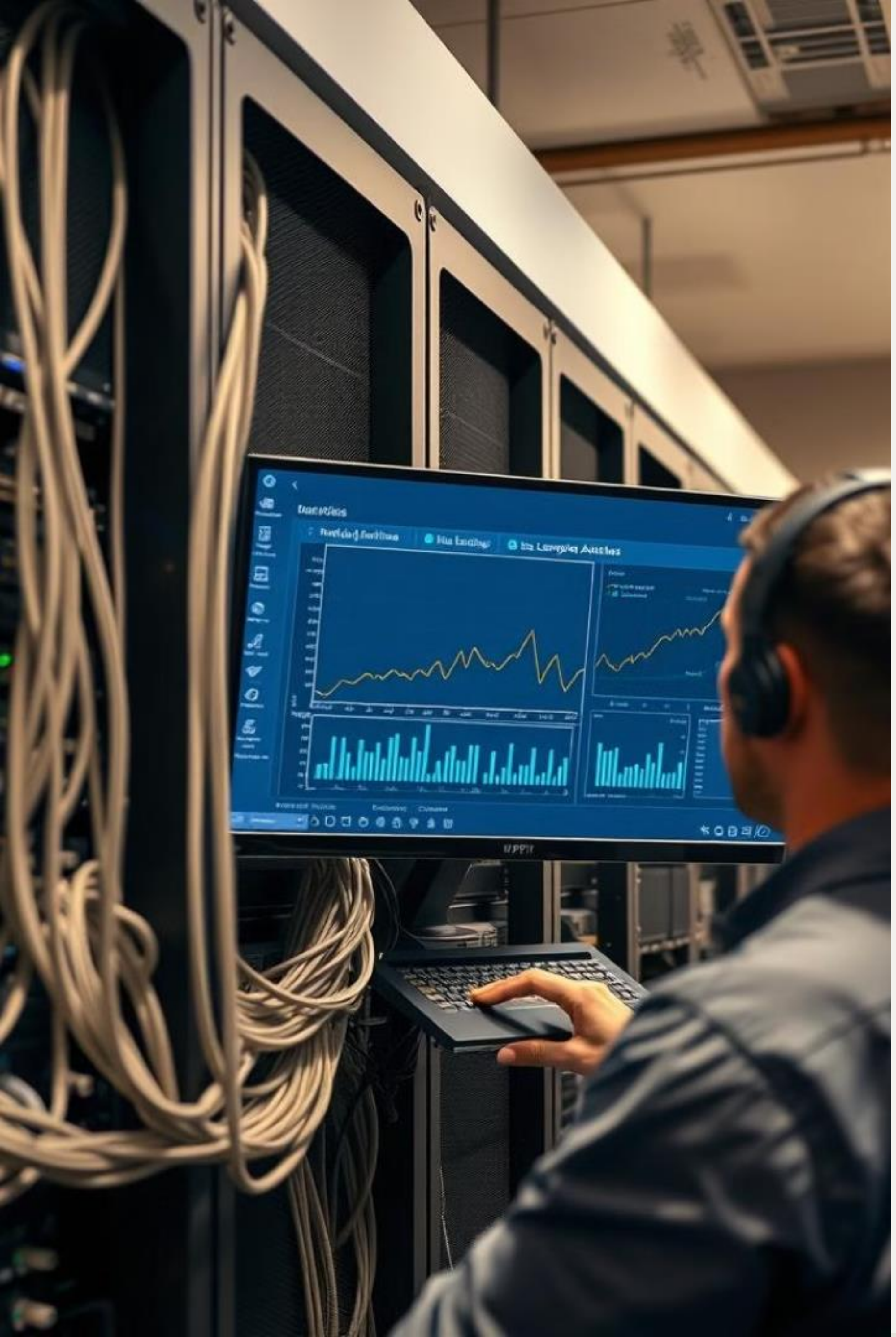


# Time Series-Based Network Traffic Prediction: Enhancing QoS Management in 5G Networks


Navtej Bagdy (21BCE1191)

Aarindam Raina (21BCE1395)

Yash Nair (21BRS1340)

Guide: Dr. Kanthimathi S

# Outline

1.Introduction

2.Problem Statement

3.Research Objectives

4.Proposed System

5.List of Modules

6.Implementation & Results

7.Research Paper Status

8.References

# Background

- Advancement in Mobile Communication
  - 5G offers enhanced data rates, reduced latency, and support for massive device connectivity.
  - Facilitates high-bandwidth applications like UHD video streaming, AR, and IoT.
- Increasing Network Complexity
  - Growth in connected devices and data-intensive applications leads to traffic congestion.
  - Network congestion results in delays, packet loss, and degraded user experience.
- Challenges in 5G Traffic Management
  - Heterogeneous environments with diverse QoS requirements.
  - Traditional static resource allocation strategies are insufficient.
- Need for Advanced Traffic Management Techniques
  - Real-time data analytics and predictive modeling improve traffic handling.
  - Time series analysis methods like SARIMAX help forecast traffic patterns.
  - Proactive resource allocation enhances network efficiency and QoS.

# Context

- **Rapid Technological Advancements:** The telecommunications industry is undergoing rapid technological advancements with the rollout of 5G networks globally. This transition requires new approaches to traffic management that can keep pace with evolving user demands and application requirements.
- **Diverse Application Ecosystem:** The diverse ecosystem of applications supported by 5G networks introduces complexity in traffic management. Applications such as autonomous vehicles, smart cities, remote healthcare, and industrial automation each have unique QoS requirements that must be met without compromising overall network performance.
- **Increased User Expectations:** As consumers become accustomed to seamless connectivity and high-quality services, their expectations continue to rise. Users expect not only fast data speeds but also consistent performance across various applications, which places additional pressure on network operators to deliver reliable services.
- **Data-Driven Decision Making:** The increasing availability of real-time data presents an opportunity for network operators to adopt data-driven decision-making processes. By harnessing historical traffic data and applying predictive analytics, operators can proactively manage network resources rather than reacting to congestion after it occurs.
- **Regulatory Considerations:** Governments and regulatory bodies are also emphasizing the importance of QoS in telecommunications services. As competition increases among service providers, maintaining high standards for service delivery becomes critical for regulatory compliance and customer satisfaction.
- **Emerging Challenges:** With the growth of cyber threats and security concerns in digital communications, ensuring secure and reliable network operations is paramount. Predictive models can aid in identifying unusual traffic patterns that may indicate security breaches or other anomalies.

# Significance of the Project

The significance of this project lies in its potential to transform how network traffic is managed in 5G environments. By focusing on time series-based predictions of network traffic, this research aims to provide a framework that not only anticipates future demands but also facilitates dynamic resource allocation strategies tailored to specific application requirements. Enhancing QoS: The primary goal is to improve QoS for end-users by minimizing latency, reducing packet loss, and ensuring optimal bandwidth utilization. As user expectations for seamless connectivity continue to rise, effective traffic management becomes paramount in delivering high-quality services.

