Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 46-53

DOI: https://doi.org/10.25195/ijci.v50i2.527
Print ISSN: 2313-190X, Online ISSN: 2520-4912

Research Article

Hybrid Machine Learning Approaches for 5G Traffic Prediction

Mohamed Burhan Mohamed
Sunni Endowment Office
mohamedalmajamie@gmail.com

ARTICLEINFO
Article History
Received: 20/08/2024
Accepted: 04/10/2024
Published:12/11/2024

ABSTRACT

Accurate traffic prediction poses great difficulties because of the continuously increasing scale and diversity of 5G network traffic, which is driven by user demands. Moreover, certain characteristics of 5G traffic are constantly changing; thus, simulations using traditional models often lead to incorrect estimations or inefficient utilization of available resources. Consequently, we propose a hybrid machine learning model that integrates support vector machine (SVM) and decision tree algorithms to enhance efficiency of 5G traffic prediction. The structure of the hybrid model dynamically adjusts by adding or removing hidden layers and units within the network to improve prediction performance. The efficacy of the proposed model is evaluated using metrics like mean squared error, mean absolute error, and root mean squared error (RMSE). Findings show that the hybrid model consistently achieves lower error rates than SVM alone. Further performance enhancement of the hybrid model in predicting 5G traffic is also supported by comparisons of R-squared values against signal-to-noise ratios. These outcomes show the potential of the proposed method to improve traffic prediction accuracy in 5G networks, serving as a powerful tool for network control.

***Keywords:* ML, SVM, DT, RMSE, MAE, AI.**

1. INTRODUCTION

The development and application of 5G technology is meeting the growing demand for mobile communication [1]. The economy and society are becoming increasingly digital, networked, and intelligent thanks to the new, rapidly expanding technological revolution [2]. A 5G network offers several advantages, including high speed, extreme reliability, and minimal latency. With global accessibility, 5G technology satisfies the substantial resource demands of extensive terminal networks. However, it also introduces exponential increases in network traffic, heterogeneity, and complexity as shown in Figure (1). To manage the considerable traffic load caused by massive, heterogeneous data streams in traditional cellular networks, 5G operators deploy numerous low-power micro- and pico-base stations around macro-base stations. This configuration serves to offload traffic and maintain load balance across macro-base stations [3, 4]. Accurate traffic prediction is essential for optimizing the deployment and allocation of 5G cellular network resources in large-scale cities and enhance the intelligence and reliability of traffic management systems [5]. Given that 5G network traffic is inherently time-series data, the prediction challenge can be framed as a time-series prediction modeling problem [6]. Past methods mostly used mathematical theories, such as statistics and probability distributions, to model and forecast traffic flow. This kind of approach rely on finite parameters rather than dataset size [7].

Several studies have explored 5G traffic prediction using machine learning and deep learning. For example, developed machine learning models based on lower-layer parameters that characterize the radio environment to predict the available throughput. These models were tested on the LTE network before they were applied to the non-standalone 5G network. also considered extending the proposed models to future standalone 5G networks. For LTE and non-standalone 5G networks, the end-user throughput was modeled with R-squared values of 93% and 84% and mean squared errors of 0.06, 0.47, and 17, respectively [8].

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 46-53
DOI: https://doi.org/10.25195/ijci.v50i2.527
Print ISSN: 2313-190X, Online ISSN: 2520-4912

