Sains Malaysiana 52(11)(2023): 3293-3306
http://doi.org/10.17576/jsm-2023-5211-20

A Hybrid Approach for Accurate Forecasting of Exchange Rate Prices using VMD-CEEMDAN-GRU-ATCN Model

(Pendekatan Hibrid untuk Ramalan Tepat Harga Kadar Pertukaran menggunakan Model VMD-CEEMDAN-GRU-ATCN)

Rehan Kausar[1], Farhat Iqbal[2,3,*], Abdul Raziq[2] & Naveed Sheikh[4]

[1]*Department of Statistics, Sardar Bahadur Khan Women's University, Quetta, Pakistan*
[2]*Department of Statistics, University of Balochistan, Quetta, Pakistan*
[3]*Department of Mathematics, Imam Abdulrahman Bin Faisal University, Saudi Arabia*
[4]*Department of Mathematics, University of Balochistan, Quetta, Pakistan*

Received: 5 May 2023/Accepted: 23 October 2023

ABSTRACT

The foreign exchange (Forex) market has greatly influenced the global financial market. While Forex trading offers investors substantial yield prospects, some risks are also involved. It is challenging to accurately model financial time series due to their nonlinear, non-stationary and noisy properties with an uncertain and hidden relationship. Thus, developing extremely precise forecasting techniques is crucial for investors and decision-makers. This study introduces a novel hybrid forecasting model, VMD-CEEMDAN-GRU-ATCN, designed to improve Forex price prediction accuracy. To begin with, our proposed model utilizes the variational model decomposition (VMD) technique for breaking down raw prices into multiple sub-components and residual terms. The complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) technique is utilized to extract features from the residual terms, which involves further decomposition and analysis of these complex information-containing terms. These sub-components are then predicted by the gated recurrent unit (GRU) model. To enhance the effectiveness of our hybrid model, we include the open, high, low, and close prices and seven Forex market technical indicators. Finally, an attention-based temporal convolutional network (ATCN) model is used to obtain the Forex price forecasts. For both one-step and multi-step ahead forecasting, our proposed VMD-CEEMDAN-GRU-ATCN model has demonstrated superior and consistent performance in predicting USD/PKR exchange rate price series.

Keywords: Attention mechanism; Forex; dual decomposition strategy; hybrid deep learning models; temporal convolutional network

ABSTRAK

Pasaran pertukaran asing (Forex) telah banyak mempengaruhi pasaran kewangan global. Walaupun perdagangan Forex menawarkan prospek hasil yang besar kepada pelabur, beberapa risiko turut terlibat. Adalah mencabar untuk memodelkan siri masa kewangan dengan tepat kerana sifatnya yang tidak linear, tidak pegun dan hingar dengan hubungan yang tidak pasti dan tersembunyi. Oleh itu, membangunkan teknik ramalan yang sangat tepat adalah penting untuk pelabur dan pembuat keputusan. Kajian ini memperkenalkan model ramalan hibrid baru, VMD-CEEMDAN-GRU-ATCN yang direka untuk meningkatkan ketepatan ramalan harga Forex. Sebagai permulaan, model cadangan kami menggunakan teknik penguraian model variasi (VMD) untuk memecahkan harga mentah kepada terma berbilang sub-komponen dan sisa. Teknik penguraian mod empirik ensembel lengkap dengan hingar suai (CEEMDAN) digunakan untuk mengekstrak ciri daripada terma sisa yang melibatkan penguraian dan analisis lanjut bagi terma yang mengandungi maklumat yang kompleks ini. Sub-komponen ini kemudiannya diramalkan oleh model unit berulang berpagar (GRU). Untuk meningkatkan keberkesanan model hibrid ini, kami memasukkan harga terbuka, tinggi, rendah dan tertutup serta tujuh penunjuk teknikal pasaran Forex. Akhir sekali, model rangkaian konvolusi temporal berasaskan perhatian (ATCN) digunakan untuk mendapatkan ramalan harga Forex. Untuk ramalan selangkah dan berbilang langkah ke hadapan, model cadangan VMD-CEEMDAN-GRU-ATCN telah menunjukkan prestasi unggul dan tekal dalam meramalkan siri harga pertukaran USD/PKR.

Kata kunci: Forex; model pembelajaran mendalam hibrid; strategi penguraian dual; mekanisme perhatian; rangkaian konvolusi temporal

Introduction

Forecasting financial time series has grown crucial for the global economy due to its ability to predict economic gains and affect countries' economic progress. As a result, academic researchers and business professionals are paying more attention to it due to its practical applications and theoretical possibilities. Unlike the stock market, the Forex market is extremely complex due to its nonlinearity, irregularity and high volatility. Additionally, the foreign exchange market is not governed by a single institution or organization, making it highly unpredictable and challenging to forecast. Therefore, predicting the Forex market is a particularly difficult problem.

The two main categories of forecasting techniques are econometric techniques and models based on artificial intelligence (AI). AI-based methods like artificial neural networks (ANNs) and especially recurrent neural networks (RNNs) have been shown to be good at handling random and nonlinear time series data (Loureiro, Miguéis & da Silva 2018; Zhang et al. 2019). RNNs establish connections between hidden layer units, allowing for a better understanding of the dependencies between data at different time points, making them particularly well-suited for forecasting time series data (Deng, He & Zeng 2018). Long Short-Term Memory (LSTM) networks overcome the challenges that are commonly encountered when dealing with long time spans, including vanishing and exploding gradients, thus enhancing the performance of traditional RNNs. LSTM has gained widespread popularity for time series forecasting and has delivered exceptional results in recent times (Karevan & Suykens 2020).

The GRU networks improve the performance of the LSTM model by combining the reset and update gates into a single three-gate mechanism. This improvement has led to the GRU outperforming the LSTM in various time series forecasting applications (Peng et al. 2020). While LSTM and GRU represent an advancement over RNNs, they still encounter some of the same challenges, including vanishing and exploding gradients, low training efficiency, and difficulty achieving good acceleration on GPUs, as noted by Hua and Zehao (2020). The TCN model was proposed by Bai, Kolter and Koltun (2018) as a response to these issues. The TCN model has a longer memory capacity than RNNs and is better suited for tasks involving long sequences. By reinforcing the impact of significant features on prediction results, the attention mechanism enhances the prediction accuracy of the model. It was first employed in deep learning for image processing to reduce redundant data and highlight important information, and it works well in various domains (Yujia et al. 2020).