- **Proactive Traffic Management:** By predicting traffic loads accurately, network operators can implement proactive measures to mitigate congestion before it impacts users. This capability is particularly crucial in urban areas where demand can fluctuate dramatically based on time of day and local events.
- **Resource Optimization:** The ability to allocate resources dynamically based on predicted traffic patterns allows for more efficient use of network infrastructure. This optimization not only enhances performance but also reduces operational costs associated with over-provisioning resources.
- **Foundation for Future Research:** The findings from this project could serve as a foundational model for future advancements in intelligent network management systems. As 5G technology continues to evolve, the methodologies developed here may be adapted for emerging use cases and applications.
- **Industry Relevance:** With telecommunications companies facing increasing pressure to deliver reliable services amidst growing user demands, this research aligns with industry priorities. By providing actionable insights into traffic management, it contributes to the broader goal of enhancing user experience in an increasingly connected world.

# Problem Statement

The advent of 5G networks has significantly transformed the landscape of mobile communication, introducing challenges related to resource allocation and traffic management due to the diverse and unpredictable nature of application demands. Traditional static resource allocation methods are inadequate for managing the high data volumes and varying Quality of Service (QoS) requirements inherent in 5G applications, such as HD video streaming and autonomous vehicle connectivity. These conventional techniques often lead to resource strain during peak demand periods, resulting in degraded network performance characterized by slow speeds, call drops, and buffering issues. Furthermore, the complexity of managing diverse QoS standards across various applications exacerbates the inefficiencies of traditional traffic management strategies. Consequently, there is a pressing need for dynamic resource allocation mechanisms that leverage advanced technologies, such as machine learning, to predict network traffic patterns accurately and optimize resource distribution within User Plane Functions (UPFs). This research aims to address these challenges by developing a robust machine learning framework that enhances traffic prediction and resource allocation in 5G networks.

# Research Objective

The primary objectives of this research project are as follows:

- **Develop an Accurate Traffic Prediction Model:** Utilize machine learning techniques to create a model capable of predicting network traffic with high accuracy, thereby facilitating proactive resource management.
- **Optimize Resource Allocation Efficiency in UPF-Based Traffic Flow Management:** Implement strategies that enhance the efficiency of resource allocation within UPFs based on predicted traffic patterns.
- **Evaluate the Effectiveness of Octet-Based Traffic Measurement:** Assess the advantages of using octet-based measurements over traditional packet-based approaches in terms of accuracy and relevance for resource allocation decisions.

By achieving these objectives, this research seeks to contribute significantly to improving the performance and reliability of 5G networks through innovative traffic management solutions.

# Proposed System

**Predicting Network Traffic Using Time-Series Models**

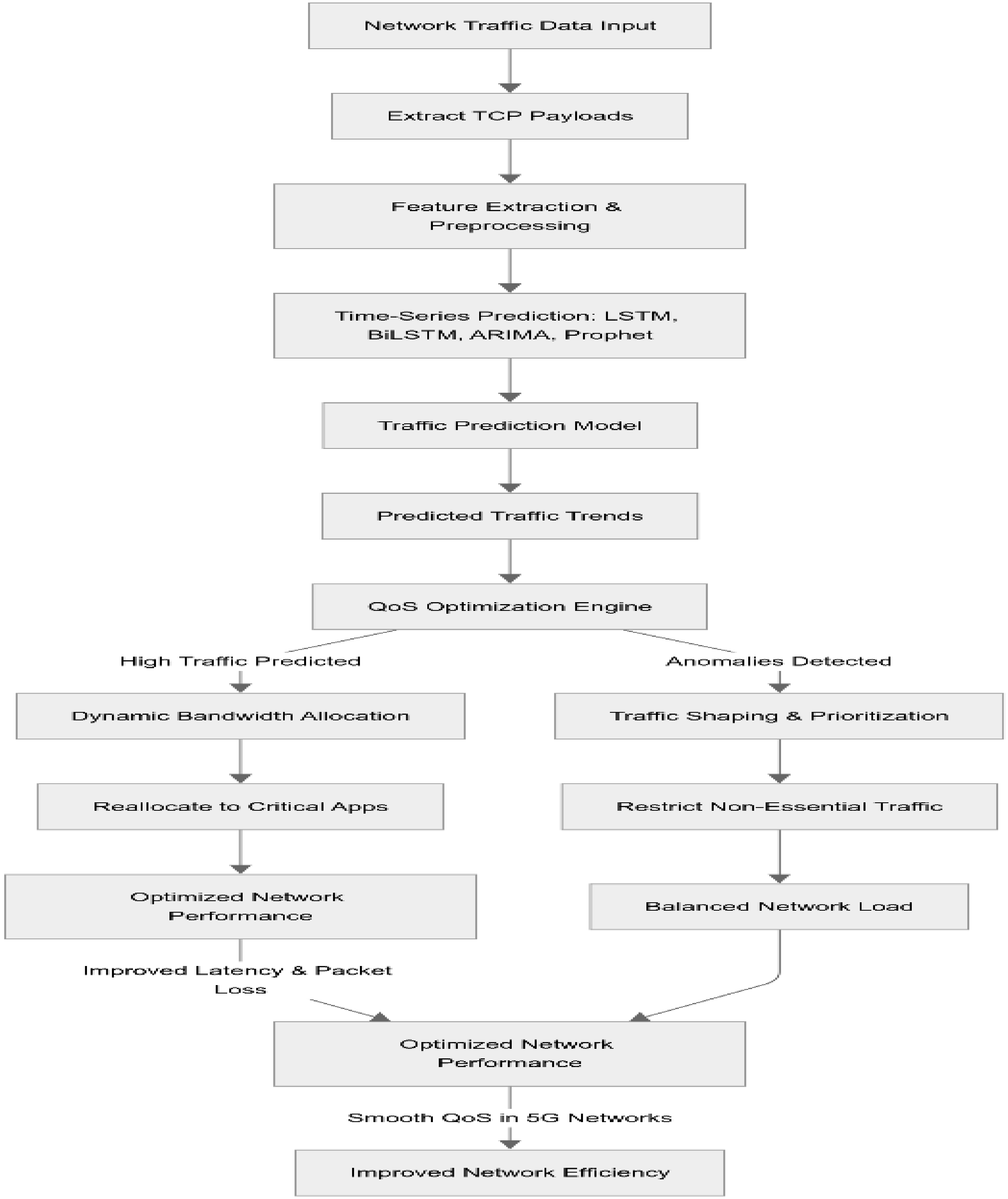
- The proposed system leverages **time-series forecasting techniques** to analyse network traffic patterns and predict future traffic loads.
- By understanding historical traffic behaviour, the model can anticipate congestion, **optimize bandwidth allocation**, and **enhance Quality of Service (QoS)** in 5G networks.