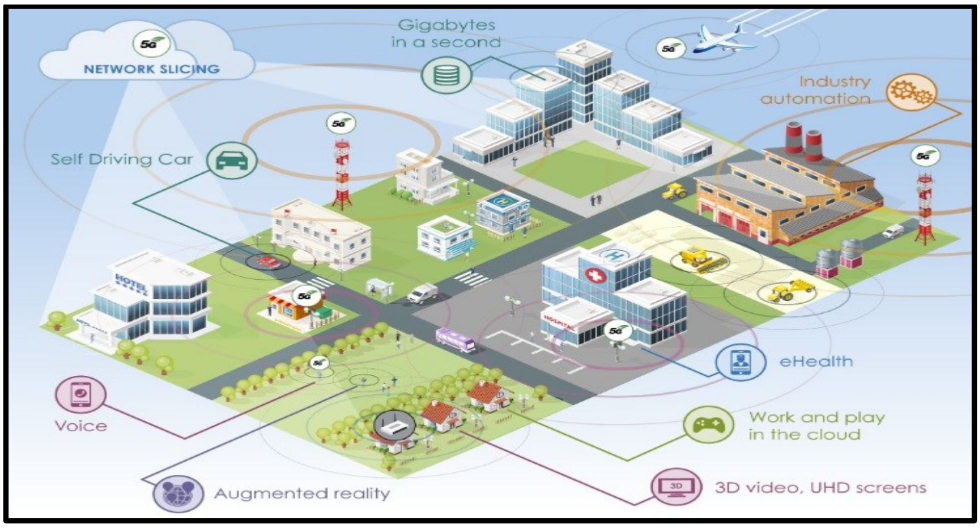

Fig. 1. 5G Network [9]

Weiwei Jiang categorized prediction difficulties into temporal and spatiotemporal after reviewing relevant studies on cellular traffic prediction. In this study, the prediction methods were divided into statistical models, machine learning, and deep learning categorized. They were then with artificial intelligence (AI)-based prediction models. The study also provided an overview of current applications of cellular traffic prediction and suggested potential future research directions [10].

[11] emphasized that resource management in networking has received much attention because these devices are characterized by large data volumes. Achieving great performance with limited resource utilization is difficult. Therefore, traditional analog monitoring techniques are inadequate for handling such data volumes. To this end, implemented deep learning techniques along with network monitoring tools. They focused on the work done on traffic prediction, which is among the essential features of network analysis. They also considered studies employing deep learning for resource management in network slicing through traffic prediction.

[12] comprehensively reviewed AI-based methods for mobile data traffic (DT) prediction. They analyzed the current machine learning models for predicting cellular DT in 5G network. The review also includes a detailed account of mobile network evolution. R. Raj Mohan et al. examined several aspects of current methods, including methodologies, advantages, primary objectives, performance metrics, and conclusions. In addition, extensive experimental evaluations identify the unique properties of each method. The study concludes with a summary of potential future directions and challenging problems.

The current study develops a hybrid system based on machine learning to achieve effective and accurate prediction of traffic in 5G networks. A reliable model for traffic forecasting can increase overall network efficiency and help operators better organize physical infrastructure planning.

The paper is structured as follows: The problem is introduced in the first section. The second section discusses the theoretical background on decision tree method (DTM) and support vector machine (SVM). The third section outlines the proposed system within the Tool Specification. The fourth section presents the findings, dissection of prior research, and evaluation indicators. The final section draws conclusions.

2. BASIC CONCEPTS

2.1 SVM

In 1995, Vapnik introduced a nonlinear SVM with soft margins that solved the handwritten character recognition problem satisfactorily, as shown in Figure (2) [13]. Text classification and human face picture recognition are two areas of pattern recognition where SVM has already been ingrained.

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 46-53
DOI: https://doi.org/10.25195/ijci.v50i2.527
Print ISSN: 2313-190X, Online ISSN: 2520-4912

