

A Two-Step Cellular Network Traffic Forecasting Method Integrating Decomposition and Deep Neural Networks based on Bayesian Joint Parameter Optimization

Pengfei Zhang, Junhuai Li, Dong Ding, Huaijun Wang, Kan Wang, *Member, IEEE*, and Xiaofan Wang

Abstract—Accurate cellular network traffic prediction is crucial for intelligent network planning and management in 6G. However, the non-stationary characteristics of cellular network traffic present significant challenges when training deep neural networks for traffic forecasting. To address this issue, we propose a two-stage deep learning framework, JO-DPNet, based on Bayesian joint parameter optimization, which integrates data decomposition techniques with Bayesian joint optimization to effectively mitigate the adverse impacts of non-stationarity and error accumulation on prediction accuracy. In the first stage, a data decomposition module uses Variational Mode Decomposition (VMD) to decompose the original data into network traffic subset series(TSS), thereby alleviating the negative effects of non-stationarity. In the second stage, a prediction and construction module leverages a bi-directional LSTM (Bi-LSTM) network to extract deep spatial-temporal features from the TSS in a bidirectional manner. A fully connected layer then captures the relationships between the TSS and reconstructs the predicted results into the final output. The JO module employs the Tree-structured Parzen Estimator based Bayesian optimization algorithm(TP-BO) simultaneously determines the optimal VMD mode number k and the hyperparameters of the Bi-LSTM network through probabilistic surrogate model. Extensive experiments on three real-world cellular traffic datasets demonstrate that the proposed method significantly mitigates the non-stationary characteristics of the traffic data. Compared to state-of-the-art methods, JO-DPNet achieves reductions in MAE by 29%, 3%, and 19% for three type prediction tasks on the Telecom Italia dataset. The source code is available to the public at: <https://github.com/VicentZhang259/JO-DPNet>.

Index Terms—Cellular network traffic prediction, Result construction, VMD, Bi-LSTM, Join optimization.

I. INTRODUCTION

This work was supported by the National Natural Science Foundation of China (No.61971347), the Scientific Research Program of Shaanxi Province (No.2022SF-353) and the Project of the Xi'an Science and Technology Planning Foundation (No.24ZDCYISGG0020), and Natural Science Project of Shaanxi Provincial Department of Education (No.23JK0562). (Corresponding author: Junhuai Li and Huaijun Wang.)

Pengfei Zhang, Junhuai Li, Dong Ding, Huaijun Wang, Kan Wang and Xiaofan Wang are with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China. (e-mail: zhangpf369@126.com; lijunhuai@xaut.edu.cn; dingdongexcellent@163.com; zhanghuaijun@xaut.edu.cn; wangkan@xaut.edu.cn; wangxfok@xaut.edu.cn).

Junhuai Li, Junhuai Wang, Kan Wang and Xiaofan Wang are with Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an, 710048, China.

Junhuai Li, Junhuai Wang, Kan Wang and Xiaofan Wang are also with Human-Machine Integration Intelligent Robot Shaanxi University Engineering Research Center, Xi'an, 710048, China.

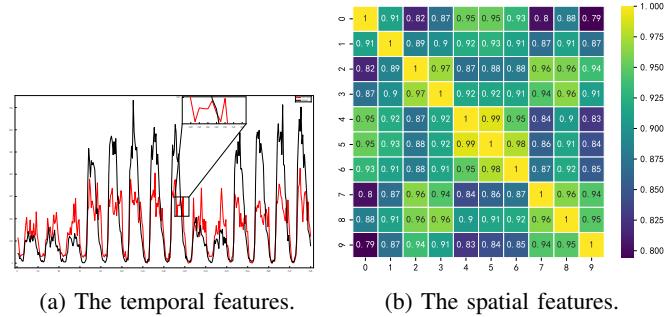


Fig. 1: (a): Internet cellular traffic in two regions of Milan, Italy, over one week. Non-linear and non-stationary patterns in the temporal domain are observed. (b): The spatial distribution of Internet cellular traffic across all regions in Milan within one hour. Complex spatial correlations between regions are shown.

6 G mobile communication networks are expected to adopt AI-driven intelligent architecture to provide higher reliability and data rates in order to meet the demands of emerging application scenarios [1]. The deployment of AI-driven cellular networks is aimed at achieving artificial intelligence for the operation and management, ensuring sufficient quality of network service (QoS), and reducing operating expenses (OPEX) [2-4]. In particular, implementing AI-driven automatic network management and optimization requires accurate traffic prediction at different time scales, which is crucial for tasks in wireless communication. For instance, the efficiency of demand-aware resource allocation, base station sleep mechanism, and network congestion control heavily depends on accurate prediction of future wireless traffic [5-7].

Nevertheless, making an accurate cellular traffic prediction is a nontrivial task for the following reasons. Firstly, mobile users exhibit diverse demands that fluctuate based on time and location, complicating traffic prediction (see Fig. 1a and Fig. 1b). Secondly, user mobility introduces spatial dependencies in communication among cells distributed across different geographical areas. Lastly, external factors such as the number of base stations, weather conditions, and holidays further complicate the spatial-temporal relationships. Recent studies have indicated that cellular network traffic displays both non-linearity and non-stationarity [7, 8]. The autocorrelation

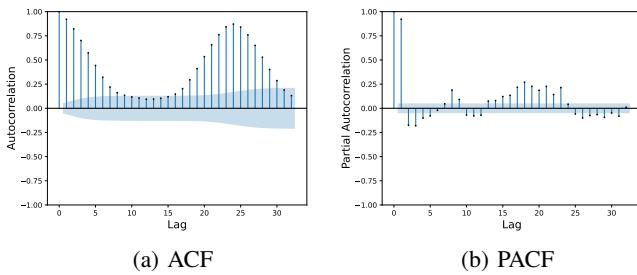


Fig. 2: Visualization the ACF and PACF of Internet cellular traffic in Duomo of Milan, Italy. The horizontal coordinate lag indicates the order of the lag, and the vertical coordinate indicates the correlation coefficient between the corresponding lag series and the original series. The blue area indicates the confidence interval.

function (ACF) and partial autocorrelation function (PACF) of Internet traffic in the Duomo area are shown in Fig. 2a and Fig. 2b, respectively. The correlations did not show obvious attenuation. This indicates that the network traffic sequence is non-stationary.

Significant efforts have been made to addressing cellular traffic prediction challenges. Early approaches employed traditional time-series methods such as exponential smoothing and autoregressive moving average (ARMA) [9-11], which are limited to capture linear temporal relationships. The rise of deep learning has shifted research focus toward neural network-based solution. LSTM has demonstrated exceptional sequence modeling capabilities by capturing long-term dependencies while mitigating gradient vanishing problem [12]. For instance, Zhang et. al [13] proposed a hybrid model that integrates convolutional neural networks (CNNs) with LSTM to jointly capture spatial and temporal dependencies. Advanced models like STCNet [2] further enhance spatial-temporal traffic prediction by incorporating external features and transfer learning techniques. Gu et al. [14] introduced a graph attention-based spatiotemporal network that effectively captures both local geographical dependencies and distant inter-region relationships. Bidirectional LSTM (Bi-LSTM) improves predictive accuracy by processing input sequences in both forward and backward directions, thereby capturing bidirectional contextual information.

Recently, another thread of research has focused on multi-step prediction methods for cellular network traffic prediction. In [15-17], Yu et al. decomposed wireless networks into the trend, seasonality, and holiday and verified the high predictability of the regularity component. Wang et al. [7] utilized Fourier analysis to extract the periodic components, while the trend and residual components were recovered by the LSTM network and Gaussian process regression model. Moreover, Liu et al. [18] developed a decomposition-based unified framework to achieve competitive results for multi-seasonal traffic prediction, which investigated the influence of different decomposition strategies and exploited common patterns. Notably, modal decomposition has proven highly effective in mitigating data non-stationarity via adaptive sig-

nal decomposition techniques [19], and has seen successful applications in electricity price forecasting [20], equipment remaining useful life prediction [21], and power load forecasting [22]. In particular, the application of variational mode decomposition technology has yielded promising results [23]. However, its use in mobile network traffic prediction remains underexplored. Moreover, conventional approaches typically rely on empirical settings for deep neural network model and VMD hyperparameters, often resulting in suboptimal solutions and limited prediction accuracy.

Motivated by the insights above, we propose a novel two-stage traffic prediction framework for cellular networks based on Bayesian joint optimization combined with decomposition and deep neural networks (JO-DPNet). First, by applying the VMD method to decompose the original cellular network traffic into simpler and more stable traffic subset series, JO-DPNet mitigates the impact of the non-linearity and non-stationarity of the original cellular network traffic data on prediction accuracy. Second, this paper presents a deep-learning model for cellular network traffic prediction. In particular, the model utilizes a Bi-LSTM capturing past and future context information at each point, thereby extracting deep spatial-temporal features from the traffic subset series bidirectionally. The FC is then designed to reconstruct the prediction results based on the linear dependencies among the subset series predictions. In addition, by employing the TPE-BO optimization algorithm to optimize the critical parameters of VMD, this paper effectively alleviates the mode aliasing and endpoint effects that can arise from manual parameter selection in the VMD method. The number of layers in the Bi-LSTM module is optimized to enable the deep learning module to capture the complex spatial-temporal features of network traffic easily. The two-stage synchronous optimization searches for the best parameter combination simultaneously, avoiding the issue of finding a local optimum when parameters are optimized separately in two stages. Consequently, JO-DPNet mitigates the impact of non-stationarity and fully extracts deep spatial-temporal features. The main contributions of this paper are summarized as follows:

- 1) We propose a novel two-stage cellular network traffic prediction framework, JO-DPNet, which integrates BO-TPE joint optimization with data decomposition and deep neural networks. JO-DPNet is designed to effectively address the non-stationary and non-linear characteristics of cellular network traffic and mitigate the problem of error accumulation, thereby achieving accurate traffic forecasting.
- 2) This article adapts VMD with joint optimization to mitigate the non-linear and non-stationary features of the original data. Furthermore, the proposed deep learning model effectively captures the spatial-temporal features within the traffic subset series and the linear relationships between these subset series. Meanwhile, the two-stage synchronous optimization strategy ensures an efficient search for the optimal parameter combination.
- 3) Extensive experiments conducted on a large-scale, real-world dataset demonstrate the superior performance of

our JO-DPNet model against compared state-of-the-art methods. Specifically, our framework has achieved significant Mean Absolute Error (MAE) reductions of 29%, 3%, and 19% across three traffic types compared to conventional modeling approaches.

