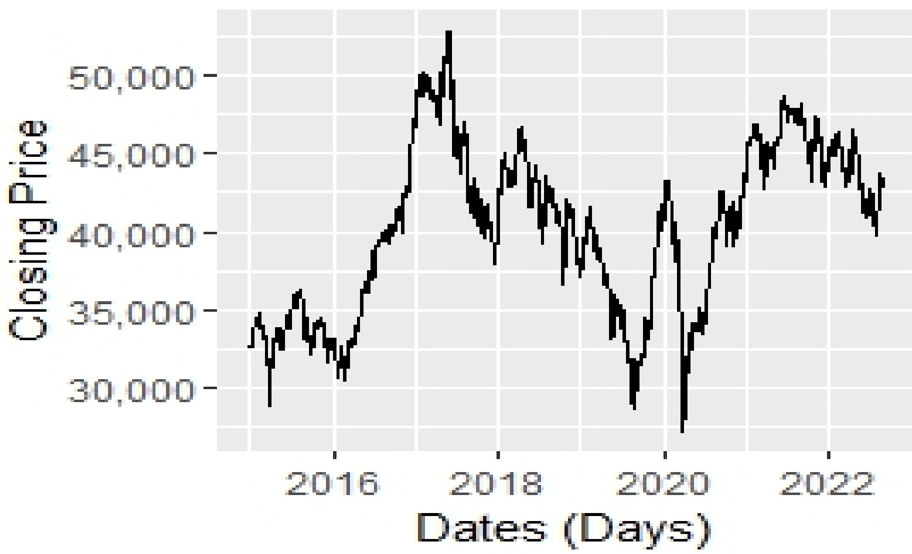

Figure 4. The daily closing price of the KSE-100 index.

Table 1. Descriptive statistics of KSE-100 index daily closing price.

Index	Count	Mean	Minimum	Maximum	St. Dev.
KSE-100	1895	40,210	27,229	52,876	5345.791

Table 2. Descriptive statistics of yearly KSE-100 index closing price.

Year	Number of Trading Days	Average Closing Price	Minimum	Maximum
2015	249	33,649	28,927	36,229
2016	248	37,617	30,565	47,807
2017	249	45,621	37,919	52,876
2018	246	42,153	36,663	46,638
2019	247	36,064	28,765	41,769
2020	251	38,311	27,229	43,767
2021	247	45,959	42,780	48,726
2022	158	43,577	39,832	46,602

It can be seen from Table 1 that the minimum value of the index is 27,229, that is recorded on 25 March 2020, whereas the maximum value was 52,876, recorded on 24 May 2017. The minimum value is because of the COVID-19 pandemic that affected the majority of the well-known stock indexes across the world [57,58]. However, the KSE-100 index of the Pakistan stock exchange recovered and soon gained potential because of the government's bold decision to keep open the business sector across the country [59].

The Pakistan stock market showed strong resistance during the period of the pandemic, which might be the government's strategies to control the outbreak while keeping an eagle eye on the positivity rate and implementing smart lockdowns in specific areas of the country to avoid large lock-downs that crippled the economy of the most developed countries. The mean value that is observed during the seven years is 40,210, whereas the largest value of 5345.791 of the standard deviation verifies that the nature of the stock market is very chaotic.

Furthermore, Table 2 shows a yearly descriptive summary, such as the average closing price, minimum and maximum closing price, and the number of trading days of the KSE-100 index. It is evident from Table 2 that the maximum average closing price from 2015 to 2022 is 45,959 recorded in the year 2021. The minimum closing price in these years is 33,649, recorded in the year 2015. It is worth mentioning here that the average closing price during 2020 was 38,311, which is more than the year 2015, signifying our claim that the KSE-100 index showed more resistance to the COVID-19 pandemic.

4.2. Decomposition with EMD, EEMD, SEMD, and Akima-EMD

The actual closing prices of the KSE-100 index are decomposed into numerous IMFs and monotone residues by implementing EMD, EEMD, SEMD, and the proposed new method. It can be seen from Figure 5 that the IMFs extracted by implementing the procedure of EMD, EEMD, and the proposed new method are the same, whereas the number of IMFs produced by the SEMD method is less than the benchmark EMD, EEMD, and the new method. The first few IMFs represent high-frequency components in the actual data and have a very low Pearson coefficient of correlation with the original stock prices. Therefore these high-frequency components are not considered to be modeled with the help of the LSTM network, as these sub-parts represent only noise in the stock price data.