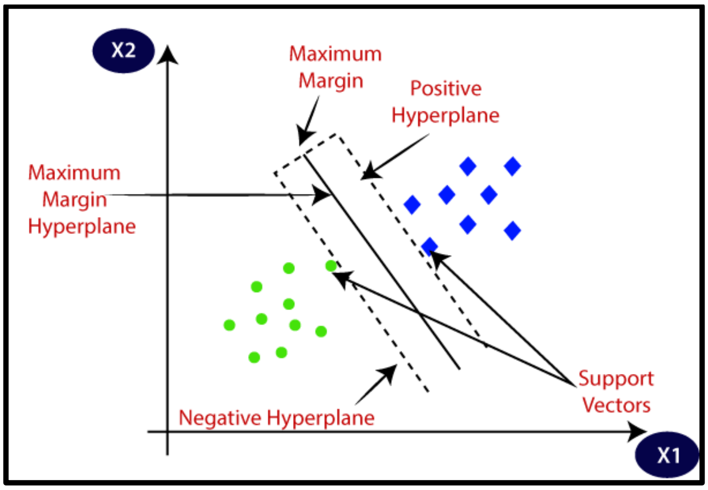

Fig. 2. . SVM

One-against-all SVM creates t binary SVM algorithms for a t class classification problem such as modulation coding system prediction. Consequently, using a training data set:

$$D = \{(X1, Y1), (X2, Y2), \ldots \ldots, (Xm, Ym)\} -----(1)$$

Where X_i is the instance vector, with $i = 1 \ldots., D$ and Y_j is the class of X_i. All of the cases in the jth class with positive markers and all others with negative markers are used to train each jth binary SVM algorithm. After solving the optimization problem, the following linear expression, provides the exact hyper plane that each jth SVM creates:

$$F_j(x) = W^T \emptyset(x) + b_j -----(2)$$

Where W is the natural vector that controls the direction of hyperplane, $\phi(x)$ is the non-linear assignment function, and b stands for bias. Next, the binary classification problem for each jth SVM can be defined as follows by applying the Lagrange multipliers and solving the issue in accordance with [14]:

$$F_j(x) = \sum_{i}^{l} c\, y_i K(x_i, x) - b_j ------(3)$$

Where c is the Lagrange multiplier and K (.,.) is the kernel function that maps data from a low-dimensional to high-dimensional space. The radial basis function is a popular kernel for multiclass classification applications. The related expression can be indicated as

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2) -----(4)$$

Where $\|.\|$ indicates the norm, and γ is the adjustable parameter that is calibrated to fit the data and controls the performance of the kernel. The prediction from the SVM method can be defined as the maximum prediction from each binary SVM classifier. The SVM classifier's decision function is defined as

$$\widehat{MCS} = \arg F_j \; j \in \{1, \ldots\ldots, t\} -----(5)$$

2.2 DTM

As indicated in Figure (3), the supervised learning method of decision trees is predominantly used to solve categorization problems. It can also solve some regression issues. However, this classifier is a multiple-level structure of decision rules where the leaves are connected to the results, the branches are connected to the decision rules, and the internal nodes are associated with the properties of given dataset. A decision tree consists of two nodes: the decision node and the leaf node. Decision nodes are used to make any form of decision and has more than one branch, whereas the leaf node signifies the results of decisions and does not have any branch. A test or a decision-making process is made on the basis of the characteristics provided in the dataset. It is a graphical program that shows all possible resolutions or choices based on specified input parameters [15].

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 46-53
DOI: https://doi.org/10.25195/ijci.v50i2.527
Print ISSN: 2313-190X, Online ISSN: 2520-4912

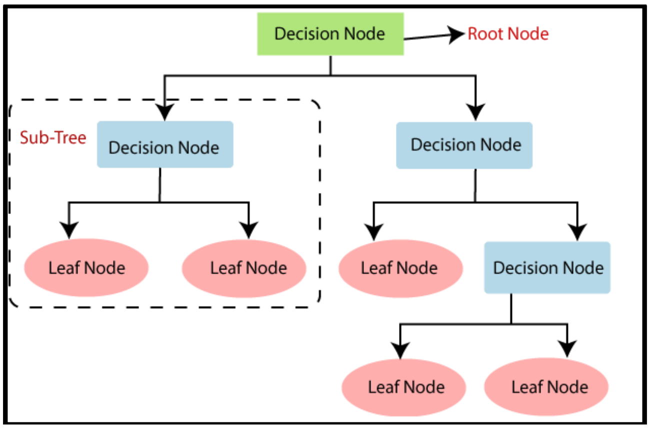

Fig. 3. DTM

The procedure in a decision tree starts at the root node to predict the class of a given dataset. This algorithm follows the branch and advances to the next node by comparing the values of the root attribute with the record (actual dataset) attribute.

