Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: 10.61784/mjet3007

PERFORMANCE COMPARISON OF CEEMDAN-LSTM AND BASIC LSTM MODELS IN PREDICTING REALIZED VOLATILITY

ZheShuo Zhang
Wenzhou-Kean University, Wenzhou 325060, Zhejiang, China.
Corresponding Email: zhanzhes@kean.edu

Abstract: This paper presents a comparative analysis of the effectiveness of hybrid CEEMDAN-LSTM models and traditional LSTM models in predicting realized volatility in financial markets. By utilizing realized volatility data from 2004 to 2024, the study highlights significant market fluctuations during the 2008 financial crisis and the 2020 COVID-19 pandemic. The findings indicate that the CEEMDAN-LSTM model, which decomposes time series data into intrinsic mode functions (IMFs) before applying LSTM networks, outperforms the basic LSTM model in terms of predictive accuracy, particularly during periods of high volatility. This enhanced performance is evidenced by lower error metrics, such as Mean Absolute Error (MAE) and Mean Squared Error (MSE). The research underscores the value of integrating advanced decomposition techniques with deep learning models to better capture the complex dynamics of financial markets.
Keywords: CEEMDAN-LSTM; Intrinsic mode functions (IMFs); Mean Absolute Error (MAE); Mean Squared Error (MSE)

1 INTRODUCTION

Long sequence time-series forecasting (LSTF) is gaining increasing attention and application across various fields. The main approaches involve establishing time series models and utilizing machine learning techniques. With the development of numerous methods, accurately distinguishing the strengths and weaknesses of different models and choosing appropriate methods for forecasting in different domains has become increasingly important. Therefore, this article aims to draw the following conclusions by comparing the different predictive performances of Decision Trees (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Autoregressive (AR), and Hybrid ARIMA and Recurrent Neural Networks (HAR): (i) which forecasting approach is more accurate in the same domain; (ii) which forecasting approach is more efficient in the same domain; and (iii) in which domains certain forecasting approaches are more suitable for application.

The financial market is characterized by its inherent complexity and volatility, posing significant challenges for accurate stock market prediction. Traditional forecasting methods often struggle due to the non-linear and non-stationary nature of financial time series data. To address these challenges, researchers have increasingly turned to hybrid models that combine advanced signal decomposition techniques with sophisticated neural networks. One promising approach that has emerged in recent years is the integration of Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Long Short-Term Memory (LSTM) networks. CEEMDAN effectively decomposes complex time series data into simpler intrinsic mode functions (IMFs), which can then be processed by LSTM networks to capture both linear and non-linear patterns. This literature review provides an in-depth analysis of recent studies employing CEEMDAN-LSTM models for stock market prediction, highlighting their methodologies, findings, and contributions to the field.

Adebiyi et al.[1] perform a comparative analysis between ARIMA (AutoRegressive Integrated Moving Average) and Artificial Neural Networks (ANNs) models for stock price prediction. Their study highlights the strengths and limitations of both approaches. The research demonstrates that while ARIMA models are effective for linear time series data due to their reliance on past values and error terms, they often fall short in capturing the non-linear patterns inherent in financial markets. Conversely, ANNs exhibit a superior capability to model complex, non-linear relationships within stock price data, owing to their flexible structure and learning algorithms. Adebiyi et al. [1] conclude that ANNs generally outperform ARIMA models in stock price prediction tasks, especially in capturing intricate market dynamics. This comparison underscores the potential benefits of integrating neural network techniques with traditional statistical methods to enhance forecasting accuracy.

Recent studies have explored the application of CEEMDAN-LSTM models in financial time series forecasting, demonstrating their superior performance compared to traditional methods. Akşehir and Kılıç [2] propose a novel denoising approach, 2LE-CEEMDAN, which enhances the accuracy of time series forecasting by effectively decomposing complex signals into simpler components, thus facilitating more accurate LSTM modeling. Their study underscores the importance of noise reduction in improving predictive performance. Similarly, Cao et al. [3] present a comprehensive financial time series forecasting model that integrates CEEMDAN with LSTM. Their model demonstrates superior performance, attributed to CEEMDAN's ability to decompose time series into intrinsic mode functions (IMFs), which are then used as inputs for LSTM networks, effectively capturing both linear and non-linear

patterns in the data.

