**Performance of machine learning-based network slicing methods in 5G and beyond communication**

***Furkan BALCI1,[[1]](#footnote-1)\****

*1 Faculty of Technology, Department of Electrical-Electronics Engineering, Gazi University, Ankara, Türkiye*

|  |
| --- |
| **Abstract** In recent years, advancements in communication technologies have given rise to needs such as high transmission speed, reliability, and low latency. Improvements in these aspects are crucial in fourth-generation (4G) communication technologies. Following 4G, the Network Slicing method introduced with 5G allows the network infrastructure to be divided to meet different service requirements, enabling flexible and efficient utilization of network resources. The performance of machine learning-based 5G network slicing methods was tested by simulating 3rd Generation Partnership Project (3GPP) compliant error-prone users and base stations. Five different machine learning methods, along with their parameter spaces, were used in tests for network slicing, employing four methods (eMBB, M1oT, V2X, and URLLC). The performance of these classifier models was analyzed using both error-prone user data and ideal user data. The simulation data were used to conduct a performance analysis of machine learning methods mentioned in the literature, investigating their usability. A 96.56% accuracy rate was achieved using the multi-layer perceptron method with error-prone user data, and a 98.73% accuracy rate was achieved with ideal user data. Additionally, the relationships between the system cycle and user count, as well as the data rate reduction system, were examined in the simulation. |
| Keywords: 5G and Beyond Communication, Machine Learning, Network Slicing |

1. **Introduction**

Since ancient times, communication technologies used by humans have continually evolved as a result of increasing user numbers and emerging security needs. Various generational advancements (1G, 2G, 3G, 2.5G, 2.75G, etc.) have been attempted to meet these needs. The fifth-generation (5G) communication currently in use brings with it high expectations and significant opportunities due to the growing diversity of mobile devices, usage variations, and application diversity. The rapid diversification of different application possibilities, the increase in user demands, and the need for various technical requirements have marked a prominent era for 5G technology. Users increasingly demand various service types that align with their growing data needs. With the introduction of 5G, these dynamic demands have necessitated the evolution of existing infrastructure to enable the flexible provision of different services within the same platform and the creation of higher-capacity systems [1]. The high-performance standards set by 5G allow this evolution to manifest its impact across various sectors. For example, advancements such as making autonomous systems smarter and more reliable, enhancing the impact and accessibility of virtual reality experiences, and enabling more efficient operations of smart factories are planned based on the infrastructure and speed advantages provided by 5G. In this context, the opportunities offered by 5G technology not only meet the increasing data demands but also pave the way for significant advancements in various sectors.

In recent years, the increases in communication demands have led communication service providers to develop programmable system solutions in response to the rising technical requirements. The methods of Software-Defined Networking (SDN) and Network Functions Virtualization (NFV) used in fourth-generation (4G) communication systems, along with the solutions developed in this context, have allowed for greater control over the system [1]. The new technologies introduced with 4G have enabled higher performance compared to previous generations [1]. In the transition to 5G, many of these technologies have been retained, and necessary enhancements have been made. The most notable among these enhancements is network slicing technology. The technologies used from 1G to 5G are summarized in Table 1. Network slicing allows the creation of multiple virtual networks within the same physical infrastructure, enabling these networks to provide services to different types of users. This approach aims to utilize systems more effectively and to implement new services in a more controllable and rapid manner. The controllability targeted by this approach plays a crucial role in the marketability of new services [2].

Since 2014, research on 5G network slicing has been underway. Network slicing technology benefits from the performance of machine learning methods in signal classification. Machine learning methods have been applied in areas such as network slicing reservation and network slice resource status [3]. In their study, Mei et al. focused on autonomous vehicles as different users from mobile communication devices and utilized inter-vehicle communication variables (such as packet delay, packet loss, data rate) [4]. Abidi et al. achieved network slicing with 93% accuracy using neural networks and deep belief networks in a simulation environment designed in accordance with 3GPP standards [5]. In their work, Toscana et al. employed a deep learning model to classify data into temporal and time-independent categories, achieving feedback-based network slicing with 75% accuracy [6]. Sun et al. rearrange unused network portions using a deep reinforcement learning model, thus reserving unused network portions for more efficient network management [7]. Thantharate et al., in their study, simulated 65,000 different users using machine learning methods in compliance with 3GPP rules, achieving a 95% accuracy in classifying network slices [8].

