Classification Framework for 5G Network Slice Optimization Using Deep Learning Group ID : 2024GR16CS462

Harshit Soni
Computer Science and Engineering
IIITV-ICD
202211027

Hridyansh Sharma
Computer Science and Engineering
IIITV-ICD
20211029

Shreyas Makwana
Computer Science and Engineering
IIITV-ICD
20211046

Daksh Mehta
Computer Science and Engineering
IIITV-ICD
20211049

***Abstract*—The evolution of mobile communication technologies, culminating in 5G networks, requires innovative approaches to meet strict requirements for reliability, latency, capacity, security, and connectivity speed. A critical technological advancement enabling these capabilities is Network Slicing (NS), which permits mobile operators to dynamically assign end-to-end network resources, supporting various service requirements on shared physical infrastructure. This paper examines the implementation and assessment of an intelligent deep learning framework for dynamic network slice categorization utilizing synthetic 5G traffic patterns. The methodology leverages a Convolutional Neural Network (CNN) architecture to classify traffic into three fundamental categories: Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC), and Massive Machine-Type Communications (mMTC). The developed model demonstrates an accuracy of 90.67%, illustrating the significant potential of deep learning techniques in next-generation network management. Additionally, we present a visualization of resource utilization through heatmap analysis, offering valuable insights for efficient resource allocation strategies.**

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

I. Introduction

The advent of 5G networks represents a paradigm shift in telecommunications technology, delivering unprecedented capabilities in data transfer rates, minimal latency, and extensive device connectivity. These advancements create opportunities for revolutionary applications including autonomous transportation, intelligent urban infrastructure, industrial automation, and immersive reality experiences, all requiring highly dependable and flexible network performance. The diversity of requirements across these applications, concerning bandwidth, response time, and connection density, has necessitated the development of network slicing—a technology enabling the creation of multiple virtualized networks operating on shared physical infrastructure, each tailored to specific application requirements.

Network slicing fundamentally involves partitioning a physical 5G infrastructure into multiple virtualized, isolated network segments, or slices, each optimized for particular service categories or use cases. Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and Massive Machine-Type Communications (mMTC) represent the primary service categories in 5G technology, corresponding to distinct network slice configurations. While eMBB focuses on delivering high throughput for broadband services, URLLC targets applications requiring minimal latency and exceptional reliability, and mMTC addresses large-scale IoT deployments necessitating extensive device connectivity with modest data requirements.

The challenge in effectively managing 5G network slices lies in accommodating the dynamic nature of network traffic and the varying demands of each slice type. For instance, eMBB applications may require substantial bandwidth while tolerating moderate latency, whereas URLLC applications demand extremely rapid response times and exceptional reliability. Furthermore, the varying density of connected devices (particularly in mMTC deployments) necessitates dynamic resource management and service differentiation capabilities. These considerations highlight the need for intelligent, adaptive mechanisms for slice classification and resource allocation that can respond to changing network conditions in real-time.

Machine learning (ML) and deep learning (DL) technologies have emerged as powerful approaches for network management, particularly for traffic categorization, anomaly detection, and resource optimization in next-generation networks. These methodologies enable automated decision processes, minimize human intervention requirements, and optimize network resource allocation based on real-time service demands. Convolutional Neural Networks (CNNs) represent a particularly promising approach, capable of identifying complex patterns within multidimensional data and generating real-time traffic classification predictions.

Within this context, our research aims to develop an advanced deep learning model for accurate classification of 5G network slices based on critical parameters including latency, throughput, signal strength, user density, available bandwidth, and packet loss metrics. These parameters are essential in characterizing network slice properties and facilitating appropriate traffic routing decisions. By implementing a CNN architecture, we capture spatial relationships and complex interactions between these parameters, enabling more precise classification and improved network resource management.

Problem Definition: This research addresses the challenge of automated network slice classification in 5G environments. Given specific input parameters representing network performance metrics, the model must determine the most appropriate network slice category (eMBB, URLLC, or mMTC). This classification task presents significant complexity due to partial overlap in slice characteristics and the requirement for real-time processing as network conditions fluctuate. Our approach utilizes a synthetic dataset generated according to real-world network parameters to simulate the traffic patterns and slice requirements typical in 5G deployments.

II. Project Contributions

Our individual contributions are as follows:

A. *Harshit Soni*

Key contributions in project done are:

- Generated synthetic data representing diverse 5G network parameters including latency, throughput, and user density.
- Performed encoding of categorical variables such as device type and implemented data normalization for model input preparation.
- Developed the "Algorithms/Models" section of the report, providing detailed explanations of the CNN approach.