The remainder of the paper is organized as follows: Section II discusses related work, Section III presents the proposed method, Section IV provides experimental results and discussions, and Section V concludes the paper.

II. RELATED WORK

In recent years, the application of machine learning to cellular network traffic prediction has garnered significant scholarly interest. Precise and timely forecasting of cellular network traffic facilitates operators in minimizing operational expenditures (OPEX) and supporting superior quality of service (QoS). As mobile networks expand and new services proliferate, the dynamics of cellular traffic have grown increasingly intricate. A plethora of research, including works by Katris and Wang, has demonstrated the non-stationary and non-linear features of cellular network traffic [7]. These intrinsic characteristics pose substantial challenges to achieving accurate traffic forecasting, thus drawing considerable research focus.

A. Cellular Traffic Prediction

Cellular traffic prediction has recently attracted increased amounts of attention since it is essential to intelligent cellular network management. In recent decades, researchers have conducted many impressive studies in cellular traffic prediction. Linear methods, such as the autoregressive (AR) model, ARMA, and ARIMA, are widely used for predicting time series data [10, 24-26]. However, the methods mentioned above primarily rely on approximate linear fitting of network traffic series, failing to capture the non-linear characteristics inherent in the data adequately [27]. To address non-linear features in cellular network traffic series, researchers have shifted their focus to models capable of handling complex and non-linear data. The effectiveness of such models in predicting network traffic series has been established. Among these, support vector machines (SVM) are commonly employed. Nikravesh et al. [27, 28] applied SVM to a dataset from commercial trial mobile networks, demonstrating that SVM exhibits good flexibility and robustness, outperforming the multi-layer perceptron model with weight decay. Xia et al. [29] combined random forest (RF) with LightGBM, achieving notable prediction performance on network traffic data. RF filters redundant features, and LightGBM served as the final prediction model. The Gaussian mixture model (GMM) has also gained popularity among researchers for network traffic prediction due to its predictive and generalization capabilities [30]. Additionally, Sun et al. [31] proposed a feature embedding kernel that has been proposed for Gaussian processes (GP). They utilized this kernel to predict wireless network traffic, achieving a balance between overall prediction accuracy and peak-trough accuracy. Nevertheless, these methods exhibit limitations in modeling complex non-linear relationships.

Deep learning approaches have revolutionized traffic prediction through end-to-end feature learning. Recurrent Neural Networks (RNNs) are particularly adept at capturing long-term dependencies in sequences, making them a natural choice for time series modeling tasks such as electricity price forecasting [4], wind energy forecasting [32], mobile demand forecasting [33], and load forecasting [34]. Among the various RNN variants, LSTM is the most widely used to address the issues of gradient explosion and vanishing gradients encountered during RNN training [12]. For instance, Ma et al. [35] employed LSTM to forecast transportation traffic flow, while another study [36] utilized it for multi-step predictions of mobile data traffic. The NLG-NLAM model is a deep learning approach based on a non-linear gradient non-local attention mechanism that considers dynamic non-local spatial correlations, self-attention, and the fusion of spatio-temporal features [37]. Gu et al. [14] proposed a graph attention spatial-temporal network that captures local geographical dependencies and distant inter-region relationships when considering spatial factors. The MVSTGN model employs an attention mechanism to analyze data from multi-perspectives [4]. However, these machine learning methods could adapt to non-stationary data. We propose a novel two-stage traffic prediction framework for cellular networks, JO-DPNet, with decomposition and deep neural networks to capture complex spatial-temporal characteristics.

B. Decomposition then Prediction

The decomposition-prediction paradigm, which reduces data complexity through data decomposition, has proven effective in road traffic prediction. However, the pattern of cellular traffic is in general much more complex than that of road traffic due to indeterminate connections between different cells. Therefore, direct decomposition followed by prediction is challenging to adapt for cellular traffic forecasting. Notable improvements in cellular network traffic prediction are as follows. Li et al. [15] decomposed cellular networks into trend, seasonality, and holiday components, then model the component by Prophet model and Gaussian process regression based respectively. Wang et al. [7] utilized Fourier analysis to extract periodic components, while the trend and residual components were reconstructed using LSTM networks and Gaussian process regression models. Additionally, Liu et al. [18] developed a decomposition-based unified framework that achieved competitive results in multi-seasonal traffic prediction by investigating the effects of different decomposition strategies and identifying common patterns within time series data. Nan et al. [38] propose a Multiple Seasonal-Trend decomposition to decompose the cellular traffic into the multiseasonal, trend, and residual components through the MSLT algorithm. In MSLT-GLTP, the Bi-LSTM, temporal convolutional network (TCN) and Gaussian process regression (GPR) model is deployed to learn the dynamic regional and local traffic. Consequently, these methods further improve the performance of load forecasting. However, those mentioned above capture the non-stationary and non-linear characteristics of cellular network traffic by simply combining decomposition

and prediction methods; thus, hidden spatial-temporal correlations cannot be efficiently captured; thus, hidden spatial-temporal correlations cannot be efficiently captured. Wang et al. proposed the Arctic Puffin Optimization (APO) algorithm, which is a meta-heuristic optimization algorithm inspired by the survival and predation behavior of the Arctic puffin, and comprehensively classified and studied the latest optimization methods [39]. Inspired by recent advances in the Bi-LSTM [40], VMD [41, 42], and BO-TPE [43], which are able to learn implicit features, we propose a novel two-stage traffic prediction framework for cellular networks, JO-DPNet, combining Bayesian joint optimization with decomposition and deep neural networks to capture complex spatial-temporal characteristics.

III. OUR METHOD

In order to mitigate the non-stationary and non-linear characteristics of cellular network traffic, we introduce a novel two-stage cellular network traffic prediction framework, JO-DPNet. The framework employs the VMD method to decompose raw cellular network traffic into TSS characterized by simpler and regular internal patterns. Subsequently, the prediction reconstruction module, designed specifically for this purpose, is utilized to extract the spatial-temporal features of TSS. Following this, a fully connected network layer, also tailored for this task, reconstructs the predicted results of TSS. Initially, Section III-A outlines the cellular network traffic prediction problem. Sections III-B and III-C then provide comprehensive introductions to the decomposition and prediction reconstruction methods, respectively. Section III-D elaborates on the TPE-BO process tailored for our two-stage model. Finally, Section III-E offers a detailed description of the overall framework of the proposed model, accompanied by a schematic illustration as shown in Fig. 3.

A. Problem Formulation

Cellular network traffic prediction can be framed as a time series prediction problem, represented by a time-series prediction model. Let the traffic data be divided into equal intervals, such as 10 minutes, 30 minutes, etc. Denote X_t as the traffic value at time t , and the cellular network traffic prediction can be expressed as

$$\hat{X}_t + l = f([X_{t-T+1}, X_{t-T}, \dots, X_t]), \quad (1)$$

where X_t represents the traffic within the time slot X_{t-T} , where T is the traffic collection period, t is the current time, and l is the prediction horizon. The problem of cellular traffic prediction is to forecast the traffic, \hat{X}_{t+l} , at the time interval $t+l$ for a given prediction horizon l . The ultimate objective is to determine a set of model parameters that minimize the error between the predicted flow \hat{X}_{t+l} and the observed value X_{t+l} expressed as follows:

$$\theta^* = \arg \min \mathcal{L}(\hat{X}_{X+l}, X_{t+l}; \theta^*), \quad (2)$$

where \mathcal{L} can be a common loss function for regression problems, such as root mean squared error and mean absolute error.

θ^* is the parameters of the prediction model. In this paper, we proposed a two-stage traffic prediction framework for cellular networks, which integrates Bayesian joint optimization with decomposition and deep neural networks. The model framework is shown in Fig. 3.

B. Decomposition

Empirical Mode Decomposition (EMD), which was developed by Huang et al., is widely used in the field of signal processing to decompose the non-stationary data into a finite number of intrinsic mode functions (IMFs), which contains local features of the original signal at different time scales [44]. However, this method is highly dependent on methods of extremal point finding and lacks physical significance. Additionally, several areas for improvement, such as sensitivity to noise and sampling, limit the further development of EMD. Therefore, Dragomiretskiy et al. [41] proposed an entirely non-recursive VMD. This algorithm uses the variational model to determine the correlation frequency band and extract the corresponding mode components, demonstrating effective anti-noise capabilities and a stronger theoretical foundation compared to EMD. The implementation process of the VMD algorithm is shown in Figure 4. VMD assumes that the original cellular network traffic data X comprises k modes, each corresponding to center frequency ω_k and a finite bandwidth A_k . The mode component $u_k(t)$ is represented as

$$u_k(t) = A_k(t) \cos(\phi_k(t)), \quad (3)$$

where $\phi_k(t)$ is the phase function at time t . Initializing the number of all modes k and the center frequency ω_k , the essence of the VMD calculation is to solve a bandwidth-constrained optimization problem, which is represented as

$$\begin{aligned} \min_{\{u_k\}, \{w_k\}} & \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j w_k t} \right\|_2^2 \right\}, \\ \text{s.t. } & \sum_{k=1}^K u_k(t) = X(t), \end{aligned} \quad (4)$$

where $\delta(t)$ is the impulse function. $\|\cdot\|^2$ denotes the Euclidean norm. To minimize the objective function, the Lagrange multiplier is introduced, converting the constrained optimization problem into an unconstrained one, described as

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \lambda) &= \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j w_k t} \right\|_2^2 \\ &+ \left\| X(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), X(t) - \sum_{k=1}^K u_k(t) \right\rangle, \end{aligned} \quad (5)$$

where α is quadratic penalty factor and λ is Lagrange multiplier. In order to optimize L , the alternating multiplier method is employed, where u_k , w_k and λ are updated alternately, taking u_k as an example, which is mapped to the frequency domain solution by the Fourier transform.