In deep learning, selecting input variables is critical for optimal performance. Researchers have utilized diverse features, such as technical indicators (TI), online data, refactored technical indicators and text sentiment analysis to forecast Forex prices. Yildirim, Toroslu and Fiore (2021) achieved good results in forecasting the direction of Forex data by training an LSTM model with two different macroeconomic and technical indicator data. Similarly, Hu, Zhao and Khushi (2021) found that deep learning models have been frequently used in Forex and stock price prediction due to their better performance and results.

Researchers have come up with hybrid models that incorporate two or more different models to increase the precision of financial time series forecasting. Due to their greater performance than single models, these hybrid models have become prominent (Cao, Li & Li 2019). To extract the primary features of time series, various signal processing techniques, including empirical mode decomposition (EMD), complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and variational modal decomposition (VMD), can be employed in the hybrid forecasting approach. The VMD has superior noise robustness and component decomposition accuracy compared to EMD, as Jun et al. (2017) demonstrated.

Torres et al. (2011) introduced a CEEMDAN algorithm that avoids mode mixing of EMD by adding adaptive white noise. This method achieves better signal decomposition and reduces noise residuals with fewer averaging times, resulting in a minor reconstruction error. Cao, Li and Li (2019) developed a hybrid model that integrates CEEMDAN with LSTM and found that CEEMDAN achieved a more thorough decomposition than EMD, resulting in better performance in predicting stock price series. However, single decomposition techniques may not effectively handle the non-stationarity of random and irregular data series. Therefore, to improve Forex price predictions' accuracy, this research suggests a two-phase decomposition strategy combined with a method based on deep learning. This study introduces a novel approach for predicting Forex prices by utilizing the VMD-CEEMDAN-GRU-ATCN model. A comparative empirical analysis of the USD/PKR dataset is carried out between the proposed VMD-CEEMDAN-GRU-ATCN model and six additional benchmark models to evaluate the suggested approach's effectiveness.

This paper presents three main contributions. First, a two-stage signal decomposition procedure is introduced for Forex price forecasting, which combines VMD and CEEMDAN techniques to extract complex features and patterns that are not easily visible in the original price sequence. This leads to a reduction in complexity and improved accuracy of price predictions. Second, the suggested method combines VMD, CEEMDAN, GRU, attention mechanism, and TCN to develop a novel hybrid approach to Forex forecasting. This approach employs GRU fitted values as input features and integrates an Attention mechanism into the TCN model to enhance the effect of important characteristics on predicting outcomes, leading to improved accuracy of forecasts. Finally, twelve predictors are utilized to obtain a comprehensive understanding of the Forex market, including technical indicators, fundamental market data and the fitted values of a GRU model on the closing price. Additionally, the proposed model is evaluated for forecasting one-step and five-step ahead and compared with multiple benchmark methods.

The results of the study demonstrate the superior forecasting accuracy of the proposed approach across various evaluation metrics, suggesting its effectiveness and robustness for forecasting Forex prices.

Materials and Methods

This section's primary goal is to thoroughly review all the modules comprising the forecasting system, namely VMD, CEEMDAN, GRU, TCN, and the Attention mechanism; subsequently, the study outlines the fundamental framework of the proposed approach.

VARIATIONAL MODE DECOMPOSITION (VMD)

VMD is a signal-processing method that decomposes a real-valued signal f into K intrinsic mode functions (IMF) based on center frequency fluctuations w_k. It was introduced by Dragomiretskiy and Zosso (2013) and utilized a multiresolution non-recursive variational structure that combines techniques such as the Hilbert transform, Wiener filtering, and Alternating Direction Method of Multipliers (ADMM). VMD involves solving variational problems, starting with a constrained problem for the input signal.

$$min(w_k, u_k)\left\{\sum_{k=1}^{k}\left\|\partial(t)\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-jwk^t}\right\|_2^2\right\}$$
$$s.t.\sum_{k=1}^{k}u_k=f \tag{1}$$

The *k-th* IMF with a restricted bandwidth is denoted by u_k. The function's partial derivative with respect to time t is represented by $\partial(t)$, while Dirac distribution is denoted by $\delta(t)$. The imaginary unit is given by j, and the convolution is denoted by the * symbol. The introduction of α quadratic penalty function term a and a Lagrange multiplier λ is necessary to achieve the best possible solution for the constrained variational mode. This can be expressed as follows:

$$L(\{u_k\},\{w_k\},\lambda)=\alpha\sum_{k=1}^{k}\left\|\partial(t)\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_kt}\right\|_2^2+$$
$$\left\|f(t)-\sum_{k=1}^{k}u_k(t)\right\|_2^2\ \langle\lambda(t),f(t)-\sum_{k=1}^{k}u_k(t)\rangle \tag{2}$$

Next, the non-constrained variational problem is solved using the ADMM technique, which can be represented as follows:

$$\hat{u}_k^{n+1}(w)=\frac{\hat{f}(w)-\sum_{i\neq k}\hat{u}_j(w)+\left(\frac{\hat{\lambda}(w)}{2}\right)}{1+2\alpha(w-w_k)^2} \tag{3}$$

$$w_k^{n+1}=\frac{\int_0^\infty w|\hat{u}_k^{n+1}(w)|^2dw}{\int_0^\infty\left|\hat{u}_k^{n+1}(w)\right|^2dw} \tag{4}$$

The unconstrained variational problem can be resolved using ADMM, and iterative procedures yield the most suitable solution for the restricted variational model. The result is K modal components, each with a narrow bandwidth, breaking down the original signal.

COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITION WITH ADAPTIVE NOISE (CEEMDAN)

CEEMDAN is an algorithm that builds on EMD and EEMD concepts. It avoids mode mixing by incorporating adaptive white noise at each stage. This leads to a more comprehensive breakdown of signal data with minimal reconstruction errors, as fewer averaging times are needed to remove noise residuals. The decomposition steps of CEEMDAN are as follows:

Step 1 CEEMDAN uses noisy signals to decompose the original signal into N sub-experiments. It then calculates the average of the N sub-experiments to obtain the first intrinsic mode component (IMF_1).

$$x_i(t)=x(t)+\varepsilon w_i(t) \tag{5}$$

$$IMF_1(t) = \frac{1}{N}\sum_{i=1}^{N} IMF_{1i}(t) \quad (6)$$

Step 2 After obtaining the first intrinsic mode component IMF_1, the residual sequence $r_1(t)$ is calculated, and a new $r_1^i(t)$ is obtained for each of the N sub-experiments. This process is repeated for each of the remaining intrinsic mode components until the EMD decomposition is completed.