The method proceeds to the next node by comparing its attribute value with those of the other sub-nodes once more. This process continues until it reaches a leaf node.

3. METHODOLOGY

3.1 Dataset

In this research, the dataset originates from a foreign network operator and encapsulates a comprehensive view of the 5G network environment. The dataset encompasses two distinctive modes: the static mode and the in-vehicle mode, both capturing diverse aspects of 5G networks. A defining feature of 5G is its substantial bandwidth, which lays the foundation for numerous groundbreaking applications and services. The dataset, stored in a CSV file, includes the following parameters:
Time: Recorded time for each sample.
Normalized Time: A modeling-specific normalized time interval between 0 and 1.
Download Bandwidth: Measured in megabits per second (Mbps).
Normalized Bandwidth: Normalized bandwidth value ranging from 0 to 1.
Data were gathered over 30 consecutive days at one-hour intervals. A normal distribution with a mean of 100 Mbps and a standard deviation of 20 Mbps is used to mimic bandwidth.

3.2 Proposed Method

Automated traffic prediction is a critical feature of 5G networks to prevent congestion, assign resources efficiently, and ensure satisfactory service delivery. In this study, a stable predictive model is developed using machine learning techniques, specifically a hybrid approach combining DTM and SVM. This hybrid strategy utilizes a technique called stacking where the strengths of both models are leveraged. The goal is to establish a decision model that acquires a high level of accuracy to predict 5G network traffic. Using this strategy in developing models can enhance the issues relating to robustness and predictability. This method is implemented using the following steps, as shown in Figure (4):

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 46-53
DOI: https://doi.org/10.25195/ijci.v50i2.527
Print ISSN: 2313-190X, Online ISSN: 2520-4912

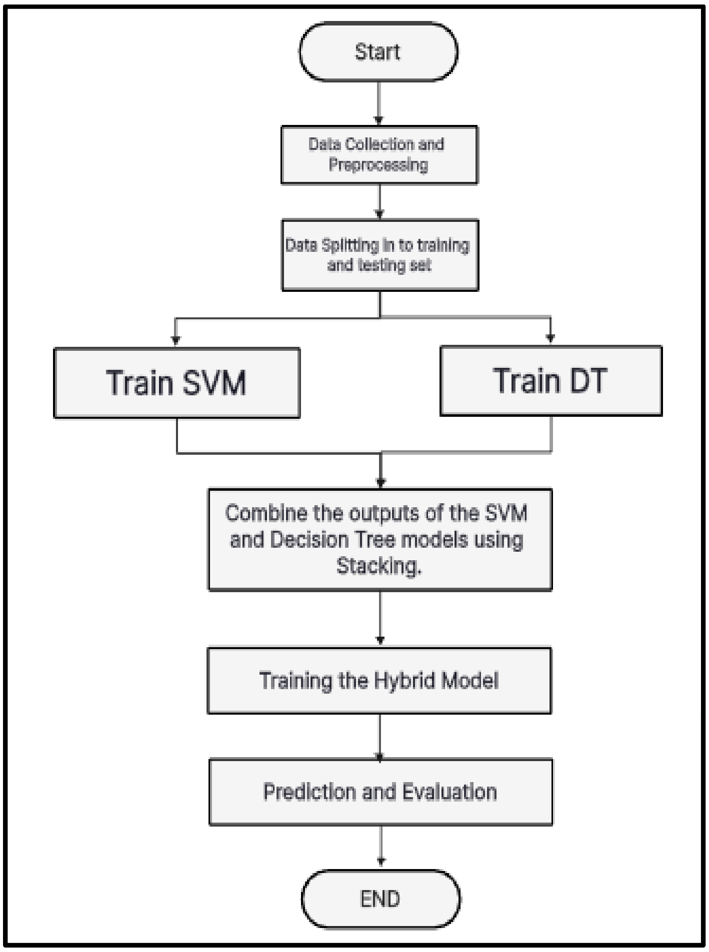

Fig. 4. Proposed Method

3.3 Evaluation

To thoroughly assess the network model's prediction performance, the root mean squared error (RMSE), mean absolute error (MAE), and accuracy are used.