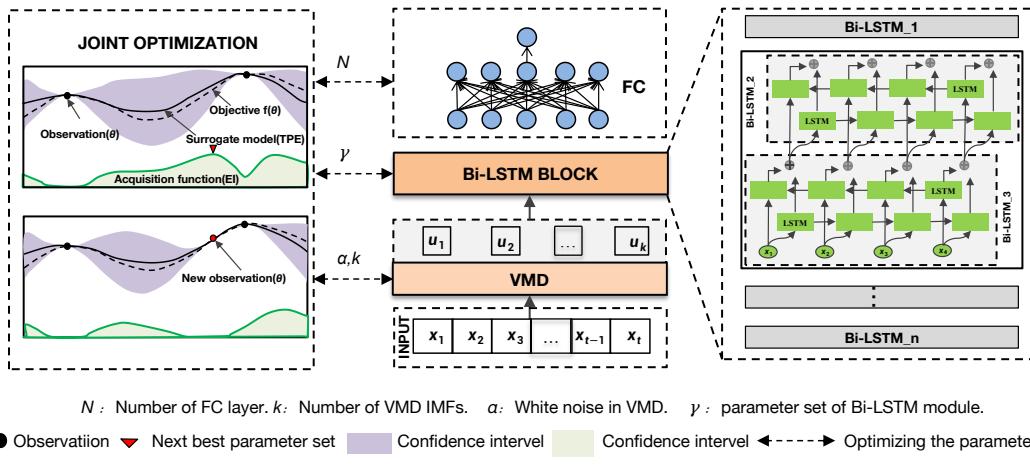


Fig. 3: An overview of our JO-DPNet model. DPNet adapts VMD with joint optimization to mitigate the non-linear and non-stationary features of the original data. The Bi-LSTM BLOCK captures the spatial-temporal features of the traffic subset series, and the FC module is utilized to extract the linear relation of the subset series.

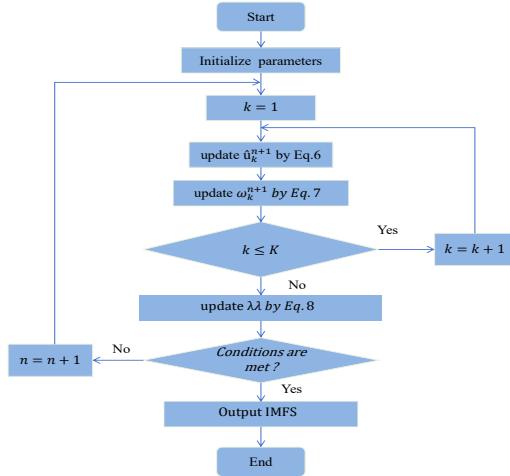


Fig. 4: VMD Algorithm Flowchart.

Next, we apply the alternating multiplier method to solve the optimal solution of the unconstrained problem. The specific procedure is as follows:

- 1) Initialize the mode function u_k , centre frequency ω_k and the Lagrangian multiplier λ .
 - 2) Transform each variable from the time domain to the frequency domain. For the $n + 1$ count, in the non-negative frequency interval, k mode component u_k is updated as follows:

$$\hat{u}_k^{n+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}, \quad (6)$$

where $u^k(\omega)$, $f(\omega)$ and $\lambda(\omega)$ are the Fourier transforms of u , k , $f(t)$ and λ , respectively.

- 3) For the $n + 1$ iteration, update the centre frequency ω_k of each mode component as Eq.7.

$$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}. \quad (7)$$

4) update the Lagrange multiplier as follows:

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right), \quad (8)$$

where τ is the iteration factor.

5) Repeat steps 2-4 until the convergence condition is satisfied:

$$\frac{\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon^*. \quad (9)$$

If Eq.9 holds true, the iteration is stopped, and the result is output. Otherwise, return to Step 2 to continue. The final output of all mode functions can represent the original signal.

During the VMD decomposition, k and α need to be preset, and their reasonable values significantly impact the prediction performance in the subsequent stage [45]. This paper employs a TPE-BO-based optimization algorithm for optimal selection, which will be discussed in Section III-D.

C. Prediction and Construction

While traditional statistical methods, such as Autoregressive Integrated Moving Average (ARIMA) and shallow learning models, including those resembling deep neural networks, can process time series data, their efficiency is often limited. This limitation arises from their inability to effectively account for the long-term temporal dependencies inherent in time series data. To address this shortcoming, LSTM emerges as a robust alternative, recognized as a powerful dynamic model for managing sequence-based tasks.

Although LSTM networks can address the long-term dependency problem, they do not utilize future information. Therefore, this paper employs Bi-LSTM model to simultaneously consider both past and future information. One layer performs forward computations, while another layer executes backward computations to obtain and retain the backward hidden state outputs at each time step. Ultimately, the final output is derived by combining the corresponding outputs from both layers

at each time step. The Bi-LSTM block processes the time series data in two directions iteratively, thereby improving computational efficiency and reducing the risk of overfitting. This paper utilizes Bi-LSTM to capture the spatial-temporal characters of cellular network traffic. The VMD decomposition generates the components $u_k = (u_k(1), u_k(2), \dots, u_k(t))$. Then, we utilize Bi-LSTM to compute forward and backward calculations to obtain the forward forgetting gate vector $f_t^{(\rightarrow)}$, the input gate vector $i_t^{(\rightarrow)}$, the output gate vector $o_t^{(\rightarrow)}$, and the candidate cell state vector $\tilde{c}_t^{(\rightarrow)}$, and the backward forgetting gate vector $f_t^{(\leftarrow)}$, the input gate vector $i_t^{(\leftarrow)}$, the output gate vector $o_t^{(\leftarrow)}$ and the candidate cell state vector $\tilde{c}_t^{(\leftarrow)}$. The computations of Bi-LSTM are as follows:

$$\begin{aligned} i_t^{(\rightarrow)} &= \sigma(W_i^{(\rightarrow)}[h_{t-1}^{(\rightarrow)}, u_k(t)] + b_i^{(\rightarrow)}), \\ f_t^{(\rightarrow)} &= \sigma(W_f^{(\rightarrow)}[h_{t-1}^{(\rightarrow)}, u_k(t)] + b_f^{(\rightarrow)}), \\ \tilde{c}_t^{(\rightarrow)} &= \tanh(W_c^{(\rightarrow)}[h_{t-1}^{(\rightarrow)}, u_k(t)] + b_c^{(\rightarrow)}), \\ o_t^{(\rightarrow)} &= \sigma(W_o^{(\rightarrow)}[h_{t-1}^{(\rightarrow)}, u_k(t)] + b_o^{(\rightarrow)}), \\ h_t^{(\rightarrow)} &= o_t^{(\rightarrow)} \odot \tilde{c}_t^{(\rightarrow)}, \end{aligned} \quad (10)$$

where \odot denotes the element-wise multiplication. $W_f^{(\rightarrow)}$, $W_i^{(\rightarrow)}$, $W_c^{(\rightarrow)}$, and $W_o^{(\rightarrow)}$ denote the weights of neural networks, while $b_f^{(\rightarrow)}$, $b_c^{(\rightarrow)}$, and $b_o^{(\rightarrow)}$ are the bias terms for the respective gates. Finally, the current hidden $h_t^{(\rightarrow)}$ is computed as:

$$\begin{aligned} f_t^{(\leftarrow)} &= \sigma(W_f^{(\leftarrow)}[h_{t+1}^{(\leftarrow)}, x_t] + b_f^{(\leftarrow)}), \\ i_t^{(\leftarrow)} &= \sigma(W_i^{(\leftarrow)}[h_{t+1}^{(\leftarrow)}, x_t] + b_i^{(\leftarrow)}), \\ \tilde{c}_t^{(\leftarrow)} &= \tanh(W_c^{(\leftarrow)}[h_{t+1}^{(\leftarrow)}, x_t] + b_c^{(\leftarrow)}), \\ o_t^{(\leftarrow)} &= \sigma(W_o^{(\leftarrow)}[h_{t+1}^{(\leftarrow)}, x_t] + b_o^{(\leftarrow)}), \\ h_t^{(\leftarrow)} &= o_t^{(\leftarrow)} \odot \tilde{c}_t^{(\leftarrow)}, \\ h_t &= h_t^{(\rightarrow)} \odot h_t^{(\leftarrow)}, \end{aligned} \quad (12)$$

where h_t represents the final hidden element of Bi-LSTM, which is the concatenated vector of the forward output $h_t^{(\rightarrow)}$ and the backward output $h_t^{(\leftarrow)}$. Through the above process, Bi-LSTM can learn both past and future features of time series data and the predictive output is generated from past and future contexts. Finally, the forward and backward hidden states are concatenated to obtain the current bidirectional hidden state h_t . For each IMF $u_k(t)$, the feature vector $f(u_k)$ is derived through a mapping function $f(\cdot)$, which is then processed through FC to obtain the prediction result X_i representing:

$$X_i = f \left(\sum_{j=1}^N w_{ij} u_{kj} + \beta_{ki} \right), \quad i = 1, 2, \dots, n. \quad (13)$$

Where u_{kj} is the output of the j -th neuron in the previous layer, w_{ij} is the weight connecting the j -th neuron in the previous layer to the i -th neuron in the current layer, β_{ki} is the bias term of the i -th neuron, $f(\cdot)$ is the activation function applied on the weighted sum of inputs, and y_i is the output of the i -th neuron in the current layer. The last layer of FC output is the prediction value of cellular network X_i . The multi-horizon forecasting adopts a recursive approach where

each prediction serves as input for subsequent time steps. As formalized in Equation 14, the l -hour ahead prediction \hat{y}_{t+l} depends on both external features x_{t+l-1} and the previous prediction \hat{y}_{t+l-1} . This chained structure necessitates joint optimization of decomposition and prediction parameters to control error propagation.

$$\hat{y}_{t+l} = f(\hat{y}_{t+l-1}, x_{t+l-1}) \quad (14)$$

D. Parameter Optimization Based Bayesian

The proposed JO-DPNet model employs a joint optimization strategy that integrates VMD and Bi-LSTM with TPE-BO. First, VMD is utilized to decompose the traffic sequence into k subsets; Subsequently, these k groups data sequences are used for training and testing the Bi-LSTM model. In this paper, the TPE-BO method is employed to optimize the selection of several model parameters, including k , α in the VMD, N in FC layer, and the parameter set γ , which is the adjustable parameter set of Bi-LSTM. The workflow of Bayesian hyperparameter optimization is illustrated in Fig. 5.