Further exploration of CEEMDAN-LSTM in financial forecasting by Guan [4] emphasizes the model's robustness in handling non-stationary time series. The study provides empirical evidence of improved prediction accuracy, highlighting the model's potential in real-world financial applications. Guresen et al. [5] also contribute to this area by exploring the use of artificial neural network (ANN) models in stock market index prediction. Their research highlights the potential of ANNs in capturing complex patterns within financial time series data, laying the groundwork for subsequent hybrid models like CEEMDAN-LSTM. Lin et al. [6] investigate the use of CEEMDAN-LSTM for forecasting stock index prices, finding that the model significantly outperforms conventional forecasting methods, particularly in capturing sudden market movements. They discuss the implications of using advanced decomposition techniques in financial modeling. Extending this research, Lin et al. [7] predict the realized volatility in stock price indices using a hybrid CEEMDAN-LSTM model. This approach combines the strengths of CEEMDAN in noise reduction and LSTM in sequence learning, resulting in highly accurate volatility forecasts.

In addition to the primary studies on CEEMDAN-LSTM, other researchers have proposed enhancements and comparative studies to further refine forecasting models. Assaad et al. [8] present a new boosting algorithm for time-series forecasting using recurrent neural networks (RNNs), improving forecasting accuracy by iteratively enhancing the model's performance on difficult-to-predict data points. The principles of boosting can be applied to CEEMDAN-LSTM models to refine their predictive capabilities, particularly in handling complex and non-linear financial data. Baek and Kim [9] introduce ModAugNet, a novel forecasting framework that addresses overfitting in LSTM models. Their approach involves an overfitting prevention module and a prediction module, significantly improving prediction accuracy and generalization capability. Insights from this study can be leveraged to enhance CEEMDAN-LSTM models by incorporating overfitting prevention techniques, ensuring robust performance in diverse market conditions.

Pin Lv et al. [10] investigate a hybrid model for stock index prediction based on modal decomposition techniques. Their research focuses on enhancing prediction accuracy by leveraging the strengths of various decomposition methods. The study demonstrates that using modal decomposition allows for the isolation of significant components within stock index data, thereby improving the inputs for predictive models. By integrating these decomposed components with advanced forecasting models, the hybrid approach provides a more robust and accurate prediction framework. The findings of Pin Lv et al. [10] highlight the importance of combining decomposition techniques with sophisticated modeling strategies to capture the complex dynamics of financial markets. Qi et al. [11] explore a variation of the hybrid model by integrating CEEMDAN with Wavelet Transform and GRU (Gated Recurrent Unit) networks for stock price prediction. This study highlights the effectiveness of combining multiple decomposition techniques with advanced neural networks, providing a comparative analysis with CEEMDAN-LSTM models and demonstrating the potential for further improvements in forecasting accuracy. Additionally, Sun and Liu [12] apply a CEEMDAN-ARMA-LSTM model for Air Quality Index (AQI) prediction, showcasing the versatility of CEEMDAN-LSTM frameworks beyond financial markets. Their findings suggest that integrating autoregressive moving average (ARMA) models with CEEMDAN-LSTM can enhance predictive accuracy for various types of time series data. Wang et al. [13] focus on predicting green bond indices using a CEEMDAN-LSTM model, illustrating the applicability of this hybrid approach in sustainable finance. The model's ability to handle the unique characteristics of green financial instruments is emphasized. Furthermore, Yanan et al. [14] delve into the prediction of chaotic time series using LSTM with CEEMDAN, providing insights into the model's capability to deal with highly irregular and complex data patterns, reinforcing the value of CEEMDAN-LSTM in diverse forecasting scenarios.