$$r_1(t) = x(t) - IMF_1(t) \quad (7)$$

Step 3 From step 2, compute the average of the resulting second intrinsic mode components:

$$IMF_2(t) = \frac{1}{N}\sum_{i=1}^{N} IMF_1\{r_1(t) + \varepsilon_1 IMF_1[w_i(t)]\} \quad (8)$$

Step 4 At the *k+1* stage, the same procedure is repeated. The residual sequence $rk(t)$ at this stage is calculated, and the IMF_{k+1th} is obtained.

$$r_k(t) = r_{k-1}(t) - IMF_k(t) \quad (9)$$

$$IMK_{k+1}(t) = \frac{1}{N}\sum_{i=1}^{N} IMF_1\{r_1(t) + \varepsilon_k IMF_k[w_i(t)]\} \quad (10)$$

Step 5 The above steps are iteratively repeated until the residual sequence has two or fewer extreme points. The EMD process is then finished, and the last intrinsic mode component IMF_k and residual sequence $R(t)$ are obtained. The original signal sequence $x(t)$ can be expressed as the sum of the intrinsic mode components and the residual sequence, given by:

$$x(t) = \sum_{i=1}^{k} IMF_k(t) + R(t) \quad (11)$$

GATED RECURRENT UNIT (GRU)

The GRU unit is comprised of three components: an update gate z_t, a reset gate r_t, and a current memory content $\hat{h}_t$, with the output h_t being saved in the final memory of the GRU. The update gate z_t determines the extent to which the input x_t and previous output h_{t-1} are transmitted to the next cell, and this is managed by the weight $W^{(z)}$. The reset gate, on the other hand, is responsible for determining how much of the past information should be disregarded. Meanwhile, the current memory content guarantees that only the pertinent information is transmitted to the next iteration, which is determined by the weight W. The following mathematical formulae regulate the GRU's basic operations:

$$z_t = \sigma\left(W^{(z)} x_t + U^{(z)} h_{t-1}\right)$$

$$r_t = \sigma\left(W^{(r)} x_t + U^{(r)} h_{t-1}\right)$$

$$h_t' = tanh(Wx_t + r_t \odot Uh_{t-1})$$

The intermediate values z_t and r_t are acquired from the update and reset gates, respectively. The hyperbolic tangent function, tanh, is utilized, as well as the sigmoid function, σ.

TEMPORAL CONVOLUTIONAL NETWORKS (TCN)

To improve its performance, the TCN neural network includes a number of approaches, including fully convolutional network, causal and dilated convolutions, and residual connections (Long, Shelhamer & Darrell 2015). The architecture makes use of zero-padding and causal convolutions in order to preserve equal layer lengths and avoid information leakage. Longer-term dependencies are captured using dilated convolutions, and their formula is written as follows:

$$F(s) = (\boldsymbol{x} *_d f)(s) = \sum_{i=0}^{k-1} f(i).\boldsymbol{x}_{s-d.i} \quad (12)$$

The * operation and dilation factor d are used in TCN to convolve x with a k-sized kernel. (s - d. i) refers to a sequence element s in the past direction. Expanding the receptive field by increasing the dilation factor enables the network to learn long-term dependencies and perform better on tasks involving sequence modeling. In deep learning, network degradation happens when the optimization process worsens as the network gets deeper. A residual block solves this by adding an extra pathway to adapt to identity mapping variations. It outputs $O = Activation(x + F(x))$, where x is the input, $F(x)$ is the residual mapping, and Activation is applied element-wise to the sum of the two. O combines the input and modified identity mapping.

TCN BASED ON ATTENTION MECHANISM (ATCN)

Inspired by the human brain's visual attention, the attention mechanism assigns weights to a neural network's output to emphasize important features. During training, the mechanism can selectively filter crucial features by increasing their weight and reducing less relevant ones. The TCN model has four main components: input layer, TCN module, attention mechanism, and output layer. They work together to process input data and generate the desired output. The next sections explain each component in detail.

Input layer: The input layer of the model is responsible for receiving the preprocessed original data, which is represented as a matrix with H time steps and L features. It is denoted as $I^{H \times L}$

TCN block: The TCN network used in the study comprises two TCN modules. Upon feeding the original sequence I as input, the f filter is processed by a Conv1D layer to generate the initial matrix of size $H \times L \times F$. The two TCN modules' respective convolution kernel widths are 64, 64, and 64, and their respective dilation factors are 1, 2, 4, 8, 16, and 32.

Attention mechanism: To enhance the TCN's ability to learn key features, an attention model is embedded between the two TCN hidden layers, which automatically assigns more attention weights to the hidden states that affect the prediction target.

Output layer: The output layer receives the output values after they have been extracted from the attention mechanism. The output layer uses a dense layer to lower the matrix's L-dimension and the resulting output sequence O represents the predicted values for the input sequence.

THE ARCHITECTURE AND WORKFLOW OF THE PROPOSED APPROACH

This study proposes a new method for predicting Forex prices, which integrates VMD, CEEMD, GRU, and ATCN techniques. This approach is referred to as VMD-CEEMDAN-GRU-ATCN. The suggested method's workflow is shown in Figure 1 and includes the following steps:

Step 1 Using the VMD decomposition method, the original time series is divided into individual mode components or VMFs. The residual series is subsequently generated by subtracting these VMFs from the original sequence.

Step 2 A two-stage decomposition process is carried out, involving the application of the CEEMDAN technique on the residual series. This step leads to the generation of a new set of subseries of residuals referred as IMFs.

Step 3 The GRU model is utilized to predict the modal components produced through decomposing the initial series and the residual term decomposition, which together form a set
of predictors.

Step 4 The final input matrix is formed by combining the open, high, low, and close (OHLC) prices, technical indicators, and the aggregate of outputs from all the sub-models. This new input matrix is then utilized for training the ATCN model, which generates the final forecasted results.

DATA DESCRIPTION AND EVALUATION CRITERIA

In this study, daily price data of USD/PKR are used to assess the effectiveness of the proposed hybrid model. Historical OHLC data from March 4, 1992, to February 20, 2023, are collected from (https://www.investing.com/). The plot in Figure 2 illustrates the daily USD/PKR prices, indicating a noticeable irregular and unstable pattern in the time series.

Additionally, various technical indicators, including Simple moving average (SMA), Momentum, Relative strength index (RSI), Average true range (ATR), Moving average convergence divergence (MACD) and Bollinger bands (BB) are computed using the closing price and appended as new features to the dataset. This work employs seven technical indicators, and their specific details are summarized in Table 1.