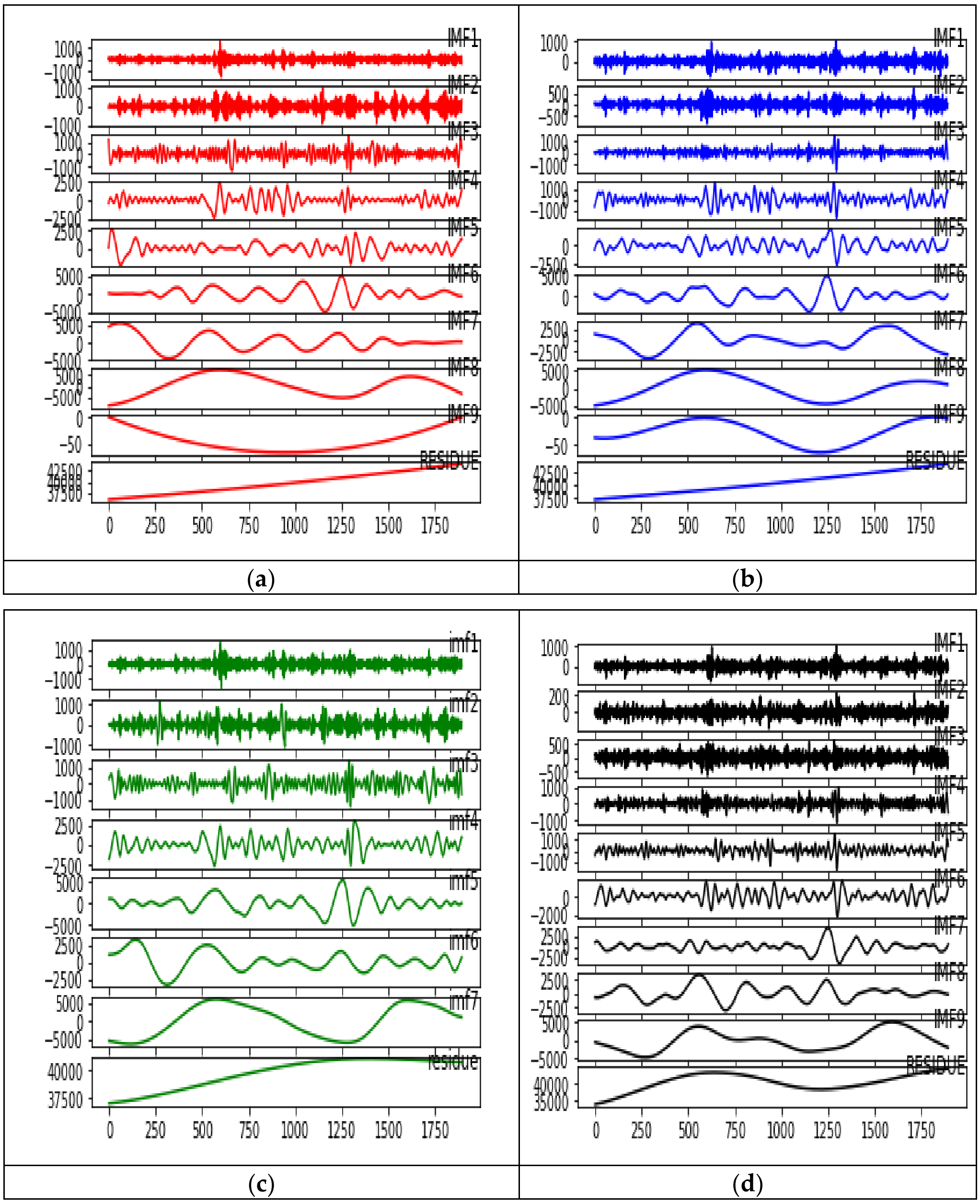

Figure 5. Decomposition results of KSE-100 index with (**a**) EMD, (**b**) EEMD, and (**c**) SEMD (**d**) Akima-EMD.

As a first step, the KSE-100 index closing prices are decomposed with the help of EMD, EEMD, SEMD, and the proposed new method. After decomposing the data, the Pearson correlation coefficient is used to separate the correlated IMF and residue from the actual stock data. To normalize the correlated IMFs in the range of '0' and '1', the well-known MinMaxScaler is used. This method of normalizing the data subtracts the minimum value from the actual data and then divides it by its range. The key characteristic of MinMaxScaler is that it preserves the shape of the original time series data. It does not meaningfully change the information embedded in the original data. However, in case of any outliers in the data, this method is not recommended, and in such scenarios, the RobustScaler method for normalization is preferred. Following is the mathematical structure of MinMaxScaler.

$$\overset{\prime}{y}(t) = \frac{y(t) - min(y(t))}{y(t) - y(t)} \tag{15}$$

where $\acute{y}(t)$ denotes the corresponding normalized IMF and monotone residue that we used to build our proposed hybrid LSTM network.

4.3. IMF's Statistics

The variance and its proportion with each IMF have been calculated, as well as the Pearson coefficient of correlation between each IMF (residue) and the real stock prices. Tables 3–6 display the relevant details regarding this decomposition. The correlations between each IMF (residue) and the observed stock prices are simultaneously measured using the Pearson product-moment correlation coefficient. We utilized the percentage of variance to illustrate how much each IMF (residue) contributed to the overall movement of the observed stock price because these IMFs (residue) are independent of one another. The variance of these IMFs and residue, however, does not necessarily equal the overall variance due to various constraints, as can be seen, for instance, from Tables 3–6 that there is a 23.262% (123.258–99.996%), −27.75% (72.495–99.99%), 2.257% (102.260–100.003%), and −17.35(82.638–99.988) differences when the actual time series data are decomposed with EMD, EEMD, SEMD, and the new method. It can also be verified from Table 3 that when the decomposition is carried out with EMD, the dominant mode is the deterministic trend in terms of IMF8 with a Pearson coefficient of correlation of 0.758. We also found out that the variance of this deterministic trend accounted for 71.883% of the total variance. The sum of the variations for these significant components, namely IMF5, IMF6, IMF7, IMF8, IMF9, and trend, accounts for 97.8491% of the total variation. Contrarily, IMF1, IMF2, IMF3, and IMF4 have very low correlation coefficients with the original stock price and only contribute 2.147 of the total variance, indicating that they have little influence on stock price and are therefore not included in the development of the proposed Akima-EMD-LSTM model.

Table 3. Statistical measures of IMFs and residue extracted by implementing EMD.

Description	Pearson Correlation	Variance	Variance as % of the Actual	Variance as % of ($\sum$*IMFs+Residue*)
Actual	-	285,774,84	-	-
IMF1	0.071	54,819.13	0.191	0.155
IMF2	0.053	69,825.12	0.244	0.198
IMF3	0.088	15,5121.3	0.542	0.440
IMF4	0.060	477,278.4	1.670	1.354
IMF5	0.122	537,974	1.882	1.527
IMF6	0.331	2,803,518	9.810	7.958
IMF7	0.102	5,288,340	18.505	15.012
IMF8	0.758	20,542,369	71.883	58.315
IMF9	−0.111	389.3048	0.0013	0.0011
Residue	0.399	5,296,823	18.53	15.036
Total		35,226,457	123.2583	99.9961

Furthermore, it can be verified from the results presented in Table 4 that the dominant modes (based on the values of the Pearson correlation coefficient) when decomposition is carried out with the help of EEMD are IMF3, IMF4, IMF5, IMF6, IMF7, IMF8, and IMF9 with moderate to high Pearson coefficient of correlation with the actual data. Similarly, the sum of the variances of IMF1, IMF2, and residue accounts for only 14.686% of the total variation (72.4955%) as well as a very low correlation with the actual stock prices, which shows that these components have very little impact on the variation in the stock prices, and shows only noise in the actual time series data.

Table 4. Statistical measures of IMFs and residue extracted by implementing EEMD.

	Pearson Correlation	**Variance**	**Variance as % of the Actual**	**Variance as % of ($\sum$*IMFs+Residue*)**
Actual	-	28,577,484	-	-
IMF1	0.049	31,847.84	0.111	0.153
IMF2	0.072	33,497.54	0.117	0.161
IMF3	0.103	67,152.56	0.234	0.324
IMF4	0.139	219,347.2	0.767	1.058
IMF5	0.374	467,575.5	1.636	2.256
IMF6	0.763	2,118,902	7.414	10.226
IMF7	0.749	3,713,021	12.992	17.921
IMF8	0.639	9,935,052	34.765	47.952
IMF9	−0.267	445.702	0.0015	0.00215
Residue	0.0977	4,131,881	14.458	19.942
Total		20,718,722	72.4955	99.995

Table 5. Statistical measures of IMFs and residue extracted by implementing SEMD.

