International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Network Traffic Prediction Using a Hybrid Deep Learning Method with CEEMDAN and Attention Mechanism

Anurag Rajput
Department of Computer Engineering
JSPM's Rajarshi Shahu College of Engineering, Pune
anuragrajput.edu@gmail.com

Dr Nihar Ranjan
Department of Information Technology
JSPM's Rajarshi Shahu College of Engineering, Pune
Nihar.pune@gmail.com

***Abstract*— Accurately predicting network traffic is critical for real-time anomaly detection, improved service quality, and effective resource management. This research delves into the most recent developments in CEEMDAN-integrated hybrid deep learning approaches for network traffic prediction. CEEMDAN breaks down data on unpredictable and non-linear network connections into IMFs, therefore isolating important data components. Deep learning architectures enhanced with attention mechanisms which give relevant features top priority for enhanced prediction accuracy are further investigated using these IMFs. By means of advanced data decomposition and feature selection, studies using real-world network traffic datasets show that these hybrid approaches exceed conventional approaches, so addressing constraints. In order to offer precise and context-aware traffic predictions, this study highlights the potential of CEEMDAN and attention driven models, leading to the development of network management systems that are stronger and more flexible.**

***Keywords—* Deep learning, Network traffic prediction, attention mechanism, gated recurrent unit, temporal convolutional network.**

I. INTRODUCTION

Due to the fact that society is becoming more and more digital, network communication has emerged as an essential component in the process of developing a smart and linked future. As of October 2022, there are 5.07 billion internet users, up 171 million from the previous year, according to the Global Cyber Statistics Report. This figure accounts for roughly 63.5 % of the worldwide population. The amount of data generated and received via the internet has risen owing to the exponential development in internet use. Simultaneously, the growth of 5G connectivity, cloud computing, and the IoT has expedited the development of network data services, necessitating more bandwidth from network base stations. The expansion has brought about several major issues, such as more network congestion, higher latency, and uneven distribution of resources.

In order to improve service quality, network traffic analysis is a must. Reliable traffic trend analysis and forecasting are cornerstones of intelligent base station scheduling and self-management. Having the ability to anticipate traffic peaks enables proactive steps to be taken, such as raising the transmission power of base stations, in order to reduce congestion. On the other hand, predicting traffic troughs permits power reductions and hibernation for certain base stations, which in turn reduces the amount of energy that particular base stations consume. Consequently, improving the quality of communication network services relies heavily on precise network traffic forecasts.

The findings of recent study suggest that it is possible to forecast network traffic and that it displays temporal connections. On the other hand, forecasting network traffic continues to be difficult because of the inherent instability and complexity of the network, as well as features such as nonlinearity and unpredictability. Recent research has concentrated on finding solutions to these problems, with special attention to the instability and time-series correlations that are present in network traffic.

Deep learning, famous for its ability to identify nonlinear features, is one of the best ways to accomplish this job. Popular models for time-series data include Gated Repetitive Units (GRUs) and Very Short-Term Memory (LSTMs), and are good at capturing long-term relationships. Moreover, TCNs,

which make use of causal convolutions, provide considerable advantages for the study of time series. The CEEMDAN-TGA technique is a hybrid deep learning approach that combines these models utilizing a device for attention. Improving the precision of feature extraction and prediction is its intended use.

Using the CEEMDAN technique, network traffic data is first deconstructed into various modes. These modes are then rebuilt into trend and noise sequences in order to enhance

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

denoising. This methodology is known as the CEEMDAN-TGA approach. The following stage is to employ a hybrid model integrating TCN and GRU and extract both short-term as long-term characteristics from the data underlying the network's traffic. From the various ways the attention function fine-tunes the prediction is by modifying the model's weights. We compare CEEMDAN-performance TGA's to that of well-established baseline methods to show that it is effective.

This study makes a number of important advances, including the following:

- Introducing the CEEMDAN method for adaptive decomposition and noise reduction in network traffic data, addressing mode mixing and white noise errors.
- Creating a hybrid deep discovering model (TCN-GRU-Attention) for helpful multi-feature extraction that can gather both quick and long-term characteristics. - - improving the efficacy and stability of model training by hyperparameter adjusting using the Bayesian Optimization Algorithm (BOA). We recommend an organizing technique that is based on this method to facilitate future work, and we conduct investigations with real base the station databases to validate the accomplishment of the CEEMDAN-TGA model. The following remains the outline for the rest of the paper after that:
- An outline of relevant research on predicting network traffic is given in the second part. In Section III, we go into detail about the CEEMDAN-TGA method.

Presented in Section IV are the findings of experiments conducted on actual datasets. In Section V, we present various solutions for energy conservation and base station scheduling. The conclusion of the study includes a summary as well as a discussion of potential future research directions.

II. RELATED WORK

With strong self-similarity and long-term interdependence, network traffic displays time-series properties [13], [14]. Numerous fields, including economics, the natural sciences, and network administration, stand to benefit greatly from solving the time-series analytic challenge of predicting network traffic [15], [16]. Network operators can improve network efficiency, optimize resource scheduling, and guarantee reliable Quality of Service (QoS) via accurate traffic forecast.

Many techniques for predicting network traffic have been developed by researchers over the years. These methods can be generally classified into two categories: those using conventional linear models and those using nonlinear neural networks.

