Performance Comparison of ML-based Throughput Prediction Mechanisms in 5G Networks

Md. Monirul Islam, Kamrul Hasan and Seong Ho Jeong*
Department of Information and Communications Engineering
Hankuk University of Foreign Studies
Seoul, South Korea
monirul@hufs.ac.kr, kamrul@hufs.ac.kr, and shjeong@hufs.ac.kr*

***Abstract*—Accurate throughput prediction in 5G networks is essential for optimizing quality of service (QoS), resource allocation, and network planning. This paper presents a performance comparison of machine learning-based throughput prediction mechanisms using a variety of features that reflect user mobility, physical link conditions, and network performance. This paper evaluates four models: Extreme Gradient Boosting Regressor (XGBR), Random Forest Regressor (RFR), Support Vector Regressor (SVR), and Multi-Layer Perceptron Regressor (MLPR). Among them, the RFR achieves the best performance, with a coefficient of determination (R^2) of 1.0000, mean absolute error (MAE) of 0.0056, mean squared error (MSE) of 0.0001, and root mean square error (RMSE) of 0.0074, followed closely by XGBoost. The SVR and MLP models also perform well, though with comparatively lower accuracy. The results demonstrate that ensemble-based models are highly effective for accurate throughput prediction in 5G environments, providing a promising approach for real-time network management and intelligent decision-making.**

Keywords*—*Throughput prediction architecture, 5G cellular network, Machine learning algorithm for regression.

I. Introduction

The advent of fifth-generation (5G) mobile networks has delivered substantial advancements in data rates, latency, reliability, and massive connectivity, enabling applications such as immersive multimedia, autonomous vehicles, and industrial automation. In dynamic and heterogeneous environments, accurate throughput prediction is essential for maintaining Quality of Service (QoS) and Reliability of Service (RoS), optimizing resource allocation, and enabling proactive network planning. Throughput in 5G networks depends on factors such as user mobility, wireless channel conditions, and network load, with parameters like user speed, movement direction, distance from the base station, bit error rate (BER), signal-to-noise ratio (SNR), received signal strength (RSS), congestion, latency, and modulation scheme playing critical roles [1]. The complex, nonlinear relationships among these parameters limit the effectiveness of traditional analytical and rule-based approaches, particularly under rapidly varying wireless conditions.

Recent advancements in machine learning (ML) have shown remarkable potential in modeling complex, nonlinear relationships in large-scale network datasets. ML-based regression models can effectively capture hidden patterns in multi-dimensional data, allowing for accurate and adaptive throughput prediction under diverse network conditions [2]. However, selecting an appropriate ML model and relevant features is critical to achieving optimal performance, as different algorithms vary in their ability to handle nonlinearities, feature interactions, and noise in real-world data.

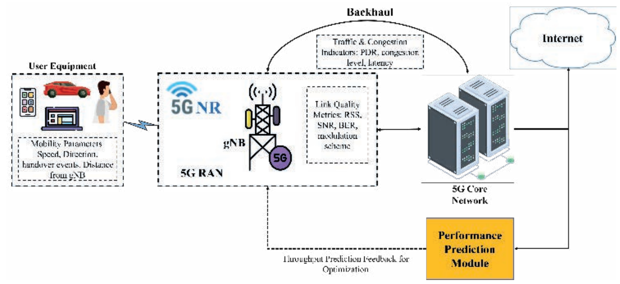


Fig. 1. ML-based throughput prediction architecture in 5G networks

In this paper, we propose a comprehensive ML-based throughput prediction framework for 5G networks, leveraging features that capture both physical layer parameters and network-level performance indicators. Fig. 1 shows the proposed architecture of ML-based throughput prediction in a 5G network. Mobility parameters, such as user speed, direction, handover events, and distance from the gNB, are collected from the user equipment (UE) layer. The 5G Radio Access Network (RAN) provides link quality metrics, including RSS, SNR, BER, and modulation scheme. Additionally, traffic and congestion indicators—such as packet delivery ratio (PDR), congestion level, and latency—are obtained from the core network. All collected features are processed by a performance prediction module, implemented at the network core or edge, to estimate the expected throughput. The predicted throughput is then fed back to the RAN for resource optimization, enabling proactive network management and improved QoS. We evaluate and compare four widely used regression algorithms—extreme gradient boosting regressor (XGBR), Random Forest Regressor (RFR), Support Vector Regressor (SVR), and Multi-Layer Perceptron Regressor (MLPR) with four performance metrics— coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), and root mean square error (RMSE) to predict the throughput.