	Pearson Correlation	**Variance**	**Variance as % of the Actual**	**Variance as % of ($\sum$*IMFs+Residue*)**
Actual	-	25,829,097	-	-
IMF1	0.0720	67,535.05	0.261	0.255
IMF2	0.090	90,198.41	0.349	0.349
IMF3	0.082	146,130.4	0.565	0.553
IMF4	0.0733	746,324.3	2.88	2.825
IMF5	0.397	3,162,753	12.244	11.973
IMF6	0.227	1,907,672	7.385	7.22
IMF7	0.859	18,470,085	71.508	69.921
Residue	0.360	1,824,661	7.064	6.907
Total		26,415,359	102.26	100.003

When the decomposition is conducted with the proposed new method, namely the Akima-EMD, we obtain the statistical measures as presented in Table 6. The dominant modes are now IMF5, IMF6, IMF7, IMF8, IMF9, and residue that contributes to 82.214% of the total variation with Pearson correlation coefficients of 0.139, 0.204, 0.495, 0.641, 0.688, and 0.523. Similarly, the sum of the variances of other components, in contrast to the dominant modes that include IMF1, IMF2, IMF3, and IMF4, accounts for only 0.2224% of the total variation (82.638%) as well as having nearly zero correlation with the actual stock prices, and therefore, has no serious impact on bringing changes in the KSE-100 index closing price. To summarize, the results presented in Tables 3–6 indicate that the low correlated components with the actual KSE-100 index closing price are considered as noise in the signal and kept aside before building our proposed hybrid Akima-LSTM model. The remaining components that have moderate to high correlation with the actual data are used to build our proposed hybrid model, and the tuning parameters of each correlated sub-components to build the LSTM network are outlined in the following sub-section.

Table 6. Statistical measures of IMFs and residue extracted by implementing a new method.

	Pearson Correlation	Variance	Variance as % of the Actual	Variance as % of $(\sum IMFs+Residue)$
Actual	-	28,577,484	-	-
IMF1	0.0492	31,804.79	0.111	0.191
IMF2	0.0340	2337.712	0.0081	0.014
IMF3	0.0510	28,818.89	0.100	0.173
IMF4	0.0882	58,590.39	0.205	0.351
IMF5	0.139	1,187,075.1	4.153	0.713
IMF6	0.204	3,232,701.8	11.312	1.99
IMF7	0.495	1,275,910	4.464	7.663
IMF8	0.641	4,527,298	15.842	9.173
IMF9	0.688	7,556,977	26.443	45.390
Residue	0.523	5,715,661	20.00	34.330
Total		16,648,875	82.638	99.988

4.4. Training Phase of the Models and Prediction Results

To achieve efficient forecasts from the sub-series (IMFs), as well as the monotone residue that is correlated with the actual data, the optimal hyper-parameters values, i.e., the number of epochs, the batch size, number of layers along with units in each hidden layer, and dropout are presented in Tables 7–10. Because of the different patterns of each IMF and residue, the values of these hyper parameters are different from each other.

Table 7. Hyper-parameters of EMD-LSTM model.

	Epochs	Batch Size	Dropout	Hidden Units	Hidden Layers
IMF5	200	32	0.2	50	04
IMF6	150	32	0.2	50	04
IMF7	100	32	0.2	50	04
IMF8	100	16	0.2	50	04
IMF9	150	16	0.2	50	04
Residue	100	16	0.2	50	04

Table 8. Hyper-parameters of the EEMD-LSTM model.

	Epochs	Batch Size	Dropout	Hidden Units	Hidden Layers
IMF3	300	64	0.2	50	04
IMF4	300	64	0.2	50	04
IMF5	250	32	0.2	50	04
IMF6	250	32	0.2	50	04
IMF7	150	32	0.2	50	04
IMF8	150	32	0.2	50	04
IMF9	100	32	0.2	50	04

Table 9. Hyper-parameters of SEMD-LSTM model.

	Epochs	Batch Size	Dropout	Hidden Units	Hidden Layers
IMF5	250	32	0.2	50	04
IMF6	250	32	0.2	50	04
IMF7	150	32	0.2	50	04
Residue	150	32	0.2	50	04

Table 10. Hyper-parameters of the Akima-EMD-LSTM model.

	Epochs	Batch Size	Dropout	Hidden Units	Hidden Layers
IMF5	250	64	0.2	50	04
IMF6	250	64	0.2	50	04
IMF7	200	32	0.2	50	04
IMF8	250	32	0.2	50	04
IMF9	200	32	0.2	50	04
Residue	100	16	0.2	50	04

The tuning phase of the hyper-parameters of the LSTM network is of extreme importance, and any wrong selection may distort the prediction accuracy. The number of epochs and batch size that play a significant role in the training of our model varies from high to low-frequency IMFs and residue. It can be seen from the results presented in the above Tables that from high to low-frequency IMFs and single monotone residue components, the number of epochs and batch size reduced systematically as it is easy to train the model for such non-complex subcomponents that have a smooth trend in an upward direction. The low-frequency IMFs and residue component shows the long-term oscillation in the data, and modeling such components with the help of a hybrid LSTM model is not a difficult task with a lower number of epochs and batch size. The reason is that the network learns very easily from the pattern of these components that are not chaotic.

4.5. Prediction Results

The prediction results of the sub-series for the KSE-100 index closing stock price are presented in Figures 6–9. The prediction accuracy of the low correlated IMFs with the actual stock price is relatively low due to the high amplitude and high frequency of the components, whereas for the high correlated IMFs, including the single monotone residue that represents the long-term trend of the index data the predicted value is close to the actual value.

According to Figures 6–9, there is a small difference between the predictive and actual values of the sub-components of the KSE-100 index closing price obtained by implementing the hybrid Akima-EMD-LSTM. The proposed new version of EMD has the potential to solve the problem of overshoot and undershoot, and wiggles at both ends using the benchmark EMD, and hence the prediction accuracy improves the prediction accuracy. It can also be seen from Table 8 that the forecast accuracy in terms of RMSE, MAE, and MAPE of our proposed hybrid Akima-EMD-LSTM model is maximum than the hybrid EMD-LSTM, EEMD-LSTM, SEMD-LSTM model and therefore used for the final prediction of the KSE-100 index closing price. The actual prediction results of the KSE-100 index are visualized in Figure 10 below, which shows that our proposed hybrid model not only predicted the stock price values precisely but also predicted its up, and down pattern in a more accurate way than the usual classical time series or any other machine learning methods without hybridization.