2 METHODS AND RESULTS

Firstly, we need to import the necessary libraries, set up the plot styles, and read data from a CSV file.Firstly, it need to import the necessary libraries. ‘pandas’ is used for data manipulation and analysis. ‘numpy’ is used for numerical calculations. ‘datetime’ is used for handling date and time. ‘matplotlib.pyplot’ is used for plotting graphs. Then, it set the plot styles, ‘plt.style.use('seaborn-v0_8')’ means that applies Seaborn's plotting style (version 0.8).‘plt.rcParams['figure.figsize']’ means that sets the plot size to 12x6 inches. ‘plt.rcParams['figure.dpi']’ means sets the plot resolution to 300 DPI. Finally, the table includes columns for the trading date, closing price, opening price, highest price, and lowest price. The overall purpose of this code is to prepare the data for further analysis or visualization.

Then code set some extract the opening and closing prices of each trading day and store them in a new data frame. ‘total_days’, ‘daily_open’ and ‘daily_close’ correspond to the date, opening price and closing price.

RV refers to realized volatility, which refers to the fluctuation range of asset price changes that have occurred, measured by calculating the standard deviation of asset prices over a period of time. The main purpose of this code is to calculate the log returns for each trading day and calculate the RV (Realized Volatility) based on these returns and store the results in a data frame for further analysis.

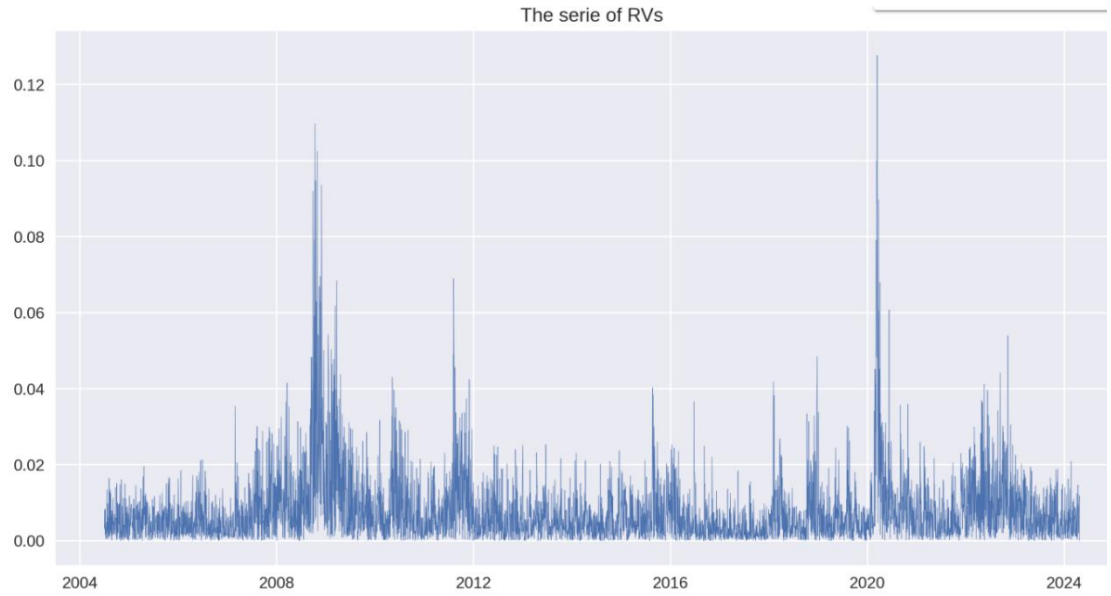

Figure 1 The Series of RVs

This represents the realized volatility from 2004 to 2024, with significant spikes occurring in 2008 due to the financial crisis and again in 2020 as a result of the COVID-19 pandemic.

