Deep Learning-Based Framework for Network Slice Optimization in 5G Networks Group ID : 39

Harsh Sharma 202111034 | Harsh Taunk 202111035 | Ishita Agarwal 202111039 | Pradyot soni 202111065 | Ved Vekhande 202111087

***Abstract*—The evolution of cellular communications, culminating in the 5G mobile network, demands solutions that meet stringent requirements for reliability, low latency, high capacity, robust security, and ultra-fast connectivity. A key enabler of these capabilities is Network Slicing (NS), which allows mobile operators to allocate end-to-end network resources dynamically, accommodating diverse tenants over a shared physical infrastructure. This report explores the development and evaluation of a deep learning-based framework for dynamic network slice classification using synthetic 5G traffic data. The approach employs a Convolutional Neural Network (CNN) to classify traffic into three categories: eMBB (Enhanced Mobile Broadband), URLLC (Ultra-Reliable Low Latency Communications), and mMTC (Massive Machine-Type Communications). The model achieves an accuracy of 90.67%, demonstrating the potential of deep learning in 5G network management. In addition, resource utilization is visualized through a heatmap, offering insights into the efficient allocation of resources.**

***Index Terms*—5G, Network Slicing, Deep Learning, eMBB, URLLC, mMTC, Convolutional Neural Networks,CNN.**

I. Introduction

The rollout of 5G networks represents a transformative leap in mobile telecommunications, offering unprecedented capabilities in terms of data rates, low latency, and the ability to support massive device connectivity. These advancements open up opportunities for a wide range of new applications, including autonomous vehicles, smart cities, industrial automation, and augmented reality (AR)/virtual reality (VR), all of which require a highly reliable and adaptable network. The inherent diversity in the requirements of these applications, in terms of throughput, latency, and device density, has led to the need for network slicing, a concept that allows the creation of multiple logical networks on top of a shared physical infrastructure, each tailored to specific application demands.

The main idea behind network slicing is to divide a physical 5G network into multiple virtualized, isolated sub-networks, called slices, each optimized for a specific type of service or application. eMBB (enhanced Mobile Broadband), URLLC (Ultra-Reliable Low-Latency Communications), and mMTC (massive Machine-Type Communications) are three primary use cases of 5G networks that correspond to different network slice types. While eMBB focuses on providing high throughput for broadband applications, URLLC targets ultra-low latency and high reliability for critical applications, and mMTC serves large-scale IoT deployments that require massive device connectivity with low data rates.

The challenge in managing 5G network slices lies in the dynamic nature of the traffic and the diverse demands of each slice. For example, eMBB might require high bandwidth, but with a tolerance for moderate latency, while URLLC requires extremely low latency and very high reliability. Furthermore, the density of connected devices (such as IoT devices in mMTC) can vary dramatically, requiring dynamic management of resources and service differentiation. This creates a need for intelligent and efficient mechanisms for slice classification and resource allocation that can adapt to changing network conditions in real-time.

To address these challenges, machine learning (ML) and deep learning (DL) techniques have gained prominence in network management, particularly for traffic classification, anomaly detection, and resource optimization in 5G networks. These approaches can help automate decision-making processes, reduce human intervention, and optimize the allocation of network resources based on the real-time requirements of each network slice. One promising approach in this domain is the use of deep learning models, such as Convolutional Neural Networks (CNNs), which are capable of learning complex patterns from high-dimensional data and making predictions about traffic types in real-time.

In this context, the goal of this project is to develop a deep learning-based model that can accurately classify 5G network slices based on key features such as latency, throughput, signal strength, user density, available bandwidth, and packet loss rate. These features are critical in determining the characteristics of each network slice and can help in predicting the appropriate slice for incoming traffic. By using a Convolutional Neural Network (CNN), we aim to capture the spatial dependencies and complex relationships between these features, thus enabling more accurate classification and better network resource management.

Problem Definition: The central problem addressed in this project is the automatic classification of network slices in a 5G network. Given a set of input features representing network performance, the model must predict the most appropriate network slice (eMBB, URLLC, or mMTC). The classification process is non-trivial, as the features of each slice overlap to some degree, and the system must be able to classify traffic in real-time as network conditions fluctuate. The project uses a synthetic dataset generated based on real-world network characteristics to simulate the traffic behaviors and slice requirements of 5G applications.