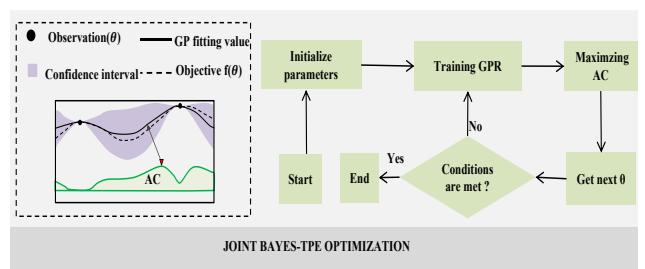


Fig. 5: Framework of Bayesian Hyperparameter Optimization. The workflow includes four core phases: (a) Surrogate model construction using Gaussian processes(GP), (b) Acquisition function (AC)maximization (e.g., Expected Improvement), (c) Objective function evaluation, and (d) Dataset updating with new observations.

In this study, we implemented a variant of the BO algorithm, namely TPE, to search for the best hyperparameter combination that yields the lowest error on the cellular network traffic dataset. A critical advantage of using TPE is that it requires fewer evaluations compared to GS and RS. This is because TPE builds a probabilistic model (a surrogate model) of the objective function and evaluates only the most promising hyperparameters [46] based on the results of previous trials. Therefore, TPE will only select the most promising combination guided by probability rather than exhaustively searching in GS and RS. This can accelerate the convergence time and lead to better hyperparameter combinations as solutions. The Bayesian optimization algorithm, with TPE as the surrogate model utilized in this paper, is delineated in Algorithm 1. Here, \mathcal{F} represents the function to be optimized, ϵ denotes the hyperparameter combination, (EI) is the acquisition function, \mathcal{M} is the probability distribution model, and \mathcal{H} is the historical record of observations, used to model the probability distribution. The acquisition function is crucial for identifying the next hyperparameter set that minimizes the loss. One of the

most commonly employed schemes is Expected Improvement (EI), defined as follows:

$$EI_{y^*}(\epsilon^*, \mathcal{M}) = \int_{-\infty}^{+\infty} \max(y^* - y, 0) p_M(y|\epsilon) dy, \quad (15)$$

where y and y^* denote the actual and threshold values, respectively. Thus, the next parameter we seek is given by:

$$\epsilon_{new} = argmax EI_{y^*}(\epsilon). \quad (16)$$

The EI function favors regions with small means and large variances.

The TPE is a parameter optimization algorithm rooted in the Bayesian approach, utilizing the Parzen window to estimate prior and posterior probability distributions and selecting the next hyperparameter to be tested based on the ratio of these distributions. TPE can identify the optimal hyperparameter combination more rapidly than traditional Gaussian process regression methods. Specifically, $p(y|\epsilon)$ can be decomposed into $p(\epsilon|y)$ and $p(y)$, and TPE divides $p(\epsilon|y)$ as follows:

$$p(\epsilon|y) = \begin{cases} l(\epsilon), & y < y^* \\ g(\epsilon), & y > y^* \end{cases} \quad (17)$$

TPE constructs distinct distributions for observations on either side of the threshold y^* . Set a hyperparameter σ , which is the quantile of σ . Thus $p(y < y^*) = \sigma$, σ takes the default value of 0.25. The Eq.17 is obtained by Eq.18.

$$p(\epsilon) = \int_{-\infty}^{y^*} p(\epsilon|y) p(y) dy + \int_{y^*}^{+\infty} p(\epsilon|y) p(y) dy = \sigma l(\epsilon) + (1 - \sigma) g(\epsilon). \quad (18)$$

Substitute Eq.18 into Eq.15 yields the following update:

$$EI_{y^*}(\epsilon) = \frac{\int_{-\infty}^{y^*} (y^* - y) p(y) dy}{\gamma + (1 - \gamma) \frac{g(\epsilon)}{l(\epsilon)}}. \quad (19)$$

$$EI_{y^*}(\epsilon) \propto \left(\gamma + (1 - \gamma) \frac{g(\epsilon)}{l(\epsilon)} \right)^{-1}. \quad (20)$$

As shown in Eq.20 show, the value of EI is proportional to the reciprocal of the denominator when σ . After determination, the size of the denominator only depends on the ratio $l(\epsilon)/g(\epsilon)$ of the two segment probabilities of ϵ . So the ϵ we are looking for is the one that maximizes the ratio $l(\epsilon)/g(\epsilon)$.

Algorithm 1 Parameters Optimization Based TPE-BO of DPNet.

- 1: **Data:** Objective function \mathcal{F} , initial parameters ϵ , search space \mathcal{E} , number of iterations \mathcal{T}
 - 2: **Result:** Best set of parameters ϵ^*
 - 3: Initialize model \mathcal{M}
 - 4: $\mathcal{H} \leftarrow \epsilon^*$
 - 5: Sample initial parameters ϵ from \mathcal{D}
 - 6: Evaluate $\mathcal{F}(\epsilon)$
 - 7: Update $\mathcal{H} \leftarrow \{(\theta, \mathcal{F}(\epsilon))\}$
 - 8: **for** $t \leftarrow 1$ to \mathcal{T} **do**
 - 9: Select parameters to evaluate based on acquisition function using \mathcal{M}
 - 10: Evaluate \mathcal{F} at selected parameters
 - 11: Update \mathcal{H} with newly observed data
 - 12: Fit model \mathcal{M} by \mathcal{H}
 - 13: **end for**
 - 14: **return** ϵ^* given \mathcal{H}
-

Using the TP-BO algorithm, we obtain the parameters combination of the proposed cellular network traffic prediction model, JO-DPNet.

E. The Whole Process of JO-DPNet

Our strategy is to make full use of the three techniques that seamlessly integrated solution owns: the TSD ability of VMD, long-term dependence capturing ability of Bi-LSTM, and parameter optimization ability of TPE-BO. The overall framework of our proposed JO-DPNet architecture is shown in Fig. 3.

The first stage involves data preprocessing. The missing values in the dataset X are filled by calculating the average of the surrounding cells and normalizing the data.

The second stage consists of decomposing the preprocessed cellular network traffic sequence into k time series segments (TSS) using the decomposition module (DP).

In the third stage, the input structure required for the decomposed TSS u_k is established. The samples are divided using a time series-based cross-validation method and fed into the N -layer Bi-LSTM of the PC for training. The output from the current layer of the Bi-LSTM serves as the input for the subsequent layer. The TSS prediction results are then fused using the FC module to produce the final prediction results.

The entire training process employs BO-TPE for synchronous joint parameter optimization. The optimal parameters k , α , and γ are obtained and substituted into the model to derive the optimal prediction model. The performance of the proposed model is evaluated using three performance indicators: root mean square error (RMSE), mean absolute error (MAE), and cumulative distribution function (CDF).

IV. EXPERIMENTAL EVALUATION

In this section, we conduct extensive experiments on real-world datasets, and the experimental results demonstrate the effectiveness of our proposed JO-DPLNet framework.

A. Experiment Setting

This section describes our experimental framework and the consequent outcomes derived from our research. The objective was to validate the efficacy of the newly proposed methodology for forecasting cellular network traffic, denoted as JO-DPLNet. Following the time-series split protocol in Table I, we implement forward-chaining validation with five temporal folds. Each validation set strictly follows its corresponding training period chronologically, with the final fold reserved for testing. The sliding window size $p = 10$ ensures temporal coherence while preventing look-ahead bias. Our experimental infrastructure was established on a platform operating on 64-bit Windows 10, enhanced by a GTX 3080 Ti GPU, and powered by CUDA version 10.0. The computational core of the system comprised an 8-core AMD Ryzen 73700X CPU. The architecture of JO-DPLNet was meticulously designed using TensorFlow version 2.1, Keras version 2.2.5, and Python version 3.7.0. This combination of hardware and software facilitated the rigorous evaluation of the proposed network traffic prediction method, enabling a comprehensive analysis of its performance characteristics.

TABLE I: RESIDUAL STATISTICS ACROSS HORIZONS

Fold	Training Hours	Validation Hours
1	1-57	58
2	1-114	115
3	1-171	171
4	1-228	229
5	1-285	286
6	1-342	343
7	1-339	340
8	1-456	457
9	1-513	514
10	1-570	571

B. Dataset

Extensive experiments were conducted on three real-world cellular traffic datasets. We utilized a dataset provided by Telecom Italia, a prominent European network operator, which is widely used in recently Reseaches [47]. The dataset records user activity, including short message service (SMS), voice call service (Call), and Internet service (Internet), for Milan from 00:00:00 on November 1, 2013, to 23:50:00 on January 1, 2014. The dataset was recorded for 62 days at 10-minute intervals, with three types and a size of approximately 19 GB. The city of Milan is divided into 100×100 square grids, each square having a size of about 235×235 meters, which we call a cell. The cellular network traffic for cell i can be expressed as follows:

$$X_i(t) = \sum_{v \in C_{map}} R_v(t) \frac{A_{v \cap i}}{A_v}, \quad (21)$$

where $X_i(t)$ is the number of records of cell i at time t , C_{map} is the coverage map of Milan city, $A_{v \cap i}$ is the intersection of the coverage area of the base station and the area of this cell, A_v is the coverage area of the base station, and $R_v(t)$ is the number of records from this base station at time t . To rigorously assess JO-DPNet's cross-environment stability,

we conducted comparative experiments on C2TM and Taiwan Cellular Traffic datasets.

