QUBAHAN ACADEMIC JOURNAL
VOL. 5, NO. 2, April 2025
https://doi.org/10.48161/qaj.v5n2a1690

Improvement of Network Traffic Prediction in Beyond 5G Network using Sparse Decomposition and BiLSTM Neural Network

Rihab Abdullah Jaber Al Hamadani [1], Mahdi Mosleh [1*], Ali Hashim Abbas Al-Sallami [1], and Rasool Sadeghi[2]

[1] Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan 81595, Iran;
[2] Department of Electrical Engineering, Dolatabad Branch, Islamic Azad University, Isfahan 81595, Iran.

* **Corresponding author:** mahdi_mosleh@yahoo.com.

ABSTRACT: Companies providing telecommunication services, especially in Beyond 5G networks, are increasingly interested in traffic forecasting to improve the services provided to their users. However, forecasting network traffic is challenging due to traffic data's dynamic and non-stationary nature. This study proposes an effective deep learning-based traffic prediction technique using BiLSTM (Bidirectional Long Short-Term Memory). The proposed method begins with preprocessing using K-SVD (K-means Singular Value Decomposition) to reduce dimensionality and enhance data representation. Next, sparse feature extraction is performed using Discrete Wavelet Transform (DWT), and a sparse matrix is constructed. A Genetic Algorithm (GA) is used to optimize the sparse matrix, which effectively selects the most significant features for prediction. The optimized sparse matrix is fed into the BiLSTM model for accurate traffic forecasting. Experimental results show that the proposed method significantly reduces Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) while achieving higher accuracy (ACC) compared to traditional neural networks. The results demonstrate that the proposed sparse matrix, integrated with BiLSTM, provides superior prediction accuracy and better generalization, making it a robust solution for network traffic forecasting in Beyond 5G networks.

Keywords: forecasting, consumer traffic, cellular networks, re-current neural networks, memory.

I. INTRODUCTION

The rapid growth of cellular networks and the increasing demand for high-speed data services have made accurate network traffic prediction a crucial challenge for operators. With global mobile data traffic expected to increase significantly by 2030, efficient forecasting methods are essential for resource allocation, congestion management, and Quality of Service (QoS) optimization [1]. End users inside the base station’s coverage area in a cellular network can access calls, messages, and data services. Cellular networks offer services according to entirely distinct user needs [2, 3]. The increase in traffic consumption in cellular networks is a main challenge for cellular network operators in managing large network traffic and, at the same time, increasing Quality of Service (QoS). Cellular network operators may anticipate total network utilization and allocate re-sources appropriately by using accurate and precise traffic forecasting. In a cellular network with varying user counts, accurate fundamental traffic forecast may significantly aid in anticipating when network congestion would develop [4]. Moreover, modeling and forecasting of mobile network traffic in cellular networks can help telecommunication Businesses should look for methods to raise network quality of service [5, 6].

In order to aid with the appropriate distribution of resources (power and spectrum) in the cellular network and to lower network operating expenses, traffic forecast is essentially based on hourly data. Operators can modify and improve network settings depending on data consumption in the network by modeling and forecasting network traffic usage [7]. Cellular network traffic prediction is performed in two categories: long-range prediction and short-range prediction [8, 9]. It offers a forecast for a considerable amount of time in long-range prediction. This type of prediction is used to validate, accurately predict, and present network traffic patterns. This type of prediction can help design networks more easily. Short-range forecasting provides forecasts for a short period of time, such as one hour, and can help improve network performance in providing services [10].

In another category of predicting data traffic consumption methods in telecommunication networks are divided into three categories: parametric, non-parametric, and machine learning methods [11]. Parametric methods depend on experimental data. Traditional methods like the autoregression integral moving average model (ARIMA) [12] are in this category. Although these methods are very stable, noise and unwanted or lost data have little effect on it. however, they cannot be used online. Non-parametric methods work without any restrictions on the data, but they need a lot of available data for prediction. Machine learning-based methods such as support vector regression (SVR) [13], Kalman filters [14], neural networks, support vector machine regression (SVMR) [15]and other machine learning methods [16] are in this category. Moreover, deep learning methods such as long-term short-term memory (LSTM) [17] are in this category, although they work with very high accuracy in forecasting, at the same time, they have high computational complexity [18].

Forecasting the traffic in cellular networks depends on the location and time. In linear methods such as ARIMA, ARMA and AR [19] both location and time dependence are considered. However, these traditional methods lose their effectiveness with the increase in the amount of data, the presence of missing data, and the noise in data. Non-linear techniques have been used to increase forecast accuracy and precision. Non-linear methods and models identify the correlation of traffic states better than linear methods. Non-linear transformations are used in a way that creates a suitable mapping between the input space and the feature space with high dimensions, such as the SVR method [20].

Besides the special features of linear and other non-linear methods, they also have major challenges and weaknesses. Determining the dimensions of the input space and the kernel function is both time-consuming and complicated [21]. Therefore, machine learning-based methods such as Support vector machines (SVM) [22] and k-nearest neighbors (KNN) [23] have also been proposed to increase the precision and accuracy of short-term traffic predictions. These methods consider the location of the data, and the prediction speed will be increased. However, it seems that neural networks and their deep learning methods, such as Bayesian networks (BN) [18], are more efficient in considering both spatial and temporal features in traffic data [19].

An artificial neural network is composed of a number of linked neurons. that use a mathematical or computational model to process information [24, 25]. Due to this unique characteristic, the neural network is used to model the complex relationship between output and input, such as nonlinear calculations, pattern recognition, voice recognition, and decision-making [26]. Recurrent Neural Networks (RNN) is a type of deep neural network that has advantages such as accurate prediction in time series, high convergence speed, and high adaptability. An RNN consists of an input layer, a feedback layer that provides state information, a hidden layer, and an output layer. In each layer, one or more neurons can transfer information from one layer to another by calculating the nonlinear functions of their weights [27]. This type of neural network has lower uncertainty and works better than artificial neural networks in modeling due to having a memory unit [28]. In contrast to the standard recurrent neural network, which rewrites the material at each time step, an LSTM recurrent neural network [24], By introducing gates, the network may decide whether to keep the present memory [25, 26]. The LSTM unit may effortlessly send information over a long distance if it intuitively recognizes a significant characteristic in the input sequence in the first steps. As a result, it can acquire and sustain potential long-term dependence. [27]. Therefore, an LSTM neural network can be used in the buffer units in a recurrent network [28].

Even though traffic prediction in cellular networks has been the subject of several studies, it is still a challenging concept. Among the challenges are the complex underlying patterns hidden in the network's historical traffic data, which can only be identified and explained by efficient prediction models that consider the data's correlation. Deep learning models have emerged in recent years. made significant strides in forecasting network traffic, but more work is still required. Another challenge that is commonly overlooked in today's literature is practical implementation. The unresolved gap between developing a high-performance prediction model and putting it into practice in a practical system [29]. Nevertheless, none of the existing research works have exploited it during traffic prediction. One of the key novelties of this study is the integration of K-SVD, Discrete Wavelet Transform (DWT), Genetic Algorithm (GA), and BiLSTM for network traffic prediction in Beyond 5G networks. Unlike conventional methods such as ARIMA, SVR, and LSTM, which individually face challenges like low accuracy, sensitivity to noise, or high computational complexity, our proposed approach enhances data preprocessing through K-SVD and DWT and optimizes sparse feature extraction using GA-based sparse matrix construction. Also, Discrete Wavelet Transform (DWT) offers several advantages in network traffic prediction. Unlike traditional methods, DWT provides multi-resolution time-frequency analysis, allowing it to effectively capture both short-term fluctuations and long-term trends in traffic data. Additionally, DWT enhances feature extraction by reducing noise and filtering irrelevant data, improving machine learning models' accuracy and efficiency, particularly in non-stationary environments like 5G networks. Subsequently, the BiLSTM model, capable of capturing both long-term and short-term temporal dependencies, is employed for improved prediction accuracy. A comparative analysis with state-of-the-art techniques demonstrates that our approach enhances predictive accuracy and reduces error metrics (MAE, MSE, RMSE) compared to traditional models, leading to improved generalization capability. This hybrid methodology significantly advances accurate and efficient network traffic modeling, contributing to optimized resource allocation and enhanced telecommunication services in next-generation networks. The following are this paper's primary contributions:

- Pre-processing step: A clustered K-means Singular Value Decomposition (K-SVD) is applied to enhance data quality and effectively remove noise.
- Sparsity-based feature extraction: The Discrete Wavelet Transform (DWT) and Genetic Algorithm (GA) extract essential features while preserving data integrity.
- BiLSTM network application: The Bidirectional Long Short-Term Memory (BiLSTM) network is a recurrent neural network technique used to capture both short-term and long-term dependencies in time-series data.
- Performance enhancement: Integrating K-SVD and the sparsity-based approach enhances the BiLSTM model's predictive performance, leading to more accurate and reliable network traffic forecasting.

The following are provided in the remainder of this work. Related work on load terrific prediction is described in Section 2. The suggested approach will be used in the prediction model in Section 3. The suggested approach and related assessment measures will be examined in Section 4. Lastly, Section 5 will offer the conclusion.

II. RELATED WORK

Various studies have been presented to predict traffic consumption in cellular networks. The authors of [29] proposed a Gradient Similarity-based Federated Aggregation for Wireless Traffic Prediction (FedGSA). In [30], decision trees (DT), random forest (RF), support vector machines (SVM), and ensemble learning (EL) are used to improve the prediction performance of the network. In [30], the authors focused on the accurate prediction of the background traffic matrix (TM) of a common local area network (LAN) for Network traffic prediction. In [7], an attention mechanism is employed to provide an end-to-end framework with two types of variables. In [30], to predict the traffic in 5G cellular networks, it has been proposed that data analysis be performed in two parts using ensemble learning (EL).