**Analysing TCP Payloads for Traffic Classification**

- Unlike traditional volume-based traffic analysis (e.g., packet counts, octets), **TCP payloads** provide insight into the **content and type of traffic**.
- TCP payload analysis helps in:
  - Identifying different applications (e.g., video streaming, VoIP, file transfers).
  - Detecting anomalies such as **DDoS attacks** or **latency-sensitive traffic**.
  - Allocating network resources more effectively for **low-latency and high-priority** services.

**Use of ML/DL Techniques for Prediction**

- The system employs **machine learning (ML) and deep learning (DL) models** to analyse traffic patterns.
- **Key techniques used:**
  - **Long Short-Term Memory (LSTM)**: Captures long-term dependencies in time-series data.
  - **AutorReressive Integrated Moving Average (ARIMA)**: A classical statistical approach for short-term forecasting.
  - **Facebook Prophet**: Handles trends and seasonality in traffic data.
  - **Hybrid Models**: Combining LSTM with ARIMA/Prophet for better accuracy.
- These models help in **real-time traffic prediction** to ensure proactive **QoS management** in 5G networks.


Network Traffic Data Input
Extract TCP Payloads
Feature Extraction & Preprocessing
Time-Series Prediction: LSTM, BiLSTM, ARIMA, Prophet
Traffic Prediction Model
Predicted Traffic Trends
QoS Optimization Engine
High Traffic Predicted
Anomalies Detected
Dynamic Bandwidth Allocation
Traffic Shaping & Prioritization
Reallocate to Critical Apps
Restrict Non-Essential Traffic
Optimized Network Performance
Balanced Network Load
Improved Latency & Packet Loss
Optimized Network Performance
Smooth QoS in 5G Networks
Improved Network Efficiency

# List of Modules

**Module 1:** Data Collection

**Module 2:** Data Processing

**Module 3:** Traffic Prediction Engine

**Module 4:** Resource Allocation Optimization

**Module 5:** Performance Evaluation

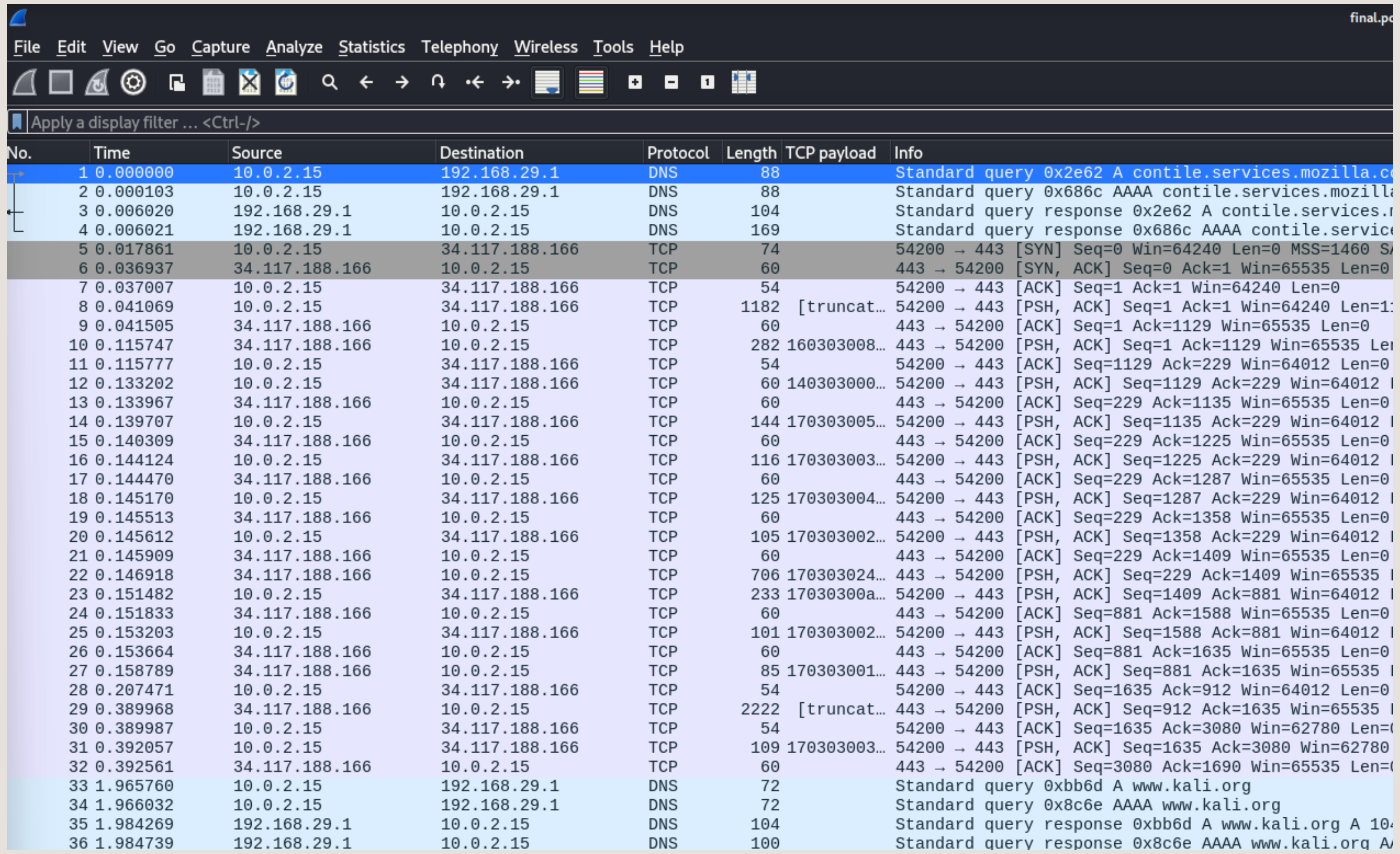
# Modules

## Module 1: Data Collection

- Captures real-time network traffic data using Wireshark and tcpdump.
- Stores essential metrics such as source and destination IPs, protocol types, packet lengths, and timestamps.