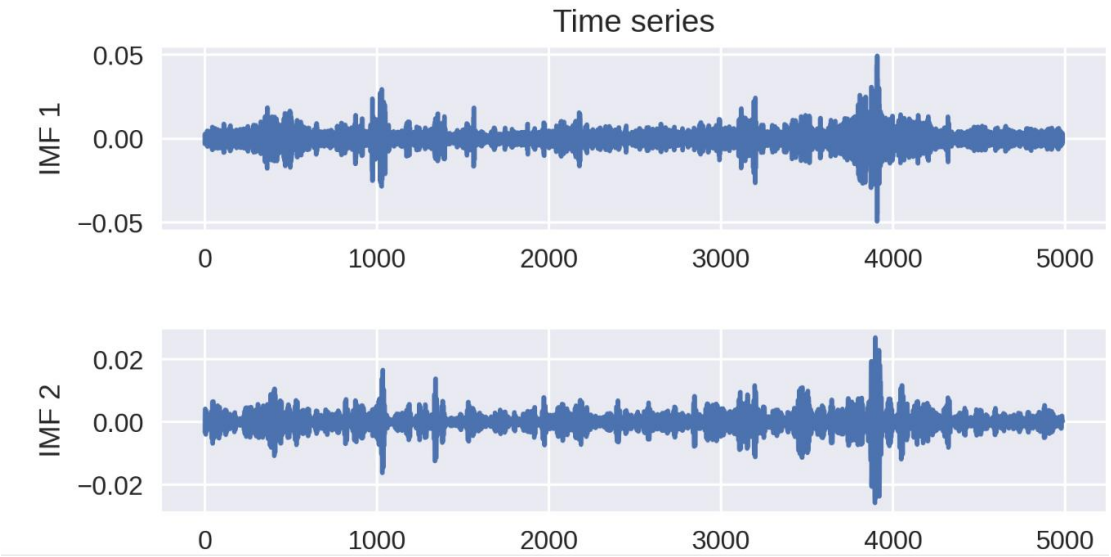

Figure 2 Time Series Diagram of IMF1 and IMF2

This image show how to use CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) to decompose time series data and visualize the decomposed results.

The picture shows the time series of the first intrinsic mode function (IMF 1). The horizontal axis represents time and the vertical axis represents the amplitude of IMF 1. This figure reflects the first level of decomposition of the original time series after CEEMDAN processing. IMF usually represents different frequency components in the time series, and the more forward IMF contains higher frequency components.

This image show how to calculate the statistical characteristics of the IMFs (Intrinsic Mode Functions) and residuals obtained from the previous decomposition, and display these characteristics in tabular form.

The table in the picture shows the statistical characteristics of each IMF and residual. Each row represents an IMF or residual, and the columns represent: count: the number of data points, mean, std (the standard deviation), skew (skewness, reflecting the symmetry of the data distribution), kurtosis (kurtosis, reflecting the sharpness of the data distribution), J-B (Jarque-Bera statistic, reflecting whether the data is close to a normal distribution), Q(10)(Ljung-Box statistic, used to test the autocorrelation of timeseries)