The C2TM dataset comprises cellular traffic records collected from 13,269 base stations in a medium-sized Chinese city between August 19 and August 26, 2012 [48]. Each entry contains the base station identifier, timestamp, number of active mobile users, transmitted packet count, and hourly traffic volume in bytes.

The Taiwan cellular traffic dataset was obtained from a large-scale cellular geo-system in Hsinchu city, Taiwan [49]. It includes records spanning January 1 to June 30, 2020, with original 5-minute granularity data aggregated into 1-hour intervals for consistency. Each record features the timestamp, GPS coordinates, location type (outdoor/indoor), and anonymized international mobile equipment identity (IMEI) counts aggregated across six major road intersections.

C. Baselines

In this paper, we compare the proposed method, JO-DPLNet, with a collection of carefully selected baseline methods:

- ConvLSTM [13]. This model integrates CNN and LSTM to capture spatial dependencies and temporal relationships holistically.
- LSTM-GPR [7]. LSTM-GPR is a hybrid model that combines the LSTM network with the Gaussian Process Regression (GPR) algorithm to capture the time dependence of cellular network traffic.
- MSTL-GLT [38]. MSTL-GLT proposes a Multiple Seasonal-Trend decomposition using Loess based Global-Local Traffic Prediction (MSTL-GLTP) framework that assures prediction accuracy while maintaining low complexity.
- STCNet [2]. STCNet is an advanced spatial-temporal cellular traffic prediction model that uses external data as features to assist traffic prediction, employing transfer learning technology to improve prediction accuracy.
- NLG-NLAM [37]. NLG-NLAM is a deep learning method based on a non-linear gradient non-local attention mechanism, considering dynamic non-local spatial correlation, self-attention, and the correlation of spatial-temporal feature fusion.
- MVSTGN [4]. MVSTGN employs an attention mechanism to model from two perspectives, space and time, using Dense CNN to capture local spatial-temporal dependencies.
- GLSTTN [14]. GLSTTN is a graph attention spatial-temporal network that can capture not only local geographical dependency but also distant inter-region relationship when considering spatial factor.

To ensure a fairer comparison, we employ identical common hyperparameters for all model trainings. These parameters are optimized using both the parameter space shown in Table II and the proposed optimization method in this paper. The common hyperparameters include: Batch size, Dropout rate, Learning rate, and epochs. The unique parameters specific to our proposed method include: $N_Bi - LSTM$, N , h ,

k , and α . The baseline models are chosen based on whether implementation resources or experimental results for the same dataset.

D. Evaluation Metrics

This study employs established metrics to assess the accuracy of our model's predictions. Specifically, the root mean squared error (RMSE), mean absolute error (MAE), and Coefficient of determination (R^2) are utilized to quantify the discrepancies between the predicted and actual values. The formulas for these metrics are as follows:

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |f(x_i) - y_i|, \quad (22)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2}, \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - f(x_i))^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (24)$$

where $f(x_i)$ and y_i are the predicted value and the corresponding ground truth, \bar{y} is the average of y_i , respectively, m is the total number of predicted values. To further enhance the interpretability of our model's predictive performance, we introduce the cumulative distribution function (CDF) of the prediction errors. This function is mathematically expressed as:

$$F(\varepsilon) = P(\varepsilon \leq a), \quad (25)$$

where ε represents the prediction error, and $F(\varepsilon)$ quantifies the probability that this error does not exceed a specified threshold a . This approach allows us to graphically represent the distribution of errors, providing a clearer visualization of the model's performance [50].

E. Prediction Results

Our experiments are conducted from various perspectives to comprehensively evaluate the performance of the proposed JO-DPNet. This paper adopts a two-stage joint synchronous optimization method based on BO-TPE to optimize and find the optimal parameter combination simultaneously, as shown in Table II. The JO-DPNet model is constructed based on the parameter settings outlined in Table II, and it is trained using the training data. The proposed JO-DPNet model consists of three primary components: first, DP is employed to decompose the original network traffic sequence into k subsets (or modes); subsequently, these k groups of data sequences serve as the input and output for the predictive component (PC). The optimal parameter set γ , which includes N , k , α , batch size, dropout rate, dense layer, epochs, and h , is identified and substituted into the model to achieve the best prediction performance. The optimization process utilizing the TPE-BO algorithm to determine model parameter values is detailed in Table II.

State-of-the-art mode decomposition methods have been applied to traffic datasets within cellular networks, with comparative results illustrated in Fig. 6. As the analysis progresses

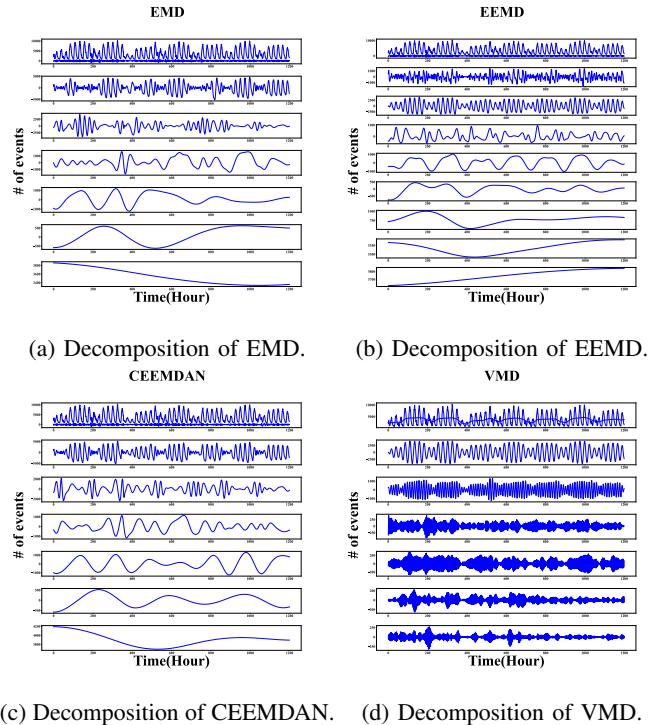


Fig. 6: Visualization of the different mode decompose methods in Internet cellular traffic on Milan's Duomo -5060.

towards smoother and more regular subsequences, the superior performance of VMD becomes evident, characterized by the absence of mode aliasing and endpoint effects. Consequently, the data processed through VMD leads to an improvement in predictive accuracy.

In Fig. 7, it is evident that the trend curves of the IMFs predicted by the proposed model closely align with the actual value curves. This observation indicates that the feature extraction module employed in this study effectively captures the characteristics of the TSS.

To evaluate the performance of JO-DPNet, we compare this novel approach with state-of-the-art algorithms while employing uniform data preprocessing methods to eliminate discrepancies arising from different methodologies. Examination of Fig. 9 reveals that most predicted values generated by JO-DPNet are much closer to the corresponding ground truths, suggesting a substantial congruence between the anticipated and actual data points. This method uses BO-TPE for joint hyperparameter optimization, the process is shown in Fig. 5. As the number of iterations increases, BO-TPE used in this paper gradually selects a more effective combination of joint parameters. When the number of decompositions is large, a mixing phenomenon occurs, while fewer decompositions lead to the loss of information from the original signal. Ultimately, $k = 6$ is chosen, and other parameters are listed in Table II. The optimal parameter combination is finally obtained after 50 iterations, and the results of each iteration are shown in Fig. 8a. The test loss of the model decreases as the joint optimization process proceed are shown in Fig. 8b. To accurately assess the efficacy of the proposed cellular traffic prediction framework,

TABLE II: OPTIMIZATION USING TPE-BO ALGORITHM TO OBTAIN MODEL PARAMETER VALUES.

Parameter	Description	Parameter space	Value
N_Bi-LSTM	number of Bi-LSTM unit	[1, 5]	3
batch_size	Batch size	[30, 80]	69
dropout_rate	Dropout rate	[0.2, 0.5]	0.3139
N	Number of fully connection layer	[1, 5]	4
learning_rate	Learning rate	[0.001, 0.01]	0.0069
epochs	Number of iterations	[1, 60]	36
h	Number of hidden units	[20, 80]	34
k	Number of vmd subsets	[3, 20]	6
α	the variance of white noise in vmd	[0, 5000]	2000

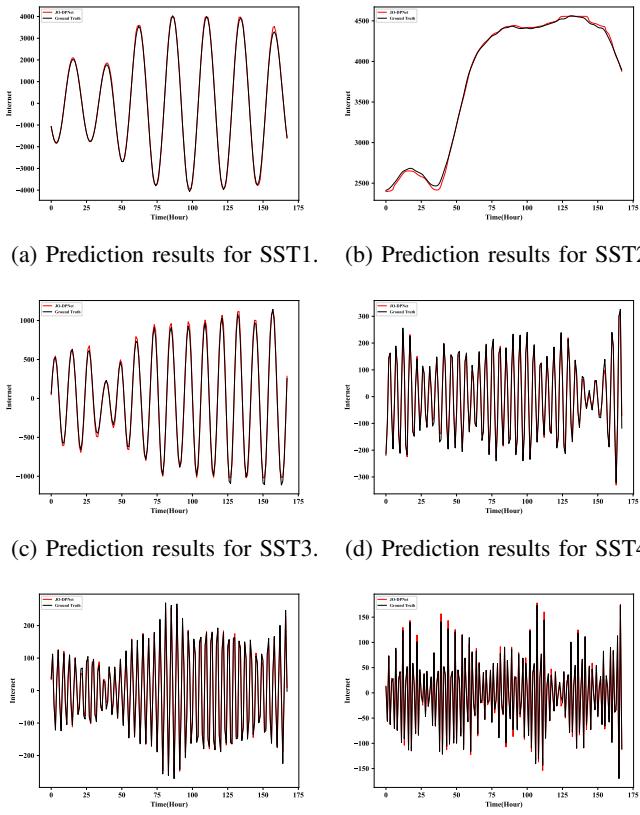


Fig. 7: Visualization of TSS prediction results in Internet cellular on Milan's Duomo-5060.