The USD/PKR price dataset is split into training and test datasets at an 80%-20% ratio. Neural networks typically require processed data to perform effectively. Standardizing the data involves scaling it to a specific range. This study uses the min-max standardization method to preprocess the data, as shown by the expression.

$$\tilde{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{13}$$

The formula shown in Equation (13) calculates the standardized subseries data, denoted by $\tilde{X}$, from the original feature data represented by X. In the original sequence, X_{min} indicates the lowest number and X_{max} the highest.

We utilized a VMD technique to identify the stationary features of USD/PKR exchange rates by analyzing the closing price time series. The resulting VMD decomposition is then used to subtract the sum of each mode component from the original sequence, producing a residual sequence. However, accurately predicting this complex residual sequence using a predictive algorithm is challenging. We adopted a secondary decomposition technique called CEEMDAN technology to gain more insight into extracting additional information from the residual series. Figure 3 illustrates the decomposition process of the USD/PKR price sequence.

In Figure 3 (left), the VMD-generated subseries are shown in low to high frequency, illustrating distinct local oscillations present in the data. These decomposed subseries appear to be smoother and more regular than the original series, leading to a reduction in dataset complexity for forecasting purposes. Notably, the subseries with high frequency components feature relatively small values, indicating short-term volatility in the original price series. On the other hand, the low frequency components have large values and represent overall changes in the daily closing prices. Additionally, the right side of Figure 3 illustrates the secondary decomposition outcomes of the residual term utilizing CEEMDAN.

To conduct a quantitative analysis of prediction performance, evaluation indicators such as the mean absolute error (MAE), symmetric mean absolute percentage error (SMAPE), root mean square error (RMSE) and mean absolute percentage error (MAPE) are used. These evaluation measures' formulas are shown in Table 2.

DIEBOLD-MARIANO (DM) TEST

The DM test is used to assess whether the hybrid approach has a statistically significant predictive accuracy compared to other models (Diebold & Mariano 2002). This test is widely used in research applications and involves testing the null hypothesis (Ho) that the expected loss from prediction errors for two methods (e_t^1 and e_t^2) are equal. The alternative hypothesis (H1) suggests a notable difference between the two methods. The test statistics are defined based on this alternative hypothesis.

$$DM = \frac{\sum_{t=1}^{T}[L(e_t^1) - L(e_t^2)]/T}{\sqrt{\lambda^2/T}} \lambda^2 \rightarrow N(0,1)$$

where λ^2 represents the estimated variance of $[L(e_t^1) - L(e_t^2)]$. To determine whether to reject the null hypothesis, compare the values of $|DM|$ and $|z_{\alpha/2}|$. If $|DM|$ exceeds $|z_{\alpha/2}|$, it indicates a significant difference in the prediction abilities of the two models. Conversely, if $|DM|$ is less than or equal to $|z_{\alpha/2}|$, the prediction capabilities of the two models are not significantly different.

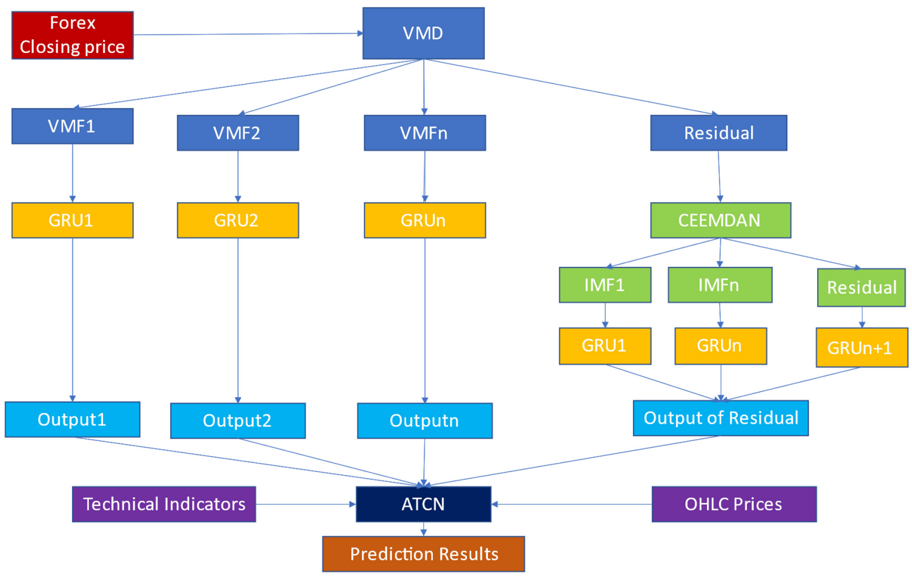

FIGURE 1. The process of VMD-CEEMDAN-GRU-ATCN model

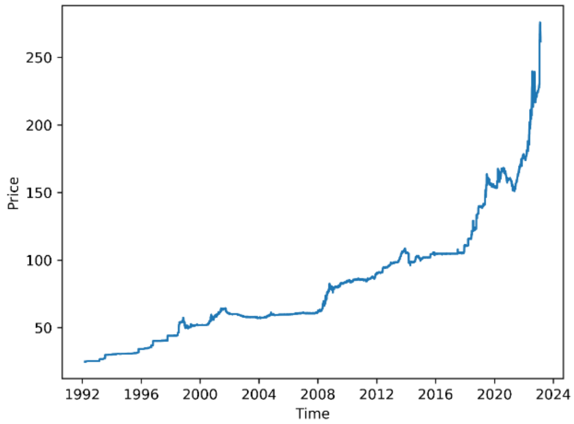

FIGURE 2. The temporal sequence of USD against PKR prices

TABLE 1. A summary of the technical indicators, along with their formulas and descriptions

Technical indicators	Formulas	Description
Simple n-day moving average (SMA)	$\frac{C_t + C_{t-1} + \cdots + C_{t-n-1}}{n}$	It computes the mean price values during a specific timeframe
Momentum	$C_t - C_{t-(n-1)}$	It depicts the rate of change in pricing
Moving average convergence divergence (MACD)	$MACD(n)_{t-1} + \frac{2}{n+1} \times (DIFF_t - MACD(n)_{t-1})$	It's a momentum indicator that follows the trend and illustrates the correlation between two moving averages
Rate of change (ROC)	$\frac{(C_t - C_{t-n})}{C_{t-n}} \times 100$	It calculates the percentage difference between the price at the moment in time and the previous price value
Average true range (ATR)	$EMA_n(\max(H_t - L_t), \lvert H_t - C_{t-1}\rvert, \lvert L_t - C_{t-1}\rvert)$	It reflects the level of market volatility
Relative strength index (RSI)	$100 - \frac{100}{1 + \left(\frac{\sum_{i=0}^{n-1} UP_{t-i}}{n}\right) / \left(\frac{\sum_{i=0}^{n-1} DW_{t-i}}{n}\right)}$	In order to determine if a price is overbought or oversold, it evaluates the size of recent price fluctuations
Bollinger bands (BB)	$MiddleBand = SMA(Close, t)$ $UpperBand = MiddleBand + SD(Close, t) * 2$ $LowerBand = MiddleBand - SD(Close, t) * 2$	It establishes a series of trendlines that are plotted at specific standard deviations from the simple moving average (SMA) of the price