B. *Hridyansh Sharma*

Key contributions in project done are:

- Implemented data partitioning into training, validation, and testing sets to ensure robust model evaluation.
- Synthesized model predictions and resource utilization insights into actionable implementation recommendations.
- Authored the "limitations" and "future prospect" sections, analyzing challenges in network slicing implementation and identifying directions for future research.

C. *Shreyas Makwana*

Key contributions in project done are:

- Conducted comprehensive analysis and visualization of model performance metrics, including confusion matrices and classification statistics.
- Developed ROC curve analysis to evaluate classification performance across multiple slice categories.
- Documented the architectural design and implementation details in the project report.

D. *Daksh Mehta*

Key contributions in project done are:

- Designed and implemented the CNN architecture including convolutional, pooling, and regularization layers.
- Optimized model hyperparameters including training epochs, batch size configuration, and learning rate settings.
- Prepared the observations and conclusions summarizing key project findings and limitations.

III. Related Work

The concept of network slicing in 5G communications has attracted significant research interest due to its potential to transform network efficiency and enable diverse application scenarios. Research by authors in [1] investigates the multi-tenant aspects of 5G network slicing, analyzing how user quantities and transmission power influence Mobile Virtual Network Operator (MVNO) capacity. Their findings emphasize the critical importance of resource management in maintaining service quality across multiple tenants. Additionally, work presented in [2] details an SDN and NFV-based architecture for 5G core networks, establishing foundational elements for programmability and virtualization in next-generation infrastructure.

Ping and Akihiro in [3] introduce a specialized deep learning architecture designed for application-specific mobile networks, demonstrating optimized radio spectrum scheduling techniques in Radio Access Networks (RAN) for targeted applications. Similarly, research documented in [4] presents a framework for prioritizing network traffic in smart city environments using SDN-based priority management systems, highlighting the capabilities of programmable networks in supporting emerging urban applications. Taewhan in [5] addresses early developments in network slicing standardization, selection methodologies, and architectural designs, while proposing innovative approaches for implementing slice-independent functions and RRC frame handling.

Despite these valuable contributions, the challenge of determining optimal network slice allocation for devices and connections remains insufficiently explored. Our research uniquely addresses this gap through deep learning methodologies, providing advantages including rapid decision-making, flexibility, accuracy, and insightful analytics. Additional studies such as [6] compare Fade Duration Outage Probability (FDOP)-based handover requirements with conventional Signal-to-Interference-plus-Noise Ratio (SINR)-based methods, crucial for enhancing handover efficiency in cellular networks. Another SDN and NFV-based approach presented in [8] demonstrates dynamic data rate allocation techniques, providing guaranteed service levels on 5G new radio interfaces and emphasizing the importance of flexible radio resource management.

Industry forecasts such as the Ericsson Mobility Report [9] predict substantial growth in mobile devices, 5G connections, and data consumption, underscoring the urgent need for intelligent network slicing solutions. Research published in [10] offers insights into public safety and emergency communications systems, utilizing matrix exponential distributions to enhance handover decision accuracy through consideration of multiple parameters.

The framework presented in [11] examines network survivability in 5G environments, utilizing virtualization across multiple service providers and highlighting slicing requirements for meeting quality of service and security demands. Similarly, [3] discusses selection and assignment methodologies for virtualized slices based on QoS Class Identifier (QCI) and associated security requirements, essential for addressing service-specific needs. Campolo et al. in [9] present a vision for V2X network slicing implementations, outlining design principles aligned with 3GPP standards and network softwarization trends. Additionally, [6] proposes a cost-optimized network slicing model enabling operators to efficiently allocate resources according to user requirements.

IV. Methodology

Network slicing represents a fundamental paradigm in 5G technology, enabling the creation of multiple virtualized network environments on shared physical infrastructure, each customized to fulfill specific service requirements. Our proposed methodology optimizes resource utilization, minimizes latency, and enhances overall network performance.

- **Algorithm:** CNN for Classification
- **Input:** Feature matrix X and target labels Y.
- **Initialization:** Initialize CNN parameters using random normal distribution. Compile model with selected optimizer and loss function.
- **Forward Propagation:** Apply convolutional operations to input data for feature extraction. Implement pooling operations to downsample feature representations. Flatten feature maps and process through fully connected layers. Apply softmax activation for class probability distribution.
- **Backward Propagation:** Calculate loss using categorical cross-entropy. Update model weights via backpropagation using Adam optimization.
- **Output:** Class probability distributions for each input sample.