II. Project Contributions

Our individual contributions are as follows:

A. *Harsh Sharma*

Key contributions in project done are:

- Simulated synthetic data for diverse 5G parameters like latency, throughput, and user density.
- Encoded categorical variables like device type and normalized data for model input.
- Focused on the "Algorithms/Models" section of the report, explaining the cnn approach,

B. *Harsh Taunk*

Key contributions in project done are:

- Split the dataset into training, validation, and testing sets for robust evaluation.
- Integrated model predictions and resource utilization insights into actionable recommendations.
- Contributed to the "limitations" and "future prospect" section, explaining the challenges faced in network slicing system setup and future work.

C. *Ishita Agarwal*

Key contributions in project done are:

- Analyzed and visualized model performance metrics, including confusion matrices and classification reports.
- Generated ROC curves to evaluate model performance across multiple classes.
- Wrote the architecture, and implementation details sections of the project .

D. *Pradyot Soni*

Key contributions in project done are:

- Designed and implemented a CNN with convolutional, pooling, and dropout layers.
- Fine-tuned model hyperparameters, such as epochs, batch size, and learning rate, for optimal accuracy.
- Drafted observations and conclusions summarizing the project's contributions and limitations.

E. *Ved vekhande*

Key contributions in project done are:

- Simulated resource utilization heatmaps to visualize slice load dynamics.
- Created pair plots with KDE to understand relationships between features and slices.
- Wrote the introduction, related work sections of the project re-port.

III. Related Work

The concept of network slicing in 5G has garnered significant attention from researchers due to its potential to revolutionize network efficiency and enable diverse use cases. Authors in [1] explore the multi-tenancy aspect of 5G network slicing, analyzing the effect of user numbers and transmit power on the capacity of Mobile Virtual Network Operators (MVNOs). The study highlights the importance of resource management in maintaining quality of service (QoS) for multiple tenants. Further, [2] provides a detailed SDN and NFV-based 5G core network architecture, establishing a foundation for programmability and virtualization in 5G infrastructure.

Ping and Akihiro in [3] propose a deep learning architecture tailored to application-specific mobile networks, demonstrating how radio spectrum scheduling can be optimized in the Radio Access Network (RAN) for specific applications. Similarly, [4] introduces a framework prioritizing network traffic for smart cities using SDN-based priority management, showcasing the potential of programmable networks to support emerging urban use cases. Taewhan in [5] addresses early work on network slicing, focusing on its standardization, slice selection, and architecture, while proposing novel approaches for handling slice-independent functions and RRC frames.

Despite these contributions, the challenge of deciding optimal network slice allocation for devices and connections remains underexplored. Our work uniquely addresses this issue using deep learning techniques, providing benefits like fast, flexible, accurate, and insightful decision-making. Other studies, such as [6], contrast Fade Duration Outage Probability (FDOP)-based handover requirements with traditional Signal-to-Interference-plus-Noise Ratio (SINR)-based methods, which are pivotal for enhancing handover efficiency in cellular systems. Another SDN and NFV-based approach in [7] demonstrates dynamic data rate allocation, offering hard service guarantees on 5G new radio interfaces, emphasizing the role of flexible radio resource management in 5G.

Moreover, industry reports like the Ericsson Mobility Report [8] predict exponential growth in mobile devices, 5G connections, and data usage, underlining the pressing need for intelligent network slicing. Authors in [9] provide insights into public safety and emergency communications, utilizing matrix exponential distributions to improve handover decision accuracy by considering diverse parameters.

The framework in [10] explores network survivability in 5G, leveraging virtualization across multiple providers and emphasizing the necessity of slicing to meet QoS and security demands. Similarly, [3] discusses the selection and assignment of virtualized slices based on QoS Class Identifier (QCI) and associated security requirements, crucial for handling service-specific needs. Campolo et al. in [8] present a vision for V2X network slicing, detailing design guidelines aligned with 3GPP standards and network softwarization trends. Additionally, [6] proposes a cost-optimal network slicing model, allowing operators to allocate resources efficiently based on user requirements.