At time t, C_t represents the closing price, L_t represents the low price, and H_t represents the high price. The difference between the exponential moving averages (*EMAs*) of time periods 12 and 26 is represented by $DIFF_t$. Upward price changes are represented by UP_t, while downward price changes are represented by DW_t. The standard deviation is denoted by *SD*. The absolute value of a number is represented by the vertical bars |…|

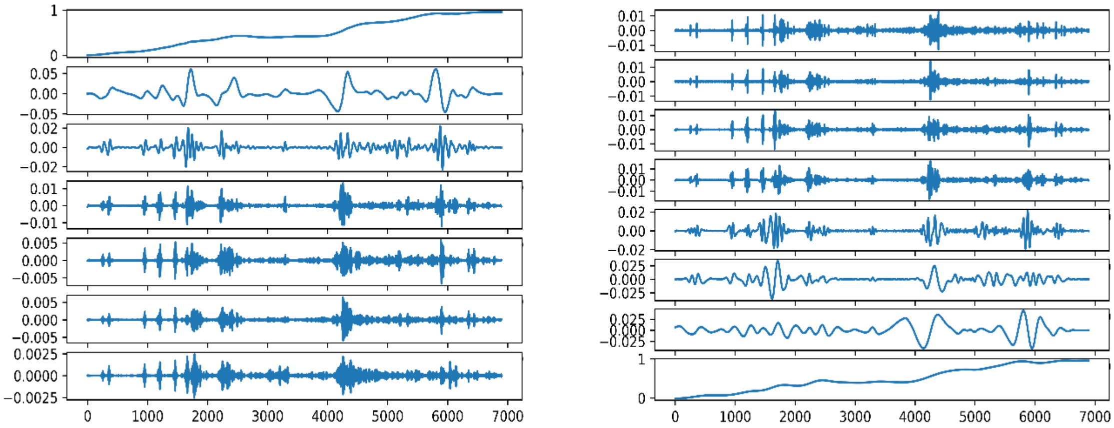

FIGURE 3. The decomposition of the prices via VMD decomposition (left) and the decomposition of the residuals via CEEMDAN (right)

TABLE 2. Relevant evaluation indicators

Evaluation metrics	Definition	Equation
MAE	Mean absolute error	$\frac{\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert}{n}$
MAPE	Mean absolute percentage error	$\frac{1}{n}\sum_{i=1}^{n}\frac{\lvert y_i - \hat{y}_i \rvert}{y_i} \times 100$
RMSE	Root mean square error	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
SMAPE	Symmetric mean absolute percentage error	$\frac{1}{n}\sum_{i=1}^{n}\frac{\lvert y_i - \hat{y}_i \rvert}{(\lvert y_i \rvert + \lvert \hat{y}_i \rvert)/2}$

The variable y_i represents the original time series, while $\bar{y}_i$ denotes its average value. The predicted time series, computed from the model, is represented by $\hat{y}_i$. The total number of observations is denoted by n

Results and Discussion

To demonstrate the VMD-CEEMDAN-GRU-ATCN model's improved performance in practical applications, this study compares it with six different benchmark models, namely GRU-ATCN, VMD-GRU-ATCN, CEEMDAN-GRU-ATCN, VMD-LSTM-ATCN, VMD-CEEMDAN-GRU-ALSTM, VMD-CEEMDAN-GRU-ATCN (without (w/o) TI), and VMD-CEEMDAN-GRU-ATCN (with TI). Furthermore, both one-step and five-step ahead predictions are made in order to test the model's robustness. Specifically, we use the data from the first ten trading days to predict the 11th and 15th trading days.

Table 3 displays the prediction outcomes of USD/PKR Forex prices. The experimental results indicate that the proposed VMD-CEEMDAN-GRU-ATCN hybrid model in this study outperforms all the other benchmark models in all the evaluated scenarios.

By examining Table 3, a notable difference between the hybrid models incorporating the decomposition method (VMD-GRU-ATCN, CEEMDAN-GRU-ATCN) and without the decomposition method (GRU-ATCN) becomes apparent. It is evident that those incorporating the decomposition technique have shown substantial enhancements in all evaluation metrics when compared to the GRU-ATCN model. Furthermore, when analyzing the models using the decomposition method separately, it is observed that the prediction results of VMD-GRU-ATCN exhibit better evaluation indices compared to those of CEEMDAN-GRU-ATCN. Therefore, the VMD decomposition technique is a more powerful tool for analyzing intricate time series, especially Forex prices, and has been shown to accurately extract important features from the data.

Moreover, Table 3 demonstrates that in terms of forecasting accuracy, the VMD-GRU-ATCN model performs better than the VMD-LSTM-ATCN model, as evidenced by its lower MAE values of 4.05746 in one-step and 4.81123 in five-step ahead forecasting. This indicates that the VMD-GRU-ATCN model has superior forecasting performance compared to the VMD-LSTM-ATCN model.

Table 3 also demonstrates that using secondary decomposition models leads to better forecasting results than single decomposition models. The effectiveness of the secondary decomposition technique in enhancing the forecasting performance of single decomposition models is clear when comparing VMD-CEEMDAN-GRU-ALSTM with VMD-GRU-ATCN and CEEMDAN-GRU-ATCN. The low MAE values of 3.51185 and 4.01279 in one-step and five-step forecasting, respectively, demonstrate the VMD-CEEMDAN-GRU-ALSTM model's exceptional predicting performance.

Between the VMD-CEEMDAN-GRU-ATCN and VMD-CEEMDAN-GRU-ALSTM models, Table 3 clearly contrasts their performance, with the former demonstrating substantially better results. Its low MAE values of 3.30695 and 3.34693 in one-step and five-step forecasting, respectively, indicate excellent forecasting performance. To assess the impact of technical indicators on the hybrid model's performance, we can compare the VMD-CEEMDAN-GRU-ATCN model with and without TIs. By doing so, we can conclude that our developed approach outperforms the benchmark model by a significant margin. This improvement is particularly noteworthy when fundamental market data is the only input, as demonstrated by the MAE values of 7.32954 and 8.62458 for one and five-step ahead, respectively. However, incorporating TIs into the input data results in a substantial enhancement of the model's performance.