The rest of the paper is organized as follows. Section II deals with the related works followed by proposed methodology in Section III. Result analysis is presented in the Section IV and finally Section V concludes the paper.

II. Related Work

Recent research has increasingly focused on leveraging machine learning techniques to predict network performance metrics in 5G environments, aiming to enhance resource allocation, QoS, and overall network efficiency. In [3], the authors proposed a deep learning-based approach for throughput prediction in 5G cellular networks. They evaluated four models—RFC, ANN, LSTM, and BiLSTM—using two performance metrics, R^2 and MSE. The highest reported performance was an MSE of 0.00009 and an R^2 of 0.36. In [4], the authors proposed an algorithm for throughput prediction in 5G cellular networks. They evaluated their algorithm on 7 datasets using two performance metrics, R^2 and MAE. The algorithm achieved the highest performance was an MAE of 0.05 and an R^2 of 0.89. In [5], the authors proposed machine learning based throughput prediction in LTE and 5G networks. They applied four models including MLPR, XGBR, RFR, and SVR with two performance metrics. They achieved the highest results of MSE of 0.06 and an R^2 of 0.93 in LTE and 0.17 and 0.96 in 5G networks. In [6], the authors proposed an ML-based framework for predicting downlink throughput in 4G and 5G cellular networks. They selected key parameters, including RSRP, SINR, RSSI, and RSRQ, based on correlation analysis to train the models. The proposed approach achieved R^2 ranging from 89% to 96% across three different network operators. In [7], authors proposed an ML-based approach for predicting downlink throughput in LTE networks. They employed SVR, LR, KNN, and DTR models using parameters with high correlation to throughput, including GPS coordinates, RSRP, RSRQ, SINR, and RSSI. Statistical analysis showed that KNN and DTR achieved the highest prediction accuracy, with R^2 values of 99%, 93%, and 98% for operator 1, operator 2, and operator 3, respectively.

III. Proposed Methodology

The workflow of ML-based throughput prediction in 5G networks is shown in Fig. 2. Mobility parameters, link quality metrics, energy, and traffic/congestion indicators are collected as input features. After that, the data undergo preprocessing, including cleaning, encoding, scaling, and splitting, before being fed into various machine learning models such as XGBR, RFR, SVR, and MLPR. The predicted throughput is then evaluated using performance metrics including R^2, MAE, MSE, and RMSE.

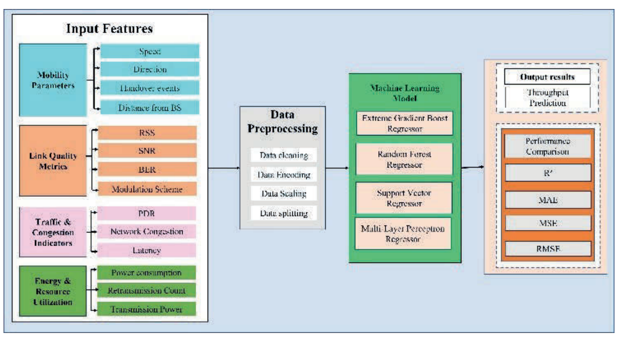


Fig. 2. Proposed methodology of throuput prediction using machine learning method

A. Dataset Description

The dataset used in this study is collected Kaggle platform [8]. It comprises 5000 samples of 5G network measurements, including mobility parameters, link quality metrics, and traffic/congestion indicators. These features, collected under varying mobility and load conditions, are used to predict throughput (Mbps) for model training and performance evaluation.

B. Data Preprocessing

In data preprocessing step, we follows four key steps: data cleaning to remove inconsistencies and missing values, data encoding to convert categorical features into numerical form, data scaling to normalize feature ranges for better model convergence, and test-train splitting by 20:80 to divide the dataset into testing and training sets for performance evaluation.