**Table 1.** Characteristics of Communication Systems Generations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1G | 2G | 3G | 4G | 5G |
| Time | 1980’s | 1990’s | 2000’s | 2010’s | 2020’s |
| Technology | FDMA | CDMATDMA | CDMAFDMATDMA | OFDMAMIMOSDNNFV | OFDMAmMIMONetwork Slicing |
| Download Speed | 2.4 kbits/s | 64 kbits/s | 2 Mbits/s | 1 Gbits/s | 10 Gbits/s |
|  Latency | - | 500 ms | 100 ms | 50 ms | <1ms |

In this study, the performance of machine learning methods on 5G network slicing classification was analyzed in detail by simulating realistic 3GPP-compliant error-prone users and base stations for 5G communication technologies. The association of users and base stations was established to ensure healthy communication without service disruption. For the smooth continuation of this binary association, accurate classification of network slices is crucial. To address the common issue of base station capacity problems, a data rate management mechanism was developed within the simulation architecture. Representing the mobile communication network, the simulation environment facilitated the realization of handovers between base stations and the internal handover processes within base stations, primarily due to changes in user-to-base station connections. Thus, alongside network slices, the simulation environment incorporated data rate management and user handover mechanisms.

In this study, Section 2 provides information about the simulation environment and the general workflow created. Section 3 offers a detailed interpretation of the performance analyses of machine learning methods on the simulation and the results obtained. Section 4 encompasses the conclusions and discussions.

1. **Generation of Simulation Environment and Synthetic Data**

In this study, a simulation has been conducted using 3GPP documents to create a scenario suitable for 5G communication, involving users and base stations. With this simulation, a scenario has been established where users communicate using 5G communication technology over 5 base stations placed in a 1 km² area. The simulation, conducted in two different scenarios, utilized error-prone environment and ideal environment scenarios.

The data presented in this study has been obtained from a synthetic simulation. User and base station parameters conforming to 3GPP documents with different configurations have been identified. In this simulation, parameters such as the position of users, users' movement speeds, source type, packet delay, packet loss, and data rate are utilized. The network slices designated for 5G communication in the simulation are considered as advanced versions of existing mobile services, including Enhanced Mobile Broadband (eMBB), which focuses on 5G communication, Ultra-Reliable Low Latency Communications (URLLC), which emphasizes packet delay and packet loss, Massive Internet of Things (MIoT), supporting communication with a high number of devices, and Vehicle-to-Everything (V2X), supporting communication technology from vehicles to other devices.

In creating the simulation environment, both lower and upper bounds specified in 3GPP documents are utilized. To ensure equal data distribution across all classes, the number of data instances in each class has been equalized. Some examples of synthetic data obtained from the simulation environment are shared in Table 2.

**Table 2.** Characteristics of Communication Systems Generations

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type | Network Slice | Source Type | Packet Delay | Packet Loss | Data Rate (byte) | Movement Speed (m/s) |
| File sharing | eMBB | Non-GBR | 245.52 | 3.2e-02 | 180674 | 1.95 |
| Smart home/car | MIoT | GBR | 203.83 | 3.4e-04 | 286 | 1.67 |
| Healthcare | URLLC | DC-GBR | 1.16 | 6.2e-07 | 367 | 1.45 |
| Autodriving | V2X | GBR | 61.25 | 6.7e-06 | 215378 | 63.74 |
| AR/VR/Gaming | eMBB | Non-GBR | 15.25 | 2.1e-04 | 2758967 | 0.06 |
| Sensor Notification | V2X | GBR | 21.67 | 9e-05 | 89547 | 0.85 |
| Electric Distribution | URLLC | GBR | 53.84 | 5.3e-06 | 567 | 0.02 |

One of the first parameters to be determined for the simulation environment is the type (macro, micro) and location of base stations. When determining the positions of base stations, care has been taken to ensure that they do not overlap in coverage areas within the simulation field. A total of 6 base stations have been positioned, with 3 being macro and 3 being micro base stations.