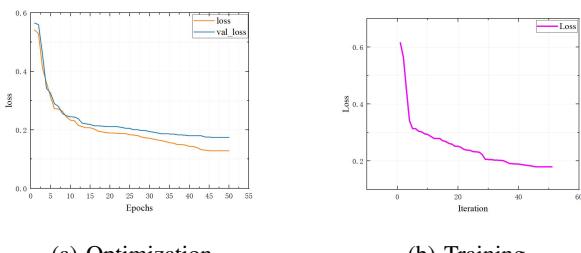


Fig. 8: The convergence of the process of model.

this study utilized a suite of metrics, including the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), facilitating a comprehensive comparison against seven contemporary analytical techniques, as elaborately presented in Table III. The results indicate that JO-DPNet outperformed its counterparts, achieving superior prediction accuracy. Specifically, when compared to GLSTTN, JO-DPNet demonstrated a reduction in MAE by 29%, 3%, and 19%, underscoring a significant improvement. These findings attest to the precision and enhanced stability of the proposed method, making it a preferable alternative to existing approaches within the realm of cellular traffic forecasting.

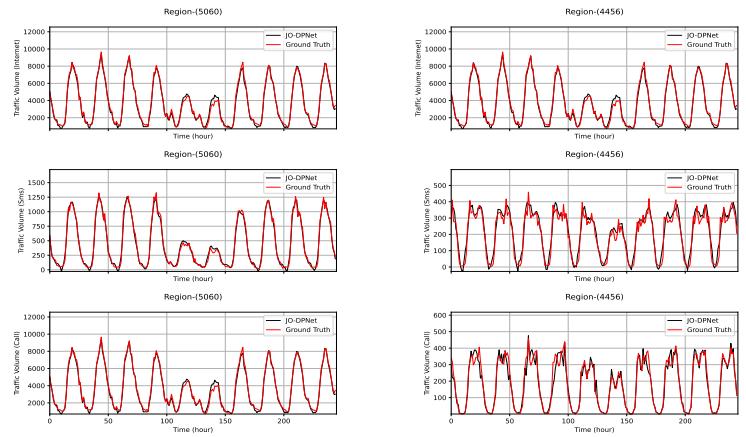


Fig. 9: Prediction results for the cellular network traffic time series with the JO-DPNet.

To rigorously assess JO-DPNet's cross-environment stability, we conducted comparative experiments on both datasets. As demonstrated in Table IV, our model achieves MAE reductions of 53% and 49% respectively compared to state-of-the-art baseline methods. These improvements confirm the method's generalizability and effectiveness across diverse operational contexts. Table V demonstrates horizon-dependent residual characteristics. While all horizons show white noise residuals (Ljung-Box $p > 0.05$), the increasing autocorrelation(AC) at lag 1, and the Standard Deviation (STD) reveals cumulative error effects in longer forecasts. This motivates our adoption of a joint optimization strategy to mitigate such error propagation. Furthermore, we performed F-tests on each baseline method and the proposed method. As presented in Table III, all p-values are below 0.05, indicating that the residual variance of JO-DPNet is significantly lower than that of baseline models

TABLE III: PERFORMANCE COMPARISON ON MILANO DATASET: JO-DPNET VS. BASELINE METHODS.

METHOD	SMS				Internet				Call			
	RMSE	MAE	R2	P	RMSE	MAE	R2	P	RMSE	MAE	R2	P
LSTM-GPR [7]	60.8192	36.9156	0.8135	0.0526	125.2986	80.2629	0.8753	0.0622	7.2713	16.7423	0.8325	0.1841
ConvLSTM [13]	90.1151	47.4972	0.7132	0.0826	265.3725	142.3629	0.8901	0.0479	47.6126	26.1825	0.7359	0.1670
MSTL-GLTP [38]	57.7284	33.4174	0.8203	0.0232	209.5665	125.493	0.926	0.0232	42.6947	23.8939	0.8675	0.0232
SCTNet [2]	56.8673	30.7026	0.8256	0.0232	172.0828	95.9306	0.9024	0.0232	38.3281	18.2881	0.8939	0.0462
NLG-NLAM [37]	52.7875	30.2396	0.8125	0.0193	167.25	93.1604	0.9162	0.0426	32.1264	17.5839	0.9104	0.0312
MVSTGN [4]	45.9331	24.2944	0.882	0.0262	163.3348	88.3751	0.954	0.0590	30.0512	15.2506	0.9339	0.0462
GLSTTN [14]	54.0937	50.1053	0.9281	0.0257	149.9281	81.3751	0.974	0.0499	27.1037	14.4773	0.9498	0.1175
JO-DPNet(OURS)	44.3438	35.6683	0.9881	1	132.9566	79.1096	0.9901	1	23.8404	11.6600	0.9892	1

TABLE IV: PERFORMANCE COMPARISON ON C2TM AND TAIWAN DATASET: JO-DPNET VS. BASELINE METHODS

METHOD	C2TM				Taiwan			
	RMSE	MAE	R2	P	RMSE	MAE	R2	P
LSTM-GPR [7]	419.8967	345.523	0.9661	0.0204	161.8464	113.0105	0.9152	0.0190
ConvLSTM [13]	433.2539	308.8965	0.9682	0.0238	94..3512	64.4878	0.9821	0.0181
MSTL-GLTP [38]	489.9661	305.5917	0.2771	0.1554	180.1292	103.7638	0.1037	0.0118
SCTNet [2]	354.3342	264.9276	0.9799	0.0337	105.6715	126.2351	0.5323	0.0183
GLSTTN [14]	390.1956	356.9328	0.3466	0.0442	98.5261	102.1356	0.9876	0.0196
MVSTGN [4]	343.4824	267.5458	0.9615	0.0421	112.5482	153.1545	0.9454	0.02163
JO-DPNet(OURS)	259.6188	174.1752	0.9885	1	76.1815	65.9451	0.9851	1

[51]. These findings underscore the enhanced stability of our approach.

TABLE V: RESIDUAL STATISTICS ACROSS HORIZONS

Metric	1h	3h	6h
MAE	0.0834	0.1300	0.1695
STD	49.9767	109.4189	155.1593
p-value	0.1328	0.1290	0.1073
AC(lag=1)	0.062	0.0508	0.0599

F. Ablation Studies

To understand the contribution of each module in JO-DPNet, we conducted ablation experiments on real-world datasets using various model variants. The same data preprocessing methods were applied to mitigate the effects of different methodologies. Fig. 10 presents the predicted values for EMD, CEEMDAN, and VMD alongside the actual values. It is evident from Fig. 10 that the accuracy of the three models is significantly higher than that of models without TSD. The

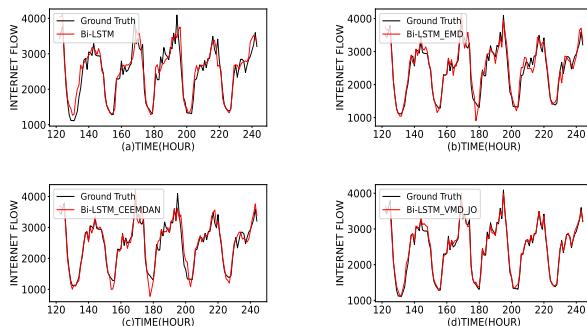


Fig. 10: Prediction results for the cellular network traffic time series with the variants of TSD on Milan's Duomo -5060.

curve fitting degree between the predicted values from VMD-Bi-LSTM and the ground truth is the highest in most cases.

Fig. 11 compares the predicted values of LSTM, GRU, and Bi-LSTM, demonstrating that the prediction results from VMD-Bi-LSTM are closer to the ground truth than those of the other two models. In conclusion, the proposed JO-DPNet achieves the highest prediction accuracy.

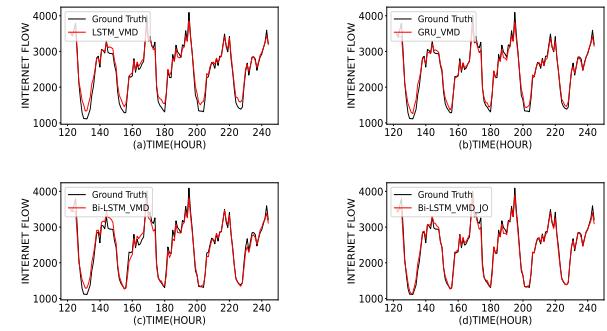


Fig. 11: Prediction results for the cellular network traffic time series with prediction model variants on Milan's Duomo - 5060.

To further validate the efficacy of the proposed cellular traffic prediction model, MAE and RMSE metrics were used for comparison with seven peer models, as shown in Table VI. To illustrate the performance differences intuitively, we provide the CDF of the prediction errors in Fig. 12a and Fig. 12b. The results indicate that 75% of the prediction errors for EMD, CEEMDAN, VMD, LSTM, GRU, and Bi-LSTM are approximately lower than 484, 458, 267, 227, and 179, respectively.