In [31], the DEEXP method has been proposed to predict traffic consumption in the first term. In [32], a detailed prediction of a 5G network was proposed to build a smooth short-term memory traffic prediction model (SLSTM). In [29], a hybrid model based on a deep convolutional neural network is proposed, which combines unidimensional smoosparseg having single-exponential smoosparseg with long short-term memory (SES-LSTM). In [33], a hybrid method based on combining long-term, short-term memory (LSTM) and Gaussian process regression (GPR) was used to achieve accurate cellular traffic prediction at the single-cell level using an open cellular traffic dataset in the city of Milan, Italy. In [34]the prediction of mobile data consumption traffic at the city level is made based on a deep learning approach for modeling the nonlinear dynamics of wireless traffic.

In [35] a spatiotemporal attention-complexity network is proposed to predict the traffic of the cellular telecommunication network in the city. Considering the temporal correlation of cellular traffic, traffic data are independently modeled by hourly, daily, and weekly traffic components. The work of [36] describes a dynamic traffic slice model based on ML (ML-TADS). This model allows the management of traffic in the network properly - to provide it in such a way that its distribution is uniform, there is no congestion in one BS, and at the same time, traffic to another is zero. The work of [37] investigates the effectiveness of different ML models in terms of prediction accuracy and computational time cost. They analyze how to identify critical factors limiting the application of ML-based predictive models to support real-time services.

In [38] proposed a method to improve the traffic prediction accuracy of cellular BS home base using ML machine learning. This system combines the Naive Bayes classifier and HoltWinters ML model to improve traffic prediction in a cloud-based platform. In [39] compared different supervised learning algorithms for traffic flow estimation. Based on the collected data, they evaluate several different prediction and classical regression algorithms, including SVR, Kernel Ridge, Decision Tree, Random Forest, and LSTM. In [40] proposes a federated learning (FL) framework for mobile traffic prediction in satellite-terrestrial networks, where AGCN and LSTM are used to train local models at base stations and aggregate a global model on a satellite edge server. This approach preserves data privacy and improves scalability but requires a distributed infrastructure and synchronization between nodes. In contrast, our method focuses on enhancing input data quality through the integration of K-SVD, DWT, and GA with BiLSTM, leading to higher prediction accuracy and reduced computational complexity. Consequently, FL is beneficial for privacy-sensitive applications, while our approach is optimized for centralized, high-accuracy forecasting in Beyond 5G networks.

In [41] introduces a multimodal deep learning framework for mobile traffic prediction by integrating CNN (ConvLSTM) and GNN (AGCN) to process SMS, call, and internet data. Their approach improves prediction accuracy by fusing grid-based and graph-based representations, outperforming ten baseline models. In contrast, our method focuses on sparse feature selection (K-SVD & DWT) and GA-based optimization to enhance data quality and reduce computational complexity. While multimodal models are beneficial for diverse applications, our approach is optimized for high-accuracy traffic forecasting in Beyond 5G networks. Table 1 shows the comparison of related work.

Table 1. Comparison of related work.

Study Number	Methods and Algorithms	Research Objective	Data Type	Key Features and Innovations
[29]	Gradient Similarity-based Federated Aggregation (FedGSA)	Traffic consumption prediction in cellular networks	Wireless network traffic data	Use of federated data and gradient similarity for traffic prediction
[30]	Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), Ensemble Learning (EL)	Improving network traffic prediction performance	Network traffic data	Use of various algorithms to enhance prediction accuracy
[42]	Ensemble Learning (EL)	Accurate prediction of background traffic matrix in local area networks (LAN)	LAN traffic data	Focus on accurate prediction of traffic matrix in local networks
[7]	Attention Mechanism	Traffic prediction for wireless networks	Wireless network data	Use of attention mechanism to model two types of variables in an end-to-end framework
[43]	DEEXP Method	Traffic consumption prediction	Network traffic data	Use of DEEXP method for traffic consumption prediction
[31]	Short-term Memory Traffic Prediction Model (SLSTM)	Accurate short-term traffic prediction in 5G networks	5G network traffic data	Use of short-term memory model (SLSTM) for smooth 5G network traffic prediction
[32]	SES-LSTM (Single-exponential smoosparseg + Long Short-Term Memory)	Traffic prediction combining CNN and LSTM	Network traffic data	Hybrid SES-LSTM model combining unidimensional smoosparseg with LSTM
[44]	LSTM + Gaussian Process Regression (GPR)	Accurate cellular traffic prediction at the cell level	Open cellular traffic dataset from Milan, Italy	Combination of LSTM and GPR for high-accuracy traffic prediction at cell level
[45]	Deep Learning Approach	Mobile data consumption prediction at the city level	City-level mobile traffic data	Use of deep learning to model nonlinear dynamics of wireless traffic
[33]	Spatio-temporal Attention-Complexity Network	Cellular network traffic prediction in cities	Urban traffic data	Spatial-temporal attention mechanism for predicting traffic in urban areas
[34]	ML-based Dynamic Traffic Slice Model (ML-TADS)	Proper management of traffic in networks	Network traffic data	Dynamic traffic slicing model using machine learning for uniform traffic distribution

[35]	Various ML Models	Prediction accuracy and computational time cost analysis	Network traffic data	Comparison of various ML algorithms in terms of prediction accuracy and computational cost
[36]	Naive Bayes + Holt-Winters	Improving cellular BS traffic prediction accuracy	BS traffic data	Hybrid Naive Bayes and Holt-Winters model for improved traffic prediction in cloud-based platforms
[37]	Supervised Learning Algorithms (SVR, Kernel Ridge, DT, RF, LSTM)	Traffic flow estimation	Collected traffic data	Comparison of supervised learning algorithms for traffic flow estimation
[40]	federated learning (FL) framework for mobile traffic prediction	Mobile traffic prediction in satellite-terrestrial integrated networks while preserving data privacy	Real-world mobile traffic dataset	Uses FL to enhance privacy, trains local models at base stations, aggregates a global model on a satellite edge server, improves scalability and privacy
[41]	multimodal deep learning framework for mobile traffic prediction by integrating CNN (ConvLSTM) and GNN (AGCN)	Mobile traffic prediction using a hybrid CNN-GNN approach for single-step prediction	Real-world mobile traffic dataset	Fuses SMS, call, and internet data, integrates grid-based and graph-based representations, outperforms ten baseline models in occur

Recent advancements in network traffic prediction have introduced a wide range of methodologies aimed at improving forecasting accuracy. However, these approaches often exhibit inherent limitations that hinder their practical applicability. Federated learning frameworks leveraging gradient similarity-based aggregation have been explored for wireless traffic prediction, demonstrating an improvement in feature learning capacity. Nevertheless, these methods frequently lack robustness in short-term forecasting due to the distributed nature of model updates and inherent constraints in data synchronization. Ensemble-based machine learning paradigms, including Random Forest (RF), Support Vector Machines (SVM), and Decision Trees (DT), have been employed to enhance prediction performance. While these models facilitate multi-perspective feature learning, they remain susceptible to overfitting, particularly in dynamic network environments with high traffic variability. Deep learning techniques such as Convolutional Neural Networks (CNN) with attention mechanisms and the Kalman filter have been adopted to mitigate these shortcomings to capture intricate spatiotemporal dependencies in traffic data. Although these methods effectively integrate auxiliary contextual information, their performance deteriorates significantly in scenarios where such metadata is either sparse or unavailable. Furthermore, Explainable Artificial Intelligence (XAI) and hybrid architectures, such as LSTM coupled with Gaussian Process Regression (GPR), have been introduced to enhance interpretability and predictive reliability. While these models offer superior adaptability for short-term forecasting, they often impose substantial computational overhead, limiting their feasibility for real-time deployment in large-scale networks. Additionally, augmented data-driven frameworks and cloud-based machine learning solutions have been proposed to improve generalization capabilities. However, these approaches are often constrained by latency issues, high memory consumption, and the necessity for extensive labeled datasets, making them less practical for high-throughput, real-time network scenarios.

In contrast, the proposed method in this study effectively overcomes these limitations by integrating K-SVD, Discrete Wavelet Transform (DWT), Genetic Algorithm (GA), and BiLSTM to enhance predictive accuracy while maintaining computational efficiency. The K-SVD and DWT-based feature extraction module significantly reduces noise and dimensional redundancy, ensuring a more refined input representation. The GA-optimized sparse matrix formulation further enhances the sparsity structure of the dataset, facilitating robust pattern recognition with reduced complexity. Finally, the BiLSTM architecture, with its capability to model bidirectional dependencies in sequential data, allows for improved forecasting accuracy in highly volatile network traffic environments. Comparative analyses indicate that this hybridized approach achieves lower error margins (MAE, MSE, RMSE) while maintaining computational scalability. It is a more viable and efficient alternative for Beyond 5G network traffic prediction and resource optimization.

III. THE PROPOSED METHOD

The primary objective of this study is to propose an efficient and robust method for traffic forecasting in 5G and Beyond networks, leveraging deep recurrent neural networks. The methodology follows a structured sequence, beginning with data preprocessing using K-SVD, which enhances data quality by reducing noise and extracting significant components. Subsequently, a sparsity-driven approach is applied to traffic data using Discrete Wavelet Transform (DWT) and Genetic Algorithm (GA). This step is crucial for transforming raw network data into a compact and meaningful feature set, optimizing the representation of temporal patterns. The final stage involves employing a Bidirectional Long Short-Term Memory (BiLSTM) network, which effectively captures both short-term fluctuations and long-term dependencies in the traffic data, ensuring higher prediction accuracy compared to conventional models. A key aspect of the proposed approach is the construction and utilization of a sparsity matrix, which plays a fundamental role in enhancing the performance of the predictive model. The sparsity matrix is formulated using GA to ensure an optimal balance between feature reduction and information preservation. This matrix refines the extracted features, eliminating redundant data while maintaining the structural integrity of the traffic patterns. By feeding the optimized sparse feature vectors into the BiLSTM network, the proposed method significantly improves prediction accuracy, reduces computational complexity, and mitigates the risk of overfitting. To further illustrate the methodology, Figure 1 provides a detailed block diagram of the proposed approach, and Tables 2 and 3 define key terms and mathematical expressions used throughout the study.