For the created simulation environment, the use of the correct network slice alone is not the sole objective. To obtain more realistic results, it is also crucial that users do not experience service interruptions against changing channel parameters while in motion. In the simulation environment, the user-base station relationship is initially established, utilizing channel parameters. Subsequently, a classifier for the network slice to be placed in the simulation is integrated. For users placed in the appropriate network slice, a data rate management mechanism based on the XGBoost model comes into play to address the issue of capacity reduction at the base station. Since user movements are considered as a parameter, a handover mechanism is implemented for possible user handover processes.

To associate users and base stations, path loss (PL) and shadow loss (SL) parameters are employed. To enhance the realism of the simulation environment, non-line-of-sight (NLOS) and three-dimensional distance calculations are used. Equation 1 shares the formula for this distance calculation, where hB represents the base station height, hU represents the user height, and x-y denotes the coordinates of the user and base station.

$d\_{3D}=\sqrt{(h\_{k}-h\_{B})^{2}+(x\_{k}-x\_{B})^{2}+(y\_{k}-y\_{B})^{2}}$ (1)

Using the distances obtained in Equation 1, path loss (PL) calculations can be performed for both macro and micro base stations. In Equation 2, path loss calculations for macro (PLUMa) and micro (PLUMi) base stations are provided. Here, the carrier frequency (fc) is chosen as 3.5GHz.

$PL\_{UMa}=10^{(32.4+20log\_{10}f\_{c}+30log\_{10}(d\_{3D}))/10}$

$PL\_{UMi}=10^{(32.4+20log\_{10}f\_{c}+31.9log\_{10}(d\_{3D}))/10}$ (2)

Using the path loss (PL) obtained in Equation 2, shadow loss (SL) calculations are demonstrated in Equation 3.

$G\left(x\right)=\frac{1}{\sqrt{2πσ^{2}}}exp\frac{π^{2}}{2σ^{2}}$

$R(x)=\left‖G\left(x\right)G(x)\right‖$

$P(x)=\frac{R(x)}{PL.SL}$ (3)

Using P(x), assignments of residual users to base stations can be carried out. In this process, base stations determine and allocate users based on which one has the stronger channel parameters. Users falling below the threshold value set in base stations cannot receive service.

To enhance the realism of the simulation environment, support is provided through auxiliary control mechanisms. These mechanisms include the data rate management system and the user handover system. The data rate management system typically comes into play when establishing a new relationship between users and a base station or when the capacity of a base station falls below a threshold. Since usage types consist of categorical data, the Extreme Gradient Boost model (XGBoost) will be utilized here. This will extend the period during which the base station provides high-quality service. As for the user handover system, a simple handover system operates based on the threshold defined according to the calculated channel parameters, facilitating user handovers as needed to maintain service quality.

1. **Simulation Results**

In the scope of this study, a 5G communication environment simulation has been conducted for a 1 km² area in accordance with the 3GPP documents. Six base stations and 300 users are positioned as depicted in Figure 1. The fundamental characteristics of the employed base stations are provided in Table 3.



**Figure 1**. The users and base stations in the simulation environment.

**Table 3.** The data of the base stations in the simulation environment.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Storage | Position | Diameter | Length |
| Macro | 5e7 byte | (260,525) | 250 m | 30 m |
| Macro | 5e7 byte | (800,250) | 250 m | 30 m |
| Macro | 5e7 byte | (300,775) | 250 m | 30 m |
| Micro | 1e7 byte | (800,750) | 100 m | 10 m |
| Micro | 1e7 byte | (800,250) | 100 m | 10 m |
| Micro | 1e7 byte | (125,125) | 100 m | 10 m |

Data was collected from users in two different scenarios based on this configuration. The first set of data consists of communication data generated in an ideal environment, while the second scenario involves data with user error margins. Machine learning methods were retested for the two simulation environments. Precision, recall, and F1-score metrics were used for the performance analysis of the models. Weighted average, considering the number of examples for each class, and macro-average, not considering the number of examples, were employed when calculating these metrics.