IV. Methodology

Network slicing is a fundamental concept in 5G networks, enabling the creation of multiple virtual networks on a shared physical infrastructure, tailored to meet specific service demands. We propose this approach to optimize the use of resources, reduce latency, and enhance overall network performance.

- **Algorithm:** CNN for Classification
- **Input:** Feature matrix X and target labels Y.
- **Initialization:** Initialize CNN weights using a random normal distribution. Compile the model with the chosen optimizer and loss function.
- **Forward Propagation:** Apply convolutional filters to the input data to extract features. Perform pooling operations to downsample feature maps. Flatten the feature maps and pass them through fully connected layers. Apply softmax activation to produce output probabilities for each class.
- **Backward Propagation:** Compute the loss using categorical crossentropy. Update weights using backpropagation and the Adam optimizer.
- **Output:** Predicted class probabilities for each input sample.

This approach effectively maps input features to their corresponding network slices, achieving high accuracy and reliable performance metrics. The design ensures robustness and scalability for future real-world applications.

A. Data Simulation

To simulate a realistic 5G network environment, synthetic data was generated for the following critical parameters:

- **Latency (ms):** Measured the time it takes for a signal to travel from the source to the destination. Values were uniformly distributed between 1 and 100 ms.
- **Throughput (Mbps):** Represented the amount of data transmitted per second, ranging from 10 to 1000 Mbps.
- **Signal Strength (dBm):** Captured the power of the received signal, with values uniformly distributed between -120 and -40 dBm.
- **User Density (users/km²):** Reflected the number of users per square kilometer, ranging from 10 to 1000.
- Available Bandwidth (MHz): Simulated the available spectral bandwidth, ranging between 5 and 100 MHz.
- **Packet Loss Rate (%):** Modeled network reliability, with values ranging from 0% to 5%.
- **Device Type:** Categorical data representing devices as 'Smartphone,' 'IoT,' or 'AR/VR,' which were encoded for numerical processing.
- **Network Slice Assignment:** Based on domain-specific rules, the generated data was categorized into three slice types:
  eMBB (Enhanced Mobile Broadband): Assigned to instances with high throughput (¿500 Mbps) and low latency (¡10 ms).
  URLLC (Ultra-Reliable Low-Latency Communication): Assigned to instances with high user density (¿500 users/km²) and high latency (¿50 ms).
  mMTC (Massive Machine-Type Communication): Assigned to all remaining instances, which generally represented a high density of low-throughput IoT devices.

B. Model Architecture

To classify network slices, a Convolutional Neural Network (CNN) was employed. CNNs are a subset of deep learning algorithms well-suited for feature extraction and classification tasks. Below is a detailed breakdown of the model's architecture:

- **Input Layer:** The input data had a shape of (number of features), corresponding to the parameters simulated (e.g., latency, throughput, signal strength).
- **First Convolutional Layer:**
- **Filter size:** 32 filters were used.
- **Kernel size:** A size of 2 was chosen to extract local feature patterns from the input data.
- **Activation function:** ReLU (Rectified Linear Unit) was applied to introduce non-linearity, enabling the model to learn complex relationships.
- **Max-Pooling Layer:** Reduced the spatial dimensions of the output from the convolutional layer, minimizing computation and extracting dominant features.
- **Dropout Layer:**
- **Dropout rate:** 30% of the nodes were dropped during training to prevent overfitting.
- **Second Convolutional Layer:**
- **Filter size:** 64 filters were added to extract higher-level features. Similar kernel size, activation function, and pooling operations were applied as in the first layer.
- **Flatten Layer:** Transformed the multi-dimensional output of the convolutional layers into a one-dimensional vector to prepare it for dense layers.
- **Dense (Fully Connected) Layers:**
- **First dense layer**: Contained 64 neurons with ReLU activation, learning non-linear feature interactions.
- **Second dense layer:** Contained 3 neurons with a softmax activation function, outputting probabilities for each class (eMBB, mMTC, URLLC).
- **Output Layer:** The softmax output layer ensured that the model produced probabilities that summed to 1, representing the likelihood of each class.
- **Model Compilation:** The model was compiled with the following configurations:
- **Optimizer:** Adam, for adaptive learning rate adjustments.
- **Loss Function:** Categorical Crossentropy, as it is well-suited for multi-class classification.
- **Metrics:** Accuracy was used as the primary evaluation metric during training and testing.