G. Complexity Analysis

To evaluate the computational feasibility of the proposed JO-DPNet framework, we analyze the complexity of each

TABLE VI: PERFORMANCE COMPARISON of DIFFERENT VARIANTS

	Duomo (5060)		Navigli(4456)	
	MAE	RMSE	MAE	RMSE
Bi-LSTM	320.0822	437.0363	246.98	357.5032
EMD+Bi-LSTM	358.6912	500.2894	100.6714	148.3482
EEMD+Bi-LSTM	130.7573	169.3412	71.375	104.2
CEEMDAN+Bi-LSTM	155.28	224.0554	105.4437	148.4666
VMD+LSTM	116.9974	142.6883	67.2619	87.4665
VMD+GRU	109.2442	134.5193	69.8825	92.1214
VMD+Bi-LSTM	107.2385	129.1423	66.3575	85.4254
JO-DPNet (OURS)	79.1096	132.9566	61.873	77.077

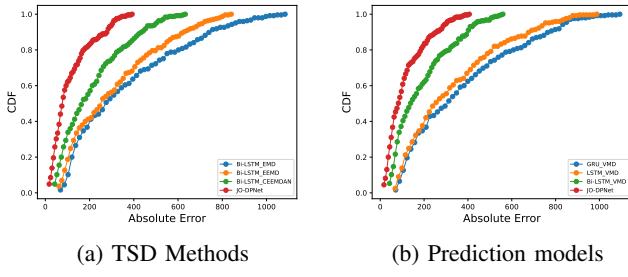


Fig. 12: CDF of the prediction errors between different variants on Milan's Duomo -5060.

component and aggregate them to derive the overall complexity. The VMD decomposes the input signal into k IMFs through iterative optimization. Let I denote the number of iterations and n represent the data length. The complexity is dominated by Fourier transforms and mode updates $\mathcal{O}(kIn \log n)$ [41]. In this paper, by replaced Gaussian process modeling with tree-structured hyperparameter space partitioning, TPE-BO reduces the per-iteration computational complexity from $\mathcal{O}(n^3)$ to $\mathcal{O}(n \log n)$ [52]. The BO module performs B optimization rounds, evaluating S samples per round. Let C_{VMD} , $C_{Bi-LSTM}$, and C_{FC} represent the costs of the VMD decomposition, Bi-LSTM model, and fully connected layer, respectively. The total complexity of the Bayesian Optimization is: $\mathcal{O}(BS(C_{VMD} + C_{Bi-LSTM} + C_{FC}))$. For an input sequence of length n with d features, the Bi-LSTM processes data in both directions. Let h be the number of hidden units per direction. The complexity for a single LSTM layer is $\mathcal{O}(n(dh + h^2))$ per direction, and hence the total Bi-LSTM complexity is: $\mathcal{O}(2n(dh + h^2))$ [12]. The fully connected (FC) layer projects the Bi-LSTM outputs (dimension $2h$) to o -dimensional predictions, with complexity: $\mathcal{O}(2ho)$. The total computational complexity of JO-DPNet, dominated by the optimization iterations, is:

$$\mathcal{O}(BS(kIn \log n + 2n(dh + h^2) + 2ho)). \quad (26)$$

After parameter optimization, the inference complexity reduces to:

$$\mathcal{O}(kIn \log n + 2n(dh + h^2) + 2ho). \quad (27)$$

Experiments demonstrate that JO-DPNet achieves high prediction accuracy with tractable computational complexity. Dominant terms scale linearly or quasilinearly with input size, ensuring applicability to large-scale cellular traffic data.

The derived complexities affirm the model's practicality for real-world deployment, balancing accuracy and computational efficiency. To tackle the challenge of identifying the global optimum during the two-stage parameter optimization process, this paper employs a two-stage joint synchronous optimization method based on BO-TPE. This approach allows for the simultaneous optimization and identification of the most effective parameter combinations. To assess the efficiency and computational complexity of this method, we conducted comparative experiments against several conventional optimization algorithms. The results of these experiments are presented in Table VII.

In comparison to the Particle Swarm Optimization (PSO) method, as well as the Random Search-Variational(RS), Grid Search-Variational(GS), and Bayesian optimization based on Gaussian Process(BO-GP), the JO-DPNet demonstrates significant advantages. JO-DPNet outperforms these models in terms of overall prediction accuracy, stability during training, and reduced training times. Specifically, JO-DPNet achieves a substantial reduction in training time compared to its counterparts. To verify the advancedness of the optimization algorithm used in this paper, we selected representative methods in Ref. [39], Artificial Rabbits Optimization (ARO), Biogeography-Based Optimization (BBO), Equilibrium Optimizer (EO), and Student Psychology based Optimization (SPBO) for research. As shown in Table VII, the optimization algorithm used in this paper is more efficient than the other four methods. This is attributed to the fact that the TPE-BO joint parameter optimization method gives priority to the use of prior knowledge in the parameter optimization process. Moreover, to verify the feasibility of the proposed method in real-time application scenarios, we conducted testing on a CPU-equipped server. The results show that our optimized model generates cellular traffic segment predictions within 10 seconds per instance, meeting the 5G network management system's real-time, minute-level threshold. This confirms the practical applicability of our method in real-time scenarios, making it feasible for use in dynamic, time-sensitive network environments.

TABLE VII: COMPARISON OF COMPUTATIONAL COMPLEXITY.

Metrics Methods	Training(m)	CPU(s)	GPU(s)	Parameters
Bi-LSTM	1.78	59	8	27,269
VMD+Bi-LSTM	10.4	147	14	70,585
GS+VMD+Bi-LSTM	45.38	154	13	70,585
RS+VMD+Bi-LSTM	40.56	137	11	70,585
PSO+VMD+Bi-LSTM	35.25	161	13	70,585
BO-GP+VMD+Bi-LSTM	83.45	167	19	70,585
ARO+VMD+Bi-LSTM	51.51	157	13	70,585
BBO+VMD+Bi-LSTM	53.56	159	12	70,585
EO+VMD+Bi-LSTM	48.38	146	12	70,585
SPBO+VMD+Bi-LSTM	43.65	137	11	70,585
JO-DPNet (OURS)	26.45	109	10	70,655

V. CONCLUSION AND FUTURE WORK

This paper introduces a novel two-stage cellular network traffic prediction framework named JO-DPNet, which integrates Bayesian joint optimization, decomposition techniques,

and deep neural networks. In the first stage, raw cellular network traffic data is decomposed into more regular and simplified TSS using the VMD method. In the second stage, the PC module utilizes a Bi-LSTM architecture to effectively capture contextual information at each data point, delving into the deep spatial-temporal features of the cellular traffic subsequences. Furthermore, the FC layer is employed to reconstruct the prediction results based on the linear dependencies among the TSS predictions. To enhance prediction accuracy, the TPE-BO optimization algorithm is employed to fine-tune critical parameters of the VMD, including the number of decompositions and the penalty factor. This effectively mitigates mode aliasing and endpoint effects that can arise from improper manual parameter selection. Additionally, the TPE-BO algorithm is used to optimize the number of layers in the Bi-LSTM architecture and the FC layer, enabling the PC module to more accurately capture the complex spatial-temporal characteristics of network traffic. The two-stage synchronous optimization strategy simultaneously searches for the optimal parameter combination, thereby avoiding the potential for local optima that can occur with separate stage-by-stage optimization. Experimental results demonstrate that the non-linearity and non-stationarity characteristics of cellular network traffic data are significantly improved after processing by the decomposition module. With the aid of synchronous joint parameter optimization, the designed PC module captures the spatial-temporal features of the data with greater precision. Consequently, JO-DPNet outperforms the current state-of-the-art cellular network traffic prediction methods in terms of predictive accuracy.

Practically, JO-DPNet can be integrated into 6G network controllers for dynamic resource allocation, reducing latency during peak traffic periods. Future work includes: Extending JO-DPNet to hybrid 5G/4G networks and satellite-terrestrial integrated communications, with validation on multi-service traffic patterns (e.g., video streaming, online gaming, and industrial IoT). Developing dynamic decomposition strategies for emerging 6G services with ultra-bursty traffic characteristics, such as holographic communications and digital twin synchronization.