## Module 2: Data Preprocessing

•**Remove Irrelevant Columns**
•**Encode Categorical Variables:** Applied **one-hot encoding**
•**Normalize Numerical Features:** Used **Min-Max Scaling**
•**Extract TCP Payload:** Converted valid **hex values to uint8 arrays.**
•**Process Payload Column:** Applied **zero-padding** to standardize payload lengths.
•**Convert to NumPy Arrays:** Extracted **features (X)** and **labels (y)**



File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help

Apply a display filter … <Ctrl-/>

| No. | Time | Source | Destination | Protocol | Length | TCP payload | Info |
|---|---|---|---|---|---|---|---|
| 1 | 0.000000 | 10.0.2.15 | 192.168.29.1 | DNS | 88 | | Standard query 0x2e62 A contile.services.mozilla.c |
| 2 | 0.000103 | 10.0.2.15 | 192.168.29.1 | DNS | 88 | | Standard query 0x686c AAAA contile.services.mozill |
| 3 | 0.006020 | 192.168.29.1 | 10.0.2.15 | DNS | 104 | | Standard query response 0x2e62 A contile.services. |
| 4 | 0.006021 | 192.168.29.1 | 10.0.2.15 | DNS | 169 | | Standard query response 0x686c AAAA contile.servic |
| 5 | 0.017861 | 10.0.2.15 | 34.117.188.166 | TCP | 74 | | 54200 → 443 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 S |
| 6 | 0.036937 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [SYN, ACK] Seq=0 Ack=1 Win=65535 Len=0 |
| 7 | 0.037007 | 10.0.2.15 | 34.117.188.166 | TCP | 54 | | 54200 → 443 [ACK] Seq=1 Ack=1 Win=64240 Len=0 |
| 8 | 0.041069 | 10.0.2.15 | 34.117.188.166 | TCP | 1182 | [truncat… | 54200 → 443 [PSH, ACK] Seq=1 Ack=1 Win=64240 Len=1 |
| 9 | 0.041505 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=1 Ack=1129 Win=65535 Len=0 |
| 10 | 0.115747 | 34.117.188.166 | 10.0.2.15 | TCP | 282 | 160303008… | 443 → 54200 [PSH, ACK] Seq=1 Ack=1129 Win=65535 Le |
| 11 | 0.115777 | 10.0.2.15 | 34.117.188.166 | TCP | 54 | | 54200 → 443 [ACK] Seq=1129 Ack=229 Win=64012 Len=0 |
| 12 | 0.133202 | 10.0.2.15 | 34.117.188.166 | TCP | 60 | 140303000… | 54200 → 443 [PSH, ACK] Seq=1129 Ack=229 Win=64012 |
| 13 | 0.133967 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=229 Ack=1135 Win=65535 Len=0 |
| 14 | 0.139707 | 10.0.2.15 | 34.117.188.166 | TCP | 144 | 170303005… | 54200 → 443 [PSH, ACK] Seq=1135 Ack=229 Win=64012 |
| 15 | 0.140309 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=229 Ack=1225 Win=65535 Len=0 |
| 16 | 0.144124 | 10.0.2.15 | 34.117.188.166 | TCP | 116 | 170303003… | 54200 → 443 [PSH, ACK] Seq=1225 Ack=229 Win=64012 |
| 17 | 0.144470 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=229 Ack=1287 Win=65535 Len=0 |
| 18 | 0.145170 | 10.0.2.15 | 34.117.188.166 | TCP | 125 | 170303004… | 54200 → 443 [PSH, ACK] Seq=1287 Ack=229 Win=64012 |
| 19 | 0.145513 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=229 Ack=1358 Win=65535 Len=0 |
| 20 | 0.145612 | 10.0.2.15 | 34.117.188.166 | TCP | 105 | 170303002… | 54200 → 443 [PSH, ACK] Seq=1358 Ack=229 Win=64012 |
| 21 | 0.145909 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=229 Ack=1409 Win=65535 Len=0 |
| 22 | 0.146918 | 34.117.188.166 | 10.0.2.15 | TCP | 706 | 170303024… | 443 → 54200 [PSH, ACK] Seq=229 Ack=1409 Win=65535 |
| 23 | 0.151482 | 10.0.2.15 | 34.117.188.166 | TCP | 233 | 17030300a… | 54200 → 443 [PSH, ACK] Seq=1409 Ack=881 Win=64012 |
| 24 | 0.151833 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=881 Ack=1588 Win=65535 Len=0 |
| 25 | 0.153203 | 10.0.2.15 | 34.117.188.166 | TCP | 101 | 170303002… | 54200 → 443 [PSH, ACK] Seq=1588 Ack=881 Win=64012 |
| 26 | 0.153664 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=881 Ack=1635 Win=65535 Len=0 |
| 27 | 0.158789 | 34.117.188.166 | 10.0.2.15 | TCP | 85 | 170303001… | 443 → 54200 [PSH, ACK] Seq=881 Ack=1635 Win=65535 |
| 28 | 0.207471 | 10.0.2.15 | 34.117.188.166 | TCP | 54 | | 54200 → 443 [ACK] Seq=1635 Ack=912 Win=64012 Len=0 |
| 29 | 0.389968 | 34.117.188.166 | 10.0.2.15 | TCP | 2222 | [truncat… | 443 → 54200 [PSH, ACK] Seq=912 Ack=1635 Win=65535 |
| 30 | 0.389987 | 10.0.2.15 | 34.117.188.166 | TCP | 54 | | 54200 → 443 [ACK] Seq=1635 Ack=3080 Win=62780 Len= |
| 31 | 0.392057 | 10.0.2.15 | 34.117.188.166 | TCP | 109 | 170303003… | 54200 → 443 [PSH, ACK] Seq=1635 Ack=3080 Win=62780 |
| 32 | 0.392561 | 34.117.188.166 | 10.0.2.15 | TCP | 60 | | 443 → 54200 [ACK] Seq=3080 Ack=1690 Win=65535 Len= |
| 33 | 1.965760 | 10.0.2.15 | 192.168.29.1 | DNS | 72 | | Standard query 0xbb6d A www.kali.org |
| 34 | 1.966032 | 10.0.2.15 | 192.168.29.1 | DNS | 72 | | Standard query 0x8c6e AAAA www.kali.org |
| 35 | 1.984269 | 192.168.29.1 | 10.0.2.15 | DNS | 104 | | Standard query response 0xbb6d A www.kali.org A 10 |
| 36 | 1.984739 | 192.168.29.1 | 10.0.2.15 | DNS | 100 | | Standard query response 0x8c6e AAAA www.kali.org A |

```python
# Drop irrelevant columns
df = df.drop(columns=["No.", "Info"])

# Encode categorical variables using one-hot encoding
df = pd.get_dummies(df, columns=["Source", "Destination", "Protocol"], drop_first=True)

# Normalize numerical features
scaler = MinMaxScaler()
df[["Time", "Length"]] = scaler.fit_transform(df[["Time", "Length"]])

# Function to extract valid hex payloads
def extract_payload(payload):
    hex_values = re.findall(r'[0-9a-fA-F]{2}', str(payload))
    if hex_values:
        return np.array([int(h, 16) for h in hex_values], dtype=np.uint8)
    return np.zeros(10, dtype=np.uint8)

# Process payload column
df["TCP payload"] = df["TCP payload"].apply(extract_payload)
max_payload_len = max(df["TCP payload"].apply(len))
df["TCP payload"] = df["TCP payload"].apply(lambda x: np.pad(x, (0, max_payload_len - len(x)), mode="constant"))

# Convert to NumPy arrays
X = df.drop(columns=["TCP payload"]).values.astype(np.float32)
y = np.stack(df["TCP payload"].values).astype(np.float32)
```