Classical linear methods, such as naive statistical approaches and parametric techniques, analyze historical data but struggle with the burstiness and nonlinear characteristics of traffic. Examples of well-liked parametric models are AR (Autoregressive) [17], MA (Moving Average) [18], and various combinations of the two, such as ARMA [19] and ARIMA [20].

While these models are theoretically robust, they fail to capture the self-similarity and long-term correlations inherent in network traffic, and they are limited in handling nonlinear dynamics [21].

On the other hand, nonlinear approaches have been more popular for network traffic prediction since the introduction of deep learning, which provides better capabilities for collecting complicated patterns. Models trained using neural networks, such as CNNs [22] and RNNs [23], have shown exceptional performance in applications involving computer vision and NLP.

For traffic prediction, researchers employ techniques like To capture long-term dependencies, architectures based on RNNs, Multilayer Perceptrons (MLP), Stacked Autoencoders (SAE), and Support Vector Regression (SVR) [24], [25]. For problems like training-time vanishing gradients, two RNN versions that stand out are the LSTM plus Gated Recurrent Units (GRU) models. Regarding the acquisition of long-term dependencies, GRU's simplified form and fewer parameters really shine. Transient Convolutional Networks (TCN), leveraging dilated causal convolutions and residual connections, have also shown excellent performance in time-series feature extraction [26], [27].

Signal decomposition methods further enhance prediction accuracy by isolating critical components of the data. Improving feature extraction is the goal of methods, including EMD [40], EEMD [28], and VMD [29], among others. These techniques transform non-stationary sequences into components with a stationary distribution. While these methods have their applications, they also come with some drawbacks, including sensitivity to parameters and mixing modes [30]-[34]. Complementary Collective Empirical Mode Decomposition using Adaptive Noise (CEEMDAN) is able to overcome these challenges by adaptively dividing sequences, which effectively isolates noise and gives clearer features. difficulties [35].

Hybrid models combining multiple approaches have emerged to address the limitations of single-method solutions. For example, integrating LSTM with attention mechanisms leverages LSTM's temporal feature extraction capabilities and attention's ability to adjust hidden state weights, improving prediction accuracy [36]. Combining CNN, LSTM, and other components has also yielded significant performance improvements [37]. Methods incorporating signal

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

decomposition, such as wavelet-based LSTM models [38] or TCN-LSTM hybrids with preprocessing filters [39], have further demonstrated the benefits of multi-method approaches.

To address noise and limited feature extraction in existing methods, we propose CEEMDAN-TGA, combining CEEMDAN, TCN, GRU, and attention mechanisms. GRU replaces LSTM for simpler structure and faster training while preserving long-term dependency modeling. TCN captures short-term patterns such as bursts and periodicity via 1D convolution. The attention mechanism optimizes hidden state weighting to minimize information loss.

III. METHODOLOGY

This part goes over the CEEMDAN-TGA in great length together with the fundamental ideas and benefits of every module.

A. CEEMDAN

Common examples of non-stationary data include sequences of network traffic, which display nonlinear, time-dependent, and random properties. The reliability of traffic forecasts for networks is greatly impacted by these characteristics. Decomposing In order to increase prediction accuracy and limit the influence of noise, this research employs the Complementary Experimental Mode Decomposition Ensemble with Adaptive Distortion (CEEMDAN) to transform nonlinear and irregular, network traffic sequences towards stationary components.

Noise is then identified based on the characteristics of the decomposed components, enabling the reconstruction of noise and trend sequences. The trend sequence is subsequently utilized for prediction tasks.

According to earlier studies, CEEMDAN is based on EMD [41] and EEMD [28], which are the bases of Empirical Mode Decomposition (EMD). By breaking them down into their Intrinsic Mode Function (IMF) components, EMD successfully captures changes in trend across different time scales for non-stationary data sets. However, mode mixing does occur often in EMD. Negative and positive Gaussian white noise pairs are introduced to the data by EEMD to address this issue and reduce modal mixing. Nonetheless, residual white noise in EEMD can still impact the data.

In order to optimize this approach even further, CEEMDAN extracts each order's IMF and then calculates the overall average by adding adaptive white noise in pairs of equal magnitude and opposite sign. In addition to fixing incomplete decomposition, this approach reduces computational load dramatically while achieving reconstruction errors close to zero [42].

Initially, let's define Ei(·) where Ci(·) represents the ith iteration of the inside maximum frequency (IMF) after EMD breakdown, wj represents the sequence of normal-distributed Gaussian white noise, j represents the iteration of white noise adding times, as well as ε represents the value of the white noise parameter. Here are the primary procedures that CEEMDAN adheres to:

Applying Gaussian white noise to the initial signal x(t) yields the function x′ (t) = x(t) + (−1)mεwj, with m is either 1 or 2. The new signal is decomposed using EMD to get the initial IMF. Calculating C1j (t) is as follows:

$$E(x'(t)) = C_1^j(t) + r^j \tag{1}$$

To acquire the first IMF, first calculate all of its components individually. Then, take the overall average and calculate C1(t). The formula for computation, CEEMDAN, is

$$\widetilde{C_1(t)} = \frac{1}{N}\sum_{j=1}^{N} C_1^j(t) \tag{2}$$

Finally, determine the first stage's residual r1(t).

$$r_1(t) = x(t) - \widetilde{C_1(t)} \tag{3}$$

Fourthly, the residual r1(t) is re-added with The initial IMF Dj1 (t) is computed by EMD after a fresh signal is generated using both sides of pair Gaussian white noise. At long last, CEEMDAN calculates the second IMF C2(t) by averaging the N IMFs that were generated.