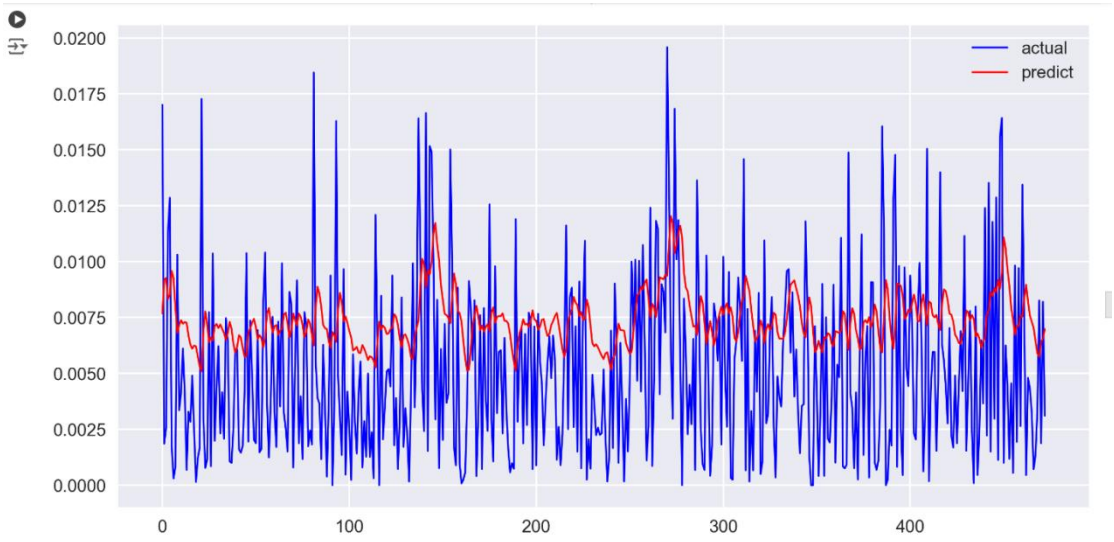

Figure 3 Actual vs Predicted Values Using Basic LSTM on Time Series Data

Table 1 Error Table of Basic LSTM on Time Series Data

	MAE	MSE	HMAE	HMSE
Basic LSTM	0.00355	1.82079	0.60725	0.53171

First, we use traditional time series models to make forecasts. The image displays a comparison between the actual values (blue curve) and the predicted values (red curve) of a time series model. These metrics indicate that there are discrepancies between the predicted and actual values, especially in certain fluctuating parts, suggesting that the prediction accuracy is not optimal.

It also uses four metrics: MAE (Mean Absolute Error), MSE (Mean Squared Error), HMAE (Harmonic Mean Absolute Error), and HMSE (Harmonic Mean Squared Error) to evaluate the model's accuracy. These values indicate the prediction effect of the model. The smaller the value, the smaller the model error.

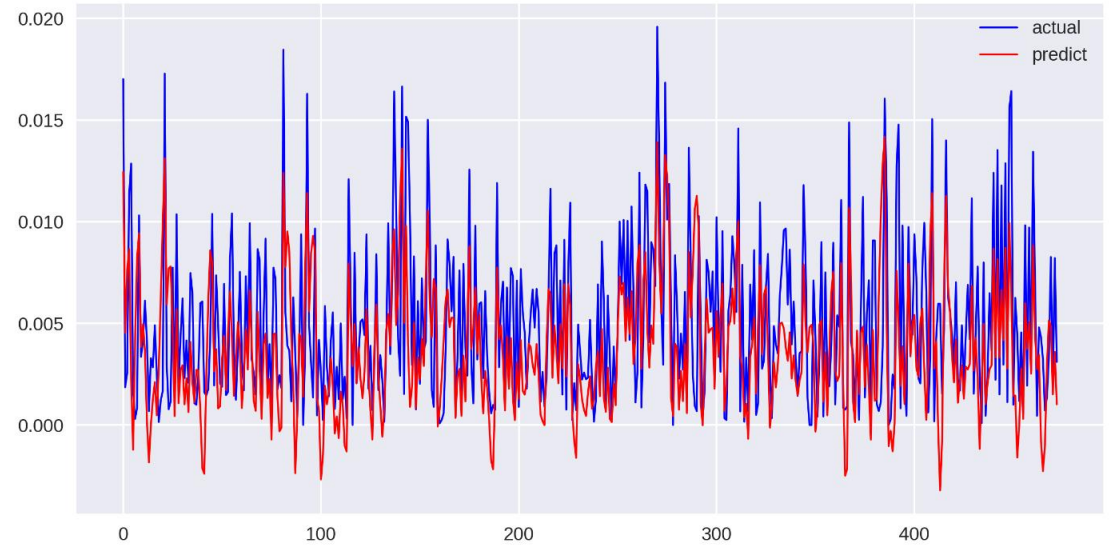

Figure 4 Actual vs Predicted Values Using CEEMDAN-LSTM Model on Time Series

The CEEMDAN-LSTM model first uses CEEMDAN to decompose the time series into several IMF components, then uses the LSTM model to predict each component, and finally recombine the prediction results to obtain the final prediction results. The prediction at this time is the sum of the predictions of all decomposition results.