Fundamental machine learning methods commonly encountered in the literature were utilized. These methods include artificial neural networks, k-nearest neighbors, support vector machines, and random forests. In the initial testing phase, simulation data without error margins was used. User data without error margins conforms to the limits specified in the 3GPP documents. Table 4 provides information about the parameter space used and the parameters with the highest accuracy.

**Table 4.** Machine learning algorithms parameter space.

|  |  |  |  |
| --- | --- | --- | --- |
| Algorihm | Parameter | Parameter Space | Selected Parameter |
| kNN | k | 3, 5, 7, 9, 11, 13, 15 | 11 |
| Weights | uniform, distance | uniform |
| Metric | euclidean, Manhattan | manhattan |
| p | 2, 3 | 2 |
| MLP | Hidden Layer | 1, 2, 3, 4, 5 | 3 |
| Neuron (Hidden) | 100, 150, 200, 250, 300 | 200 |
| Activation Function (Hidden) | ReLU, Sigmoid, Tanh | Sigmoid |
| Activation Function (Output) | Softmax, Sigmoid, Linear | Softmax |
| Learning Rate | 0.0001, 0.001, 0.01 | 0.001 |
| Epochs | 50, 100, 150 | 100 |
| Dropout Rate | 0.2, 0.5, 0.7 | 0.2 |
| RF | Number of Trees | 50 |  |
| Tree Depth | None, 10, 20, 30 | 20 |
| XGBoost | Number of Trees | 50, 100, 200 | 100 |
| Maximum Depth | 3, 6, 9 | 9 |
| Learning Rate | 0.01, 0.1, 0.3 | 0.01 |
| Subsample | 0.8, 0.9, 1.0 | 0.9 |
| L1 | 0, 0.1, 1.0 | 0.1 |
| L2 | 0, 0.1, 1.0 | 0.1 |
| SVM | Kernel | Linear, Poly, RBF, Sigmoid | Poly |
| Degree of Poly | 2, 3, 4 | 4 |

In tests conducted with user data without error margins, where the training data and test data exhibit similarity, the metric results generally show high performance. The results of the performance metrics for the models are shared in Table 5. The XGBoost algorithm achieved the highest performance.

**Table 5.** Results of data without error margins.

|  |  |  |
| --- | --- | --- |
|  | Weighted Average | Macro Average |
|  | Sensitivity | Precision | F1-Score | Sensitivity | Precision | F1-Score |
| kNN | 97% | 97% | 97% | 97% | 97% | 97% |
| MLP | 95% | 96% | 95% | 95% | 96% | 96% |
| RF | 94% | 94% | 94% | 94% | 94% | 94% |
| XGBoost | 98% | 98% | 98% | 98% | 97% | 98% |
| SVM | 89% | 89% | 89% | 89% | 89% | 89% |

Error amounts in data with error margins were calculated by taking the average of error rates mentioned in the literature. These error margins were determined as 7% for delay-critical users and 25% for non-delay-critical users. These error rates were randomly added to the test portions of the user data obtained from the simulation. User data with error margins conforms to the limits specified in the 3GPP documents.

In tests conducted with user data containing error margins, where there is a difference between training data and test data due to the added error margins, the metric results generally show high performance compared to the tests without error margins. The results of the performance metrics for the models are shared in Table 6. The XGBoost algorithm achieved the highest performance.

**Table 6.** Results of data containing error margins.

|  |  |  |
| --- | --- | --- |
|  | Weighted Average | Macro Average |
|  | Sensitivity | Precision | F1-Score | Sensitivity | Precision | F1-Score |
| kNN | 96% | 96% | 96% | 96% | 96% | 96% |
| MLP | 93% | 94% | 94% | 93% | 94% | 94% |
| RF | 92% | 93% | 93% | 92% | 93% | 93% |
| XGBoost | 97% | 97% | 97% | 97% | 97% | 97% |
| SVM | 85% | 84% | 85% | 85% | 84% | 85% |

When examining both tables, it is observed that XGBoost methods adapt best to this simulation, while SVM method classifies with the worst performance. The cyclic system and data rate reduction system in the simulation also work in conjunction with the network language classifier. Parameters such as user turnover rate, data rate reduction rate, and total user connection rate are associated with the increase in the number of users. With the increase in the number of users, the data rate reduction rate is directly proportional, while the user turnover rate and data rate reduction rate show an inversely proportional change.