C. Machine Learning Model

The throughput prediction framework employs four supervised regression models named XGBR, RFR, SVR, and MLPR. XGBR leverages boosting-based decision trees to capture complex feature interactions. RFR uses an ensemble of decision trees with bagging to improve prediction stability and reduce overfitting. SVR applies kernel-based learning to map input features into higher-dimensional spaces for accurate nonlinear regression. MLP, a feedforward artificial neural network, captures nonlinear dependencies through multiple hidden layers. These models are trained on the preprocessed dataset, and their performance is evaluated using performance metrics.

D. Throughput Prediction

The throughput prediction is obtained by the ML models based on the test dataset and model performance is assessed using the statistical metrics of R^2, MAE, MSE, and RMSE.

IV. Performance Analysis

To comprehensively assess the predictive accuracy of the regression models, four widely adopted statistical performance metrics are employed: R^2, MAE, MSE, and RMSE. The R^2 metric evaluates the proportion of variance in the dependent variable that is explained by the model. The MAE measures the average magnitude of prediction errors without considering their direction. The MSE evaluates the average squared difference between actual and predicted values. The RMSE is the square root of the MSE, provides an interpretable error measure in the same unit as the target variable. They are shown in mathematically in (1) to (4) where y_i denotes the actual value, $\hat{y}_i$represents the predicted value, and $\bar{y}$ is the mean of the actual values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \mid y_i - \hat{y}_i \mid \quad (2)$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

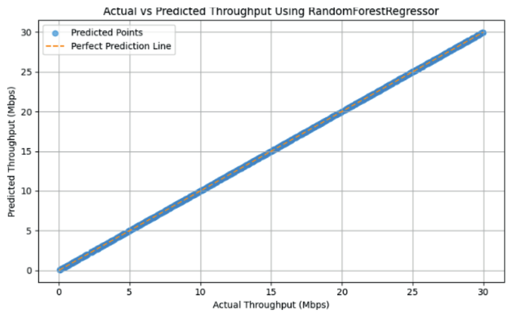
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

TABLE I. Performance Comparison of Regression Models for Throughput Prediction in 5G Networks

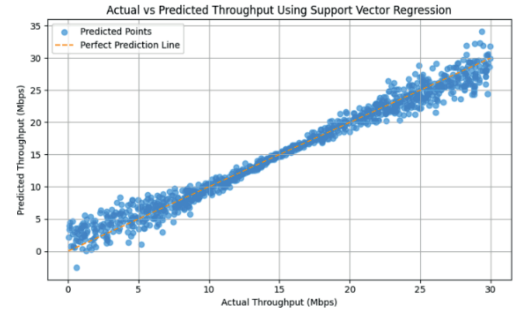
Model Name	Performance Metrics			
	R²	***MAE***	***MSE***	***RMSE***
XGBR	1.0000	0.0407	0.0026	0.0506
RFR	1.0000	0.0056	0.0001	0.0074
SVR	0.9743	0.9922	1.8903	1.3749
MLPR	0.9998	0.0953	0.0163	0.1276

Table I presents the comparative performance of ML models for throughput prediction in 5G networks. Among the models, XGBR and RFR achieved perfect R^2 scores of 1.0000, indicating an excellent fit, with RFR demonstrating the lowest error values (MAE = 0.0056, MSE = 0.0001, RMSE = 0.0074). The MLPR also exhibited high accuracy (R^2 = 0.9998) but with slightly higher error metrics compared to

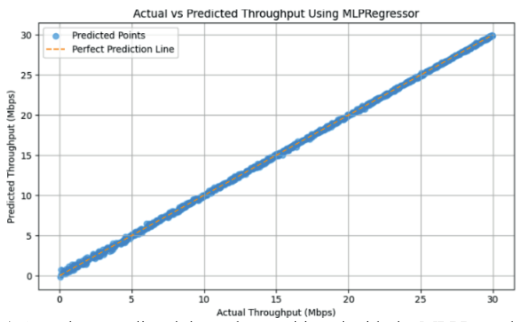
RFR and XGBR. In contrast, SVR showed a relatively lower performance ($R^2 = 0.9743$) and higher error values compared to the other models.