1. PREPROCESSING

Sparse decomposition (SD) is a relatively recent signal processing technique that has gained increased interest in mechanical defect diagnostics since it may capture the essential characteristics of the examined signal without assuming orthogonal base expansion. One of the main components of SD is dictionary generation, and there are two methods for creating an overcomplete lexicon: self-learning and predetermined. The self-learning SD dictionary approach typically has low interference resilience, whereas the preset dictionary SD technique often requires prior knowledge of the investigated signal. In recent years, a vast amount of literature has appeared to address the aforementioned issues. The Empirical Wavelet Transformation is enhanced To extract features, the dispersion-guided experimental wavelet transform method is proposed and applied based on the SD idea. Furthermore, many self-learning SD approaches are computationally inefficient. To address the abovementioned issues, this study proposes a sparse display technique of K-means singular value decomposition (K-SVD) based on the classic K-SVD method, Applying the chase algorithm for sparse matching and an iterative method based on minimal atomic structure similarity. Through simulation and experiment verification, the suggested technique successfully captures the hidden aspects of the studied signal and has a greater computing efficiency.

In recent years, a feature extraction and data compression technique called k-means singular value decomposition (K-SVD) has applications in a wide range of domains, including image processing and language recognition. K-SVD, which is also considered an extension of K-means, gives the sample data a sparse linear form. Much like a dictionary, a collection of overcomplete basis vectors is searched to get this data. [46], to improve the convergence rate of the dictionary atom and sparse coding, optimal matching and SVD are used together in KSVD. It causes them to keep pace with the update of dictionary atoms and sparse coefficients. To show the sparse signal, the KSVD algorithm is used to sparse the S coefficient as much as possible by finding the dictionary D as the following:

$$Y = DS \qquad (1)$$

Where D is the complete matrix, either the over-complete matrix (where There are more columns than rows in this instance) or the matrix with an equal number of columns, and Y is the trained signal. The column vectors $(s_1, s_2, \ldots . s_N)$ in S correspond to Y. This means that Y is formed linearly under si from the columns of D and in the form of an equation that has been converted into a mathematical model [46] and it can be expressed as:

$$\min_{D,X}\{\|S_i\|_{l_0}\} \text{ s.t. } \|Y - DS\|_{l_2}^2 \leq \varepsilon \qquad (2)$$

This equation could be rewritten as follows:

$$\min_{D,X} \|Y - DS\|_{l_2}^2 \; s.t. \forall i, \|S_i\|_{l_0} \leq T_0 \qquad (3)$$

Equations (2) and (3) describe the sparse representation problem in the context of dictionary learning and signal approximation, which is a crucial step in K-SVD-based feature extraction. Equation (2) represents an optimization problem where the objective is to minimize the ℓ_0 norm of the sparse representation S_i, ensuring that only a limited number of nonzero elements are retained in the representation. The constraint $\|Y - DS\|_{l_2}^2$ enforces that the reconstructed data DS should approximate the original data Y within an allowable error margin ε, measured using the ℓ_2 norm (which represents the Euclidean distance between the original and reconstructed data). Equation (3) reformulates this problem by shifting the optimization focus to minimizing the ℓ_2 norm of the reconstruction error while maintaining a strict sparsity constraint $\|S_i\|_{l_0} \leq T_0$ for all columns i. Here, T_0 represents the maximum allowed number of nonzero elements in each sparse representation vector S_i, ensuring that the solution remains computationally efficient while preserving essential structural information. This formulation is particularly useful in sparse coding and compressed sensing, where maintaining a balance between accuracy and sparsity is essential for effective feature extraction and data compression.

Table 2. Description of abbreviation.

Term in use	Description
ACC	Accuracy
ADAM	Adaptive Moment
ARIMA	Autoregression integral moving average model
BiLSTM	Bidirectional LSTM
BN	Bayesian networks
DCT	discrete cosine transforms
DWT	Discrete wavelet transforms
FNN	Fuzzy neural networks
GPR	Gaussian process regression
HPF	High pass filter
KNN	K-nearest neighbors
LAN	Local area network
LPF	Low pass filter
LSTM	Long-term short-term memory
MAE	Mean absolute Error
MSE	Mean square error
QoS	Quality of Service
RBFNN	Radial basis function neural networks
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
SD	Sparse decomposition
SVM	Support vector machines
SVMR	Support vector machine regression
SVR	Support vector regression
TM	Traffic matrix
WNN	Wavelet neural networks

Table 3. Description of terms in formula.

Term in use	Description
D	complete matrix
Y	trained signal
S	column vectors (s_1, s_2, ...s_i, …, s_N)
T_0	Threshold of optimization
l_0	Zero norm
s_i^T	Transpose column vectors

d_i	represents the ith column of D
E_K	Error
$W_{X(S,U)}$	Output of wavelet transform
$X(t)$	Input data $x \in R^{N+1}$
$\Psi^*_{s,u}$	Mother wavelet
a_0^j	transfer parameter
$CD_j(k)$	detail coefficients
CA_j	approximation coefficients
Ek	residual error
φ	Sparse matrix
λ	Regulation parameter
Φ_x	Random Gaussian, random sparse binaryand random Bernoulli matrix
SR	sparse rate
d_t	predicted data
y_t	target data
TN	True negative
TP	True posetive
FN	False negative
FP	False posetive

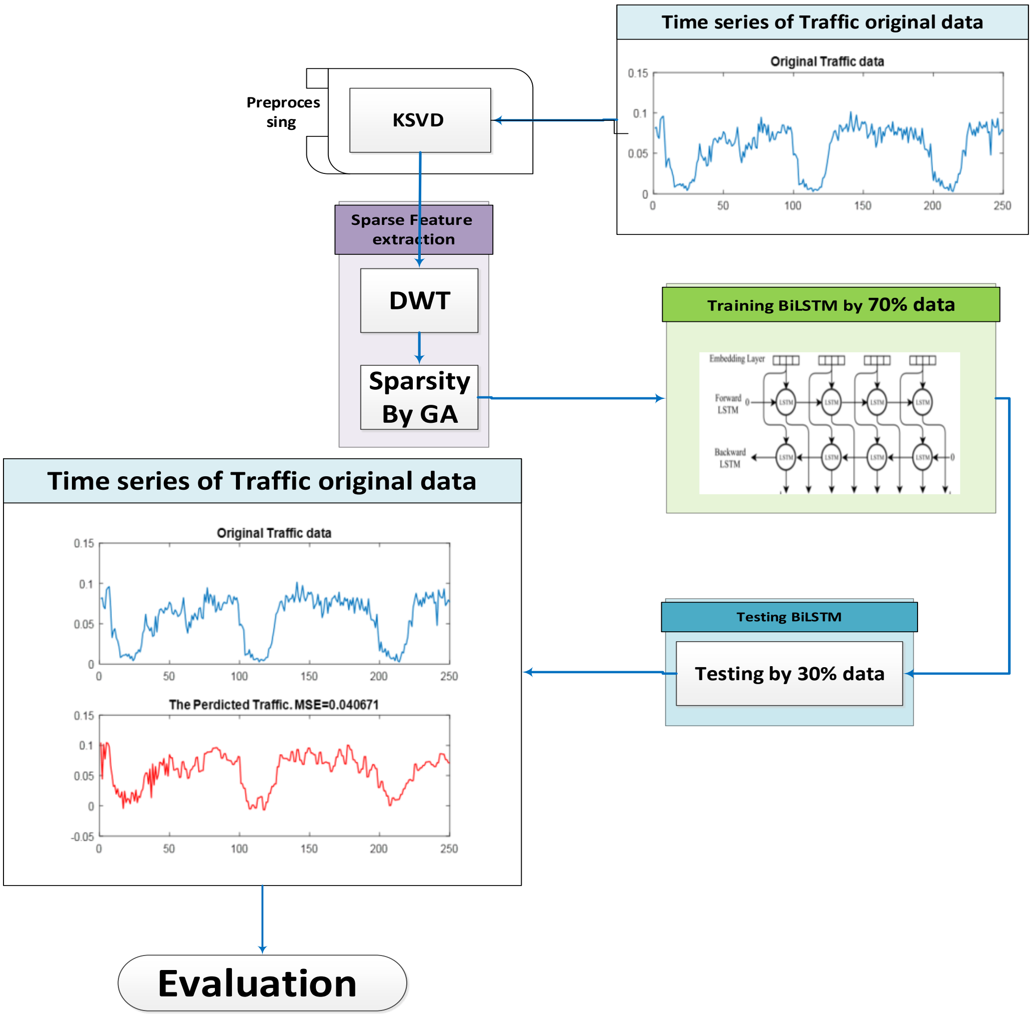

FIGURE 1. Block diagram of proposed method.

One of the non-deterministic-time polynomial-hard (NP-Hard) problems are to find the global optimal dictionary D, If the ideal answer is just gradually approached, giving the trained data Y. l_0 is zero norm. Generally, there are two stages of dictionary update and sparse coding in KSVD.

We choose a dictionary D for sparse coding in order to show the sparse data Y. The sparse coefficients S of the training data Y in the dictionary D may be found by using optimum matching tracking. As a result, DS may be thought of as the total of the product of each matching row in S and the matching column in D, which is stated as follows:

$$DS = \sum_{i=1}^{K} d_i s_i^T \tag{4}$$

Where s_i^T symbolizes the ith row of S, and di symbolizes the ith column of D. Dictionary D is updated by going over each column individually. Assuming S and D are constant when updating the kth column dk in D and the kth row s_i^T in S, the following equation applies.