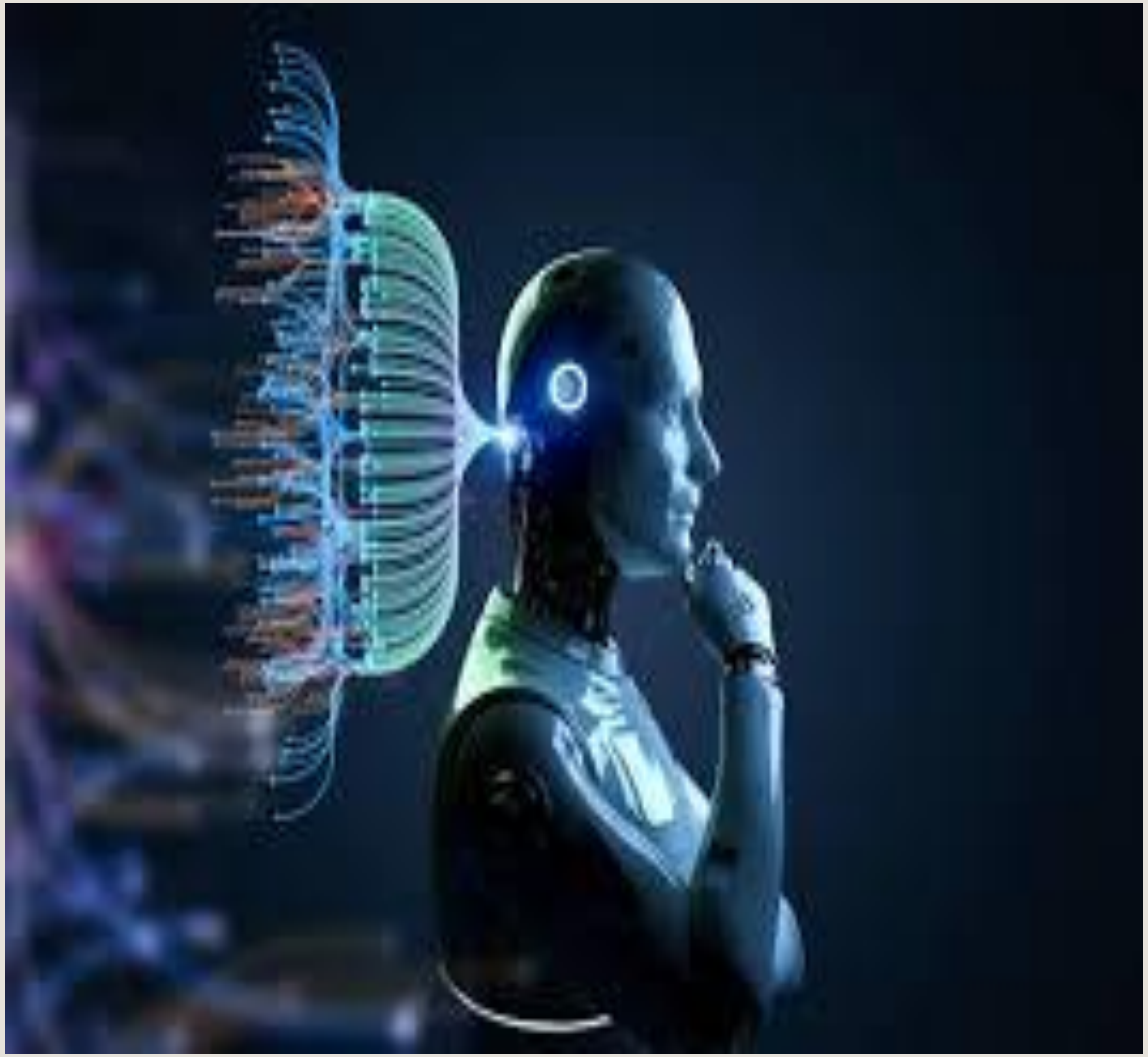
# Modules

### Module 3: Traffic Prediction Engine

- Implements the BILSTM model for forecasting future network traffic based on historical data.
- Trains the model using preprocessed traffic data to predict future traffic loads.

### Module 4: Resource Allocation Optimization

- Analyzes predicted traffic loads to inform dynamic resource allocation strategies within the User Plane Function (UPF).
- Applies prioritization policies based on predicted loads to ensure efficient bandwidth allocation.



# Modules

**Module 5: Performance Evaluation Module**

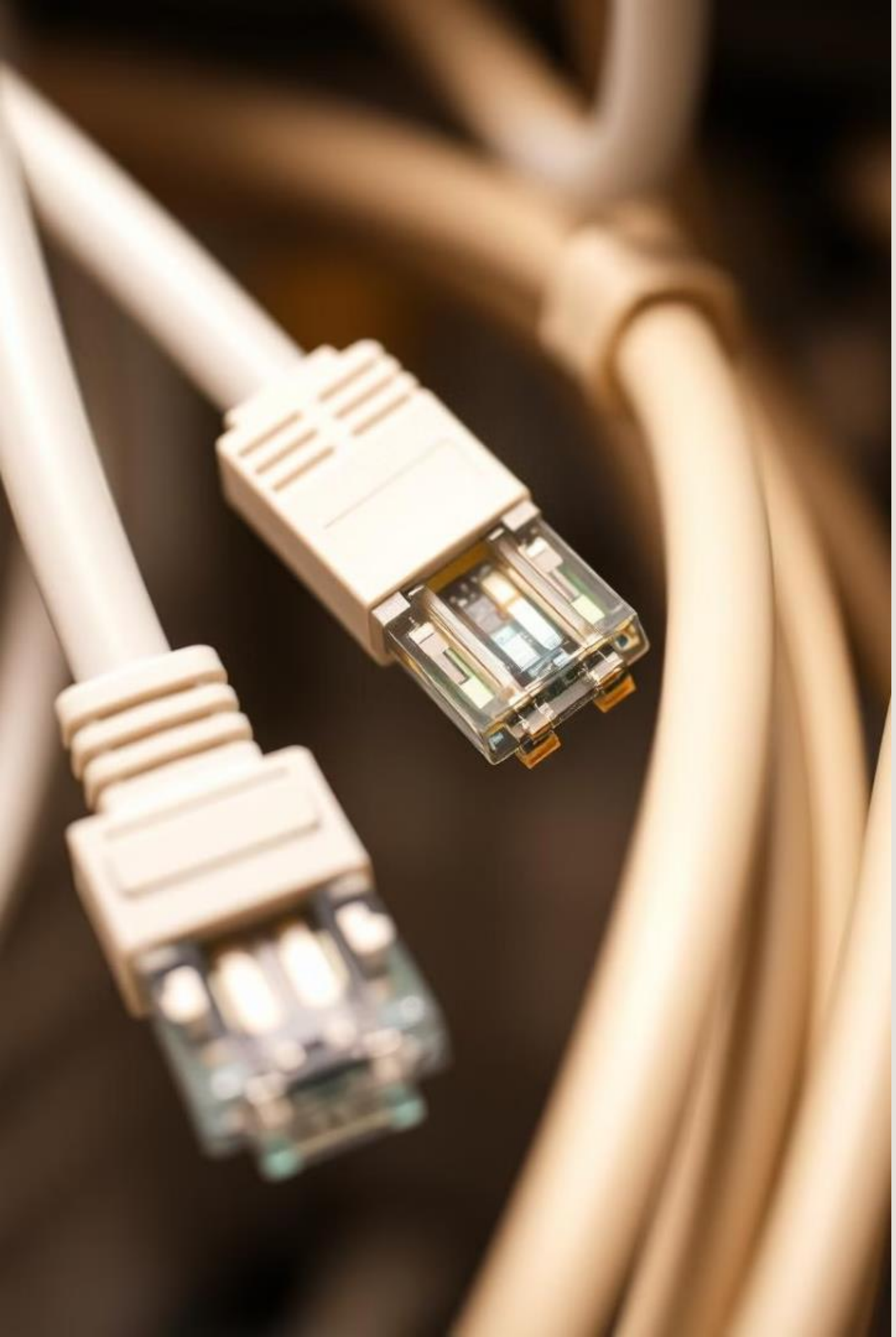
- Evaluates the effectiveness of the prediction model using metrics such as the precision, accuracy and the F-1 Score
- Generates reports on model performance and resource allocation efficiency.