$$\widetilde{C_2(t)} = \frac{1}{N}\sum_{j=1}^{N} D_1^j(t) \tag{4}$$

Determine the second stage's residual r2(t).

$$r_2(t) = x(t) - \widetilde{C_2(t)} \tag{5}$$

Keep going until you have a monotonic function as the residual signal, and then stop decomposing. We can decompose the original signal x(t) as (6) if the number of IMFs acquired is M.

$$x(t) = \sum_{m=1}^{M} \widetilde{C_m(t)} + r_m(t) \tag{6}$$

Reconstructing the trend component using the properties of each of the signal's stationary components follows CEEMDAN decomposition, which first isolates the non-stationary signal into numerous stationary components. The denoised network traffic stream is obtained from the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

reconstructed signal. When utilizing CEEMDAN for denoising, you may improve prediction accuracy and lessen the effect of noise.

B. TCN

Time sequence prediction makes heavy use of TCN due as it outperforms other recursive structures in terms of memory capacity [27] and time sequence modeling. An new one-dimensional convolutional network called Tri-Convolutional Network (TCN) combines dilated convolution, residual block, and causal convolution.

Figure 1 depicts the TCN architecture, which utilizes network traffic history to its fullest potential.

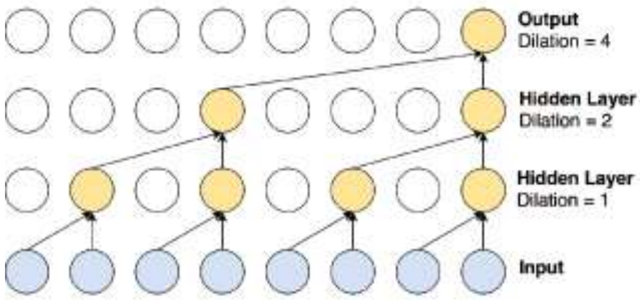


Fig. 1. TCN structure diagram.

A. Causal convolution

Layer values at time t in a causal convolution network are simply dependent on the values of the layers that came before and after it.Because the causation relationship is perfectly followed by the link as well as the data transfer across the various levels of the network, the latter takes use of data stored in the past. Increasing the number of convolutional neural network layers to adjust for gradient disappearance, complicated training, and inadequate fitting degree leads to issues. for enough historical data.

B. Dilated convolution

If you find that causal convolution has too many layers, diluted convolution may be able to help. A bigger By increasing the convolution region of view by diluted convolution, we may collect more historical information with fewer layers of networking and less computation, expanding the receptive field. Here is the formula for determining dilated convolution:

$$F(s) = (x \odot f)(s) = \sum_{i=0}^{k-1} f(i) x_{s-d\cdot i} \qquad (7)$$

The parameters for the convolution process are ⨀, Both the dilation factor (d) and the size (k) of the gaussian kernel are important parameters.

C. Residual block

To avoid the issues of data form and dimension change caused by convolution, as well as the disappearance and explosion of gradients caused by deep networks, one-dimensional convolution is suggested as the residual block for deeper models. After two rounds of dilated causal convolution, weight normalization, activation function, and dropout layer, the final output result is communicated to the next layer using one-dimensional convolution as the residual module for skip connection. The construction of the residual block and the formula for the residual connection are shown in Figure 2.

$$f_{output} = ReLU(x + f(x)) \qquad (8)$$

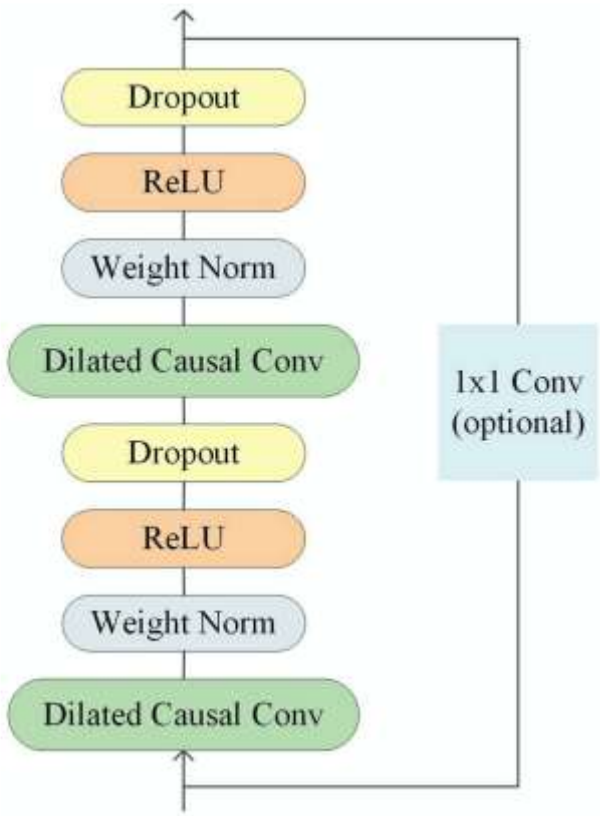


Fig. 2. Residual block structure diagram

D. GRU

To improve upon RNN and LSTM, the GRU algorithm was developed. The GRU builds upon the LSTM's gate control structure by combining the input and forgetting gates into a revision gate (zt%) and repurposing the output gate as a reset gate (rt%) [43]. The update gate measures the quantity of

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

data retained by ram from the previous instant to the present moment, whereas the reset gate measures the total quantity of historical data required to be wiped. You can see the GRU's structural diagram in Figure 3.