In (5), it is only to adjust dk and s_k^T to make the error between EK and $d_k s_k^T$ as small as possible [47] . In the suggested approach, Y data is sparsely shown using a simple dictionary D that is initially established. The optimum matching pursuit might be applied to get the sparse coefficients S of the dictionary D's training data Y. The following equation may be used to define DS, which is now the total of the product of the corresponding row in S and each column in D:

$$\|Y - DS\|_{l_2}^2 = \|Y - \sum_{i=1}^{K} d_i s_i^T\|_{l_2}^2 = \|(Y - \sum_{i \neq k} d_i s_i^T) - d_i s_i^T\|_{l_2}^2 = \|E_K - d_k s_k^T\|_{l_2}^2 \tag{5}$$

$$S = \sum_{i=1}^{K} d_i s_i^T \tag{6}$$

Where *di* stands for D's ith column, and s_i^T rsymbolizes S's ith row. Dictionary D is updated column by column. Assume first that S and D are constant, and that the kth row of D's column dk and the s_k^T in S are modified, the following equation is present:

$$\|Y - ds\|_{l2}^2 = \left\|Y - \sum_{i=1}^{K} d_i s_i^T\right\|_{l2}^2 = \left(Y - \sum_{i \neq k} d_i s_i^T\right) - d_k s_k^T\|_{l2}^2 = \|E_k - d_k s_k^T\|_{l2}^2 \tag{7}$$

In equation (12), only d_k and s_k^T should be adjusted so that the error between E_k and $d_k s_k^T$ is as tiny as feasible. The main stages of KSVD are as follows:

- **Step 1:** Initially, the initialization matrix is set. $D^0 \in R^{N \times K}$ and the initial iteration value J = 1
- **Step 2:** Sparse coding involves solving S with the optimum matching tracking technique, obtaining a sparse coefficient si to represent yi, and calculating the error ε that satisfies equation (12). Modify it.
- **Step 3:** Update the dictionary
  - Definition of the sample used $w_k = \{i | 1 \leq i \leq N, s_k^T(i) \neq 0\}$
  - Calculation of residual error Ek and estimate $Ek = Y - \sum_{i \neq k} d_i s_i^T$.
  - Restricting by choosing the columns for wk, Ek, then balancing the non-zero values in Ek using s_i^T, then the new EK is obtained, denoted by E_K^R .
  - Decomposing $E_K^R = \Delta V^T$ using SVD and updating the dictionary atom $\tilde{s}_k^T = \tilde{d}_k = u_1, \Delta[1,1]$, where D is an SVD of E_K^R , the largest value The singular is denoted by Δ[1,1], v1, v1 is the right vector V's singular matrix's first column, while the left vector U's singular matrix's first column is u1.
  - Update J = J + 1.
- **Step 4**: Output
  - D^J learned dictionary obtained.

2. SPARSE FEATURE SEXTRACTION

The combination of Sparse Decomposition (K-SVD), Genetic Algorithm (GA), and BiLSTM was chosen due to its complementary strengths in feature extraction, optimization, and sequential modeling, particularly for non-stationary network traffic data. Traditional methods like Fourier Transform fail to capture localized variations in such dynamic datasets, whereas Discrete Wavelet Transform (DWT) enables multi-resolution time-frequency analysis, making it an ideal preprocessing step. DWT decomposes traffic data into different frequency components, allowing for more effective extraction of both short-term fluctuations and long-term trends. In this study, orthogonal wavelet filters from the Daubechies family were selected for their optimal smoothness and regularity, ensuring improved denoising and feature preservation in preprocessed network data.

Once the essential features are extracted, the sparse matrix representation, generated through K-SVD and optimized using GA, plays a crucial role in enhancing network traffic prediction accuracy. Unlike conventional feature selection methods that rely on predefined statistical measures, the sparse matrix adaptively retains the most relevant traffic patterns while eliminating redundant information, thereby reducing data dimensionality without loss of critical temporal dependencies. GA was specifically chosen for its superior capability in discrete optimization tasks, effectively tuning the sparse representation to maximize feature efficiency compared to traditional methods like PSO or Simulated Annealing. Finally, BiLSTM serves as the predictive model, leveraging its bidirectional structure to learn complex dependencies in sequential traffic data, significantly improving forecasting accuracy and model stability. Comparative experiments demonstrate that this integration not only enhances predictive performance by reducing Mean Squared Error (MSE) but also lowers computational complexity, making the model more robust and efficient for real-world 5G network traffic forecasting.

Sparse matrix construction: If $x \in R^{N+1}$ it is a vector and is considered as recorded data after pre-processing, where the number of N samples is desired in a time interval, the X vector is thinned by the matrix using the CS method. which is the injection matrix or projection/sensing matrix $\varphi \in R^{M+1}$. which φ is expressed as follows:

$$y = \varphi x \tag{8}$$

Which $y \in R^{M\times 1}$ is the same data collected. The sparse rate will be defined as the following relationship.

$$SR = 1 - \left(\frac{M}{N}\right) \tag{9}$$

Data X is a sparse or t sparse able data. In the time domain, if it can be reconstructed with a high probability (with the smallest line), in fact the data is thinned, y will be as follows:

$$x = {}_{x}\underset{X}{min}\ \|y - \varphi x\|_2^2 + \lambda\|x\|_1 \tag{10}$$

which λ in this relation is the Regularization matrix and $\|.\|_P (P \geq 1)$ is a real number which is defined as follows.

$$\|a\|_P = \left(\sum_{i=1}^{n} |ai|^p\right)^{1/p} \tag{11}$$

Traffic data is inherently dense and highly dynamic, meaning it does not exhibit natural sparsity in the time domain. However, sparse representations can be effectively obtained by transforming the data into alternative domains where underlying patterns become more distinguishable. Techniques such as Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT) are widely used for this purpose, as they enable multi-resolution decomposition and capture localized variations in the data. In this study, DWT is chosen as the primary transformation technique due to its superior ability to separate noise from meaningful traffic patterns while preserving both high-frequency fluctuations and long-term dependencies. Once the traffic data is transformed into the wavelet domain, sparsification is applied to remove redundant information, enhancing the efficiency of subsequent processing steps. If the sparse matrix is constructed as the base representation, the relationship between the signal vector and the reconstructed data in matrix form X can be formulated as follows:

$$X = \min_{x} \|y - \Phi_x\|_2^2 + \lambda \|\psi_x\|_1 \quad (12)$$

Random Gaussian matrices, random sparse binary matrix and random Bernoulli matrix are three commonly used matrices for Φ_x and λ is Regulation parameter. In order for the retrieval accuracy to increase, the correlation between and should be low. random matrices with independent definite linear distribution i.i.d. Like Gaussian distribution or bivariate, they have the most. Although generating the Gaussian matrix and applying it to traffic data is computationally complex [48]. Also, its optimal selection is very effective in forecasting, and by choosing a random matrix, it is not possible to get a good result in traffic forecasting. Thus, a genetic method is used in this study to determine the matrix.

In this study, the Genetic Algorithm (GA) was chosen for sparse matrix optimization due to its superior performance in discrete and combinatorial optimization problems compared to Particle Swarm Optimization (PSO) and Simulated Annealing (SA). GA efficiently explores the search space using mutation and crossover operations, preventing premature convergence and enhancing feature selection for traffic prediction. Unlike PSO, which often suffers from early convergence to local optima, and SA, which relies on a localized search trajectory, GA maintains population diversity and ensures a more comprehensive exploration of possible solutions. Additionally, GA is particularly suitable for multi-objective optimization, which is essential in balancing sparsity constraints and predictive accuracy in this study [49]. By leveraging GA, the sparse matrix representation enhances BiLSTM's learning efficiency, leading to improved accuracy, reduced computational cost, and greater generalization capability in real-world network traffic forecasting. In the genetic algorithm, Genes are elements that contain codes for variables. Chromosomes, which are collections of genes, are the solutions to the issue. The elements of this paper are the number of ones that should be placed in a sparse matrix in such a way that the traffic prediction is done with the lowest RMSE. In each repetition of the genetic algorithm, the values of 1 are moved along the sparse matrix, as a result, the values of the genes can change, this value can be zero and one in this paper. Gene values are altered and a new chromosome is created by using mutation and crossover. (new sparse matrix) is created. Table 4 shows an example of chromosomes to create a sparse matrix. The length of the sparse matrix is variable and can change based on the conditions. Since the purpose of this article is to provide a practical example, the length of chromosome 8 is considered. In Table 4 the chromosomal representation used for constructing a sparse matrix in the proposed method. Each row represents a chromosome, where binary values (0s and 1s) indicate the presence or absence of specific elements in the sparse matrix. The genetic algorithm optimizes these chromosomes to achieve the most efficient sparse representation, ensuring that the selected structure enhances feature extraction and network traffic prediction. By iteratively evolving these chromosomes, the algorithm minimizes reconstruction error while maintaining sparsity constraints.

Table 4. Chromosomes to create a sparse matrix.

C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
0	0	1	1	0	0	1	0
0	1	0	1	0	0	1	0
1	0	0	0	0	1	0	1

To accelerate convergence and accomplish the in-tended outcomes, it is very important to consider the initial population. In thin matrices, a parameter called SR thinness rate is defined. The sparseness rate of an operation indicates the number of ones in a sparse matrix.

$$SR = \frac{\#\ of\ one}{lengths\ of\ matrix} \quad (13)$$

To ensure the optimal selection of the sparse matrix, the genetic algorithm evaluates candidate solutions based on their impact on predictive accuracy, prioritizing matrices that yield the lowest MSE values. The regularization parameter λ plays a crucial role in balancing sparsity and feature retention, where overly sparse matrices may result in information loss, whereas insufficient sparsity may introduce redundancy. The evolutionary nature of the algorithm allows it to dynamically adjust SR (sparsity rate) values, ensuring that the final matrix structure aligns with the intrinsic patterns of the traffic data. Experimental analysis reveals that tuning SR in a controlled manner leads to a significant reduction in overfitting while enhancing generalization performance, making this approach particularly suitable for real-time network traffic forecasting in Beyond 5G environments .It is defined by considering the thinness rate of the initial population. In Table 1, $SR = \frac{3}{8}$ which means the ratio of number of genes with a value of zero to the number of genes with a value of one. The starting population is produced using the length of the sparse matrix. In the genetic algorithm, two generations are combined to produce the next generation. The roulette cycle is used in this article to select parents. Then the Crossover operator is randomly created in the first step in selecting or generating the sparse matrix. For this purpose, half of the matrix is removed from each of the parents in this research. In other words, half the length of the chromosome is removed. Parents are two parental chromosomes which are separated into sections on the left and right. The right part of one chromosome and the left part of another chromosome together forms a child chromosome. The mutation operator has been used to ensure random changes in the chromosome. Since the generated chromosomes should not be similar after the Crossover operator, a thin graph has been used. Finally, fitness is defined taking into account the Mean Square Error (MSE) of the least square distance. The lower the MSE value. That is, the genetic algorithm has been more successful in producing the sparse graph matrix. The MSE value is obtained directly from the predicted values of the data volume and it can be expressed as:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(d_t - y_t)^2 \quad (14)$$

Where dt, yt and N denote the predicted data, the target data and all the data in a window respectively. uring the sparsification process, the genetic algorithm iteratively adjusts the sparse matrix configuration to minimize the MSE, ensuring that the transformed and thinned data retain sufficient structural integrity for accurate forecasting. A lower MSE value indicates that the sparsified dataset still effectively represents the original traffic patterns, thereby preventing significant information loss while reducing the dimensionality of input data. By incorporating MSE as the fitness function in the optimization process, the algorithm ensures that only the most relevant features are retained, contributing to enhanced predictive performance and computational efficiency in the BiLSTM-based forecasting model.