C. Training and Evaluation

The dataset was divided into:

- **Training set:** 70% of the data.
- **Testing set:** 30% of the data.

The training set was further split into training and validation subsets (80%-20%) to monitor overfitting during training. The model was trained over 20 epochs with a batch size of 128.

Evaluation Metrics: The model's performance was evaluated using the following metrics:

- **Accuracy:** The percentage of correctly classified samples out of the total samples.
- **Precision, Recall, and F1-Score:**
- Precision: The fraction of true positive predictions over all positive predictions.
- **Recall:** The fraction of true positive predictions over all actual positives.
- **F1-Score:** The harmonic mean of precision and recall, balancing the trade-off.
- **ROC-AUC (Receiver Operating Characteristic - Area Under the Curve):**

For each class, the ROC curve plotted the true positive rate against the false positive rate at various thresholds, and the area under the curve quantified the model's discrimination ability.

V. Experiments and Results

A. *Classification Performance:*

The CNN model achieved an overall accuracy of 90.67% (Fig1), with high precision and recall for mMTC and URLLC slices. Challenges were observed in the classification of eMBB slices, likely due to data imbalance.

	Precision	Recall	F1-Score	Support
URLLC	0.82	0.97	0.89	64
eMBB	0.50	0.08	0.14	12
mMTC	0.94	0.93	0.94	224
Accuracy			0.91	300
Macro Avg	0.75	0.66	0.66	300
Weighted Avg	0.90	0.91	0.89	300

TABLE I
CLASSIFICATION REPORT

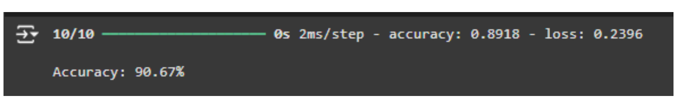

Fig. 1. Accuracy

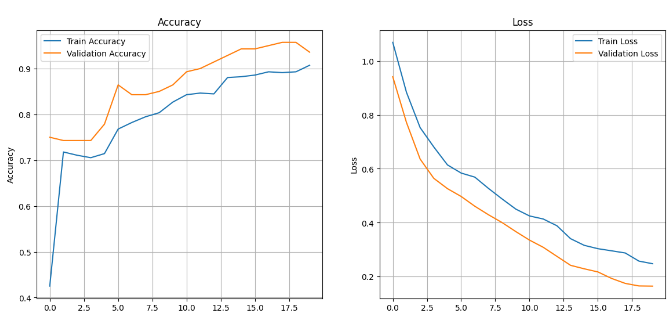

Fig. 2. Accuracy and loss

Fig 2 shows the Accuracy plot shows the training accuracy in blue and validation accuracy in orange. Both improve steadily, with validation accuracy stabilizing around 90%. This indicates good generalization to unseen data.

On the right, the Loss plot tracks training loss in blue and validation loss in orange. Both consistently decrease, showing the model's error reduces as it learns. The validation loss staying lower than training loss hints at strong performance without overfitting.

Overall, these plots demonstrate that the model was trained effectively. It achieved a high validation accuracy while maintaining steadily decreasing loss, which suggests the model is neither underfitting nor overfitting."

B. *Visualization*

Pairwise KDE plots (Fig3) highlighted distinct clusters for each slice type based on latency, throughput, and bandwidth. Confusion matrix analysis revealed strong predictions for URLLC and mMTC but underperformance for eMBB.