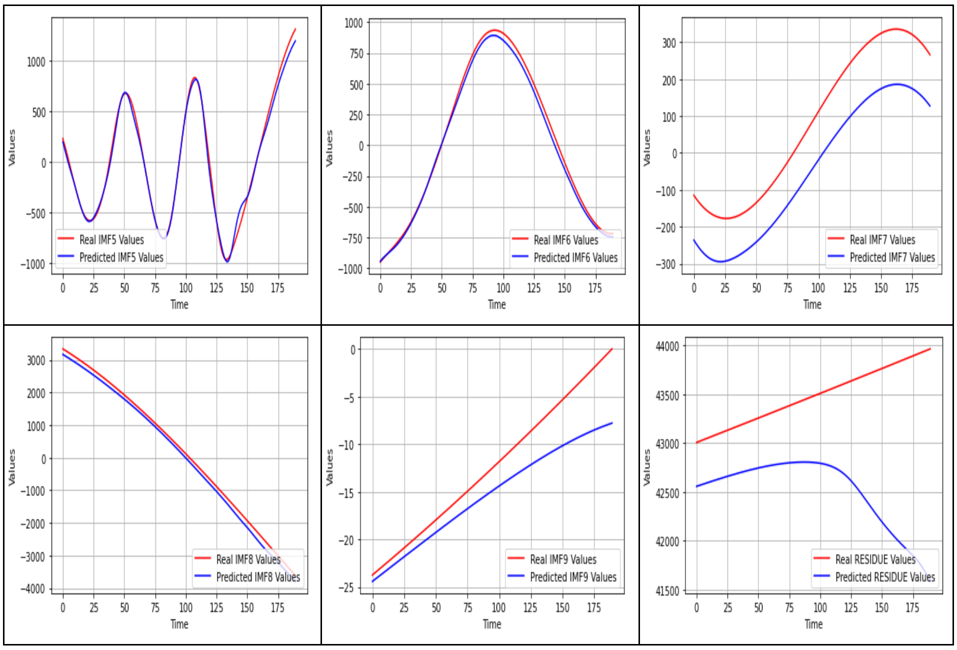

Figure 6. Predicted results of correlated subcomponents obtained from the EMD-LSTM model.

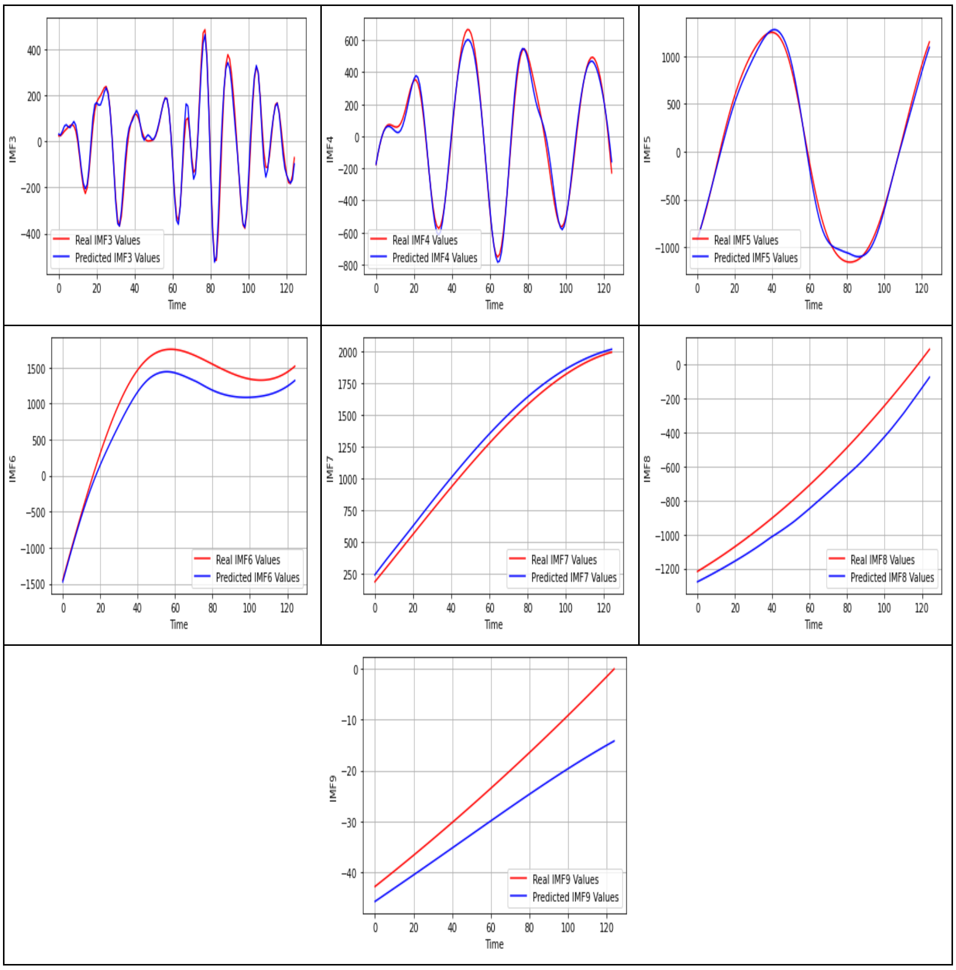

Figure 7. Predicted results of correlated subcomponents obtained from the EEMD-LSTM model.

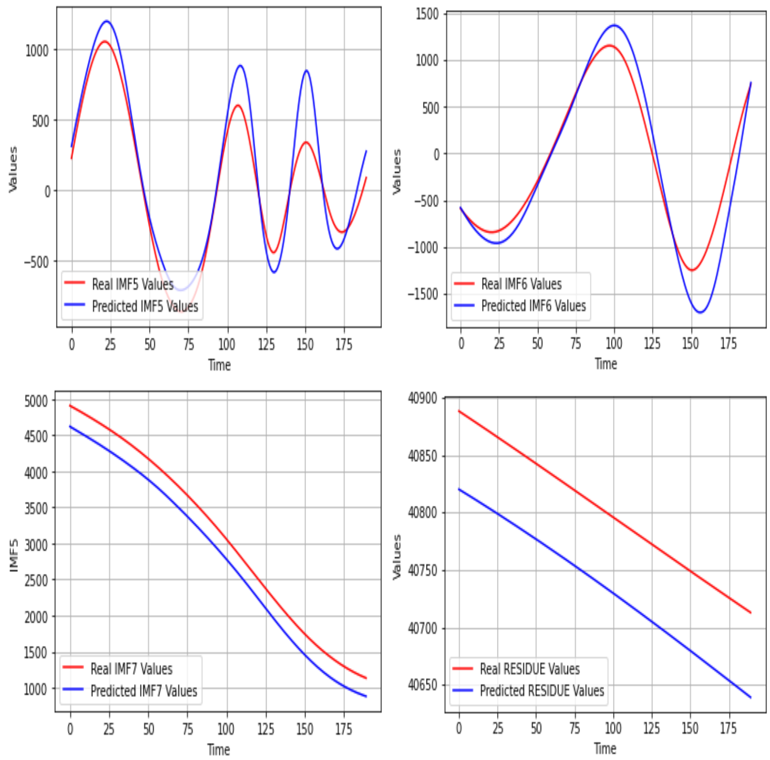

Figure 8. Predicted results of correlated subcomponents obtained from SEMD-LSTM model.