REFERENCES

- [1] Chamitha De Alwis, Anshuman Kalla, Quoc-Viet Pham, Pardeep Kumar, Kapal Dev, Won-Joo Hwang, and Madhusanka Liyanage. Survey on 6g frontiers: Trends, applications, requirements, technologies and future research. 2:836–886.
- [2] Chuanting Zhang, Haixia Zhang, Jingping Qiao, Dongfeng Yuan, and Minggao Zhang. Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data. 37(6):1389–1401.
- [3] Chuanting Zhang, Shuping Dang, Basem Shihada, and Mohamed-Slim Alouini. Dual attention-based federated learning for wireless traffic prediction. In *IEEE INFOCOM 2021-IEEE conference on computer communications*, pages 1–10. IEEE.
- [4] Yang Yao, Bo Gu, Zhou Su, and Mohsen Guizani. MVSTGN: A multi-view spatial-temporal graph network for cellular traffic prediction. 22(5):2837–2849.
- [5] Amin Azari, Fateme Salehi, Panagiotis Papapetrou, and Cicek Cavdar. Energy and resource efficiency by user traffic prediction and classification in cellular networks. 6(2):1082–1095.
- [6] Min Chen, Yiming Miao, Hamid Gharavi, Long Hu, and Iztok Humar. Intelligent traffic adaptive resource allocation for edge computing-based 5g networks. 6(2):499–508.
- [7] Wei Wang, Conghao Zhou, Hongli He, Wen Wu, Weihua Zhuang, and Xuemin Shen. Cellular traffic load prediction with LSTM and gaussian process regression. In *ICC 2020 - 2020 IEEE international conference on communications (ICC)*, pages 1–6.
- [8] J. Shi, Y. B. Leau, K. Li, and H. Chen. Optimal variational mode decomposition and integrated extreme learning machine for network traffic prediction. 9:51818–51831. Type: Journal Article.
- [9] Thanh Tran, Li Hao, and Quang Trinh. Cellular network traffic prediction using exponential smoothing methods. 18:1–18.
- [10] Yantai Shu, Minfang Yu, Jiakun Liu, and O.W.W. Yang. Wireless traffic modeling and prediction using seasonal ARIMA models. In *IEEE international conference on communications, 2003. ICC '03.*, volume 3, pages 1675–1679.
- [11] Xiaming Chen, Yaohui Jin, Siwei Qiang, Weisheng Hu, and Kaida Jiang. Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale. In *2015 IEEE international conference on communications (ICC)*, pages 3585–3591.
- [12] Hongxin Shao and Boon-Hee Soong. Traffic flow prediction with long short-term memory networks (lstms). In *2016 IEEE region 10 conference (TENCON)*, pages 2986–2989.
- [13] Chaoyun Zhang and Paul Patras. Long-term mobile traffic forecasting using deep spatio-temporal neural networks. In *Proceedings of the eighteenth ACM international symposium on mobile ad hoc networking and computing, MobiHoc '18*, pages 231–240. Association for Computing Machinery. Number of pages: 10 Place: Los Angeles, CA, USA.
- [14] Bo Gu, Junhui Zhan, Shimin Gong, Wanquan Liu, Zhou Su, and Mohsen Guizani. A spatial-temporal transformer network for city-level cellular traffic analysis and prediction. pages 1558–2248.
- [15] L. I. Yu, M. A. Ziang, Zhiwen Pan, Nan Liu, and Xiaohu You. Prophet model and gaussian process regressionbased user traffic prediction in wireless networks. 63(4):8.
- [16] José Henrique Kleintübing Larcher, Stefano Frizzo Stefenon, Leandro dos Santos Coelho, and Viviana Cocco Mariani. Enhanced multi-step streamflow series forecasting using hybrid signal decomposition and optimized reservoir computing models. 255:124856. Publisher: Elsevier.
- [17] Tian-jian Luo. Selective multi-view time-frequency decomposed spatial feature matrix for motor imagery EEG classification. 247:123239. Publisher: Elsevier.
- [18] Chunsheng Liu, Tao Wu, Zhifei Li, and Bin Wang. Individual traffic prediction in cellular networks based on tensor completion. 34(16):4952.
- [19] Hui Hu and Wenquan Xu. Time series forecasting based on empirical mode decomposition and the varying-coefficient DBN-AR model. 10:105169–105181. Publisher: IEEE.
- [20] R. Guo, Y. Wang, H. Zhang, and G. Zhang. Remaining useful life prediction for rolling bearings using EMD-RISI-LSTM. PP(99):1–1. Type: Journal Article.
- [21] Qiu Sun and Huafeng Cai. Short-term power load prediction based on VMD-SG-LSTM. 10:102396–102405. Type: Journal Article.
- [22] Jawaria Nasir, Muhammad Aamir, Zahoor Ul Haq, Shehzad Khan, Muhammad Yusuf Amin, and Muhammad Naeem. A new approach for forecasting crude oil prices based on stochastic and deterministic influences of LMD using ARIMA and LSTM models. 11:14322–14339. Type: Journal Article.
- [23] Xiwen Qin, Dingxin Xu, Xiaogang Dong, Xueteng Cui, and Sisi Zhang. EEG signal classification based on improved variational mode decomposition and deep forest. 83:104644. Publisher: Elsevier.
- [24] Hanyu Yang, Xutao Li, Wenhao Qiang, Yuhuan Zhao, Wei Zhang, and Chang Tang. A network traffic forecasting method based on SA optimized ARIMA-BP neural network. 193(3):108102. Type: Journal Article.
- [25] S. Gowrishankar and P. S. Satyanarayana. A time series modeling and prediction of wireless network traffic. 3(1).
- [26] Johann Leithon, Teng Joon Lim, and Sumei Sun. Renewable energy management in cellular networks: An online strategy based on ARIMA forecasting and a markov chain model. In *2016 IEEE wireless communications and networking conference*, pages 1–6.
- [27] Shalini Shanmugam and Selvathi Dharmar. Implementation of a non-linear SVM classification for seizure EEG signal analysis on FPGA. 131:107826. Publisher: Elsevier.
- [28] A. Y. Nikravesh, S. A. Ajila, C. H. Lung, and W. Ding. Mobile network traffic prediction using MLP, MLPWD, and SVM. In *IEEE international congress on big data*, pages 402–409.
- [29] Huiwei Xia, Xin Wei, Yun Gao, and Haibing Lv. Traffic prediction based on ensemble machine learning strategies with bagging and LightGBM. In *2019 IEEE international conference on communications workshops (ICC workshops)*, pages 1–6.

- [30] Qianqian Zhang, Mohammad Mozaffari, Walid Saad, Mehdi Bennis, and Merouane Debbah. Machine learning for predictive on-demand deployment of uavs for wireless communications. In *2018 IEEE global communications conference (GLOBECOM)*, pages 1–6.
- [31] Schyler C. Sun and Weisi Guo. Forecasting wireless demand with extreme values using feature embedding in gaussian processes. In *2021 IEEE 93rd vehicular technology conference (VTC2021-spring)*, pages 1–6.
- [32] A novel framework for ultra-short-term interval wind power prediction based on RF-WOA-VMD and BiGRU optimized by the attention mechanism. 269:126738.
- [33] Luoyang Fang, Xiang Cheng, Haonan Wang, and Liuqing Yang. Mobile demand forecasting via deep graph-sequence spatiotemporal modeling in cellular networks. 5(4):3091–3101.
- [34] A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with bayesian optimization algorithm. 237:103–116. Publisher: Elsevier.
- [35] Xiaolei Ma, Houyue Zhong, Yi Li, Junyan Ma, and Yinhai Wang. Forecasting transportation network speed using deep capsule networks with nested LSTM models. PP(99):1–12.
- [36] V. E. Mitrokhin and I. N. Bashkov. Mobile traffic prediction from raw data using lstm networks. In *V international scientific and technical conference "radio engineering, electronics and communication"*, pages 117–123.
- [37] Zheheng Rao, Yanyan Xu, Shaoming Pan, Jiabao Guo, Yuejing Yan, and Zhiheng Wang. Cellular traffic prediction: A deep learning method considering dynamic nonlocal spatial correlation, self-attention, and correlation of spatiotemporal feature fusion. 20(1):426–440.
- [38] Haihan Nan, Xiaoyan Zhu, and Jianfeng Ma. MSTL-GLTP: A global-local decomposition and prediction framework for wireless traffic. 10(6):5024–5034.
- [39] Wen-chuan Wang, Wei-can Tian, Dong-mei Xu, and Hong-fei Zang. Arctic puffin optimization: A bio-inspired metaheuristic algorithm for solving engineering design optimization. 195:103694.
- [40] Shengdong Du, Tianrui Li, Yan Yang, and Shi-Jinn Horng. Deep air quality forecasting using hybrid deep learning framework. 33(6):2412–2424.
- [41] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. 62(3):531–544.
- [42] Sinvaldo Rodrigues Moreno, Laila Oriel Seman, Stefano Frizzo Stefenon, Leandro dos Santos Coelho, and Viviana Cocco Mariani. Enhancing wind speed forecasting through synergy of machine learning, singular spectral analysis, and variational mode decomposition. 292:130493.
- [43] H. Yi and Khn Bui. An automated hyperparameter search-based deep learning model for highway traffic prediction. PP(99):1–10.
- [44] 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. 454(1971):903–995.
- [45] C. Li, Z. Xiao, X. Xia, W. Zou, and C. Zhang. A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting. 215:131–144.
- [46] Weimin Yue, Qingrong Liu, Yingjun Ruan, Fanyue Qian, and Hua Meng. A prediction approach with mode decomposition-recombination technique for short-term load forecasting. 85:104034.
- [47] Gianni Barlacchi, Marco De Nadai, Roberto Larcher, Antonio Casella, Cristiana Chitic, Giovanni Torrisi, Fabrizio Antonelli, Alessandro Vespiagnani, Alex Pentland, and Bruno Lepri. A multi-source dataset of urban life in the city of milan and the province of trentino. 2(1):150055. Type: Journal Article.
- [48] Xiaming Chen, Yaohui Jin, Siwei Qiang, Weisheng Hu, and Kaida Jiang. Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale. In *2015 IEEE international conference on communications (ICC)*, pages 3585–3591. IEEE.
- [49] Chung-Yi Lin, Hung-Ting Su, Shen-Lung Tung, and Winston H Hsu. Multivariate and propagation graph attention network for spatial-temporal prediction with outdoor cellular traffic. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 3248–3252.
- [50] Chi-Hua Chen, Fangying Song, Feng-Jang Hwang, and Ling Wu. A probability density function generator based on neural networks. 541:123344.
- [51] Dong-mei Xu, Xiao-xue Hu, Wen-chuan Wang, Kwok-wing Chau, Hong-fei Zang, and Jun Wang. A new hybrid model for monthly runoff prediction using ELMAN neural network based on decomposition-integration structure with local error correction method. 238:121719.
- [52] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for hyper-parameter optimization. In *International conference on neural information processing systems*, pages 2546–2554.



Pengfei Zhang received the B.S. degree in computer science and technology from Yan'an University, Yan'an, China in 2011 and received the M.S. degree in computer technology from the Xi'an University of Technology, Xi'an, China, in 2018, respectively, where he is currently pursuing a Ph.D. degree in computer science and technology. His research interests include intelligent optimization algorithms, machine learning, and 5G cellular networks.



Junhuai Li received the B.S. degree in electrical automation from the Shaanxi Institute of Mechanical Engineering, Xi'an, China, in 1992, the M.S. degree in computer application technology from the Xi'an University of Technology, Xi'an, in 1999, and the Ph.D. degree in computer software and theory from Northwest University, Xi'an, in 2002.

He is currently a Professor with the School of Computer Science and Engineering, Xi'an University of Technology, China. His research interests include the Internet of Things technology and network

computing.



Dong Ding received a B.S. degree in software engineering and an M.S. degree in computer system architecture from Xi'an University of Technology, Xi'an, China, in 2014 and 2018, respectively, where he is currently pursuing a Ph.D. degree in electrical engineering.



Huajun Wang received the B.Sc. and M.Sc. degrees in computer science from the Xi'an University of Technology, in 2005 and 2010, respectively, and the Ph.D. degree from Northwest University, Xi'an, China, in 2014. He is currently a Lecturer with the Xi'an University of Technology.

His research interests include application and security of CPS and modeling of effectiveness evaluation of security.



Kan Wang received the B.S. degree in broadcasting and television engineering from the Zhejiang University of Media and Communications, Hangzhou, China, in 2009, and the Ph.D. degree in military communications from the State Key Laboratory of ISN, Xidian University, Xi'an, China, in 2016. From October 2014 to October 2015, he was with Carleton University, Ottawa, ON, Canada, as a Visiting Scholar funded by the China Scholarship Council (CSC). Since March 2017, he has been with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an.

His current research interests include 5G cellular networks, resource management, and massive IoT.



Xiaofan Wang received the B.S. and M.S. degrees from the Xi'an University of Technology, Xi'an, China, in 1999 and 2003, respectively, and the Ph.D. degree in computer science and engineering from Xidian University, Xi'an, in 2012. He is currently an Professor with the Faculty of Computer Science and Engineering, Xi'an University of Technology.

His current research interests focused in data mining, machine learning, and intelligent information processing.