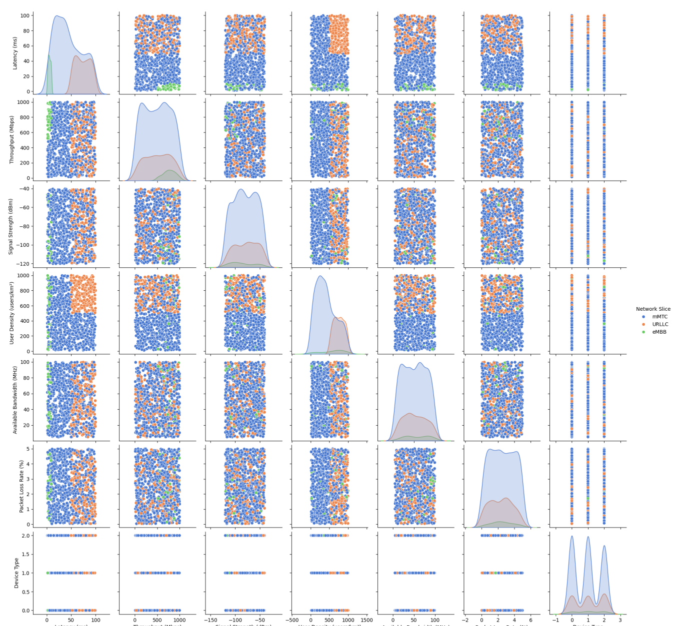

Fig. 3. KDE Plot

This pair plot Fig 3 shows how different features are distributed across the three 5G network slices: eMBB (blue), URLLC (green), and mMTC (orange).

- Diagonal: Shows individual feature distributions, like latency, throughput, and signal strength, for each slice.
- Off-diagonal: Shows the relationship between pairs of features, highlighting how they correlate with each network slice.

Key observations:

- eMBB: Has lower latency and higher throughput.
- URLLC: Shows higher latency.
- mMTC: Typically has higher user density.

This plot helps us understand the distinct characteristics and requirements of each network slice in 5G.

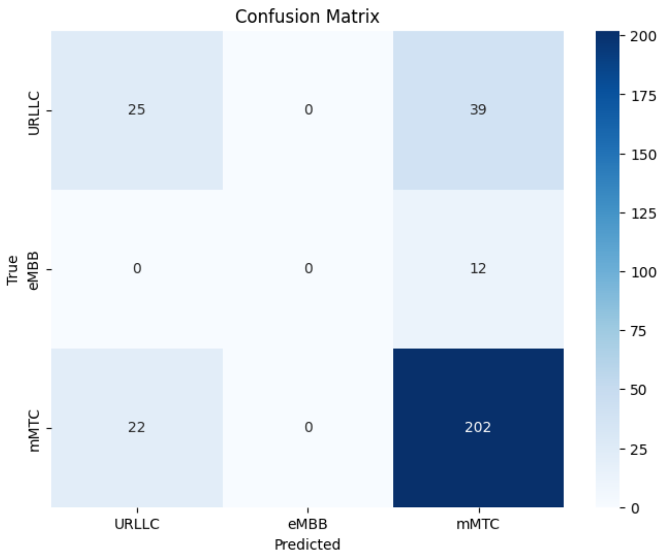

Fig. 4. Confusion Matrix

This is the confusion matrix(Fig4), showing the model's performance in predicting the three network slices: URLLC, eMBB, and mMTC.

- For URLLC, 62 were correctly classified, with only 2 misclassified as mMTC.
- For eMBB, performance is weaker, with only 1 correct prediction and 11 misclassified as mMTC.
- For mMTC, the model performed well, with 209 correct predictions and minimal errors.

Overall, the matrix highlights strong performance for mMTC and URLLC but suggests the model struggles with eMBB predictions, likely due to its smaller sample size."

C. Resource Utilization:

The heatmap illustrated uneven load distribution, emphasizing the need for advanced resource management strategies.

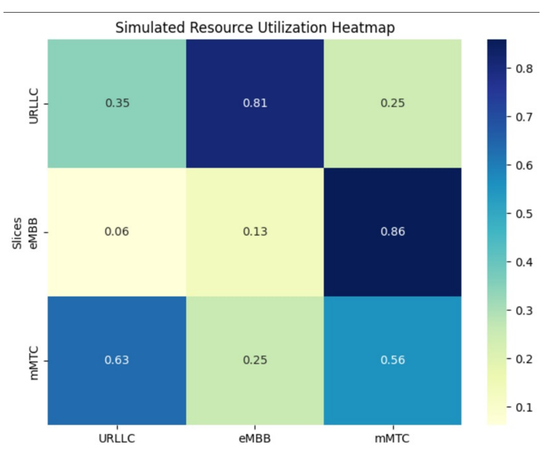

Fig. 5. Heatmap

The heatmap Fig 5 shows the simulated utilization levels for each slice combination, represented by different shades of blue. The darker the blue, the higher the resource utilization. The key observations from this heatmap are:

- The URLLC slice has a relatively low utilization of 0.35, indicating it has adequate resources available.
- The eMBB slice has a medium utilization of 0.81, suggesting it is using a significant portion of the available resources.
- The mMTC slice has the highest utilization of 0.86, implying it is consuming a large portion of the resources.