3. FORECASTING WITH LSTM

Now, the bidirectional LSTM (BiLSTM) model is suggested to enhance the forecasting capabilities of LSTM. This article forecasts traffic using a developed BiLSTM. One network receives the input sequence in regular chronological order, while another network receives it in reverse chronological order, thanks to a deep learning technique. At every time step, the two networks' outputs are consecutive. BiLSTM's stacked layer design provides great prediction accuracy by enabling the acquisition of both forward and back-ground Each time step provides information about the sequence. Memory units in the BiLSTM model play a crucial role in accurate traffic prediction for Beyond 5G networks. These units enable the model to capture long-term dependencies in network traffic data and retain important information from past time steps for use in future predictions. In Beyond 5G networks, where traffic patterns are constantly changing and complex dependencies between data points exist, memory units allow BiLSTM to store and process long-term information. This capability helps the model make accurate predictions by effectively simulating both short-term and long-term traffic variations. As a result, BiLSTM, with its memory units, is able to provide better predictions for highly complex networks like 5G and beyond.

IV. DATA ANALYSIS

In this study, the hyperparameters of the BiLSTM model were systematically optimized using the Grid Search method to enhance prediction accuracy and model stability. BiLSTM is specifically chosen for its ability to capture long-term dependencies within sequential data by processing information in both forward and backward directions, making it particularly effective for network traffic forecasting. To ensure optimal performance, five layers were employed for training the classifier. The Grid Search approach was utilized to determine the most effective hyperparameters by evaluating multiple parameter combinations and selecting those that yielded the best performance in terms of MSE, RMSE, MAE, and ACC. The following configurations were identified as optimal:

- Maximum number of iterations: Set to 350, allowing the model to refine weight updates over multiple training cycles while maintaining computational efficiency.
- Batch size: A mini-batch size of 80 was selected to strike a balance between computational cost and training stability, ensuring effective gradient updates.

- Initial Learning Rate: Optimized at 0.01, accelerating convergence while preventing divergence or unstable updates.
- Gradient Threshold: Adjusted to 1 to prevent gradient explosion and maintain numerical stability during training.
- Training Monitoring: "Plots" was set to "Training Progress", enabling real-time visualization of the loss function and validation metrics to assess model convergence and detect overfitting.

These optimized hyperparameters, derived from the Grid Search process, significantly improved the model's predictive performance by ensuring stability, faster convergence, and enhanced generalization to unseen data. The final settings are summarized in Table 5, detailing the classifier's configuration. This model was trained using an ADAM-based ap-proach. A deep learning model may be trained more quickly by using ADAM, an alternate optimization technique for stochastic gradient ratios. ADAM was selected for this work because it combines the benefits of the RMSProp and AdaGrad algorithms. ADAM facilitates better management of noise, random gradients, and sparse data. Large data or parameter difficulties can be solved with the optimizer's computational efficiency in BiLSTM model training. It has also been used to simulate and predict the traffic of other neural networks. These networks are listed in table 6. The training time of BiLSTM is higher compared to traditional models like RNN and LSTM due to its bidirectional processing and increased number of parameters. While this results in a longer computational time, it significantly enhances the model's ability to capture long-term dependencies and complex temporal patterns, leading to higher predictive accuracy. Compared to simpler models such as MLFNN and WNN, which have lower computational complexity, BiLSTM requires more processing power and memory due to its recurrent nature. However, the trade-off between computational cost and prediction accuracy is justified, as BiLSTM consistently outperforms other models in terms of error reduction (MSE, RMSE, MAE) and stability in forecasting performance.

Table 5. Used BiLSTM settings.

Parameter	Value
Educational rate	0.01
Optimization method	ADAM
The highest number of ipak	350
The smallest package size	80
The number of hidden units	100
Gradient threshold	1
Executive environment	Automatically
hidden layers	1*5
The length of the sequence	The longest
shuffle	once
Activity function	Sigmoid

Table 6. Neural network type and setting up.

Refer-ence	Neural network	Abbrevia-tion	Setting up
[50]	Multilayer feed for-ward neural networks	MLFNN	Consideration is given to multi-layer perceptron feed-forward neural networks having one hidden layer and fifteen hidden layer neurons.
[5]	Recurrent neural net-works	RNN	15 hidden layer neurons are thought to exist.
[51]	long short-term memory	LSTM	the number of hidden layer neurons is considered 15
[52]	Radial basis functions Neural networks.	RBFNN	Gaussian function and the educational method is supervised
[53]	Wavelet neural net-works	WNN	wavelet neural networks and the Mexican hat. The wavelet function is employed for the activity function of the neurons with a hidden layer of 15 neurons.
[54]	Fuzzy neural networks	FNN	fuzzy neural networks use the Gaussian membership function and the number of 15 neurons in a hidden layer

The 5G Mobile Network Data (5G-ND) dataset is a real-world traffic dataset specifically designed for analyzing and evaluating 5G mobile networks [55]. It contains comprehensive traffic data collected from telecom operators, base stations, and wireless network infrastructures, making it a valuable resource for studying network performance, traffic patterns, and optimization strategies in 5G environments. This dataset includes key network metrics such as data transmission rates, end-to-end latency, packet loss, network congestion levels, and Quality of Service (QoS) parameters, enabling researchers to assess network efficiency under various real-world conditions. It also supports diverse scenarios, including urban, rural, and industrial 5G deployments, allowing for in-depth analysis of user behavior, resource distribution, and adaptive network management. Additionally, time-stamped and location-based data enable the examination of spatial and temporal traffic variations, facilitating better insights into network scalability and dynamic load balancing. The 5G-ND dataset is particularly suitable for machine learning-based traffic prediction, as it provides rich data for training and evaluating deep learning models such as BiLSTM, CNN, and Transformer architectures. By leveraging this dataset, researchers can develop and validate intelligent traffic forecasting models, improving real-time traffic management and resource allocation in next-generation 5G networks. 70% of the entire data is utilized for training, with the remaining 30% for testing. The gathered findings determine if an algorithm or approach is suitable for prediction. The following quantitative comparison was done to assess these algorithms. Mean Square Error, Root Mean Square Error, and Mean Absolute Error (MAE) [30]. Root means square error (RMSE): The smaller this square root error number is, it indicates that the pre-diction result is more successful and it can be ex-pressed as:

$$MSE = \sqrt{\frac{1}{N}\sum(d_t - y_t)^2} \quad (15)$$

Where *Ir* is the initial available value and d_t is the obtained value. *N* dimensions are the vector of desired values. Mean Square Error (MSE): The smaller the squared error number, the more successful the prediction result is and it can be expressed as:

$$MSE = \frac{1}{N}\sum(d_t - y_t)^2 \quad (16)$$

Where, d_t is the initial available value and $y_t$is the obtained value. *N* dimensions are the vector of desired values. Mean absolute error (MAE): This standard calculates the absolute error in the corresponding values in If and *Ir*: If is the obtained values and *Ir* is the original values calculated in the proposed method.

$$MAE = \sum_{i=1}^{m}|d_t - y_t| \quad (17)$$

In this regard, *Ir* is the initial available value and if is the obtained value. Obviously, the smaller this number is, the better the result. Prediction accuracy: The prediction accuracy criterion is a common criterion in traffic data volume prediction.

$$\text{ACC}=\frac{TN+TP}{TN+FN+FP+TN} \quad (18)$$

1. COMPARISON OF THIN MATRICES

In order to evaluate the sparse graph matrix, the results of BiLSTM were evaluated with random distribution matrix graph and Gaussian distribution in the dis-cussed criteria of MAE absolute error, ACC accuracy, MSE square root mean error, and RMSE root mean square error. Tables 2 to 4 show this comparison for all three signals. As it can be seen from the comparison of the results, the LSTM method has predicted much better results in the selection of the proposed thin feature based on the genetic algorithm. The reason for this is the use of the MSE objective function in the genetic algorithm in order to select the best values for prediction, as well as taking into account the spatial and temporal dependence of the data. Also, Gaussian distribution function has better results than random distribution. This superiority can be due to the normal distribution of the recorded data of the signals.

Table 7. Comparison of evaluation criteria in BiLSTM in the first signal.

METHOD	MAE	RMSE	MSE	ACC
Random	14.2325	22.1525	4.7014	94.25
Gaussian	13.2587	19.0225	4.1041	95.74
proposed	12.8987	18.2369	4.2596	97.25

Table 8. Comparison of evaluation criteria in BiLSTM in the second signal.

METHOD	MAE	RMSE	MSE	ACC
Random	15.9742	21.2987	4.23698	94.25
Gaussian	13.4573	19.8523	4.4589	95.36
proposed	12.0289	18.1489	4.5558	98.00

Table 9. Comparison of evaluation criteria in BiLSTM in the third signal.

METHOD	MAE	RMSE	MSE	ACC
Random	15.2563	19.2631	4.2013	95.23
Gaussian	13.6359	19.2798	4.4236	95.63
proposed	12.2531	18.2589	4.2896	97.91

Three signals in the intended networks have the re-quired predictions given by the suggested technique. Figure 2 shows the MAE in the prediction and evaluation of all seven neural networks and in three signals. The MAE performance comparison across different neural network models demonstrates the superior predictive accuracy of the proposed BiLSTM-based approach. The results indicate a notable reduction in absolute error when employing BiLSTM, highlighting its capability to capture complex temporal dependencies more effectively than conventional models. Additionally, networks such as RNN and LSTM exhibit moderate error levels, while traditional architectures like MLFNN and WNN display significantly higher error values, suggesting limitations in handling dynamic traffic fluctuations. The inclusion of error bars further emphasizes the stability and robustness of the BiLSTM model, indicating lower variance and improved generalization compared to other methods.