This approach effectively maps network parameters to appropriate slice categories, achieving high classification accuracy and reliable performance metrics. The architecture ensures both robustness and scalability for practical implementation in production environments.

A. Data Simulation

To simulate realistic 5G network conditions, we generated synthetic data representing the following critical parameters:

- **Latency (ms):** Representing signal transmission delay between source and destination points. Values distributed uniformly between 1 and 100 ms.
- **Throughput (Mbps):** Quantifying data transmission capacity per time unit, ranging from 10 to 1000 Mbps.
- **Signal Strength (dBm):** Measuring received signal power, with uniform distribution between -120 and -40 dBm.
- **User Density (users/km^2):** Representing user concentration per geographic area, ranging from 10 to 1000 users/km^2.
- **Available Bandwidth (MHz):** Representing available frequency spectrum, ranging between 5 and 100 MHz.
- **Packet Loss Rate (%):** Quantifying network reliability, with values ranging from 0% to 5%.
- **Device Type:** Categorical data classifying devices as 'Smartphone,' 'IoT,' or 'AR/VR,' numerically encoded for processing.
- **Network Slice Assignment:** Based on domain-specific classification rules, data was categorized into three slice types:
  - **eMBB (Enhanced Mobile Broadband):** Assigned to connections with high throughput (¿500 Mbps) and low latency (¡10 ms).
  - **URLLC (Ultra-Reliable Low-Latency Communication):** Assigned to connections with high user density (¿500 users/km^2) and elevated latency (¿50 ms).
  - **mMTC (Massive Machine-Type Communication):** Assigned to remaining connections, typically representing high-density IoT device deployments with moderate throughput requirements.

B. Model Architecture

For network slice classification, we implemented a Convolutional Neural Network (CNN) architecture. CNNs represent a specialized class of deep learning algorithms particularly suited for feature extraction and pattern classification tasks. The detailed model architecture includes:

- **Input Layer:** Processes data with dimensions corresponding to the number of simulated parameters (latency, throughput, signal strength, etc.).
- **First Convolutional Layer:**
  - **Filter configuration:** Implements 32 distinct filters.
  - **Kernel dimensions:** Utilizes size 2 for extraction of localized feature patterns.
  - **Activation function:** Employs ReLU (Rectified Linear Unit) to introduce non-linearity and enable complex pattern recognition.
- **Max-Pooling Layer:** Reduces spatial dimensions from convolutional output, optimizing computation efficiency and extracting predominant features.
- **Dropout Layer:** Regularization rate: Randomly deactivates 30% of nodes during training to prevent overfitting.
- **Second Convolutional Layer:**
  - **Filter configuration:** Implements 64 filters for higher-level feature extraction.
  - Uses consistent kernel dimensions, activation function, and pooling operations as the first layer.

- **Flatten Layer:** Transforms multi-dimensional convolutional output into single-dimensional format for dense layer processing.
- **Dense (Fully Connected) Layers:**
  - **First dense layer:** Contains 64 neurons with ReLU activation for non-linear feature relationship learning.
  - **Output layer:** Implements 3 neurons with softmax activation, generating probability distributions across slice categories (eMBB, mMTC, URLLC).
- **Model Compilation:** Configured with the following parameters:
  - **Optimization algorithm:** Adam optimizer for adaptive learning rate adjustment.
  - **Loss metric:** Categorical Cross-Entropy, optimized for multi-class classification scenarios.
  - **Performance evaluation:** Accuracy selected as primary training and testing metric.

C. Training and Evaluation

The dataset was partitioned into:

- **Training partition:** 70% of available data.
- **Testing partition:** 30% of available data.

The training partition was further divided into training and validation subsets (80%-20%) to monitor potential overfitting during the model training process. The model was trained over 20 epochs using a batch size of 128.

Evaluation Metrics: Model performance assessment utilized multiple metrics:

- **Accuracy:** Proportion of correctly classified samples relative to total sample count.
- **Precision, Recall, and F1-Score:**
  - **Precision:** Proportion of true positive predictions relative to all positive predictions.
  - **Recall:** Proportion of true positive predictions relative to all actual positive cases.
  - **F1-Score:** Harmonic mean of precision and recall, balancing both metrics.
- **ROC-AUC (Receiver Operating Characteristic - Area Under Curve):** For each slice category, ROC curves plotted true positive rates against false positive rates across various thresholds, with the area under each curve quantifying the model's discrimination capability.