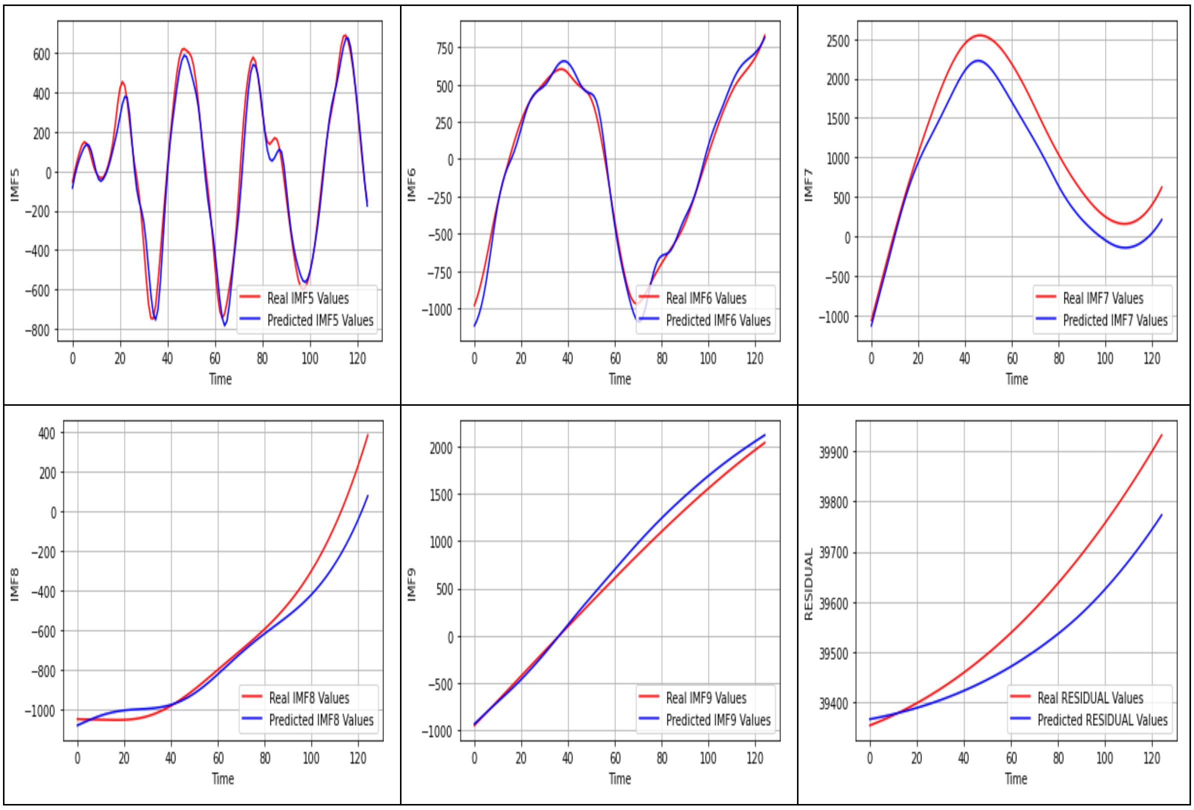

Figure 9. Predicted results of correlated subcomponents obtained from the proposed Akima-EMD-LSTM model.

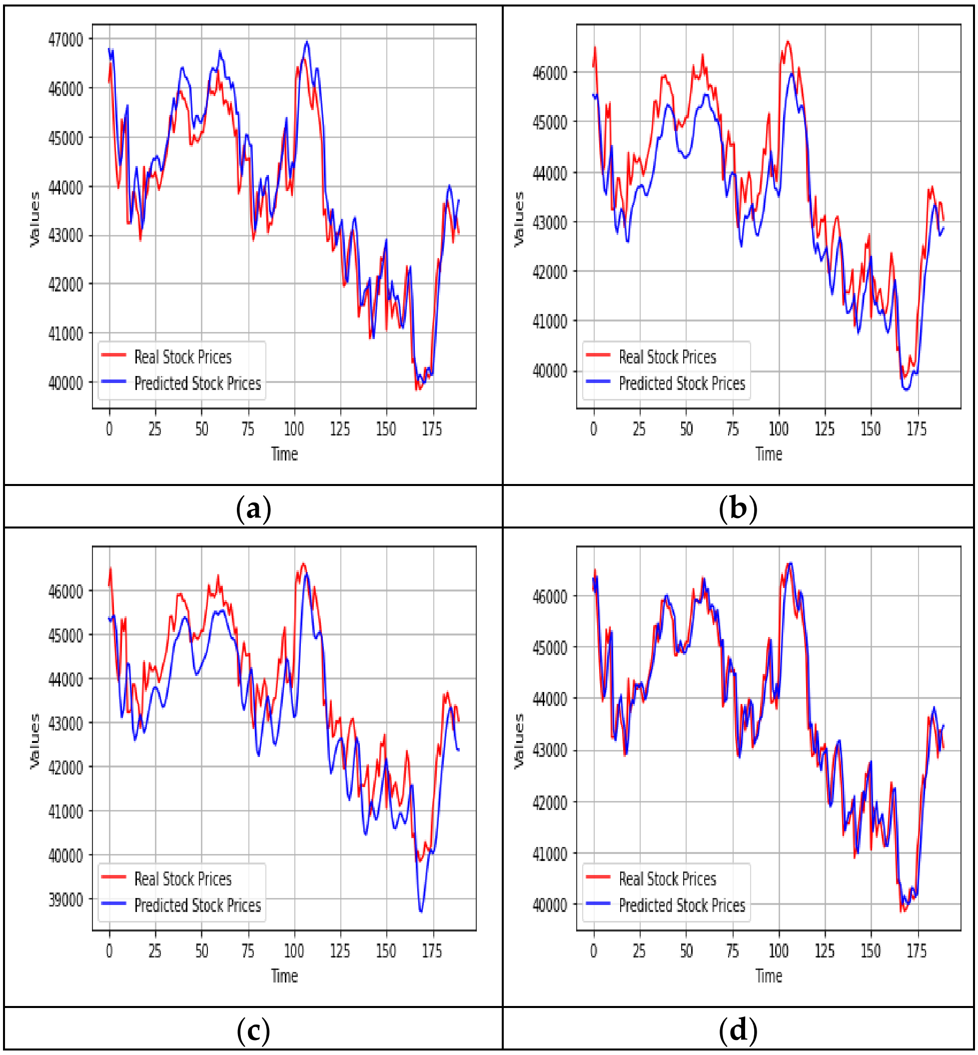

Figure 10. Prediction results of (**a**) EMD-LSTM (**b**) EEMD-LSTM (**c**) SEMD-LSTM (**d**) Proposed Akima-EMD method.