Compared to LSTM, GRU offers faster calculation speed, fewer parameters, a simpler structure, and less internal unit redundancy; it is also more in line with the timeliness criterion in the field of network traffic prediction.

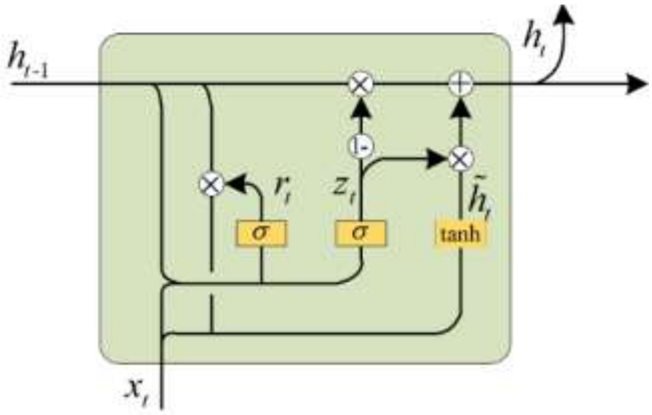


Fig. 3. GRU structure diagram.

The phrase that describes how the GRU works is:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (9)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (10)$$

$$\widetilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \quad (12)$$

At time t, the input is x, the output is ht−1, the activation function is σ, the weight matrix is W, the reset gate is rt, and huntu is the candidate hidden state.

E. Attention mechanism

Passing lengthy sequences of traffic data from the network via GRU may result in diminished impact since all relevant information cannot be completely considered and stored. The attention mechanism, when combined with GRU, may efficiently filter inputs to increase prediction accuracy. Improved prediction accuracy may be achieved by the use of an attention mechanism, which regularly evaluates the relevance of information and focuses on essential components [46].

The attention process filters away superfluous information, among other things. Reducing the weight actively removes data with poor correlation, while increasing it improves the data with strong correlation. Due to the strong daily, weekly, and monthly periodicity in network traffic data series, it is crucial to record the periodicity and supplementary information in order to enhance the accuracy of predictions. Suppressing features with poor or irrelevant correlation—a consequence of bursty information and noise—improves the approach's overall performance and efficiency.

A. CEEMDAN-TGA

We introduce the CEEMDAN-TGA, a hybrid model that integrates the denoising strength of CEEMDAN with TCN's short-term feature extraction, GRU's long-term dependency capturing, and attention mechanism's weight assignment capability, to optimize hidden feature extraction and reduce noise effects. The attention mechanism, CEEMDAN, TCN, and GRU are all seamlessly integrated in the CEEMDAN-TGA design. Signal decomposition was the first step that CEEMDAN used to identify and eliminate noise. Next, we use the attention mechanism to fine-tune the weights, and then we merge the benefits of the TCN and GRU nonlinear models to accomplish sufficient feature extraction. The fully linked layer generates accurate network traffic forecasts by mapping collected long- and short-term contextual data to real traffic values. Looking at Figure 4, you can see the CEEMDAN-TGA technique.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

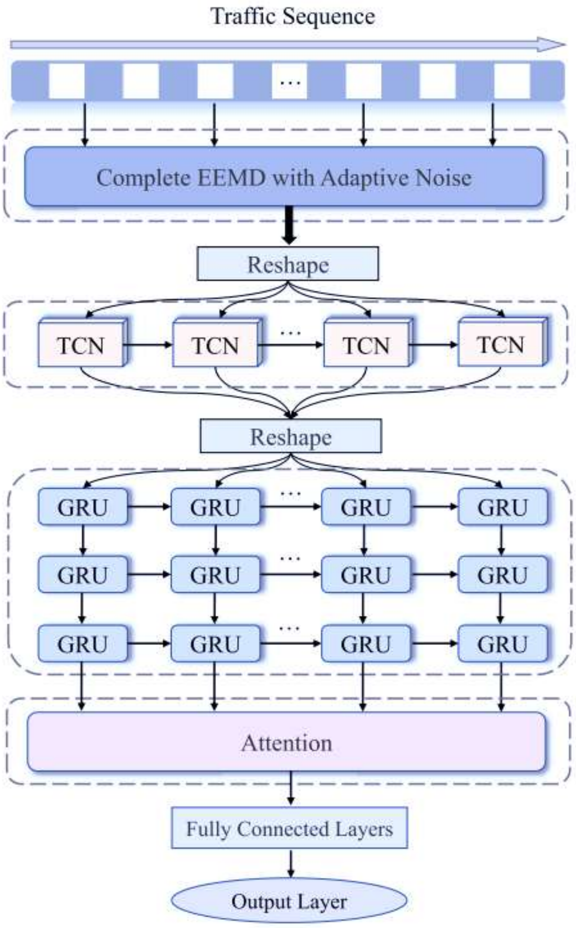


Fig. 4. CEEMDAN-TGA structure diagram

Our feature extraction method is a hybrid one, since network traffic features exhibit both time-dependent long-term dependencies and time-dependent short-term local dependencies. The hidden elements of various methods can be investigated with the use of appropriate training and learning procedures, since different approaches have different features and benefits. Because it uses convolutional networks, which are good at capturing short-term local data information, TCN is suitable for network traffic prediction in most cases[45] and its receptive field may be modified dynamically. Due to their features, GRU's small parameters and rapid convergence make it easier to manage the timely and unexpected nature of network data. The network traffic forecast problem is best handled by a mixture of TCN and GRU, since they can extract both characteristics concurrently.