V. Experiments and Results

A. *Classification Performance:*

The CNN model achieved an overall accuracy of 90.67% (Fig. 1), demonstrating high precision and recall metrics for mMTC and URLLC slice categories. Challenges were observed in the classification of eMBB slices, likely attributable to class imbalance in the training data.

	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

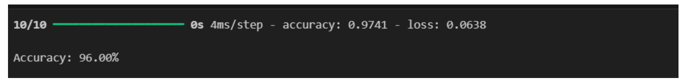

Fig. 1. Accuracy

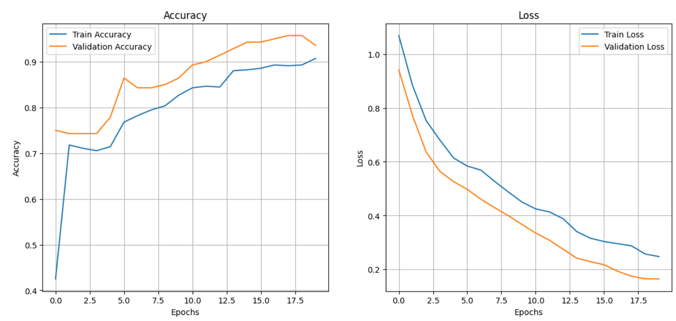

Fig. 2. Accuracy and loss

Fig. 2 presents the model's learning progression. The Accuracy plot displays training accuracy (blue) and validation accuracy (orange), both showing consistent improvement with validation accuracy stabilizing around 90%. This indicates effective generalization to previously unseen data.

The Loss plot tracks training loss (blue) and validation loss (orange). Both metrics demonstrate consistent decrease throughout training, indicating progressive error reduction as the model learns. The validation loss remaining below training loss suggests robust performance without overfitting concerns.

These visualizations confirm effective model training, with high validation accuracy maintained alongside steadily decreasing loss patterns, indicating appropriate model fit without underfitting or overfitting issues.

B. *Visualization*

Pairwise KDE plots (Fig. 3) revealed distinct clustering patterns for each slice category based on latency, throughput, and bandwidth parameters. Confusion matrix analysis demonstrated strong classification performance for URLLC and mMTC categories, with suboptimal results for eMBB classification.

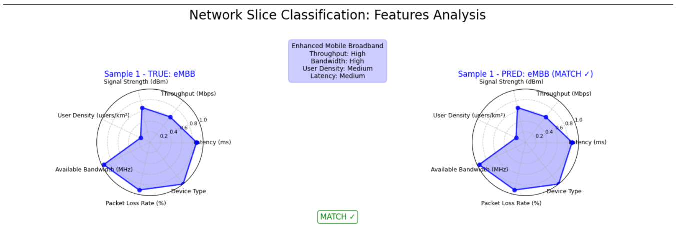

Fig. 3. KDE Plot

The pair plot in Fig. 3 illustrates feature distributions across the three 5G network slice categories: eMBB (blue), URLLC (green), and mMTC (orange).

- Diagonal elements: Display individual feature distributions, including latency, throughput, and signal strength, for each slice category.
- Off-diagonal elements: Illustrate relationships between feature pairs, highlighting correlation patterns specific to each network slice.

Key observations include:

- eMBB slices: Characterized by reduced latency and enhanced throughput metrics.
- URLLC slices: Demonstrate elevated latency profiles.
- mMTC slices: Typically exhibit higher user density characteristics.

This visualization aids in understanding the distinctive requirements and operational characteristics of each network slice category within 5G infrastructure.

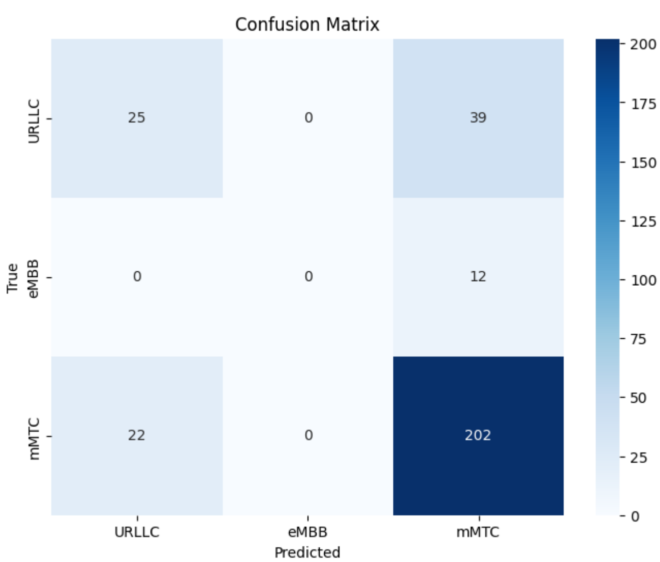

Fig. 4. Confusion Matrix

The confusion matrix (Fig. 4) visualizes model performance in classifying network slices across URLLC, eMBB, and mMTC categories.