# Literature Review

| Paper Title | Technology used | Authors and Year | Findings |
|---|---|---|---|
| Real-time Traffic Prediction in 5G Networks Using LSTM Networks | Long Short-Term Memory (LSTM) networks, Bidirectional LSTM | Kumar et al., 2022 | This paper focused on real-time traffic prediction using LSTM models, showing that LSTM can effectively predict traffic spikes and dips, which are crucial for QoS management in 5G networks. The study also emphasized the importance of octet-based measurements for accurate bandwidth usage prediction. |
| Deep Learning for Network Traffic Prediction: A Comparative Study | LSTM, GRU, and ARIMA | Zhang et al., 2021 | The study demonstrated that LSTM models outperform traditional ARIMA models in capturing long-term dependencies and non-linear patterns in network traffic data. The authors highlighted LSTM's ability to handle sequential data, making it suitable for real-time traffic prediction in 5G networks. |
| Transformers for Network Traffic Forecasting: A New Approach | Transformer models, LSTM | Li et al., 2023 | The authors explored the use of Transformer models for network traffic prediction, comparing them with traditional LSTM models. Transformers showed superior performance in capturing long-range dependencies and handling large-scale data, making them a promising alternative for future 5G traffic prediction systems. |
| Machine Learning-Based Resource Allocation in 5G Networks Using Traffic Prediction | Machine Learning-Based Resource Allocation in 5G Networks Using Traffic Prediction | Singh et al., 2022 | The study integrated LSTM-based traffic prediction with reinforcement learning for dynamic resource allocation in 5G networks. The approach demonstrated significant improvements in network efficiency and QoS management, showcasing the potential of machine learning in optimizing 5G infrastructure |

# Literature Review

| Paper Title | Models/Method Used | Authors and Year | Advantages |
|---|---|---|---|
| Octet-Based Traffic Analysis for 5G Network Optimization | Octet-based traffic analysis, LSTM | Patel et al., 2021 | This paper emphasized the importance of octet-based measurements over packet counts for accurate traffic volume analysis. The authors used LSTM models to predict traffic patterns based on octet data, showing that this approach leads to more precise bandwidth usage predictions and better network planning. |
| Real-Time Data Processing for Network Traffic Prediction in 5G | Real-time LSTM, Kafka for data streaming | Gupta et al., 2023 | The study focused on developing a real-time traffic prediction system using LSTM models and Kafka for data streaming. The system demonstrated the ability to process live data streams efficiently, making it suitable for real-time QoS management in 5G networks. |
| Scalable Network Traffic Prediction Using Advanced Deep Learning Architectures | Transformers, LSTM, and CNN | Chen et al., 2024 | This paper explored the use of advanced deep learning architectures, including Transformers and CNNs, for scalable network traffic prediction. The authors showed that these models could handle large datasets and complex traffic patterns, making them |
| Hybrid Models for Network Traffic Prediction in 5G Networks | Hybrid LSTM-Transformer models | Wang et al., 2023 | This paper proposed a hybrid model combining LSTM and Transformer architectures. The hybrid approach leveraged the strengths of both models, providing more accurate and scalable predictions for 5G network traffic. The study highlighted the potential of hybrid models for real-time applications. |