The RMSE is a measure of the training model's degree of dispersion from actual values. A lower RMSE number signifies a higher degree of model agglutination, indicating that the model is more precise. The RMSE equation is

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \text{ -----------(6)}$$

The MAE is calculated by subtracting the experiment's actual outcome from its projected result. Additionally, the mean value is computed after the absolute value has been used. A lower MAE value corresponds to a smaller model error.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| ------(7)$$

R-squared is commonly used to evaluate the goodness of fit for linear regression models.

R-squared converts the prediction results into an accuracy value within the 0–1 range, which intuitively shows the model's accuracy. Its value is less than 1 if the model fits badly, and its value is relatively close to 1 if the model fits very well. The equation of R-squared is

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^3}{\sum_{i=1}^{n}(y_i - \overline{y})} ------(8)$$

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 46-53
DOI: https://doi.org/10.25195/ijci.v50i2.527
Print ISSN: 2313-190X, Online ISSN: 2520-4912

4. RESULTS AND DISCUSSION

The results of the proposed method are shown in Figure (5), showing the RMSE, MSE and MAE values for the SVM and hybrid model. The results show that the hybrid model achieves lower MSE, RMSE, and MAE values. This outcome means this model is more precise than the SVM alone.

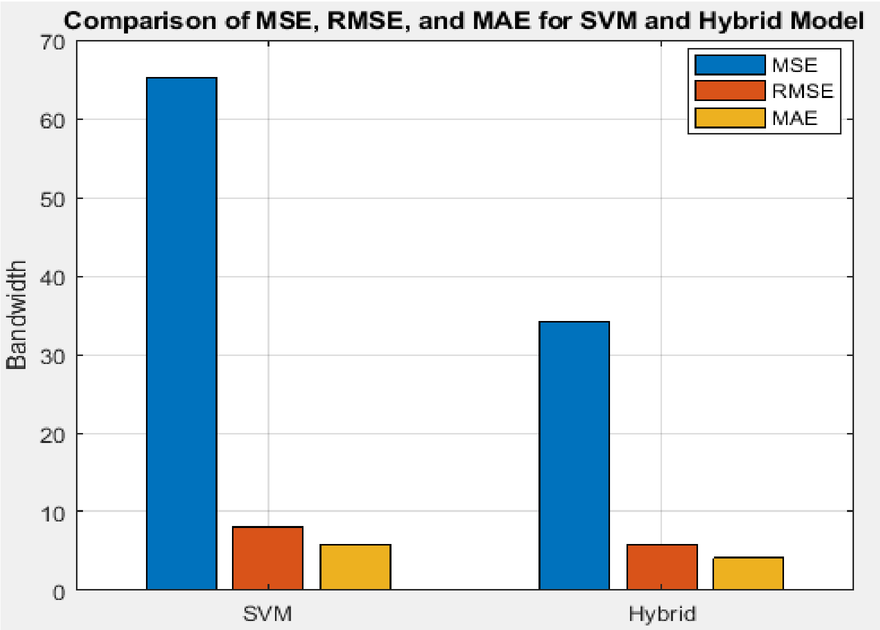

Fig. 5. Comparison of MSE, MAE, and RMSE Values for SVM and Hybrid Model

Table (1) shows the result of SVM and Table (2) illustrated the results of hybrid model.