Figure 4 depicts the forecasting errors, including MAE, MAPE, SMAPE, and RMSE, for all models considered. According to the figure, the models that use the secondary decomposition strategy perform better than the ones that use the single decomposition strategy and models that do not incorporate any decomposition. This is evident from the lower values of forecasting errors for the models utilizing the secondary decomposition strategy, demonstrating their greater prediction ability compared to the other models taken into account.

Figure 5 presents the prediction results of GRU-ATCN, VMD-GRU-ATCN, VMD-CEEMDAN-GRU-ATCN, and VMD-CEEMDAN-GRU-ATCN (without TIs) on the USD/PKR test data. It is evident that all the models are able to capture the overall changing trend of the exchange rate, indicating the strong capability of deep learning models to extract the underlying relationship of the Forex time series. However, it is noticeable that the VMD-CEEMDAN-GRU-ATCN model provided the closest prediction results to the actual exchange rate price series.

Furthermore, Figures 4 and 5 illustrate a decline in the predictive performance of all the models as the forecasting horizon increases. The increasing complexity and instability of the Forex price series as the forecasting horizon expands could be the main reason for this observation. Hence, predicting the Forex price series five days in advance is more challenging than predicting it one day in advance. Additionally, the model may not capture some data information during actual Forex forecasting, leading to a gradual deterioration of the model's predictive ability over time.

The results of the DM test, a statistical technique used to evaluate the comparative precision of the forecasting models, are shown in Table 4. The DM test determines whether variations in the model performances are statistically significant by comparing the forecast errors between various models. As a result, Table 4 contains substantial information on the relative effectiveness of the various forecasting models. The findings show that, at a 99% level of significance, the VMD-CEEMDAN-GRU-ATCN model consistently outperforms the benchmark model. This indicates that compared to other models, the VMD-CEEMDAN-GRU-ATCN model is more effective at enhancing prediction accuracy.

TABLE 3. Evaluating the precision of various models for multi-step forecasting

	One-step ahead forecasting				Five-step ahead forecasting			
Model	MAE	MAPE	SMAPE	RMSE	MAE	MAPE	SMAPE	RMSE
GRU-ATCN	5.0003	2.9765	0.7615	7.4060	6.5427	4.1847	1.03916	9.3068
VMD-GRU-ATCN	4.0574	2.6034	0.6542	5.6434	4.8112	2.8042	0.7183	7.4659
CEEMDAN-GRU-ATCN	4.4164	2.9155	0.7206	5.9530	5.0188	2.8563	0.7277	8.9182
VMD-LSTM-ATCN	4.2585	2.5464	0.6440	6.9557	4.8658	2.9535	0.7499	7.9318
VMD-CEEMDAN-GRU-ALSTM	3.5118	2.2107	0.5563	4.8550	4.0127	2.5005	0.6296	5.9189
VMD-CEEMDAN-GRU-ATCN (w/o TIs)	7.3295	4.1784	1.0797	10.4847	8.6245	4.7873	1.2508	13.280
VMD-CEEMDAN-GRU-ATCN	**3.3069**	**2.0256**	**0.5134**	**5.4573**	**3.3469**	**2.1036**	**0.5231**	**4.6825**

The smallest values are boldfaced

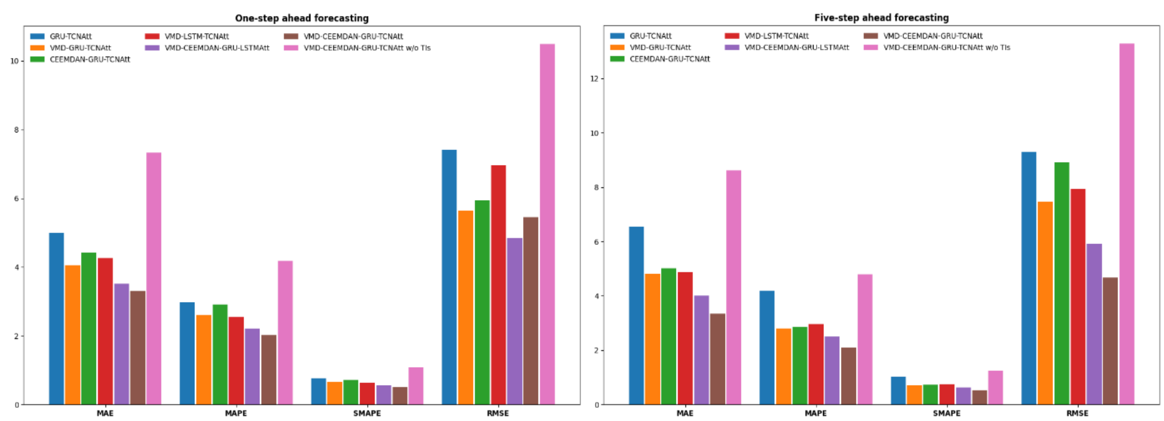

FIGURE 4. Comparing forecasting performance of different models for USD/PKR prices: one-step ahead (left) vs. five-step ahead (right) results

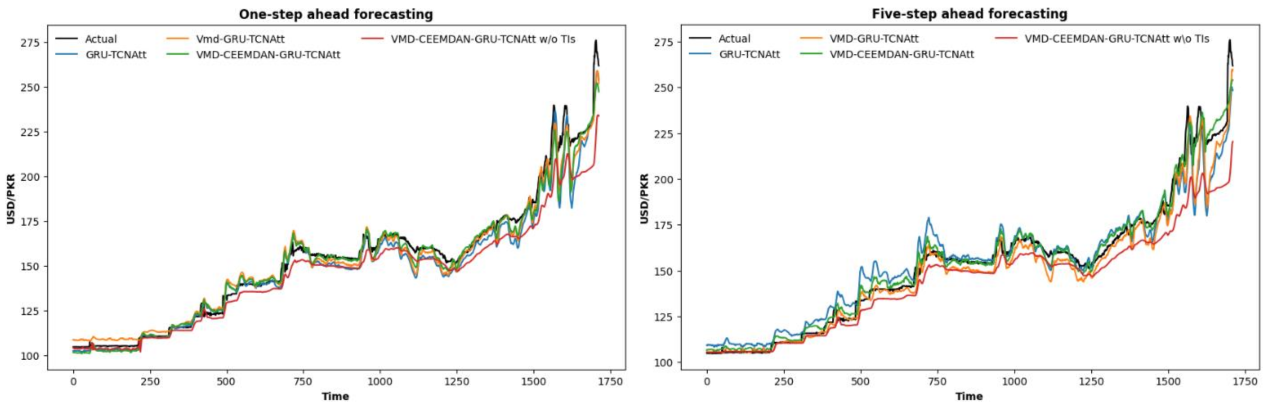

FIGURE 5. Visualization of one-step (left) and five-step (right) ahead forecasting