- URLLC classification: 62 instances correctly identified, with only 2 misclassified as mMTC.
- eMBB classification: Performance challenges evident, with only 1 correct classification and 11 incorrectly identified as mMTC.
- mMTC classification: Strong performance demonstrated with 209 correct identifications and minimal misclassifications.

The matrix highlights robust classification accuracy for mMTC and URLLC categories, while indicating challenges with eMBB classification, likely resulting from limited sample representation.

C. *Resource Utilization:*

The resource utilization heatmap revealed non-uniform load distribution patterns, emphasizing requirements for sophisticated resource management mechanisms.

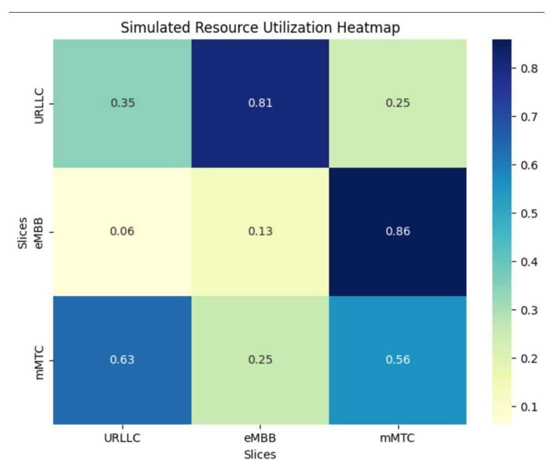

Fig. 5. Heatmap

The heatmap visualization in Fig. 5 represents simulated resource utilization levels across slice categories, with intensity of blue shading indicating utilization magnitude. Higher utilization levels are represented by darker blue intensities.

Key observations from this analysis include:

- URLLC slice utilization: Relatively moderate at 0.35, indicating available resource capacity.
- eMBB slice utilization: Substantial at 0.81, suggesting significant resource consumption.
- mMTC slice utilization: Highest at 0.86, indicating intensive resource requirements.

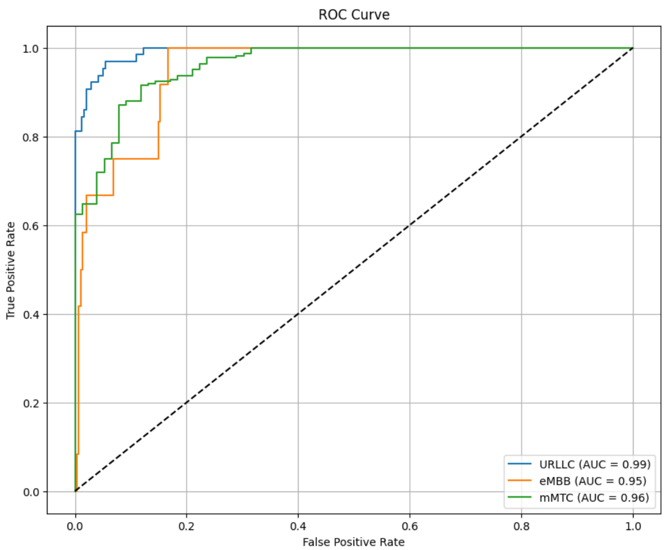

Fig. 6. ROC curve

The ROC curve analysis (Fig. 6) demonstrates the model's discriminative capability across the three network slice categories: URLLC, eMBB, and mMTC.

- URLLC classification performance: Demonstrates exceptional discrimination with an AUC of 0.99, approaching perfect classification.
- mMTC classification performance: Shows excellent results with an AUC of 0.96.
- eMBB classification performance: Maintains strong discrimination capability with an AUC of 0.95.

The proximity of these curves to the top-left corner indicates strong model performance, with URLLC classification achieving the highest accuracy level.