IV. RESULT ANALYSIS

The CEEMDAN-TGA method is tested on real base station data to evaluate its performance.

A. Data Set Description

The dataset used in this study comes from the "AIIA Home Network Competition: Network Traffic Forecasting," co-organized by China Mobile and the China Artificial Intelligence Industry Development Alliance [47]. It includes hourly traffic data from three anonymized base stations between January 2017 and November 2018. This experiment uses a subset—data from January 1 to March 31, 2017—totaling 2160 hourly entries. To capture temporal patterns, a 24-hour input window is set for forecasting, with an 80:20 split into training (1728 samples) and testing (432 samples). The TGA model is trained and tuned using Bayesian optimization, and evaluation is conducted on the test set.

The traffic exhibits strong daily periodicity, peaking between 11:00–23:00 and dipping from 03:00–08:00. Burstiness appears in short-term spikes and during events such as the Chinese New Year (samples 620–790), where traffic surges due to increased usage. The data is inherently nonlinear, reflecting complex and dynamic network behavior [47].

B. Data Preprocessing

Data preprocessing involves normalization and denoising. Due to the complexity of network traffic, Min-Max normalization is applied to scale the data to [0,1], improving training convergence and mitigating gradient vanishing issues [1]. The transformation maps each value Xi using the formula:

$$X_i^* = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{13}$$

Next, CEEMDAN decomposes the normalized series into Intrinsic Mode Functions (IMFs) and a residual. In this study, 8 IMFs and 1 residual are generated. These IMFs represent distinct characteristics: IMF1 shows noise, IMF2–4 show periodicity, and IMF4–8 along with the residual reveal trend components [2].

To identify and remove noise, combinations of IMFs are tested by eliminating suspected noisy ones, reconstructing the remaining components, and evaluating prediction performance using a hybrid deep learning model. The optimal outcome is achieved when IMF1 is excluded—confirming it as the primary noise carrier—while IMF2–8 and the residual are retained as trend signals [3].

The denoised data, reconstructed from the retained components, is then used for training and forecasting with the Temporal Graph Attention (TGA) model.

C. Evaluation metrics

The effectiveness of network traffic prediction methods is typically assessed using error metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

R-squared coefficient (R^2). Lower RMSE and MAE values indicate more accurate predictions, while a higher R^2 reflects better model fit [4].

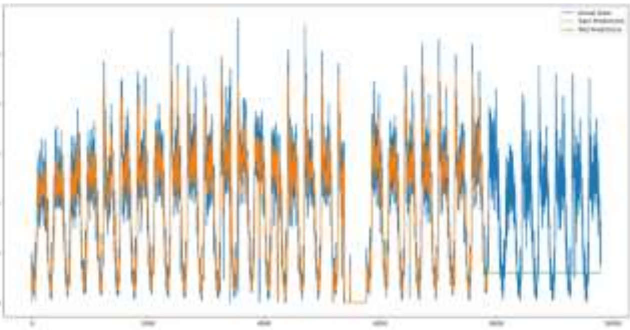
$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|x_i - x_i'\right|^2} \tag{14}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - x_i'\right| \tag{15}$$

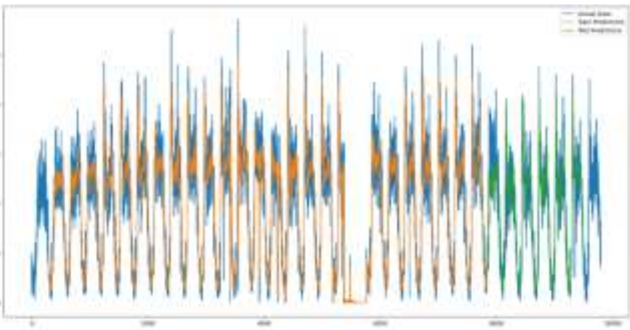
$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - x_i')^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{16}$$

D. Prediction result analysis

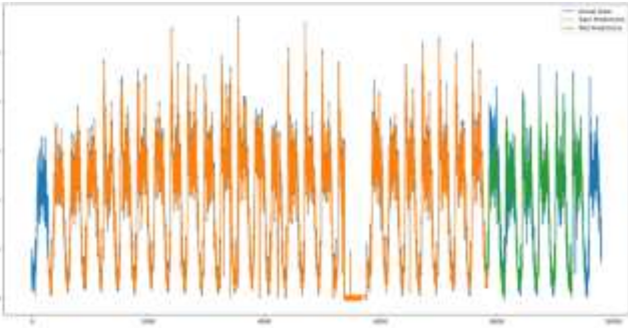
The experiment evaluates prediction performance through both visual (Figure 5) and quantitative analyses. Visually, the CEEMDAN-TGA model closely follows the real traffic trends, outperforming other models with clearer accuracy and better fit [1]. Quantitatively, CEEMDAN-TGA yields the lowest RMSE and MAE and highest R^2, indicating superior accuracy [2].