TABLE I. RESULTS OF SVM

	SVM algorithm		
	RMSE	*MSE*	*MAE*
1	8.0789	65.2683	5.9097

TABLE II. RESULTS OF HYBRID MODEL

	Hybrid Model		
	RMSE	MSE	MAE
1	5.8510	34.2339	4.1685

The results indicate that the proposed hybrid model is better than SVM. The values of the RMSE, MSE, and MAE for the hybrid model were lower than those of the SVM method, so they were utilized to assess the performance of the proposed hybrid system. For instance, the hybrid model's RMSE value (5.8510) is less than that of the SVM model (8.0789), indicating that the hybrid model is more accurate because it has fewer mistakes. The MSE scores of the hybrid model and SVM model are 34.239 and 65.2683, respectively. This outcome indicates that the hybrid model provides more accurate predictions. Additionally, comparing the MAE of the proposed hybrid model (4.1685) to that of the SVM model (5.9097) shows that the degree of accuracy of the proposed hybrid model is higher than that of the SVM model. For instance, the theory held in this paper is that a given model is conceptually better if it yields lower values of error metrics because the values suggest that the model is likely to produce appropriate predictions that reflect

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 46-53

DOI: https://doi.org/10.25195/ijci.v50i2.527
Print ISSN: 2313-190X, Online ISSN: 2520-4912

the actual population data. The hybrid model makes fewer mistakes than SVM, so the former is more effective and efficient in predicting traffic in 5G networks.

Typically, to test the prediction performance of each model, the R-squared value is used as the main evaluation metric. This metric is important because it gives a good indication of the model's ability to explain the variance of the data.

Figure (6) shows the comparison of R-squared values between the SVM model and the hybrid model. The closer to 1 the R-squared values, the better the model fits the actual data. From Figure (6), the hybrid model achieves an R-squared value approximately 0.1 higher than the that of the SVM model. This outcome indicates that the hybrid model performs well because it predicts network traffic data accurately.

Figure (6) shows how noise (data interference or randomness) affects the network traffic forecasting process. The figure shows that the hybrid model is more robust when exposed to higher levels of noise, which means that even if noise is present, the proposed model still performs better than the SVM model.

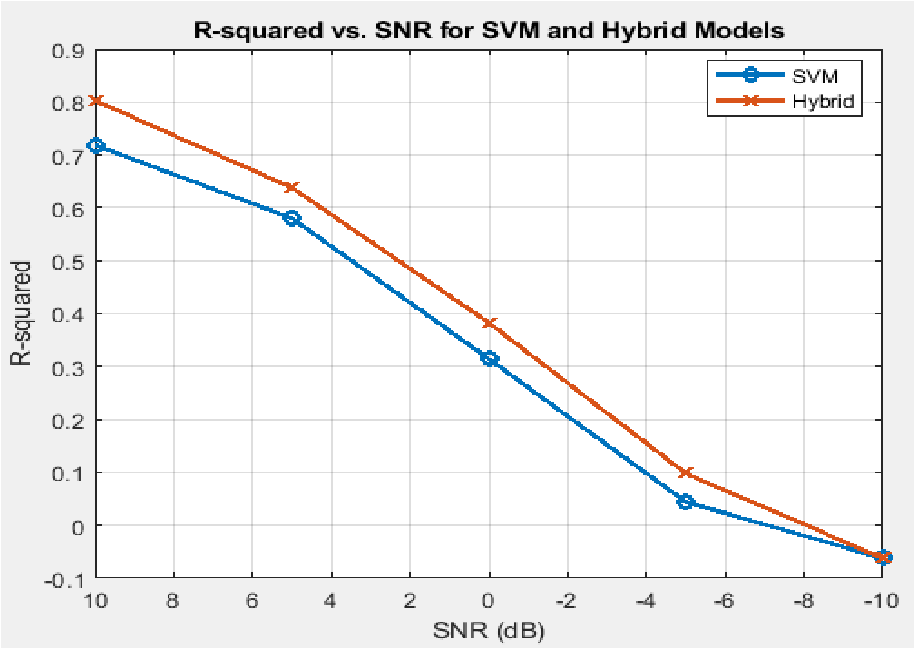

Fig. 6. R-Squared Values of Hybrid Model and SVM

5. CONCLUSION

Precise traffic forecasting remains vital in the management and usage of the network resources in 5G systems. This study proposes a novel traffic prediction model which integrates SVM and DTM. This integrated model shows improved performance than simple SVM model. The proposed hybrid model has lesser RMSE, MSE, and MAE values than the SVM, which suggest improved precision. Moreover, the performance of the hybrid model is stable in predicting 5G traffic even with added noise.