Similar to Figure 3 of MSE diagram, the MSE performance evaluation highlights the significant reduction in squared error achieved by the BiLSTM model compared to other neural network architectures. The results demonstrate that BiLSTM consistently outperforms traditional models such as MLFNN, WNN, and FNN, which exhibit higher error levels, indicating limitations in capturing complex temporal dependencies in traffic data. Additionally, while LSTM and RNN achieve relatively lower errors, BiLSTM further enhances predictive accuracy by leveraging its bidirectional processing capability. The presence of error bars emphasizes the stability and lower variance of BiLSTM, reinforcing its robustness and generalization ability in real-world scenarios.

Figure 4 shows the RMSE in the simulations. The RMSE performance comparison reveals the superiority of the BiLSTM model in minimizing root mean squared error compared to other neural networks. The results indicate that traditional models such as MLFNN, WNN, and FNN exhibit higher RMSE values, suggesting lower predictive reliability when handling network traffic fluctuations. While LSTM and RNN show relatively improved performance, BiLSTM achieves the lowest RMSE, reinforcing its ability to capture both short-term variations and long-term dependencies in time-series data. Additionally, error bars demonstrate lower variance in BiLSTM's predictions, confirming its robustness and stability over multiple evaluation scenarios.

Figure 5 shows the pre-diction accuracy of the proposed method in different networks. The accuracy comparison highlights the superior classification performance of the BiLSTM model compared to other neural networks. The results indicate that traditional models such as MLFNN, WNN, and FNN exhibit lower accuracy, reflecting their limitations in effectively capturing temporal dependencies. While CNN, RNN, and LSTM demonstrate improved accuracy, BiLSTM achieves the highest accuracy across all evaluation scenarios, reinforcing its ability to learn complex patterns in network traffic data. Additionally, the error bars indicate minimal variance in BiLSTM's performance, confirming its stability and robustness in predictive modeling.

The lower the numerical value of the MAE absolute mean error, the better the prediction result. Based on the results presented in Figure 2-4, the MAE value of the BiLSTM neural network has recorded a lower value compared to other methods. Although convolutional neural network LSTM has recorded similar results to BiLSTM, albeit weaker, but it is su-perior compared to RNN. Perhaps the superiority of LSTM method compared to RNN can be ignored considering the computational complexity. It should be mentioned that the amount of computing involved in BiLSTM is acceptable compared to LSTM. The smaller the RMSE and MSE values are, the better the prediction result. Based on the results presented in Figure 4, the RMSE value of the BiLSTM neural network has recorded a lower value compared to other methods. Although LSTM has recorded similar results to BiLSTM and although it is weaker compared to BiLSTM, they are superior compared to other networks. The memory unit in the BiLSTM technique is what makes it better than other approaches. In similar neural network methods including MLFNN,

https://doi.org/10.48161/qaj.v5n2a1690

FNN, WNN and RBFNN, the activity function plays an essential role in the prediction output. Therefore, the RBFNN method is better than other similar methods. The reason for the superiority is the type of Gaussian radial basis function used in prediction in the neural network. As expected, the proposed BiLSTM neural network has obtained the best result in accuracy by processing the data with the help of the proposed method. The reason for the superiority of combining two neural networks and strengthening the memory unit is also the use of thin matrix in feature selection for prediction. The LSTM neural network is also at the next stage because of the memory. Other methods based on neural network are in later stages.

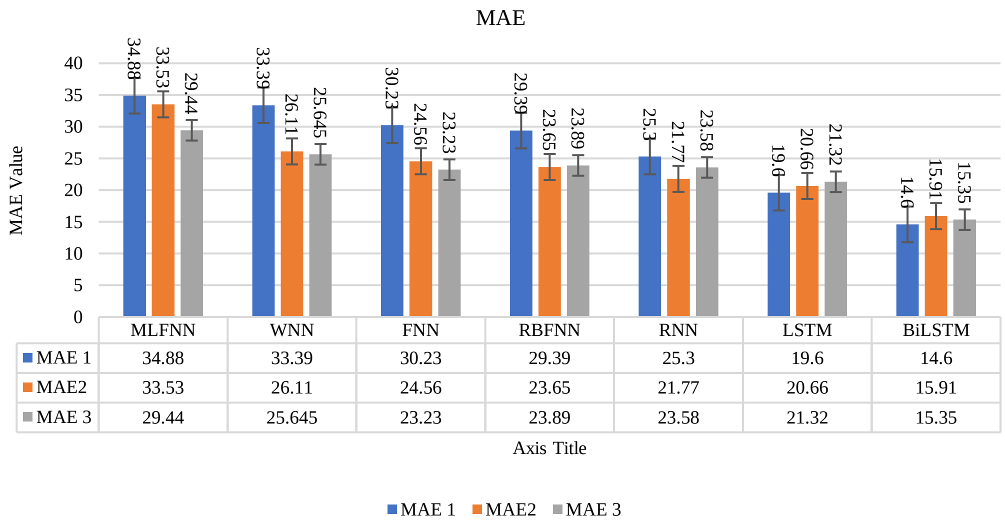

FIGURE 2. Comparison of MAE in various networks and prediction of all three traffic data signals.

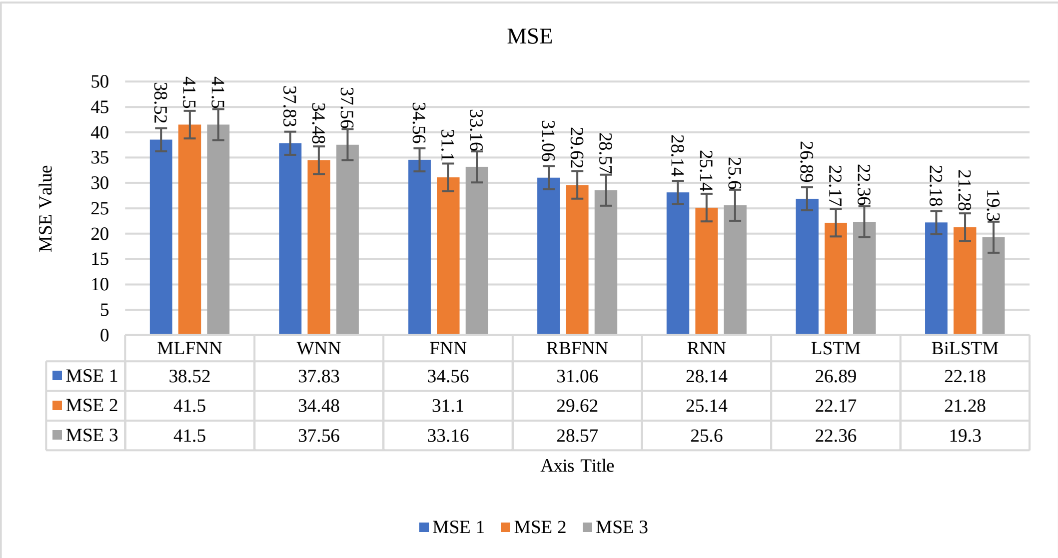

FIGURE 3. Comparison of MSE across networks and prediction in all three traffic data streams.

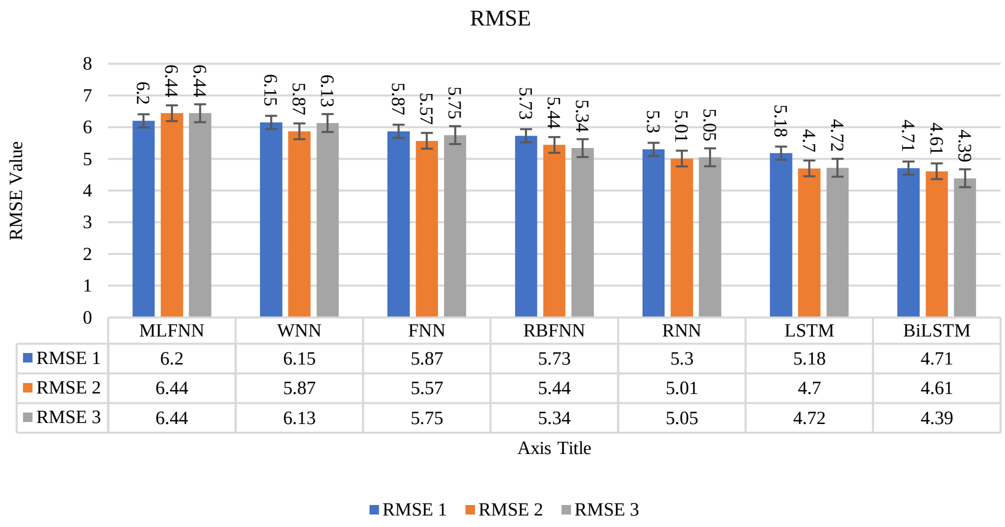

FIGURE 4. RMSE comparison across several networks and forecasting for each of the three traffic data signals.

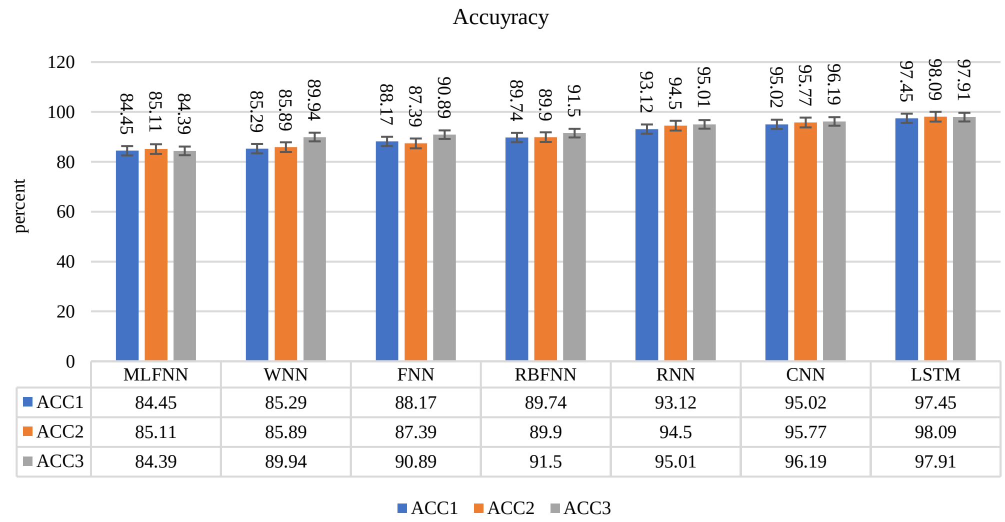

FIGURE 5. Comparison of the three traffic data signals' predictions and accuracy (ACC) across several networks.