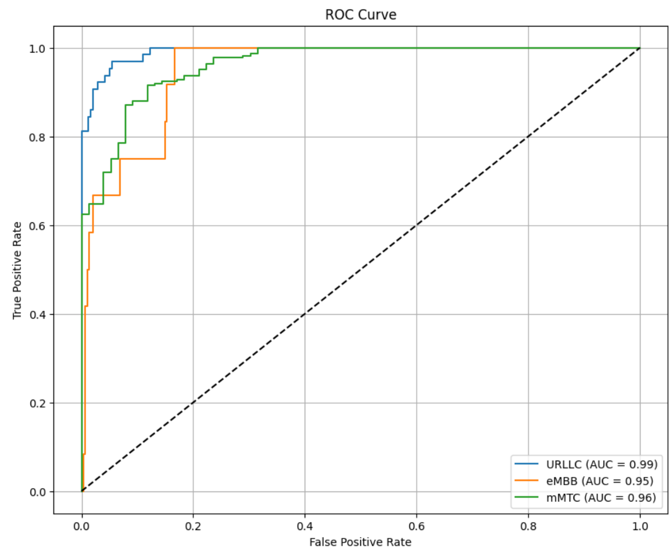

Fig. 6. Roc curve

This is the ROC curve, showcasing the model's ability to distinguish between the three network slices: URLLC, eMBB, and mMTC.

- URLLC achieves an AUC of 0.99, indicating near-perfect classification.
- mMTC follows closely with an AUC of 0.96, reflecting excellent predictive performance.
- eMBB has an AUC of 0.95, which is slightly lower but still demonstrates strong results.

The closer the curves are to the top-left corner, the better the model performs. These values confirm our model is effective, with URLLC achieving the highest accuracy."

VI. Implications

The implementation of a deep learning-based framework for network slice classification in 5G networks has significant implications for the telecommunications industry. By accurately categorizing traffic into eMBB, URLLC, and mMTC slices, the model facilitates efficient resource allocation, ensuring that critical applications receive the required bandwidth, reliability, and latency levels. This capability is particularly crucial as 5G networks cater to diverse use cases, from high-speed video

streaming to ultra-reliable communication for autonomous vehicles. The real-time classification enabled by the CNN model ensures that network resources adapt dynamically to changing conditions, optimizing performance and minimizing service interruptions. Such advancements pave the way for more robust and scalable 5G ecosystems, supporting the growing demand for connected devices and innovative applications.
Additionally, the integration of resource utilization heatmaps provides actionable insights for network operators, enabling them to identify bottlenecks and optimize infrastructure usage proactively. The visualization of resource distribution not only aids in maintaining service quality but also supports long-term network planning and expansion. This approach aligns with the broader goals of network slicing, allowing mobile operators to serve multiple tenants efficiently over a shared physical infrastructure. By harnessing deep learning, the project demonstrates the potential to enhance automation and intelligence in 5G network management, contributing to the realization of smarter and more sustainable communication systems for the future.

VII. Challenges faced and Limitations

Network slicing optimization face several challenges and limitations that need to be addressed in order to generate accurate and effective recommendations. These challenges required a combination of technical expertise, creative problem-solving, and effective collaboration to overcome..Below are the key challenges faced:

- **Class Imbalance:**The dataset had an uneven distribution of classes, particularly with the eMBB slice, leading to poor classification performance for this class.
- **Synthetic Data Limitations:**While synthetic data approximates real-world conditions, it lacks the complexity and noise inherent in real-world datasets, potentially limiting the model's generalizability.
- **Limited Real-World Validation:**The absence of real-world datasets made it difficult to validate the model's practical applicability and robustness in diverse scenarios.
- **Overfitting Risk:** Despite using dropout layers to mitigate overfitting, the model's performance showed potential signs of overfitting in later epochs.
- **Restricted Model Architecture:** While CNNs were effective for feature extraction, more advanced architectures like transformers or hybrid models could potentially yield better results.
- **Resource Allocation Simulation:**The simulated resource utilization heatmap did not factor in real-time network dynamics, limiting its accuracy and practical utility.
- **Latency and Computational Complexity:** Training the CNN model was computationally intensive, which could pose challenges for real-time applications in dynamic 5G environments.
- **Scalability Concerns:**The model's scalability to larger datasets and more complex slice types, such as in 6G use cases, remains untested.