TABLE 4. DM test comparison outcomes for forecasting models

Models	One-step ahead forecasting	Five-step ahead forecasting
GRU-ATCN	21.61206*	9.01316*
VMD-GRU-ATCN	15.47076*	4.29338*
CEEMDAN-GRU-ATCN	13.094515*	4.14530*
VMD-LSTM-ATCN	17.59365*	4.42546*
VMD-CEEMDAN-GRU-ALSTM	2.60313*	3.27727*
VMD-CEEMDAN-GRU-ATCN (w/o TI)	28.24315*	9.00195*
VMD-CEEMDAN-GRU-ATCN	-	-

*Represents significance level at 1%

Overall, the proposed approach outperforms all other benchmark models, including those without decomposition models, single decomposition models, and other secondary decomposition models. We employ the dual decomposition VMD-CEEMDAN method, which possesses superior decomposition efficiency and effectiveness as well as robustness for parameters, to preliminarily weaken the non-stationarity of the original data. Previous studies on Forex price forecasting have only used a single decomposition-based combined model (Ulina, Purba & Halim 2020; Wei et al. 2019).

Zhang et al. (2021) proposed a hybrid model incorporating dual decomposition, EEMD (ensemble empirical mode decomposition), to break down the residual term after VMD and combine it with an optimized machine learning model. The model was tested on crude oil prices to check its effectiveness, and the outcomes demonstrated that it improved the reliability of predictions for crude oil prices. Compared to the one-level decomposition model, the proposed model was determined to be much superior. Based on our findings, our proposed model demonstrates superior performance compared to the single decomposition models. These results align with a previous study we reviewed, which supports the usefulness of our suggested model. Several other studies have also demonstrated the superiority of dual decomposition over single-level decomposition in hybrid models (Yang et al. 2019; Zhou & Wang 2021). The study by Guo et al. (2021) employed causal multi-head attention with the TCN model for cryptocurrency price forecasting, and the proposed model outperformed other benchmark models.

Another study by Cheng et al. (2021) employed the TCN architecture to forecast chaotic time series, benefiting from its remarkable stability during training, adjustable perception field and high level of parallelism. When compared to the traditional LSTM network and the hybrid convolutional neural network CNN-LSTM, the TCN model exhibited better performance. Our findings are consistent with the two studies mentioned above; our ATCN model, when combined with the decomposition method, provides outstanding results compared to the hybrid LSTM model. This comparison highlights the effectiveness of attention-based models and the ability of TCN to capture long-term dependencies in time series data. Additionally, some other studies show that integrating the attention mechanism into the TCN model enhances its prediction performance (Fan et al. 2021; Zhen et al. 2022). In order to forecast exchange rates, Das, Mishra and Rout (2019) developed a novel hybrid machine-learning system that combines statistical measures and technical indicators. The authors conducted a performance analysis to compare the effectiveness of these measures. The study showed that technical indicators exhibited superior performance compared to statistical measures and the hybrid approach through simulation graphs and error comparison metrics. Our research findings are consistent with these results, indicating that our proposed model substantially improves when incorporating technical indicators.

Aryal et al. (2019) employed various deep learning models (namely, LSTM, CNN, and TCN) to forecast the USD/LKR exchange rate. They then assessed and compared the models' performance, finding that the CNN model outperformed the others. In contrast, our experimental findings suggest that TCN-based models, which combine features from both RNN and CNN architectures, outperform LSTM models in terms of accuracy. Notably, this study is the first to utilize a hybrid deep learning approach with a multi-step forecasting horizon to estimate the USD/PKR exchange rate price. Compared to Yasir et al. (2019), who examined the USD/PKR currency exchange rate using machine learning and a single deep learning model with sentiment data, our evaluation metrics values have shown improvement.

Conclusions

The study introduces a novel hybrid forecasting model for predicting exchange rate prices. The model considers the highly nonlinear and intricate features of the Forex price time series, which arise from various influencing factors. Likewise, the individual models' limitations led to the development of the proposed VMD-CEEMDAN-GRU-ATCN hybrid model. Currently, data processing methods that involve dividing data into multiple modal components are widely used. However, some of these methods remain intricate and challenging to forecast accurately. To address this issue, a new secondary decomposition strategy has been developed that combines the technical advantages of VMD and CEEMDAN to successfully decrease the complexity of the Forex time series. Furthermore, the proposed method benefits greatly from a GRU model's powerful nonlinear mapping ability, which is employed to predict all the decomposed components. Building upon the TCN backbone, by exploiting the inherent parallelism of the TCN model, we can speed up the training process while avoiding the gradient problems that are often encountered with RNNs. Moreover, integrating the attention mechanism within TCN networks significantly improves their capability to choose input features. To boost the effectiveness of the hybrid model, fundamental data and seven indicators are included, in addition to the historical data of Forex market closing prices. The methodology of complex systems has been utilized in the empirical analysis, leading to the following conclusions. The preliminary results indicate that the proposed hybrid model substantially improves the prediction accuracy of the original sequence, in comparison to the non-decomposition model. Secondly, VMD technology outperforms CEEMDAN in terms of accurately decomposing the Forex price time series for prediction purposes. The hybrid model utilizing VMD technology in combination with GRU-ATCN exhibits superior prediction accuracy than the one using CEEMDAN technology combined with GRU-ATCN. Finally, the VMD-CEEMDAN-GRU-ATCN hybrid model, which incorporates multiple decomposition levels, exhibits superior prediction accuracy compared to the hybrid model with one-level decomposition.

The suggested approach consists of a well-designed framework that integrates multiple technical modules and has shown outstanding performance. Empirical results show that the developed approach exhibits favorable error values in both one-step and five-step forecasting in comparison to the other competing models. Considering the effectiveness of this proposed method, the study is likely to be attractive to policymakers and investment institutions and individual investors interested in PKR currency exchange rates.