As can be seen from the figure, the red forecast curve closely follows the blue actual curve. Although there are errors in some fluctuations, the accuracy of traditional time series forecasting has been greatly improved.

Table 2 Error table of Basic LSTM and CEEMDAN-LSTM Model on Time Series

	MAE	MSE	HMAE	HMSE

TimeSeries	0.00355	1.82079	0.60725	0.53171
CEEMDAD LSTM	0.00273	1.16912	0.73335	0.84252

3 CONCLUSION

This study evaluates and compares the performance of two predictive models—CEEMDAN-LSTM and Basic LSTM—in forecasting realized volatility. Realized volatility data from 2004 to 2024 was analyzed, capturing significant market fluctuations during the 2008 financial crisis and the 2020 COVID-19 pandemic.

The results indicate that the CEEMDAN-LSTM model, which combines the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Long Short-Term Memory (LSTM) networks, outperforms the Basic LSTM model. The CEEMDAN-LSTM approach decomposes the time series data into various Intrinsic Mode Functions (IMFs) and residuals, then uses LSTM to predict each component, leading to more accurate predictions, especially in periods of high volatility.

This model's effectiveness is reflected in lower error metrics, such as Mean Absolute Error (MAE) and Mean Squared Error (MSE), when compared to the Basic LSTM model. The CEEMDAN-LSTM model offers a more refined prediction of volatility, better capturing market dynamics and reducing prediction discrepancies, particularly in turbulent market conditions.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCE

[1] Adebiyi A A, Adewumi A O, Ayo C K. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. Journal of Applied Mathematics, 2014: 1–7.

[2] Akşehir Z D, Kılıç E. A new denoising approach based on mode decomposition applied to the stock market time series: 2LE-CEEMDAN. PeerJ. Computer Science, 2024, 10, e1852.

[3] Cao J, Li Z, Li J. Financial time series forecasting model based on CEEMDAN and LSTM. Physica A: Statistical Mechanics and Its Applications, 2019, 519: 127–139.

[4] Guan Y. Financial time series forecasting model based on CEEMDAN-LSTM. 2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC), 2022.

[5] Guresen E, Kayakutlu G, Daim T U. Using artificial neural network models in stock market index prediction. Expert Systems with Applications, 2011, 38(8): 10389–10397.

[6] Lin Y, Yan Y, Xu J, Liao Y, Ma F. Forecasting stock index price using the CEEMDAN-LSTM model. The North American Journal of Economics and Finance, 2021, 57: 101421.

[7] Lin Y, Lin Z, Liao Y, Li Y, Xu J, Yan Y. Forecasting the realized volatility of stock price index: A hybrid model integrating CEEMDAN and LSTM. Expert Systems with Applications, 2022, 206: 117736.

[8] Assaad M, Boné R, Cardot H. A new boosting algorithm for improved time-series forecasting with recurrent neural networks. Information Fusion, 2008, 9(1): 41–55.

[9] 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 Systems with Applications, 2018, 113: 457–480.

[10] Pin Lv, Shu Y, Xu J, Wu Q. Modal decomposition-based hybrid model for stock index prediction. Expert Systems with Applications, 2022, 202: 117252–117252.

[11] Qi C, Ren J, Su J. GRU Neural Network Based on CEEMDAN–Wavelet for Stock Price Prediction. Applied Sciences, 2023, 13(12): 7104–7104.

[12] Sun Y, Liu J. AQI Prediction Based on CEEMDAN-ARMA-LSTM. Sustainability, 2022, 14(19): 12182.

[13] Wang J, Tang J, Guo K. Green Bond Index Prediction Based on CEEMDAN-LSTM. Frontiers in Energy Research, 2022, 9.

[14] Yanan G, Xiaoqun C, Bainian L, Kecheng P. Chaotic Time Series Prediction Using LSTM with CEEMDAN. Journal of Physics: Conference Series, 2020, 1617: 012094.