The paired t-test analysis was conducted to statistically validate the superiority of the proposed BiLSTM model, which integrates K-SVD, DWT, GA, and BiLSTM for enhanced network traffic prediction. The results across four evaluation metrics (MAE, MSE, RMSE, and Accuracy) demonstrate that BiLSTM consistently outperforms all competing models, including LSTM, RNN, RBFNN, FNN, WNN, and MLFNN, with statistically significant improvements. For the MAE metric (Figure 6 A), all p-values are below 0.05, confirming that BiLSTM significantly reduces absolute prediction errors compared to other models. In the MSE metric (Figure 6 B), BiLSTM maintains a statistically significant advantage over all models, including LSTM, ensuring superior error minimization. Similarly, in the RMSE metric (Figure 6 C), the p-values for all comparisons remain below 0.05, validating BiLSTM's robust performance in reducing squared error deviations. Finally, in the Accuracy metric (Figure 6 D), BiLSTM achieves significantly higher accuracy than all other models, with all p-values below 0.05, confirming its superior generalization capability. These statistical results reinforce that the proposed BiLSTM-based approach delivers a substantial and statistically significant improvement in network traffic prediction, making it a more efficient, accurate, and reliable choice compared to traditional neural network models.

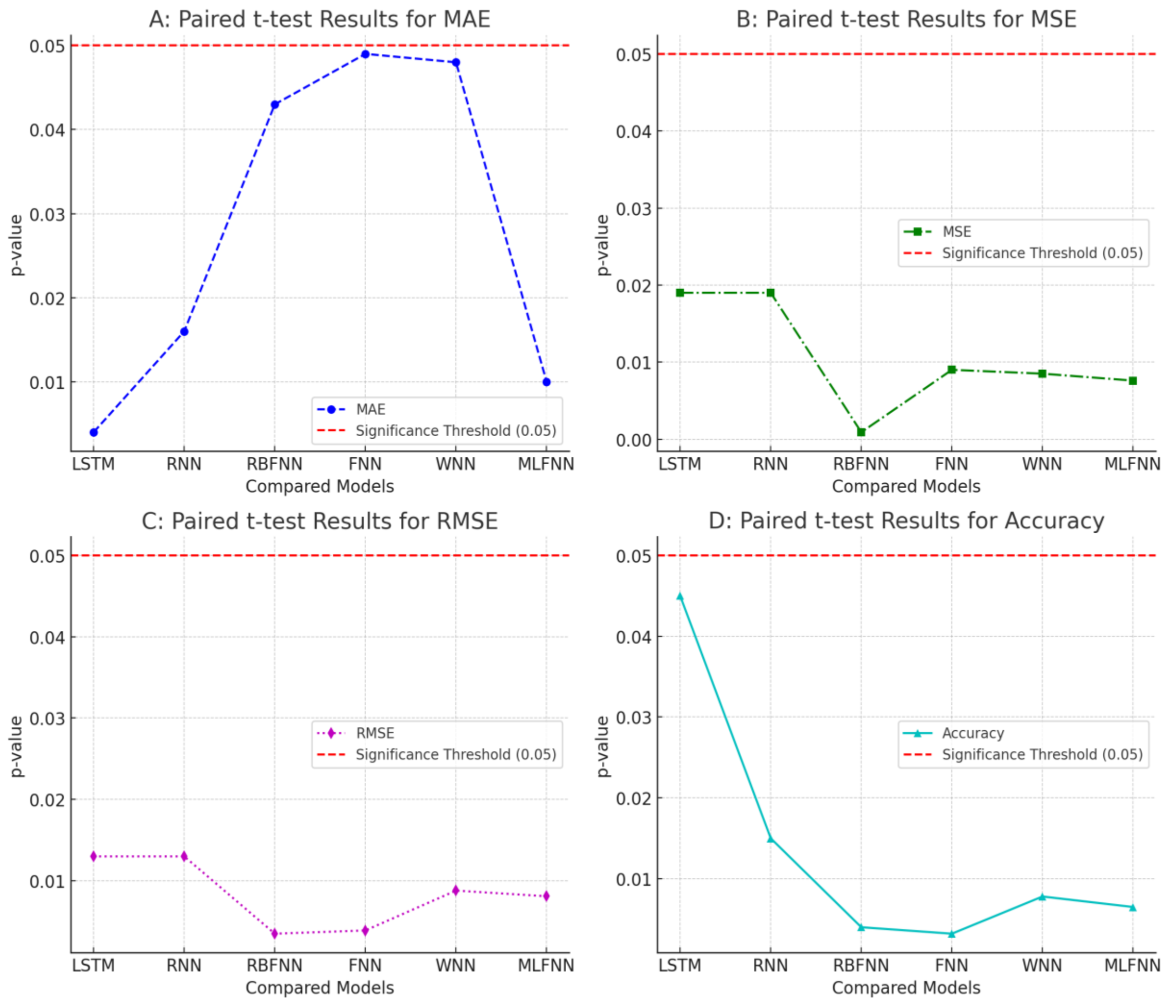

FIGURE 6. Paired t-test results for MAE (A), MSE (B), RMSE (C), and Accuracy (D). The proposed BiLSTM model shows statistically significant improvements over all competing models ($p < 0.05$).

V.CONCLUSION

In this study, a memory-enhanced deep learning approach using Bidirectional Long Short-Term Memory (BiLSTM) is proposed for cellular network traffic prediction. The proposed method integrates K-SVD for data preprocessing, Discrete Wavelet Transform (DWT) for feature extraction, and a Genetic Algorithm (GA) for sparse matrix construction, ensuring optimal feature selection. The sparse matrix representation is evaluated across multiple neural network architectures, including MLFNN, WNN, FNN, RBFNN, RNN, LSTM, and BiLSTM, where BiLSTM demonstrates superior predictive performance. The paired t-test analysis confirms that the performance improvements achieved by BiLSTM over conventional models are statistically significant ($p < 0.05$) across multiple evaluation metrics, including MAE, MSE, RMSE, and Accuracy. The enhanced predictive capability of BiLSTM is attributed to its bidirectional memory units, which effectively capture long-term dependencies and complex temporal patterns in network traffic. Furthermore, the practical applicability of the proposed model in real-world 5G networks is discussed, emphasizing its potential for adaptive traffic management and resource allocation. The study also highlights key deployment challenges, such as computational complexity, real-time processing constraints, and data availability, providing directions for future research in scalable and distributed deep learning-based traffic forecasting. The results validate BiLSTM as an efficient, accurate, and statistically reliable solution for next-generation cellular traffic prediction.

Funding Statement

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Author Contributions

Rihab and Mahdi were responsible for the conceptualization, methodology, and writing of the main manuscript text. Ali and Rasool contributed by pre-paring Figures and performing data analysis. All authors reviewed and edited the manuscript before submission.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data are available from the authors upon request.

Acknowledgments

The authors would like to acknowledge the assistance of the Editor and Reviewers in the preparation of the article for publication.

REFERENCES

1. Nabi, S. T., Ahmed, M., Alshehri, S., & Khan, M. A. (**2023**). Deep learning based fusion model for multivariate LTE traffic forecasting and optimized radio parameter estimation. *IEEE Access, 11*, 14533–14549.
2. Elsayed, E. E., Kassem, A. M., & Farag, A. A. (**2024**). Coding techniques for diversity enhancement of dense wavelength division multiplexing MIMO-FSO fault protection protocols systems over atmospheric turbulence channels. *IET Optoelectronics, 18*(1–2), 11–31.
3. Jubair, M. A., Mostafa, S. A., Zebari, D. A., Hariz, H. M., Abdulsattar, N. F., Hassan, M. H., ... & Alouane, M. T. H. (2022). A QoS aware cluster head selection and hybrid cryptography routing protocol for enhancing efficiency and security of VANETs. *IEEE Access*, *10*, 124792-124804.
4. Koursioumpas, N., Magoula, L., Barmpounakis, S., & Stavrakakis, I. (**2022**). Network traffic anomaly prediction for beyond 5G networks. In *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)* (pp. 589–594). IEEE.
5. Jaffry, S., & Hasan, S. F. (**2020**). Cellular traffic prediction using recurrent neural networks. In *2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT)* (pp. 94–98). IEEE.
6. Ahmed, F. Y., Masli, A. A., Khassawneh, B., Yousif, J. H., & Zebari, D. A. (2023). Optimized Downlink Scheduling over LTE Network Based on Artificial Neural Network. *Computers*, *12*(9), 179.
7. Ma, H., Yang, K., & Pun, M.-O. (**2023**). Cellular traffic prediction via deep state space models with attention mechanism. *Computer Communications, 197*, 276–283.
8. Nikravesh, A. Y., Ajila, S. A., Lung, C.-H., & Ding, W. (**2016**). Mobile network traffic prediction using MLP, MLPWD, and SVM. In *2016 IEEE International Congress on Big Data (BigData Congress)* (pp. 402–409). IEEE.
9. Sadeghi, R., Barraca, J. P., & Aguiar, R. L. (**2017**). A survey on cooperative MAC protocols in IEEE 802.11 wireless networks. *Wireless Personal Communications, 95*, 1469–1493.
10. Shi, J., Leau, Y. B., Li, K., & Obit, J. H. (**2021**). A comprehensive review on hybrid network traffic prediction model. *International Journal of Electrical and Computer Engineering, 11*(2), 1450.
11. Harouni, M., Karimi, M., Nasr, A., Mahmoudi, H., & Najafabadi, Z. A. (**2022**). Health monitoring methods in heart diseases based on data mining approach: A directional review. In *Prognostic Models in Healthcare: AI and Statistical Approaches* (pp. 115–159). Springer.
12. Xu, S., & Zeng, B. (**2014**). Network traffic prediction model based on auto-regressive moving average. *Journal of Networks, 9*(3), 653.
13. Wei, D., & Liu, H. (**2013**). An adaptive-margin support vector regression for short-term traffic flow forecast. *Journal of Intelligent Transportation Systems, 17*(4), 317–327.
14. Cai, Y., Cheng, P., Ding, M., Chen, Y., Li, Y., & Vucetic, B. (**2020**). Spatiotemporal Gaussian process Kalman filter for mobile traffic prediction. In *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications* (pp. 1–6). IEEE.
15. Elsisi, M., Rusidi, A. L., Tran, M.-Q., Su, C.-L., & Ali, M. N. (**2024**). Robust indoor positioning of automated guided vehicles in Internet of Things networks with deep convolution neural network considering adversarial attacks. *IEEE Transactions on Vehicular Technology*.
16. Aouedi, O., Le, V. A., Piamrat, K., & Ji, Y. (**2025**). Deep learning on network traffic prediction: Recent advances, analysis, and future directions. *ACM Computing Surveys, 57*(6), 1–37.
17. Kavehmadavani, F., Nguyen, V.-D., Vu, T. X., & Chatzinotas, S. (**2023**). Intelligent traffic steering in beyond 5G open RAN based on LSTM traffic prediction. *IEEE Transactions on Wireless Communications, 22*(11), 7727–7742.