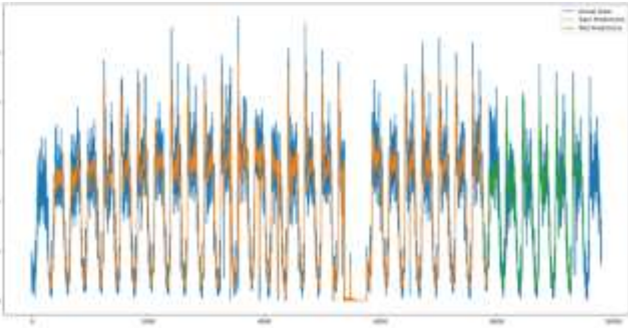
(a) ARIMA



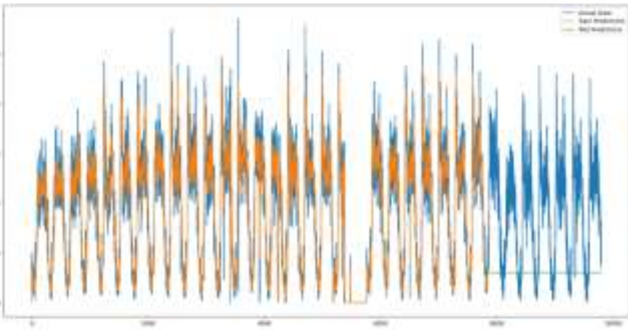
(b) XGBoost



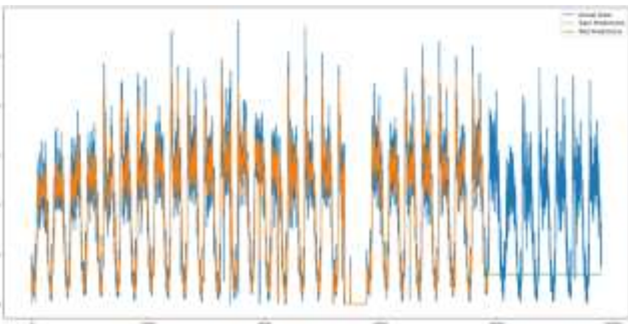
(c) SVM



(d) LSTM



(e) TCN Model



(f) GRU Model

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Fig. 5. Predicted values of CEEMDAN-TGA compared to the baseline method.

With CEEMDAN denoising, CEEMDAN-TGA achieves RMSE of 0.05149, MAE of 0.04207, and R^2 of 0.9828. Without denoising, TGA's RMSE and MAE increase to 0.05856 and 0.04306, and R^2 drops to 0.9777—showing 12.07% RMSE and 2.3% MAE reductions, and a 0.52% R^2 improvement [3]. This confirms CEEMDAN's effectiveness in noise reduction.

Compared to advanced models like 3SG+TCN+LSTM, SSA+LSTM, and Wavelet+LSTM, CEEMDAN-TGA shows a 7.72% lower RMSE, 2.12% lower MAE, and 0.31% higher R^2. Against traditional models (XGBOOST, SVR, ARIMA), its advantage is even more pronounced [4].

CEEMDAN-TGA's time complexity is $O(n^2)$, mainly due to CEEMDAN and GRU-Attention, while TCN contributes linear complexity. Optimized with Bayesian tuning, training and prediction times are 280.11s and 0.1086s, respectively—suitable for practical use [5].

The model's strength lies in TCN's ability to capture short-term spikes, GRU's effectiveness with long-term patterns, CEEMDAN's noise suppression, and the attention mechanism's feature filtering. Together, these enhance both accuracy and robustness [6].

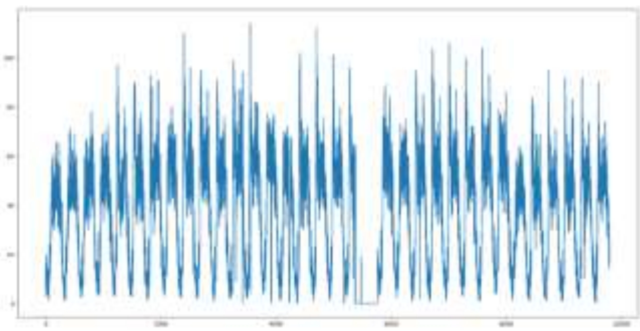


Fig. 6. Traffic Flow Over Time

In Figure 5 plot shows the variation in traffic volume over time. It highlights clear patterns of peak and off-peak hours, indicating regular daily traffic cycles—useful for time-series forecasting models.

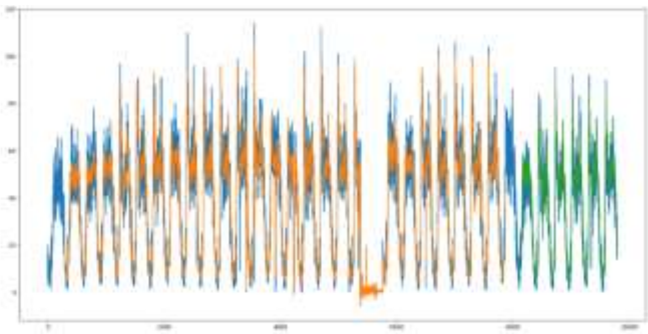


Fig. 7. Actual vs Predicted Traffic Flow (Train and Test)

In Figure 6 plot visualizes the actual traffic volume alongside model predictions for both training and testing periods. The actual data is shown as a continuous line, while the predictions are overlaid with proper time shifts to match their real-time positions. This helps assess the model's accuracy across the entire time series.