REFERENCES

Aryal, S., Nadarajah, D., Kasthurirathna, D., Rupasinghe, L. & Jayawardena, C. 2019. Comparative analysis of the application of Deep Learning techniques for Forex Rate prediction. *International Conference on Advancements in Computing (ICAC)*. pp. 329-333. https://doi.org/10.1109/ICAC49085.2019.9103428

Bai, S., Kolter, J.Z. & Koltun, V. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *ArXiv Preprint ArXiv*: 1803.01271. https://doi.org/10.48550/arXiv.1803.01271

Cheng, W., Wang, Y., Peng, Z., Ren, X., Shuai, Y., Zang, S., Liu, H., Cheng, H. & Wu, J. 2021. High-efficiency chaotic time series prediction based on time convolution neural network. *Chaos, Solitons & Fractals* 152: 111304. https://doi.org/10.1016/j.chaos.2021.111304

Cao, J., Li, Z. & Li, J. 2019. Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications* 519: 127-139. https://doi.org/10.1016/j.physa.2018.11.061

Das, S.R., Mishra, D. & Rout, M. 2019. A hybridized ELM using self-adaptive multi-population-based Jaya algorithm for currency exchange prediction: An empirical assessment. *Neural Computing and Applications* 31(11): 7071-7094. https://doi.org/10.1007/s00521-018-3552-8

Deng, T., He, X. & Zeng, Z. 2018. Recurrent neural network for combined economic and emission dispatch. *Applied Intelligence* 48: 2180-2198. https://doi.org/10.1007/s10489-017-1072-3

Diebold, F.X. & Mariano, R.S. 2002. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 20(1): 134-144. https://doi.org/10.1198/073500102753410444

Dragomiretskiy, K. & Zosso, D. 2013. Variational mode decomposition. *IEEE Transactions on Signal Processing* 62(3): 531-544. https://doi.org10.1109/TSP.2013.2288675

Fan, J., Zhang, K., Zhu, Y. & Chen, B. 2021. Parallel spatio-temporal attention-based TCN for multivariate time series prediction. *Neural Computing and Applications* 35: 13109-13118. https://doi.org/10.1007/s00521-021-05958-z

Guo, H., Zhang, D., Liu, S., Wang, L. & Ding, Y. 2021. Bitcoin price forecasting: A perspective of underlying blockchain transactions. *Decision Support Systems* 151: 113650. https://doi.org/10.1016/j.dss.2021.113650

Hu, Z., Zhao, Y. & Khushi, M. 2021. A survey of Forex and stock price prediction using deep learning. *Applied System Innovation* 4(1): 9. https://doi.org/10.3390/asi4010009

Hua, Y. & Zehao, C. 2020. Short-term traffic flow prediction based on temporal convolutional networks. *Journal of South China University of Technology (Natural Science Edition)* 48: 8. https://doi.org/10.1109/ITSC48978.2021.9564803

Jun, W., Lingyu, T., Yuyan, L. & Peng, G. 2017. A weighted EMD-based prediction model based on TOPSIS and feed forward neural network for noised timeseries. *Knowledge-Based Systems* 132: 167-178. https://doi.org/10.1016/j.knosys.2017.06.022

Karevan, Z. & Suykens, J.A.K. 2020. Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks* 125: 1-9. https://doi.org/10.1016/j.neunet.2019.12.030

Long, J., Shelhamer, E. & Darrell, T. 2015. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

Loureiro, A.L., Miguéis, V.L. & da Silva, L.F. 2018. Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems* 114: 81-93. https://doi.org/10.1016/j.dss.2018.08.010

Peng, Z., Peng, S., Fu, L., Lu, B., Tang, J., Wang, K. & Li, W. 2020. A novel deep learning ensemble model with data denoising for short-term wind speed forecasting. *Energy Conversion and Management* 207: 112524. https://doi.org/10.1016/j.enconman.2019.112524

Torres, M.E., Colominas, M.A., Schlotthauer, G. & Flandrin, P. 2011. A complete ensemble empirical mode decomposition with adaptive noise. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 4144-4147. https://doi.org/10.1109/ICASSP.2011.5947265

Ulina, M., Purba, R. & Halim, A. 2020. Foreign exchange prediction using CEEMDAN and improved FA-LSTM. *Fifth International Conference on Informatics and Computing (ICIC)*. pp. 1-6. https://doi.org/10.1109/ICIC50835.2020.9288615

Wei, Y., Sun, S., Ma, J., Wang, S. & Lai, K.K. 2019. A decomposition clustering ensemble learning approach for forecasting foreign exchange rates. *Journal of Management Science and Engineering* 4(1): 45-54. https://doi.org/10.1016/j.jmse.2019.02.001

Yang, W., Wang, J., Niu, T. & Du, P. 2019. A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting. *Applied Energy* 235: 1205-1225. https://doi.org/10.1016/j.apenergy.2018.11.034

Yasir, M., Durrani, M.Y., Afzal, S., Maqsood, M., Aadil, F., Mehmood, I. & Rho, S. 2019. An intelligent event-sentiment-based daily foreign exchange rate forecasting system. *Applied Sciences* 9(15): 2980. https://doi.org/10.3390/app9152980

Yildirim, D.C., Toroslu, I.H. & Fiore, U. 2021. Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. *Financial Innovation* 7(1): 1-36. https://doi.org/10.1186/s40854-020-00220-2

Yujia, Z., Zhicheng, D., Songwei, G. & Jirong, W. 2020. Dynamic personalized search based on RNN with attention mechanism. *Chinese Journal of Computers* 43: 812-826. https://doi.org/10.11897/SP.J.1016.2020.00812

Zhang, C., Pan, H., Ma, Y. & Huang, X. 2019. Analysis of Asia Pacific stock markets with a novel multiscale model. *Physica A: Statistical Mechanics and its Applications* 534: 120939. https://doi.org/10.1016/j.physa.2019.04.175

Zhang, T., Tang, Z., Wu, J., Du, X. & Chen, K. 2021. 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* 229: 120797. https://doi.org/10.1016/j.energy.2021.120797

Zhen, Y., Fang, J., Zhao, X., Ge, J. & Xiao, Y. 2022. Temporal convolution network based on attention mechanism for well production prediction. *Journal of Petroleum Science and Engineering* 218: 111043. https://doi.org/10.1016/j.petrol.2022.111043

Zhou, J. & Wang, S. 2021. A carbon price prediction model based on the secondary decomposition algorithm and influencing factors. *Energies* 14(5): 1328. https://doi.org/10.3390/en14051328

*Corresponding author; email: Farhat.iqbal@um.uob.edu.pk