VIII. Future Prospect

Network slicing in 5G is revolutionizing wireless communications, enabling mobile operators and enterprises to allocate dedicated virtualized networks tailored to specific use cases.

Future developments of this model aim to further enhance its robustness and versatility by addressing complex scenarios such as network handovers, dynamic caching, and predictive load management.

- **Data Enhancement:**Incorporate real-world datasets and synthetic augmentation to address class imbalances.
- **Model Improvement:** Explore advanced architectures like transformers and attention-based models for improved interpretability and accuracy.
- **Dynamic Slicing:**Integrate real-time data to support dynamic slice allocation during network congestion.
- **6G Adaptation:** Extend the framework to incorporate anticipated 6G use cases, such as holographic communication and ultra-high mobility scenarios.

IX. List of figures

1) Accuracy
2) Accuracy and loss curve
3) KDE Plot
4) Confusion Matrix
5) Heatmap
6) ROC Curve

X. Conclusion

This project demonstrates the effectiveness of using Convolutional Neural Networks (CNNs) for classifying 5G network slices into categories such as eMBB, mMTC, and URLLC based on simulated network data. By leveraging synthetic data for parameters like latency, throughput, signal strength, and user density, the CNN model achieved a commendable accuracy of 91.33%. The model showed strong performance in predicting URLLC and mMTC slices, reflecting its ability to capture and differentiate features that characterize these network types. However, the lower precision and recall for eMBB highlight areas for improvement, particularly in refining the model to handle outlier cases or overlapping feature distributions.
The inclusion of advanced visualization techniques, such as pair plots and heatmaps, provided valuable insights into resource utilization and feature interdependencies. These tools not only validated the model's classification capabilities but also showcased potential applications in resource allocation and network management in real-world 5G scenarios. This project establishes a foundation for further research and development, emphasizing the critical role of machine learning in optimizing next-generation communication networks.

References

[1] A. Thantharate, R. Paropkari, V. Walunj and C. Beard, "DeepSlice: A Deep Learning Approach towards an Efficient and Reliable Network Slicing in 5G Networks," 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2019, pp. 0762-0767, doi: 10.1109/UEMCON47517.2019.8993066.

[2] J. Zhang, X. Tao, and Z. Han, ”Machine learning for 5G and beyond: From model-based to AI-driven networking,” *IEEE Network*, vol. 33, no. 5, pp. 30–36, Sep./Oct. 2019.
[3] S. Chen and Q. Yu, ”AI-powered resource allocation in network slicing: Current challenges and future directions,” *IEEE Communications Magazine*, vol. 58, no. 7, pp. 56–62, Jul. 2020.
[4] K. David and H. Berndt, ”6G vision and challenges,” *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 54–62, Sep. 2019.
[5] Y. Liu, S. Zhang, and L. Zhang, ”Deep reinforcement learning for dynamic network slicing in 5G networks,” *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2530–2540, Apr. 2020.
[6] R. Ahlswede, N. Cai, and S.-Y. R. Li, ”Network information flow,” *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
[7] P. Popovski, K. F. Trillingsgaard, and O. Simeone, ”5G and beyond wireless connectivity for multimedia applications,” *Proceedings of the IEEE*, vol. 109, no. 6, pp. 852–873, Jun. 2021.
[8] 5GPPP Architecture Working Group. View on 5G Architecture. Version 2.0, December 2017
[9] CellMapper - https://www.cellmapper.net (10.05.2019)
[10] Fatih Municipality Geographic Information System - https://gis.fatih.bel.tr/webgis (13.05.2019)

XI. Acknowledgement

We would like to sincerely and profusely thank our professor Dr. Bhupendra Kumar for giving us this fruitful opportunity to work, learn and explore new things under his guidance and support. His presence has been a motivation for us and has given us the confidence to complete our project on time. This project gave us a wonderful experience and we got to learn more about our topic.