V. CONCLUSION

Base station anomaly detection, intelligent scheduling, and congestion control all rely on accurate prediction of network traffic. Extensive feature extraction and effective denoising are necessary for precise prediction because of the complexity of network traffic patterns. Our CEEMDAN-TGA is a deep learning hybrid that incorporates an attention mechanism, GRU, and TCN; it has been evaluated in this way. Improving prediction accuracy and resilience is achieved by denoising and decomposing the data using CEEMDAN, extracting local features using TCN, capturing dependencies using GRU, and adjusting information weights using the attention method. The results of the experiments demonstrate that CEEMDAN successfully reduces noise.

and that the CEEMDAN-TGA model has substantial performance benefits according to quantitative and qualitative assessments. These precise forecasts serve as a springboard for further investigation into base station scheduling and energy savings. There will be two primary areas of future study. Finding a reliable metric to differentiate between trend sequences and noise sequences is the first step towards eliminating all noise. Secondly, taking a more methodical approach to the problem by delving further into the topic of network traffic scheduling in order to integrate prediction and scheduling thoroughly.

REFERENCES

[1] K. Sundaresan, ''5G: An evolution towards a revolution,'' in Proc. 24th Annu. Int. Conf., Oct. 2018, p. 659.

[2] A. Imran, A. Zoha, and A. Abu-Dayya, ''Challenges in 5G: How to empower SON with big data for enabling 5G,'' IEEE Netw., vol. 28, no. 6, pp. 27–33, Nov. 2014.

[3] W. Xu, G. Zhang, and B. Bo, ''Ten key ICT challenges in the post-Shannon era,'' Scientia Sinica Mathematica, vol. 51, no. 7, pp. 1095–1138, May 2021.

[4] M. Kang, J. Song, and P. Fan, ''Survey of network traffic forecast based on deep learning,'' Comput. Eng. Appl., vol. 57, no. 10, pp. 1–9, Oct. 2021.

[5] W. Jiang, ''Cellular traffic prediction with machine learning: A survey,'' Exp. Syst. Appl., vol. 201, Sep. 2022, Art. no. 117163.

[6] C. Pan, Y. Wang, H. Shi, J. Shi, and R. Cai, ''Network traffic prediction incorporating prior knowledge for an intelligent network,'' Sensors, vol. 22, no. 7, p. 2674, Mar. 2022.

[7] N. Jiang, Y. Deng, O. Simeone, and A. Nallanathan, ''Online supervised learning for traffic load prediction in framed-ALOHA networks,'' IEEE Commun. Lett., vol. 23, no. 10, pp. 1778–1782, Oct. 2019.

[8] Y.Ji,Y.Wu,D.Zhang,Y.Yuan,S.Liu,R.Zarei,andJ.He,''Anovelflash P2P network traffic prediction algorithm based on ELMD and garch,'' Int. J. Inf. Technol. Decis. Making, vol. 19, no. 1, pp. 127–141, Jan. 2020.

[9] C. Katris and S. Daskalaki, ''Dynamic bandwidth allocation for video traffic using FARIMA-based forecasting models,'' J. Netw. Syst. Manage., vol. 27, no. 1, pp. 39–65, Jan. 2019.

[10] A. Sang and S.-Q. Li, ''A predictability analysis of network traffic,'' Comput. Netw., vol. 39, no. 4, pp. 329–345, Aug. 2002.

[11] F. Xu, Y. Lin, and J. Huang, ''Big data driven mobile traffic understanding and forecasting: A time series approach,'' Services Comput., vol. 9, no. 5, pp. 796–805, Sep. 2016.

[12] L. Nie, X. Wang, S. Wang, Z. Ning, M. S. Obaidat, B. Sadoun, and S. Li, ''Network traffic prediction in industrial Internet of Things backbone networks: A multitask learning mechanism,'' IEEE Trans. Ind. Informat., vol. 17, no. 10, pp. 7123–7132, Oct. 2021.

[13] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, ''Long-range dependence in variable-bit-rate video traffic,'' IEEE Trans. Commun., vol. 43, no. 2, pp. 1566–1579, Feb. 1995.

[14] W.E.Leland,M.S.Taqqu,W.Willinger,andD.V.Wilson,''Ontheself-similar nature of Ethernet traffic (extended version),'' IEEE/ACM Trans. Netw., vol. 2, no. 1, pp. 1–15, Feb. 1994.

[15] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, ''Financial time series forecasting with deep learning : A systematic literature review: 2005–2019,'' Appl. Soft Comput., vol. 90, May 2020, Art. no. 106181.

[16] S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, ''Time series forecasting of COVID-19 using deep learning models: India-USA comparative case study,'' Chaos, Solitons Fractals, vol. 140, Nov. 2020, Art. no. 110227.

[17] Z.Zhang,X.Li,andS.Yang,''Short-termpredictionofsignificantwave height based on AR-SVR model,'' Acta Energiae Solaris Sinica, vol. 42, no. 7, pp. 15–20, Aug. 2021.

[18] G.U.Yule,''Onamethodofinvestigatingperiodicitiesindisturbedseries, with special reference to Wolfer's sunspot numbers,'' Philos. Trans. Roy. Soc. London A, Containing Papers Math. Phys. Character, vol. 226, no. 1927, pp. 267–298, Apr. 1927.

[19] Z.Duan,E.Sun,andY.Zhang,''Thepredictionofnetworktrafficbased on ARMA,'' J. China Acadewy Electoronics Inf. Technol., vol. 4, no. 4, pp. 352–356, Aug. 2009.

[20] R.ZhangandC.Zhao,''Applicationresearchonnetworktrafficprediction base on ARIMA,'' Comput. Simul., vol. 28, no. 2, pp. 171–174, Feb. 2011.