1. **Conclusion an Discussions**

In this study, a simulation environment designed based on 3GPP resources aims to classify network slicing. Various ML models have been employed in the simulation, which includes users and base stations, to place users in the most suitable network slice. The performance of these models is compared to select the most suitable one. The simulation environment is designed to resemble a real environment, so not only the network slicing architecture but also architectures like user handover and data rate regulation have been added. In the combination of these seemingly complex structures, the data rate reduction architecture and the network slicing architecture use the same training dataset. These data feed the ML models. The XGBoost model is used in the data rate reduction architecture to classify usage types. This allows customization and optimization of communication networks for specific usage types. The RF model, by detecting intensive video streams or inter-object communication, can manage network resources more efficiently. In the network slicing mechanism, the integration of various ML models aims to classify based on the majority decision of these models. These ML models are designed to classify network traffic in a more detailed and accurate way. Diversifying the selected models is important to focus on different features and understand the strengths of each. Classifying decisions based on the majority principle of the models will be crucial for stability and reliability in the system. The integration of these two mechanisms can make communication networks more flexible and scalable, responding more quickly and effectively to user demands. Additionally, the performance comparison of ML models will play a critical role in understanding the advantages and limitations of each model and creating a roadmap for future improvements. Network slicing classification in 5G technology is realized in two different scenarios: an ideal environment and a user error-prone environment. Precision, sensitivity, and F1-score metrics are used when analyzing the performance of ML models. Examining these metrics, it is observed that simulations involving error lead to a performance decrease for some models and an increase for others.

In conclusion, in future studies, developing a mechanism that can predict the required capacity for a specific network slice within a certain time frame using past user data is considered. This predictive mechanism can make simulation studies smoother and more efficient. In these future studies, using memory-based Machine Learning (ML) models and making accurate predictions will be crucial. These models, fed with past user data, can more accurately predict future capacity needs.

**References**

1. Matias, J., Garay, J., Toledo, N., Unzilla, J., & Jacob, E. (2015). Toward an SDN-enable NFV architecture. *IEEE Communications Magazine*, 53(4):187-193.
2. Gawas., A. U. (2021). An overview on evolution of mobile wireless communication network:1G-6G. *International Journal on Recent and Innovation Trends in Computing and Communication*, 3(5):3130-3133.
3. Borcoci, E., Dragulinescu, A., Li, F., Vochin, M. C., & Kjellstadli, K. (2021). An overview of 5G slicing operational business models for internet of vehicles, maritime IoT applications and connectivity solutions. *IEEE Access.*
4. Myung, H., Lim, J., & Goodman, D. (2006). Single carrier FDMA for uplink wireless transmission. *IEEE vehicular technology magazine, 1(3):30-38.*
5. Nelson, R., & Kleinrock, L. (1985). Spatial TDMA: A collision-free multihop channel access protocol. *IEEE Transactions on communications*, 33(9):934–944.
6. Foukas, X., Patounas, G., Elmokashfi, A., & Marina, M. (2017). Network slicing in 5G: Survey and challenges. *IEEE communications magazine*, 55(5):94–100.
7. Lucena, J., Ameigeiras, P., Lopez, D., Ramos-Munoz, J., Lorca, J., & Folgueira, J. (2017) Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges. *IEEE Communications Magazine*, 55(5):80–87.
8. Abidi, M. H., Alkhalefah, H., Moiduddin, K., Alazab, M., Mohammed, M. K., Ameen, W., & Gadekallu, T. R. (2021). Optimal 5G network slicing using machine learning and deep learning concepts. *Computer Standards & Interfaces*, 76:103518.
1. \* Corresponding author. *e-mail address: furkanbalci@gazi.edu.tr* [↑](#footnote-ref-1)