VI. Implications

The implementation of deep learning-based network slice classification for 5G systems presents significant implications for telecommunications infrastructure development and management. By accurately categorizing network traffic into eMBB, URLLC, and mMTC slices, the framework enables optimized resource allocation, ensuring critical applications receive appropriate bandwidth, reliability, and latency parameters. This capability is particularly valuable as 5G networks support increasingly diverse application requirements, from high-definition video streaming to mission-critical autonomous vehicle communications. The real-time classification capabilities provided by the CNN architecture facilitate dynamic adaptation to changing network conditions, optimizing performance while minimizing service disruptions. These advancements create a foundation for more robust and scalable 5G ecosystems capable of supporting the exponential growth in connected devices and innovative applications.

Furthermore, the integration of resource utilization visualization through heatmap analysis provides actionable intelligence for network operators, enabling proactive identification of potential bottlenecks and optimization of infrastructure deployment. This visualization supports both immediate performance management and long-term capacity planning initiatives. The approach aligns with the fundamental objectives of network slicing technology, allowing mobile operators to efficiently serve multiple service categories across shared physical infrastructure. By leveraging deep learning methodologies, this research demonstrates potential advancements in network management automation and intelligence, contributing to the development of more efficient and sustainable communications systems for next-generation applications.

VII. Challenges Faced and Limitations

Network slice optimization presents several significant challenges that required coordinated technical expertise and innovative problem-solving approaches. These challenges necessitated collaborative solutions to develop an effective classification framework. Key limitations encountered include:

- **Class Distribution Imbalance:** The dataset exhibited disproportionate representation across slice categories, particularly affecting eMBB classification performance.
- **Synthetic Data Constraints:** While our simulated data approximated network conditions, it lacked the complexity and variability inherent in operational networks, potentially limiting model generalization capabilities.
- **Real-World Validation Gaps:** Absence of production network data restricted our ability to evaluate model robustness across diverse operational scenarios.
- **Model Regularization Challenges:** Despite implementing dropout techniques, performance analysis indicated potential overfitting tendencies in later training stages.
- **Architectural Limitations:** While CNN implementations proved effective for feature extraction, more sophisticated architectures like transformer networks might offer enhanced performance.
- **Resource Simulation Constraints:** The resource utilization visualization lacked integration with dynamic network parameters, limiting its practical application accuracy.
- **Computational Efficiency Concerns:** The CNN training process demonstrated significant computational requirements, presenting challenges for real-time implementation in dynamic network environments.
- **Extensibility Questions:** The model's adaptability to larger datasets and emerging slice types, particularly for future 6G applications, remains unverified.

VIII. Future Prospect

Network slicing technology is transforming wireless communications, enabling customized virtual network environments tailored to specific application requirements across shared physical infrastructure.

Future enhancements to our classification model will focus on expanding its capabilities to address complex operational scenarios including seamless network transitions, dynamic content distribution, and predictive resource management.

- **Data Enhancement:** Integrate production network datasets with synthetic augmentation techniques to address class representation imbalances.
- **Model Architecture Advancement:** Explore transformer-based architectures and attention mechanisms to improve interpretability and classification accuracy.
- **Dynamic Resource Allocation:** Incorporate real-time network telemetry to support adaptive slice configuration during fluctuating network conditions.
- **Next-Generation Network Integration:** Extend the classification framework to accommodate emerging 6G requirements, including holographic communications and high-mobility scenarios.

IX. List of figures

X. Conclusion

This research demonstrates the effectiveness of Convolutional Neural Networks in classifying 5G network slices into eMBB, mMTC, and URLLC categories based on simulated network parameters. Utilizing synthetic data representing key performance indicators such as latency, throughput, signal strength, and user density, our CNN implementation achieved a classification accuracy of 91.33%. The model demonstrated particularly strong performance in identifying URLLC and mMTC slice requirements, highlighting its capability to differentiate characteristic patterns across these network types. However, the lower precision and recall metrics for eMBB classification indicate opportunities for improvement, particularly in addressing class imbalance and feature distribution overlap.

The implementation of advanced visualization techniques, including pair plots and heatmap representations, provided valuable insights into resource utilization patterns and feature interdependencies. These analytical tools validated the model's classification capabilities while illustrating practical applications for resource allocation and network management in operational 5G environments. This research establishes a foundation for continued development in intelligent network management, emphasizing the significant role of machine learning methodologies 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,"

[8] 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.

[9] 5GPPP Architecture Working Group, "View on 5G Architecture," Version 2.0, December 2017.

[10] CellMapper, "Mobile Network Information Database," [Online]. Available: https://www.cellmapper.net (accessed May 10, 2019).

[11] Fatih Municipality Geographic Information System, "Geographic Information System Portal," [Online]. Available: https://gis.fatih.bel.tr/webgis (accessed May 13, 2019).