18. Crone, S. F., Hibon, M., & Nikolopoulos, K. (**2011**). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting, 27*(3), 635–660.
19. Wang, L., Hu, J., Jiang, D., Zhang, C., Jiang, R., & Chen, Z. (**2024**). Deep learning models for spectrum prediction: A review. *IEEE Sensors Journal*.
20. Chowdhury, S. (**2024**). LSTM-oriented handover decision-making with SVR-based mobility prediction in 5G vehicular networks. *(Unpublished manuscript)*.
21. Alizadeh Aliabadi, M., Karimi, M., Karimi, Z., & Soheili Fard, M. (**2025**). The effect of photoplethysmography signal denoising on compression quality. *Iranian Journal of Electrical and Electronic Engineering, 21*(1), 3277.
22. Mathur, S., Chaba, Y., & Noliya, A. (**2023**). Performance analysis of support vector machine learning based carrier aggregation resource scheduling in 5G mobile communication. *Procedia Computer Science, 218*, 2776–2785.
23. Kaci, A., & Rachedi, A. (**2021**). Named data networking architecture for Internet of Vehicles in the era of 5G. *Annals of Telecommunications, 76*(9), 717–729.
24. Rehman, A., Harouni, M., Zogh, F., Saba, T., Karimi, M., & Jeon, G. (**2023**). Detection of lung tumors in CT scan images using convolutional neural networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.
25. Asgarnezhad, R., Alhameedawi, K. A. M., & Mahfoud, H. A. (**2023**). An effective model of autism spectrum disorder using machine learning. *Indonesian Journal of Electrical and Electronic Engineering (IJEEI), 11*(2), 389–401.
26. Emadi, M., Jafarian Dehkordi, Z., & Iranpour Mobarakeh, M. (**2023**). Improving the accuracy of brain tumor identification in magnetic resonance imaging using super-pixel and fast primal dual algorithm. *International Journal of Engineering, 36*(3), 505–512.
27. Islam, M. A., Siddique, H., Zhang, W., & Haque, I. (**2022**). A deep neural network-based communication failure prediction scheme in 5G RAN. *IEEE Transactions on Network and Service Management, 20*(2), 1140–1152.
28. Raftarai, A., Mahounaki, R. R., Harouni, M., Karimi, M., & Olghoran, S. K. (**2021**). Predictive models of hospital readmission rate using the improved AdaBoost in COVID-19. In *Intelligent Computing Applications for COVID-19* (pp. 67–86). CRC Press.
29. Li, L., Zhao, Y., Wang, J., & Zhang, C. (**2023**). Wireless traffic prediction based on a gradient similarity federated aggregation algorithm. *Applied Sciences, 13*(6), 4036.
30. Haider, U., Waqas, M., Hanif, M., Alasmary, H., & Qaisar, S. M. (**2023**). Network load prediction and anomaly detection using ensemble learning in 5G cellular networks. *Computer Communications, 197*, 141–150.
31. Moghadas Gholian, S., Fiandrino, C., Collet, A., Attanasio, G., Fiore, M., & Widmer, J. (**2023**). Spotting deep neural network vulnerabilities in mobile traffic forecasting with an explainable AI lens. In *IEEE International Conference on Computer Communications*.
32. Gao, Z. (**2022**). 5G traffic prediction based on deep learning. *Computational Intelligence and Neuroscience, 2022*, Article 2022.
33. Wang, W., Zhou, C., He, H., Wu, W., Zhuang, W., & Shen, X. (**2020**). Cellular traffic load prediction with LSTM and Gaussian process regression. In *ICC 2020—2020 IEEE International Conference on Communications (ICC)* (pp. 1–6). IEEE.
34. Zhang, C., Zhang, H., Yuan, D., & Zhang, M. (**2018**). Citywide cellular traffic prediction based on densely connected convolutional neural networks. *IEEE Communications Letters, 22*(8), 1656–1659.
35. Zhao, N., Ye, Z., Pei, Y., Liang, Y.-C., & Niyato, D. (**2020**). Spatial-temporal attention-convolution network for citywide cellular traffic prediction. *IEEE Communications Letters, 24*(11), 2532–2536.
36. Song, C., et al. (**2018**). Machine learning enabling traffic-aware dynamic slicing for 5G optical transport networks. In *CLEO: Science and Innovations* (p. JTu2A.44). Optica Publishing Group.
37. Sun, P., Aljeri, N., & Boukerche, A. (**2020**). Machine learning-based models for real-time traffic flow prediction in vehicular networks. *IEEE Network, 34*(3), 178–185.
38. Clemente, D., Soares, G., Fernandes, D., Cortesao, R., Sebastiao, P., & Ferreira, L. S. (**2019**). Traffic forecast in mobile networks: Classification system using machine learning. In *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)* (pp. 1–5). IEEE.
39. Yaghoubi, F., Catovic, A., Gusmao, A., Pieczkowski, J., & Boros, P. (**2021**). Traffic flow estimation using LTE radio frequency counters and machine learning. *arXiv preprint arXiv:2101.09143*.
40. Jiang, W., Mu, J., Han, H., Zhang, Y., & Huang, S. (**n.d.**). Federated learning-based mobile traffic prediction in satellite-terrestrial integrated networks. *Software: Practice and Experience*.
41. Jiang, W., Zhang, Y., Han, H., Huang, Z., Li, Q., & Mu, J. (**2024**). Mobile traffic prediction in consumer applications: A multimodal deep learning approach. *IEEE Transactions on Consumer Electronics, 70*(1), 3425–3435.
42. Lai, J., et al. (**2023**). Deep learning based traffic prediction method for digital twin network. *Cognitive Computation*, 1–19.
43. Amiri, P. A. D., & Pierre, S. (**2023**). An ensemble-based machine learning model for forecasting network traffic in VANET. *IEEE Access, 11*, 22855–22870.
44. Alsaade, F. W., & Al-Adhaileh, M. H. (**2021**). Cellular traffic prediction based on an intelligent model. *Mobile Information Systems, 2021*, Article 2021.
45. Wang, Z., Hu, J., Min, G., Zhao, Z., & Wang, J. (**2020**). Data-augmentation-based cellular traffic prediction in edge-computing-enabled smart city. *IEEE Transactions on Industrial Informatics, 17*(6), 4179–4187.

46. Wang, H., & Du, W. (**2020**). A new K-means singular value decomposition method based on self-adaptive matching pursuit and its application in fault diagnosis of rolling bearing weak fault. *International Journal of Distributed Sensor Networks, 16*(5), 1550147720920781.
47. Wu, Q., & Zhao, X. (**2018**). Traffic time series prediction based on CS and SVR. In *2018 Chinese Control And Decision Conference (CCDC)* (pp. 3427–3432). IEEE.
48. Papazoglou, G., & Biskas, P. (**2023**). Review and comparison of genetic algorithm and particle swarm optimization in the optimal power flow problem. *Energies, 16*(3), 1152.
49. Huang, Y., Qian, L., Feng, A., Yu, N., & Wu, Y. (**2019**). Short-term traffic prediction by two-level data driven model in 5G-enabled edge computing networks. *IEEE Access, 7*, 123981–123991.
50. Hassan, M. K., Ariffin, S. H. S., Syed-Yusof, S. K., Ghazali, N. E., Kanona, M. E. A., & Rava, M. (**2022**). Smoothing-aided long-short term memory neural network-based LTE network traffic forecasting. *International Journal of Electrical and Computer Engineering, 12*(6), 6859.
51. Abhishek, & Verma, S. (**2021**). Machine learning in 5G wireless networks. In *5G and Beyond Wireless Systems: PHY Layer Perspective* (pp. 391–410).
52. Lohrasbinasab, I., Shahraki, A., Taherkordi, A., & Jurcut, A. D. (**2022**). From statistical- to machine learning-based network traffic prediction. *Transactions on Emerging Telecommunications Technologies, 33*(4), e4394.
53. Perveen, A., Abozariba, R., Patwary, M., & Aneiba, A. (**2023**). Dynamic traffic forecasting and fuzzy-based optimized admission control in federated 5G-Open RAN networks. *Neural Computing and Applications, 35*(33), 23841–23859.
54. Idowu-Bismark, O., Kennedy, O., Husbands, R., & Adedokun, M. (**2019**). 5G wireless communication network architecture and its key enabling technologies. *International Journal of Distributed Sensor Networks, 12*, 70–82.
55. Wang, H., & Du, W. (**2020**). A new K-means singular value decomposition method based on self-adaptive matching pursuit and its application in fault diagnosis of rolling bearing weak fault. *International Journal of Distributed Sensor Networks, 16*(5), 1550147720920781.