In the future work. Exploring other machine learning approaches may be beneficial in improving the ability of predictive models. These enhancements aim to develop a more adaptive and resilient model that meets the design requirements of 5G networks.

References

[1] N.-N. Dao *et al.*, "A review on new technologies in 3GPP standards for 5G access and beyond," *Computer Networks,* p. 110370, 2024.

[2] A. M. Ahmed, S. A. Hasan, and S. A. Majeed, "5G mobile systems, challenges and technologies: A survey," *J. Theor. Appl. Inf. Technol,* vol. 97, no. 11, pp. 3214-3226, 2019.

Iraqi Journal for Computers and Informatics
Information Technology and Communications University
Vol. 50, No. 2, 2024, pp. 46-53
DOI: https://doi.org/10.25195/ijci.v50i2.527
Print ISSN: 2313-190X, Online ISSN: 2520-4912

[3] D. Raca, D. Leahy, C. J. Sreenan, and J. J. Quinlan, "Beyond throughput, the next generation: A 5G dataset with channel and context metrics," in *Proceedings of the 11th ACM multimedia systems conference*, 2020, pp. 303-308.

[4] A. Azab, M. Khasawneh, S. Alrabaee, K.-K. R. Choo, and M. Sarsour, "Network traffic classification: Techniques, datasets, and challenges," *Digital Communications and Networks,* vol. 10, no. 3, pp. 676-692, 2024.

[5] P. Chakraborty, M. Corici, and T. Magedanz, "A comparative study for Time Series Forecasting within software 5G networks," in *2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2020: IEEE, pp. 1-7.

[6] D. Kutuzov, A. Osovsky, and D. Starov, "Properties of 5G Video Conference Traffic and Short-Term Forecasting of Input Data Rate," in *2024 International Russian Smart Industry Conference (SmartIndustryCon)*, 2024: IEEE, pp. 286-293.

[7] L. Tang, J. Li, H. Du, L. Li, J. Wu, and S. Wang, "Big data in forecasting research: a literature review," *Big Data Research,* vol. 27, p. 100289, 2022.

[8] D. Minovski, N. Ögren, K. Mitra, and C. Åhlund, "Throughput prediction using machine learning in LTE and 5G networks," *IEEE Transactions on Mobile Computing,* vol. 22, no. 3, pp. 1825-1840, 2021.

[9] "5G.", 2024, ETS https://www.etsi.org/technologies/mobile/5g?jjj=1723889716483.

[10] W. Jiang, "Cellular traffic prediction with machine learning: A survey," *Expert Systems with Applications,* vol. 201, p. 117163, 2022.

[11] D. Kulkarni, M. Venkatesan, and A. V. Kulkarni, "Traffic Prediction with Network Slicing in 5G: A Survey," in *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, 2023: IEEE, pp. 1360-1365.

[12] R. R. Mohan, K. Vijayalakshmi, P. J. Augustine, R. Venkatesh, M. G. Nayagam, and B. Jegajohi, "A comprehensive survey of machine learning based mobile data traffic prediction models for 5G cellular networks," in *AIP Conference Proceedings*, 2024, vol. 2816, no. 1: AIP Publishing.

[13] C. Campbell and Y. Ying, *Learning with support vector machines*. Springer Nature, 2022.

[14] Y. Duan, B. Zou, J. Xu, F. Chen, J. Wei, and Y. Y. Tang, "OAA-SVM-MS: A fast and efficient multi-class classification algorithm," *Neurocomputing,* vol. 454, pp. 448-460, 2021.

[15] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Information Technology and Engineering Journal,* vol. 10, no. 07, pp. 3897-3904, 2023.