[21] Y. Shu, Z. Jin, and L. L. Zhang, ''Traffic prediction using FARIMA models,'' in Proc. IEEE Int. Conf. Commun. Vancouver, BC, Canada, Jun. 1999, pp. 891–895.

[22] H. Shin, H. Roth, and M. Gao, ''Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,'' IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1285–1298, May 2016.

[23] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, ''Learning phrase representations using RNN encoder–decoder for statistical machine translation,'' 2014, arXiv:1406.1078.

[24] X.Chen,Y.Liu,andJ.Zhang,''TrafficpredictionforInternetofThings through support vector regression model,'' Internet Technol. Lett., vol. 5, no. 3, p. 336, May 2022.

[25] T. P. Oliveira, J. S. Barbar, and A. S. Soares, ''Computer network traffic prediction: A comparison between traditional and deep learning neural networks,'' Int. J. Big Data Intell., vol. 3, no. 1, pp. 28–37, 2016.

[26] W. Zhao, Y. Gao, T. Ji, X. Wan, F. Ye, and G. Bai, ''Deep temporal convolutional networks for short-term traffic flow forecasting,'' IEEE Access, vol. 7, pp. 114496–114507, 2019.

[27] S. Bai, J. Z. Kolter, and V. Koltun, ''An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,'' 2018, arXiv:1803.01271.

[28] Z. Wu and N. E. Huang, ''Ensemble empirical mode decomposition: A noise-assisted data analysis method,'' Adv. Adapt. Data Anal., vol. 1, no. 1, pp. 1–41, Jan. 2009.

[29] K. Dominique and D. Zosso, ''Variational mode decomposition,'' IEEE Trans. Signal, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[30] Y. Zhao, Nurbol, and S. Whshour, ''Prediction method of network traffic based on timing EEMD,'' Comput. Simul., vol. 35, no. 11, pp. 466–469, Nov. 2018.

[31] Y. Li, B. Tang, and S. Jiao, ''Optimized ship-radiated noise feature extraction approaches based on CEEMDAN and slope entropy,'' Entropy, vol. 24, no. 9, p. 1265, Sep. 2022.

[32] Y. Li, B. Tang, X. Jiang, and Y. Yi, ''Bearing fault feature extraction method based on GA-VMD and center frequency,'' Math. Problems Eng., vol. 2022, pp. 1–19, Jan. 2022.

[33] M. Civera and C. Surace, ''A comparative analysis of signal decomposition techniques for structural health monitoring on an experimental bench- mark,'' Sensors, vol. 21, no. 5, p. 1825, Mar. 2021.

[34] Z. Li, L. Gao, W. Lu, D. Wang, H. Cao, and G. Zhang, ''A novel noise suppression and artifact removal method of mechanomyography based on RLS, IGWO-VMD, and CEEMDAN,'' J. Sensors, vol. 2022, Sep. 2022, Art. no. 4239211.

[35] Y. Li, B. Tang, and Y. Yi, ''A novel complexity-based mode feature representation for feature extraction of ship-radiated noise using VMD and slope entropy,'' Appl. Acoust., vol. 196, Jun. 2022, Art. no. 108899.

[36] M.Li,Y.Wang,Z.Wang,andH.Zheng,''Adeeplearningmethodbasedon an attention mechanism for wireless network traffic prediction,'' Ad Hoc Netw., vol. 107, Oct. 2020, Art. no. 102258.

[37] H. Xiong, J. Liu, and J. Wang, ''Network traffic prediction based on neural network and autoregressive model,'' J. Comput. Appl., vol. 41, pp. 180–184, Jun. 2021.

[38] N. W. Branco, M. S. M. Cavalca, S. F. Stefenon, and V. R. Q. Leithardt, ''Wavelet LSTM for fault forecasting in electrical power grids,'' Sensors, vol. 22, no. 21, p. 8323, Oct. 2022.

[39] J. Bi, X. Zhang, and H. Yuan, ''A hybrid prediction method for realistic network traffic with temporal convolutional network and LSTM,'' IEEE Trans. Autom. Sci. Eng., vol. 99, no. 3, pp. 1–11, May 2021.

[40] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, ''A complete ensemble empirical mode decomposition with adaptive

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

noise,'' in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2011, pp. 4144–4147.

[41] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, ''The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,'' Proc. Roy. Soc. London A, Math., Phys. Eng. Sci., vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[42] F. Zhou, Z. Huang, and C. Zhang, ''Carbon price forecasting based on CEEMDAN and LSTM,'' Appl. Energy, vol. 311, Apr. 2022, Art. no. 118601.

[43] L. Yin and Y. Wu, ''Traffic flow combination prediction model based on improved VMD-GAT-GRU,'' J. Electron. Meas. Instrum., vol. 36, no. 7, pp. 62–72, Jul. 2022.

[44] Y.Bin,Y.Yang,F.Shen,N.Xie,H.T.Shen,andX.Li,''Describingvideo with attention-based bidirectional LSTM,'' IEEE Trans. Cybern., vol. 49, no. 7, pp. 2631–2641, Jul. 2019.

[45] S. Zhou, C. Song, T. Wang, X. Pan, W. Chang, and L. Yang, ''A short-term hybrid TCN-GRU prediction model of bike-sharing demand based on travel characteristics mining,'' Entropy, vol. 24, no. 9, p. 1193, Aug. 2022.