5. Comparison of the Proposed Model

To check the prediction accuracy and resilience of the proposed models, three different statistical metrics, i.e., RMSE, MAE, and MAPE, are utilized. The mathematical expressions of these performance metrics are given as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(A_t - P_t)^2} \tag{16}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|A_t - P_t| \tag{17}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - P_t}{A_t}\right| \tag{18}$$

whereas A_t, and P_t denote the actual and predicted values, and N is the total number of trading days used in the testing data set. The smaller values of these three statistical metrics signify that there is less variation between the actual and predicted values. Furthermore, among these three statistical measures, it is easy to interpret and understand the MAPE. Hence, it is used as a key evaluating metric to check the prediction performance of our proposed model. In this research article, we used EMD, EEMD, and SEMD, proposed a new method to decompose the actual signal, and then built our model for the correlated subcomponents using the LSTM network. We calculated these three main statistical measures to compare all four hybrid models and select the best prediction model in terms of minimum values of RMSE, MAE, and MAPE. It can be seen from the investigational results presented in Table 11 that the best model to predict the KSE-100 index closing price is our proposed hybrid Akima-EMD-LSTM model having minimum values of RMSE, MAE, and MAPE as compared to the other hybrids models. After selecting the best model among the four proposed hybrid models, our next goal is to compare it with other models. The following subsection will describe the details of such a comparison.

Table 11. Performance metrics of different hybrid models.

Model	RMSE	MAE	MAPE
Hybrid EMD-LSTM	490.664	383.899	0.941
Hybrid EEMD-LSTM	567.813	491.061	0.79
Hybrid Akima-EMD-LSTM	299.541	234.34	0.578
Hybrid SEMD-LSTM	987.98	756.985	2.328

It is evident from the empirical results presented in Table 11 that the most efficient prediction model among the four hybrid models is our proposed Akima-EMD-LSTM. Now the performance of the proposed new model is compared with other models, including the single LSTM, EMD-SVM, EEMD-SVM, SEMD-SVM, Akima-EMD-SVM, EMD-Random Forest, EEMD-Random Forest, SEMD-Random Forest, Akima-EMD-Random Forest, EMD-Decision Tree, EEMD-decision tree, SEMD-decision tree, and Akima-EMD-decision tree. The deep learning LSTM network does not require any input features, and the actual stock price data can be used directly to build the single LSTM network, whereas the other ensemble models use the correlated subcomponents (IMFs), and the residue obtained by decomposing it with the help EMD, EEMD, SEMD and proposed new model.

The prediction performance of the proposed model is similarly effective in all scenarios, i.e., for each IMF and residue obtained from the KSE-100 index actual closing price, which is a key strength of this research. The investigational results shown in Table 12 make it simple to confirm that the three statistical measures of the suggested model are minimal; for example, the value of RMSE is 299.5416, which is minimal among other models. Akima-EMD-SVM, which has an RMSE of 309.345, is the second-best model in this competition, and the same SVM model with the EEMD ensemble, which has an RMSE of 322.675, is in third place. Similarly, the MAE and MAPE values of our proposed hybrid model are 234.340 and 0.578, respectively, which are also minimum as compared to the other models. Interestingly, the ensemble Akima-EMD-SVM model stands second in comparison with the proposed model, but the performance of this model is much better than the other ensemble random forest and decision tree models.

A snapshot will more accurately depict the scope of this research work rather than presenting the actual and forecasted values, which are not presented here. The actual and predicted values are schematically shown in Figures 6–10 to make the prediction simple and comprehensible for each subseries and the actual closing KSE-100 price. Yet, there is a slight difference between the actual and predicted values, but the direction accuracy of our prediction is more than any other model.

In a nutshell, the hybrid model that is a combination of LSTM and the proposed new version of EMD, shortly written as Akima-EMD, consistently achieves the highest accuracy in terms of RMSE, MAE, and MAPE, with less variation between the predicted and actual KSE-100 index closing price.

Table 12. Performance metrics of different hybrid models.

Model	RMSE	MAE	MAPE
LSTM	473.319	354.75	0.87
EMD-SVM	317.504	246.675	0.711
EEMD-SVM	322.675	280.937	0.65
SEMD-SVM	324.494	285.268	0.724
Akima-EMD-SVM	309.345	242.879	0.726
EMD-Random Forest	406.554	289.891	0.748
EEMD-Random Forest	389.023	253.851	0.65
SEMD-Random Forest	521.028	338.445	1.611
Akima-EMD-Random Forest	367.987	256.987	0.712
EMD-Decision Tree	480.217	352.165	0.896
EEMD-Decision Tree	742.619	411.442	1.067
SEMD-Decision Tree	630.078	377.376	0.973
Akima-EMD-Decision Tree	443.91	345.1	0.709
Hybrid Akima-EMD-LSTM	299.541	234.34	0.578

6. Conclusions

For data scientists and practitioners, predicting the stock market indices is a fascinating and difficult area of research. Precise forecasting of the stock market is of prime importance as it will help not only the investors to maximize their profits, but the governments as well because the stock markets are the backbone of every country's economy. This section summarizes and presents the major findings of this research study. In the previous literature, the researchers employed a single time series or machine-learning model to predict the accurate values of the stock prices and hence, were criticized for their low prediction ability. The primary goal of this study is to suggest a unique approach for effectively and correctly forecasting the daily closing prices of the KSE-100 index. The fact that we presented a hybrid technique that combines the best features of both the Akima-EMD and the LSTM network. As a result, the proposed approach is extremely effective for prediction with nonlinear and nonstationary data. The suggested model is not an ensemble model since we used subcomponents after decomposing the data into various IMFs and a single monotone residue using the Akima-EMD approach. The Pearson coefficient of correlation was utilized to find out which IMFs (residue) is correlated with the actual data and then construct the proposed hybrid model. Comparison is made with other ensemble models that are based on other machine learning techniques such as SVM, random forest, and decision tree. The findings of four statistical measures, namely RMSE, MAE, and MAPE, show that the suggested model surpassed the other models, implying that it is a useful addition to stock market prediction. Interestingly, the hybrid ensemble SVM model stands second in this competition and outperformed the individual LSTM, random ensemble forest, and decision tree models even though the single LSTM model, which is a type of RNN, performs better than all the ensemble machine learning models except the proposed one.

In future work, the proposed method can be used for other countries' stock markets data, and comparisons could be made with XGBoost, SVM, and ANN. Whereas the SVM and ANN models can be constructed using different variables such as investor sentiments, interest rates, the political climate of the country, news, and exchange rates as inputs. Furthermore, we will also find out the prediction accuracy of our proposed model on other nonlinear and non-stationary time series, such as exchange rates, crude oil prices, wind speed, temperature, rainfall, earthquakes, and tourist arrival. However, for the current research work, we proved that the proposed hybrid Akima-EMD-LSTM model performs better in predicting the daily closing price of the KSE-100 index.