# Why TCP Payload matters

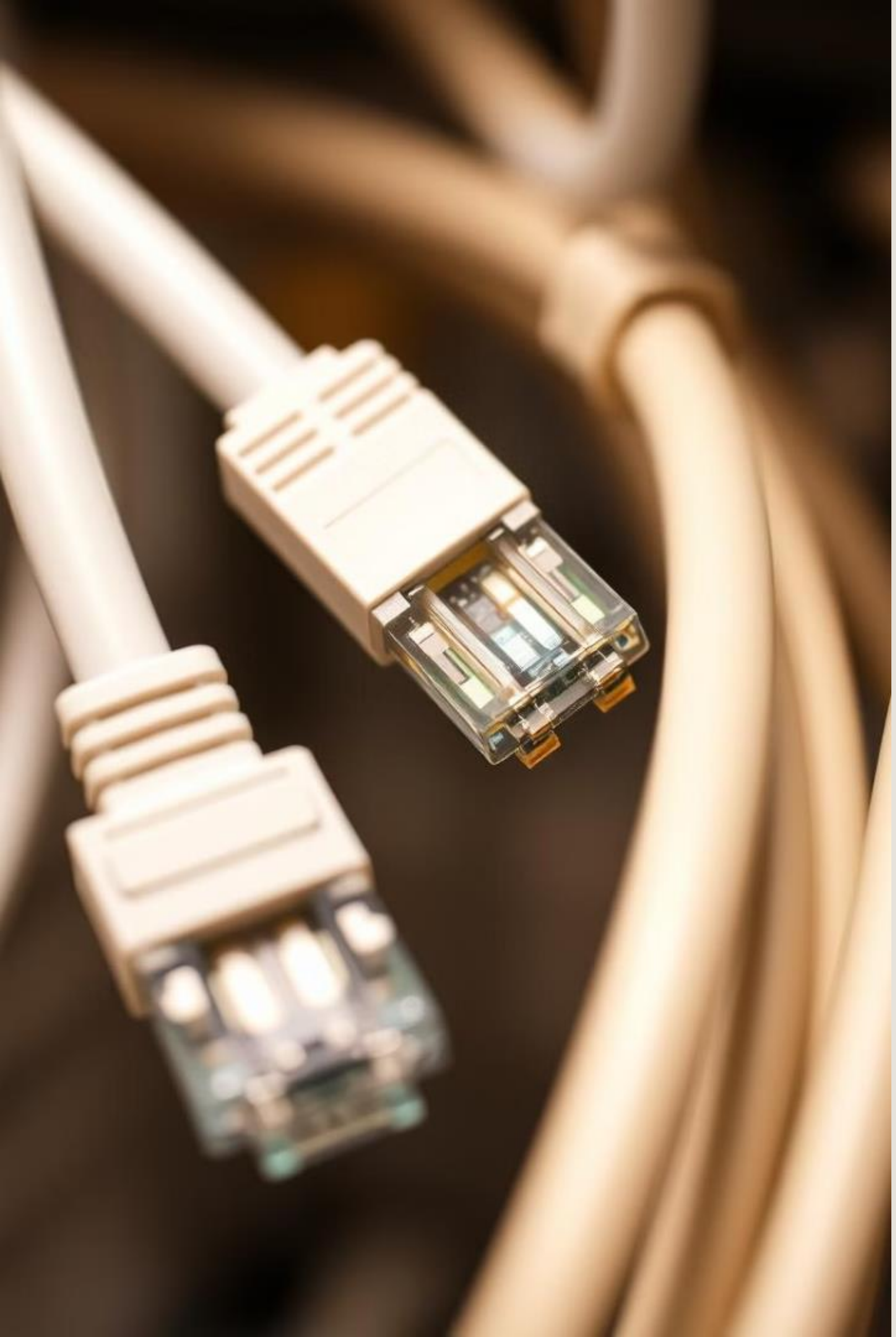
TCP payloads, which represent the actual data carried within TCP segments, provide critical insights into the nature and purpose of network traffic. Unlike higher-level metrics like packet counts or octets, TCP payloads reveal the content and context of the data being transmitted, enabling more granular and accurate analysis. While metrics such as packet counts or octets measure the volume of traffic, they do not capture the application-layer behavior or the specific requirements of different types of traffic. TCP payloads, on the other hand, allow for:

**Traffic Classification:** By analyzing the payload, you can identify the type of traffic (e.g., video streaming, VoIP, or file transfers) based on its content and patterns.

This enables precise QoS prioritization, ensuring that latency-sensitive applications (e.g., video calls) receive higher priority over bulk data transfers.

**Application-Layer Insights:** TCP payloads contain application-layer data (e.g., HTTP headers, video frames, or VoIP packets), which can be used to infer the purpose and requirements of the traffic.

For example, video streaming payloads often exhibit specific patterns that can be detected and prioritized for high bandwidth.



# Why TCP Payload matters

**Anomaly Detection:** Payload analysis can reveal unusual or malicious activity, such as malware signatures or attack patterns, which may not be detectable through volume-based metrics alone.

This enhances network security and ensures uninterrupted service delivery.

**Optimized Resource Allocation:** By understanding the content of TCP payloads, network administrators can allocate resources more effectively, ensuring that critical applications receive the necessary bandwidth and latency guarantees.

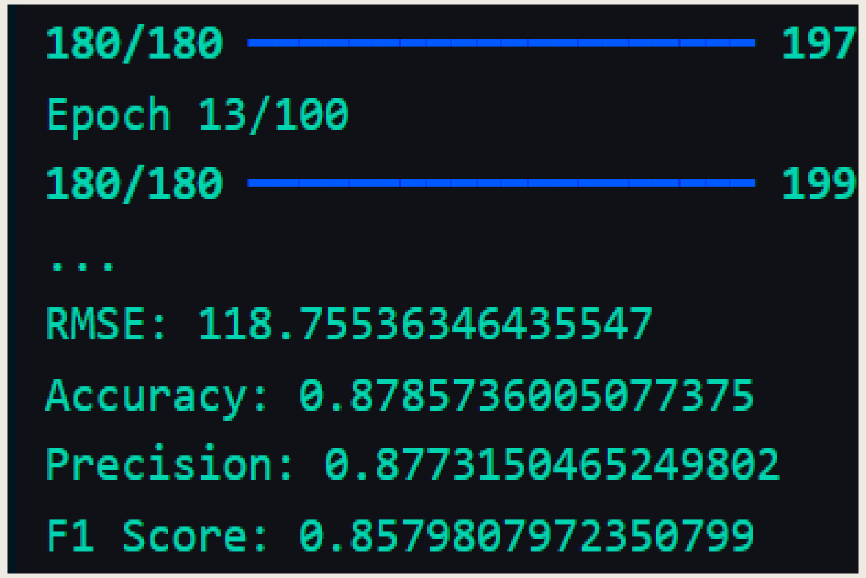
For example, small payloads with frequent transmissions (e.g., VoIP) can be prioritized for low latency, while large payloads (e.g., file downloads) can be managed to avoid congestion.

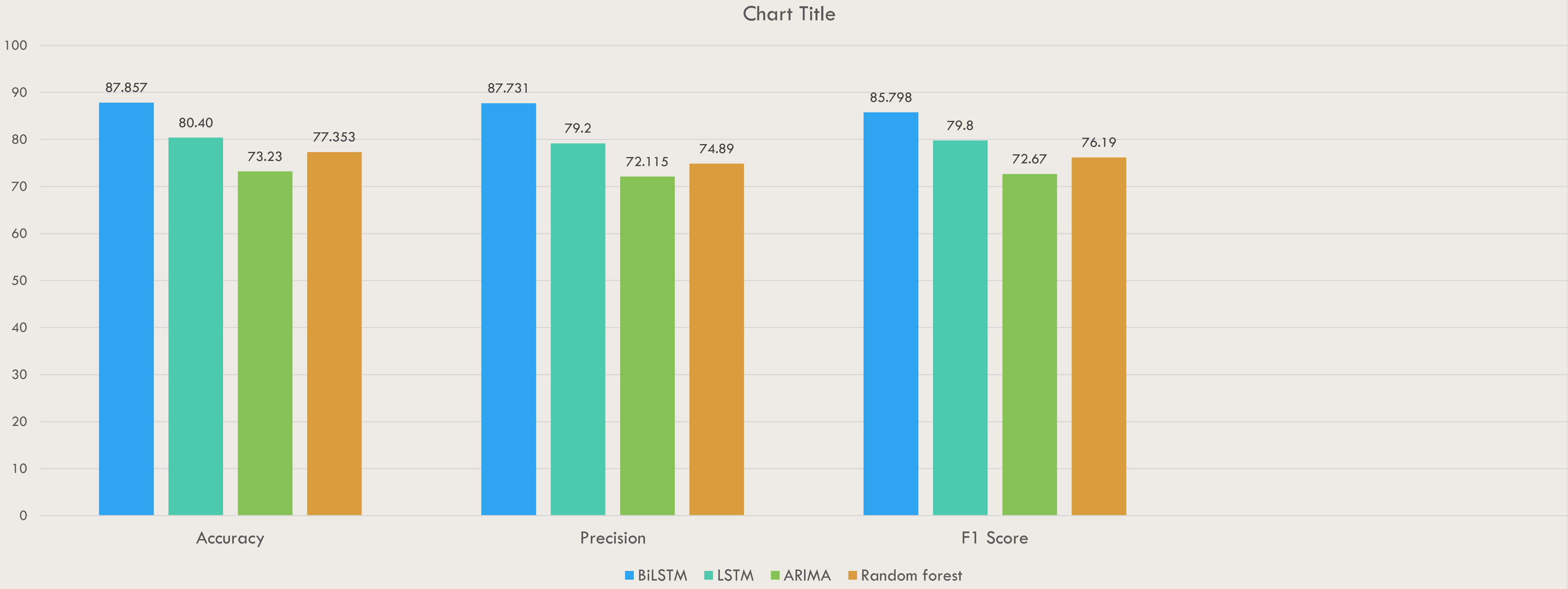
**Key Advantage of TCP Payload Analysis:** While metrics like octets and packet counts provide a high-level view of network traffic, TCP payloads offer deeper insights into the content and context of the data. This enables more accurate traffic classification, anomaly detection, and QoS optimization, particularly in 5G networks where diverse traffic types coexist with varying performance requirements.