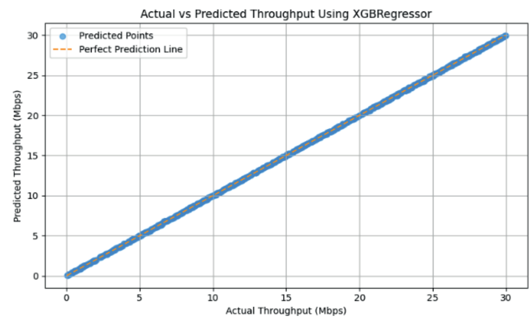
(a) Actual vs. predicted throughput achieved with the RFR model



(b) Actual vs. predicted throughput achieved with the SVR model



(c) Actual vs. predicted throughput achieved with the MLPR model



(d) Actual vs. predicted throughput achieved with the XGBR model

Fig. 3. Comparative results of actual vs. predicted throughput for RFR, SVR, MLPR, and XGBR models, respectively

Fig. 3 shows the ccomparative results of actual vs. predicted throughput for RFR, SVR, MLPR, and XGBR models, respectively. The predicted values closely follow the diagonal line, indicating a strong correlation and high prediction accuracy. This demonstrates that the model effectively captures the impact of physical, mobility, and network features on throughput in 5G networks.

V. Conclusion

In this paper, we proposed a comprehensive machine learning-based framework for throughput prediction in 5G networks, incorporating features from user mobility, physical-layer conditions, and network-level performance indicators. Our evaluation of four widely used regression models—XGBR, RFR, SVR, and MLPR—demonstrated the effectiveness of ensemble-based approaches, with RFR achieving the highest accuracy across all metrics. These results highlighted that machine learning models, particularly ensemble methods, can provide highly accurate throughput predictions, facilitating resource allocation, QoS, RoS, and efficient network management. This study underscored the potential of leveraging ML-driven analytics in 5G networks and laid the groundwork for future research in adaptive, real-time network optimization using predictive intelligence.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (RS-2024-00436887) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

References

[1] J. Hu, L. Wang, J. Wu, Q. Pei, F. Liu, and B. Li, “A comparative measurement study of cross-layer 5G performance under different mobility scenarios,” *Comput. Netw.*, vol. 257, p. 110952, Feb. 2025, doi: 10.1016/j.comnet.2024.110952.

[2] H. Yuliana, Iskandar, and Hendrawan, “Comparative Analysis of Machine Learning Algorithms for 5G Coverage Prediction: Identification of Dominant Feature Parameters and Prediction Accuracy,” *IEEE Access*, vol. 12, pp. 18939–18956, 2024, doi: 10.1109/ACCESS.2024.3361403.

[3] I. Batool, M. M. Fouda, and Z. M. Fadlullah, “Deep Learning-Based Throughput Prediction in 5G Cellular Networks,” in *2024 International Conference on Smart Applications, Communications and Networking (SmartNets)*, May 2024, pp. 1–6. doi: 10.1109/SmartNets61466.2024.10577702.

[4] M. Biswas, A. Chakraborty, and B. Palit, “A Kalman Filter Based Low Complexity Throughput Prediction Algorithm for 5G Cellular Networks,” *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 7089–7101, May 2024, doi: 10.1109/TVT.2023.3339649.

[5] D. Minovski, N. Ögren, K. Mitra, and C. Åhlund, “Throughput Prediction Using Machine Learning in LTE and 5G Networks,” *IEEE Trans. Mob. Comput.*, vol. 22, no. 3, pp. 1825–1840, Mar. 2023, doi: 10.1109/TMC.2021.3099397.

[6] A. Al-Thaedan *et al.*, “A machine learning framework for predicting downlink throughput in 4G-LTE/5G cellular networks,” *Int. J. Inf. Technol.*, vol. 16, no. 2, pp. 651–657, Feb. 2024, doi: 10.1007/s41870-023-01678-w.

[7] A. Al-Thaedan *et al.*, “Downlink throughput prediction using machine learning models on 4G-LTE networks,” *Int. J. Inf. Technol.*, vol. 15, no. 6, pp. 2987–2993, Aug. 2023, doi: 10.1007/s41870-023-01358-9.

[8] “AI-Driven Cross-Layer Wireless Network Dataset.” Accessed: Aug. 13, 2025. [Online]. Available: https://www.kaggle.com/datasets/ziya07/ai-driven-cross-layer-wireless-network-dataset