Author Contributions: Conceptualization, D.M.K. and M.A.; methodology, D.M.K. and H.M.A.; software, M.A. and D.M.K.; validation, H.M.A. and A.A.-A.H.E.-B.; formal analysis, M.A. and D.M.K.; investigation, H.M.A. and A.A.-A.H.E.-B.; resources, M.A. and D.M.K.; data curation, D.M.K.; writing—original draft preparation, M.A.; writing—review and editing, M.A. and D.M.K.;

visualization, H.M.A. and A.A.-A.H.E.-B.; supervision, D.M.K.; project administration, D.M.K. and H.M.A.; funding acquisition, H.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R299), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Petram, L.O. The World's First Stock Exchange: How the Amsterdam Market for Dutch East India Company Shares Became a Modern Securities Market, 1602–1700. Ph.D. Thesis, Universiteit van Amsterdam, Amsterdam, The Netherlands, 2011.
2. Fama, E.F. Market Efficiency, Long-Term Returns, and Behavioral Finance. *J. Financ. Econ.* **1998**, *49*, 283–306. [CrossRef]
3. Wang, B.; Huang, H.; Wang, X. A Novel Text Mining Approach to Financial Time Series Forecasting. *Neurocomputing* **2012**, *83*, 136–145. [CrossRef]
4. De Gooijer, J.G.; Hyndman, R.J. 25 Years of IIF Time Series Forecasting: A Selective Review. Tinbergen Institute Discussion Papers. *SSRN Electron. J.* **2005**, *22*, 5–68.
5. Neild, S.A.; McFadden, P.D.; Williams, M.S. A Review of Time-Frequency Methods for Structural Vibration Analysis. *Eng. Struct.* **2003**, *25*, 713–728. [CrossRef]
6. Krunz, M.M.; Makowski, A.M. Modeling Video Traffic Using M/G/∞ Input Processes: A Compromise between Markovian and LRD Models. *IEEE J. Sel. Areas Commun.* **1998**, *16*, 733–748. [CrossRef]
7. Farina, L.; Rinaldi, S. *Positive Linea Systems: Theory and Applications*; Wiley: New York, NY, USA, 2000.
8. Contreras, J.; Espinola, R.; Nogales, F.J.; Conejo, A.J. ARIMA Models to Predict Next-Day Electricity Prices. *IEEE Power Eng. Rev.* **2002**, *22*, 57. [CrossRef]
9. Smola, A.J.; Schölkopf, B. A Tutorial on Support Vector Regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
10. Cavalcante, R.C.; Brasileiro, R.C.; Souza, V.L.F.; Nobrega, J.P.; Oliveira, A.L.I. Computational Intelligence and Financial Markets: A Survey and Future Directions. *Expert Syst. Appl.* **2016**, *55*, 194–211. [CrossRef]
11. Haq, M.A.; Jilani, A.K.; Prabu, P. Deep learning based modeling of groundwater storage change. *Comput. Mater. Contin.* **2022**, *70*, 4599–4617.
12. Haq, M.A.; Rahaman, G.; Baral, P.; Ghosh, A. Deep learning based supervised image classification using UAV images for forest areas classification. *J. Indian Soc. Remote Sens.* **2021**, *49*, 601–606. [CrossRef]
13. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
14. Daubechies, I.; Lu, J.; Wu, H.-T. Synchrosqueezed Wavelet Transforms: An Empirical Mode Decomposition-like Tool. *Appl. Comput. Harmon. Anal.* **2011**, *30*, 243–261. [CrossRef]
15. Wang, Q.; Xu, W.; Huang, X.; Yang, K. Enhancing Intraday Stock Price Manipulation Detection by Leveraging Recurrent Neural Networks with Ensemble Learning. *Neurocomputing* **2019**, *347*, 46–58. [CrossRef]
16. Berradi, Z.; Lazaar, M. Integration of Principal Component Analysis and Recurrent Neural Network to Forecast the Stock Price of Casablanca Stock Exchange. *Procedia Comput. Sci.* **2019**, *148*, 55–61. [CrossRef]
17. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
18. Sundermeyer, M.; Ney, H.; Schluter, R. From Feedforward to Recurrent LSTM Neural Networks for Language Modeling. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 517–529. [CrossRef]
19. Mahmoudi, N.; Docherty, P.; Moscato, P. Deep Neural Networks Understand Investors Better. *Decis. Support Syst.* **2018**, *112*, 23–34. [CrossRef]
20. Kraus, M.; Feuerriegel, S. Decision Support from Financial Disclosures with Deep Neural Networks and Transfer Learning. *Decis. Support Syst.* **2017**, *104*, 38–48. [CrossRef]
21. Deep Learning for Decision Making and the Optimization of Socially Responsible Investments and Portfolio. *Decis. Support Syst.* **2019**, *124*, 113097. [CrossRef]
22. Wang, K.; Qi, X.; Liu, H. Photovoltaic Power Forecasting Based LSTM-Convolutional Network. *Energy* **2019**, *189*, 116225. [CrossRef]
23. Karevan, Z.; Suykens, J.A.K. Transductive LSTM for Time-Series Prediction: An Application to Weather Forecasting. *Neural Netw.* **2020**, *125*, 1–9. [CrossRef]
24. Ding, G.; Qin, L. Study on the Prediction of Stock Price Based on the Associated Network Model of LSTM. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 1307–1317. [CrossRef]