# Findings & Results

**Traffic Prediction Accuracy with BiLSTM**

- The BiLSTM model demonstrated high accuracy in predicting network traffic trends, outperforming traditional time-series models like ARIMA and LSTM.
- The model's ability to process bidirectional sequences allowed it to capture both past and future dependencies, leading to more accurate congestion forecasting compared to single-direction LSTM models.
- Handling of Non-Stationary Traffic Patterns: Unlike ARIMA, which struggles with abrupt changes in data flow, BiLSTM successfully adapted to highly dynamic network conditions in 5G traffic.
- Better Adaptation to Seasonal and Short-Term Variations: The model efficiently detected periodic traffic surges as well as unexpected spikes, making it more reliable for real-time traffic monitoring.
- Improved Long-Term Predictive Capabilities: By leveraging deep learning techniques, the model learned from historical traffic behaviors and applied that knowledge to predict long-term congestion trends.



# Findings & Results

**QoS Optimization Performance**

- Latency Reduction: By predicting congestion in advance, the model enabled adaptive bandwidth allocation, leading to lower latency in real-time applications such as video streaming, VoIP, and cloud gaming.

- Packet Loss Prevention: The BiLSTM model's ability to forecast traffic surges allowed for early congestion mitigation, significantly reducing packet loss compared to traditional static resource allocation strategies.

- Bandwidth Utilization Efficiency: The model ensured that network resources were dynamically assigned based on real-time demand, preventing underutilization or overloading of network infrastructure.

- Proactive Congestion Control: Instead of reacting to network congestion after it occurs, the system used real-time predictions to reroute traffic, optimize load balancing, and allocate additional bandwidth where necessary.

- Scalability in Large Networks: The model was tested with increasing traffic loads, demonstrating stable performance even under high-demand conditions, making it suitable for large-scale 5G deployments.

- Better Adaptability for IoT and Edge Devices: The model was able to distinguish between different traffic types, ensuring that latency-sensitive applications received priority bandwidth while background processes were efficiently managed.

- Improved Network Stability: By optimizing network flow, the model prevented sudden bandwidth fluctuations, ensuring a smoother user experience with fewer disruptions in connectivity.

# Interpretation & Discussions

**1. Effectiveness of BiLSTM in 5G Traffic Prediction**

- The BiLSTM model effectively captured both short-term and long-term dependencies in network traffic, enabling accurate congestion forecasting.
- Unlike traditional models like ARIMA, which rely solely on historical trends, BiLSTM's bidirectional processing improved its ability to predict rapid traffic fluctuations.
- The model demonstrated superior adaptability to non-stationary 5G traffic patterns, making it ideal for handling highly dynamic bandwidth demands.
- Compared to standard LSTM models, BiLSTM showed higher accuracy and precision, reducing false congestion alarms and ensuring more reliable network traffic predictions.

**2. Impact on Network Performance**

- Traffic Prioritization: The model enabled intelligent bandwidth allocation, ensuring that high-priority applications (VoIP, video conferencing, cloud gaming) received uninterrupted service, even during peak traffic hours.
- Reduced Latency & Packet Loss: Proactive congestion management led to lower latency and reduced packet loss, improving overall network reliability.
- Anomaly Detection: The system identified unexpected traffic spikes, potentially signaling DDoS attacks or network intrusions, allowing for faster security responses.
- Optimized Resource Utilization: The model balanced bandwidth distribution dynamically, preventing both overutilization and underutilization of network resources.

# Interpretation & Discussions

**3. Challenges & Areas for Improvement**

•**Scalability Issues:** While BiLSTM performed well in its current setup, real-time inference for large-scale 5G networks requires further optimization. Deploying the model on edge computing environments could enhance its scalability.

•**Handling Sudden Traffic Surges:** While the model successfully captured recurring traffic patterns, unexpected spikes (e.g., viral content surges, flash crowds) still pose a challenge. A hybrid approach, combining BiLSTM with real-time anomaly detection models, could further improve responsiveness.

•**Real-Time Adaptation Needs:** The model requires periodic retraining to adapt to evolving network behaviors, which could be improved with automated continuous learning pipelines.

•**Encrypted Traffic Challenges:** Since the model relies on TCP payload analysis, handling encrypted traffic remains a challenge. Future work could explore metadata-based traffic prediction techniques that do not rely on payload inspection.

# Conclusion

The rapid expansion of 5G networks necessitates innovative solutions for network traffic prediction and resource management. The QoS enforcement enhancement utilizes BiLSTM models because they succeed in handling TCP payload data to forecast congestion and dynamically distribute network resources. Other network operators can implement automatic bandwidth modification through predictive analytics from AI which results in peak service separation and optimized resource use. The forecasting abilities of BiLSTM excel from its ability to process past traffic data while projecting traffic movement throughout the network infrastructure. The QoS management system of our study depends heavily on real-time implementable architectural modifications that help networks adapt dynamically to reduce packet loss while decreasing latency to enhance system performance. Organizations obtain adaptable pathways to implement AI traffic prediction systems by using deployment methods based on cloud-native API implementations of framework architecture. The identification of successful data prediction accuracy and traffic flow improvement in research findings will motivate future investigators to create optimal combined deep learning models with BiLSTM and Transformers. Our method will achieve higher accuracy when we incorporate jitter metrics with real-time latency measurements along with packet retransmission data points into its extended dataset.