25. Chen, S.; Ge, L. Exploring the Attention Mechanism in LSTM-Based Hong Kong Stock Price Movement Prediction. *Quant. Finance* **2019**, *19*, 1507–1515. [CrossRef]
26. Baek, Y.; Kim, H.Y. ModAugNet: A New Forecasting Framework for Stock Market Index Value with an Overfitting Prevention LSTM Module and a Prediction LSTM Module. *Expert Syst. Appl.* **2018**, *113*, 457–480. [CrossRef]
27. Liang, X.; Ge, Z.; Sun, L.; He, M.; Chen, H. LSTM with Wavelet Transform Based Data Preprocessing for Stock Price Prediction. *Math. Probl. Eng.* **2019**. [CrossRef]
28. Xu, P.; Xie, F.; Su, T.; Wan, Z.; Zhou, Z.; Xin, X.; Guan, Z. Automatic Evaluation of Facial Nerve Paralysis by Dual-Path LSTM with Deep Differentiated Network. *Neurocomputing* **2020**, *388*, 70–77. [CrossRef]
29. Elsheikh, A.H.; Shanmugan, S.; Sathyamurthy, R.; Thakur, A.K.; Issa, M.; Panchal, H.; Muthuramalingam, T.; Kumar, R.; Sharifpur, M. Low-cost bilayered structure for improving the performance of solar stills: Performance/cost analysis and water yield prediction using machine learning. *Sustain. Energy Technol. Assess.* **2022**, *49*, 101783. [CrossRef]
30. Elsheikh, A.H.; Panchal, H.; Ahmadein, M.; Mosleh, A.O.; Sadasivuni, K.K.; Alsaleh, N.A. Productivity forecasting of solar distiller integrated with evacuated tubes and external condenser using artificial intelligence model and moth-flame optimizer. *Case Stud. Therm. Eng.* **2021**, *28*, 101671. [CrossRef]
31. Moustafa, E.B.; Hammad, A.H.; Elsheikh, A.H. A new optimized artificial neural network model to predict thermal efficiency and water yield of tubular solar still. *Case Stud. Therm. Eng.* **2022**, *30*, 101750. [CrossRef]
32. Elsheikh, A. Bistable Morphing Composites for Energy-Harvesting Applications. *Polymers* **2022**, *14*, 1893. [CrossRef]
33. Elsheikh, A.H.; Abd Elaziz, M.; Vendan, A. Modeling ultrasonic welding of polymers using an optimized artificial intelligence model using a gradient-based optimizer. *Weld. World* **2022**, *66*, 27–44. [CrossRef]
34. Elsheikh, A.H.; Shehabeldeen, T.A.; Zhou, J.; Showaib, E.; Abd Elaziz, M. Prediction of laser cutting parameters for polymethylmethacrylate sheets using random vector functional link network integrated with equilibrium optimizer. *J. Intell. Manuf.* **2021**, *32*, 1377–1388. [CrossRef]
35. Moustafa, E.B.; Elsheikh, A. Predicting Characteristics of Dissimilar Laser Welded Polymeric Joints Using a Multi-Layer Perceptrons Model Coupled with Archimedes Optimizer. *Polymers* **2023**, *15*, 233. [CrossRef] [PubMed]
36. Elsheikh, A.H.; Abd Elaziz, M.; Das, S.R.; Muthuramalingam, T.; Lu, S. A new optimized predictive model based on political optimizer for eco-friendly MQL-turning of AISI 4340 alloy with nano-lubricants. *J. Manuf. Process.* **2021**, *67*, 562–578. [CrossRef]
37. Khoshaim, A.B.; Moustafa, E.B.; Bafakeeh, O.T.; Elsheikh, A.H. An optimized multilayer perceptrons model using grey wolf optimizer to predict mechanical and microstructural properties of friction stir processed aluminum alloy reinforced by nanoparticles. *Coatings* **2021**, *11*, 1476. [CrossRef]
38. Kim, H.Y.; Won, C.H. Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-Type Models. *Expert Syst. Appl.* **2018**, *103*, 25–37. [CrossRef]
39. Cao, J.; Li, Z.; Li, J. Financial Time Series Forecasting Model Based on CEEMDAN and LSTM. *Phys. A Stat. Mech. Its Appl.* **2019**, *519*, 127–139. [CrossRef]
40. Jun, W.; Lingyu, T.; Yuyan, L.; Peng, G. A Weighted EMD-Based Prediction Model Based on TOPSIS and Feed Forward Neural Network for Noised Time Series. *Knowl.-Based Syst.* **2017**, *132*, 167–178. [CrossRef]
41. Chen, L.; Chi, Y.; Guan, Y.; Fan, J. A hybrid attention-based EMD-LSTM model for financial time series prediction. In *2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*; IEEE: Piscatvie, NJ, USA, 2019; pp. 113–118.
42. Jin, Z.; Yang, Y.; Liu, Y. Stock Closing Price Prediction Based on Sentiment Analysis and LSTM. *Neural Comput. Appl.* **2020**, *32*, 9713–9729. [CrossRef]
43. Li, C.; Qian, G. Stock Price Prediction Using a Frequency Decomposition Based GRU Transformer Neural Network. *Appl. Sci.* **2023**, *13*, 222. [CrossRef]
44. Aldhyani, T.H.; Alkahtani, H. Cyber Security for Detecting Distributed Denial of Service Attacks in Agriculture 4.0: Deep Learning Model. *Mathematics* **2023**, *11*, 233. [CrossRef]
45. Liu, Y.; Liu, X.; Zhang, Y.; Li, S. CEGH: A Hybrid Model Using CEEMD, Entropy, GRU, and History Attention for Intraday Stock Market Forecasting. *Entropy* **2023**, *25*, 71. [CrossRef]
46. Lazcano, A.; Herrera, P.J.; Monge, M. A Combined Model Based on Recurrent Neural Networks and Graph Convolutional Networks for Financial Time Series Forecasting. *Mathematics* **2023**, *11*, 224. [CrossRef]
47. Fang, T.; Zheng, C.; Wang, D. Forecasting the crude oil prices with an EMD-ISBM-FNN model. *Energy* **2023**, *263*, 125407. [CrossRef]
48. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
49. Kang, A.; Tan, Q.; Yuan, X.; Lei, X.; Yuan, Y. Short-term wind speed prediction using EEMD-LSSVM model. *Adv. Meteorol.* **2017**, *2017*, 6856139. [CrossRef]
50. Wang, W.; Chau, K.; Xu, D.; Chen, X.-Y. Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition. *Water Resour. Manag.* **2015**, *29*, 2655–2675. [CrossRef]
51. 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]

52. Wu, Z.; Huang, N.E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [CrossRef]
53. Kim, D.; Kim, K.O.; Oh, H.S. Extending the scope of empirical mode decomposition by smoothing. *EURASIP J. Adv. Signal Process.* **2012**, 1–17. [CrossRef]
54. Mandic, D.P.; Chambers, J.A. Exploiting Inherent Relationships in RNN Architectures. *Neural Netw.* **1999**, *12*, 1341–1345. [CrossRef] [PubMed]
55. Deng, T.; He, X.; Zeng, Z. Recurrent Neural Network for Combined Economic and Emission Dispatch. *Appl. Intell.* **2017**, *48*, 2180–2198. [CrossRef]
56. Guo, T.; Xu, Z.; Yao, X.; Chen, H.; Aberer, K.; Funaya, K. Robust online time series prediction with recurrent neural networks. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*; IEEE: Piscatvie, NJ, USA, 2016; pp. 816–825.
57. Ali, M.; Khan, D.M.; Aamir, M.; Khalil, U.; Khan, Z. Forecasting COVID-19 in Pakistan. *PLoS ONE* **2020**, *15*, e0242762. [CrossRef] [PubMed]
58. Khan, D.M.; Ali, M.; Iqbal, N.; Khalil, U.; Aljohani, H.M.; Alharthi, A.S.; Afify, A.Z. Short-Term Prediction of COVID-19 Using Novel Hybrid Ensemble Empirical Mode Decomposition and Error Trend Seasonal Model. *Front. Public Health* **2022**, 1652. [CrossRef] [PubMed]
59. Ali, M.; Khan, D.M.; Aamir, M.; Ali, A.; Ahmad, Z. Predicting the direction movement of financial time series using artificial neural network and support vector machine. *Complexity* **2021**, *2021*, 1–13. [CrossRef]