Empowering Mobile Network Planning through Deep Learning: A Path to Democratization

by

Syed Tauhidun Nabi
20366009

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
M.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
September 2023

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

tauhid nabi

Syed Tauhidun Nabi
20366009

Approval

The thesis titled "Empowering Mobile Network Planning through Deep Learning: A Path to Democratization" submitted by

1. Syed Tauhidun Nabi (20366009)

Of Summer, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of M.Sc. in Computer Science on September 08, 2023.

Examining Committee:

Supervisor:
(Member)

Prof. Md. Golam Rabiul Alam, Ph.D.
Professor
Department of Computer Science and Engineering
Brac University
rabiul.alam@bracu.ac.bd

Examiner:
(External)

Dr. Ashis Talukder
Associate Professor
Department of Management Information Systems
University of Dhaka
ashis@du.ac.bd

Examiner:
(Internal)

Dr. Muhammad Iqbal Hossain
Associate Professor
Department of Computer Science and Engineering
Brac University
iqbal.hossain@bracu.ac.bd

Examiner:
(Internal)

Dr. Md. Ashraful Alam
Associate Professor
Department of Computer Science and Engineering
Brac University
ashraful.alam@bracu.ac.bd

Program Coordinator:
(Member)

Prof. Amitabha Chakrabarty, Ph.D.
Professor
Department of Computer Science and Engineering
Brac University
amitabha@bracu.ac.bd

Chairperson:
(Chair)

Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University
skazi@bracu.ac.bd

Abstract

In the realm of cellular network internet data traffic assessment, the imperative task of forecasting and comprehending traffic patterns assumes pivotal significance for the effective management of network-designed Quality of Service (QoS) benchmarks. Conventional methodologies employed for predicting data traffic often suffer from inaccuracies. These traditional traffic forecasts, typically conducted at a higher-level or within generously sized regional cluster contexts, tend to exhibit limitations in terms of accuracy. Furthermore, the absence of readily accessible eNodeB-level utilization data in conjunction with traffic forecasting exacerbates these challenges. This, in turn, may lead to compromised user experiences or unwarranted network expansion decisions based on outdated methodologies. This research embarks upon an ambitious journey encompassing an extensive dataset encompassing 6.2 million real network time series data points derived from Long-Term Evolution (LTE) networks. It also delves into associated parameters, including eNodeB-wise Physical Resource Block (PRB) utilization. The core objective revolves around the development of a traffic forecasting model that harnesses multivariate feature inputs and cutting-edge deep learning algorithms. Various advanced deep learning algorithms, including Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Gated Recurrent Unit (GRU), have been separately tested for training purposes, with the most suitable model being chosen among the three for eNodeB-level predictions. This state-of-the-art deep learning model not only enables highly granular eNodeB-level traffic forecasting but also provides insights into anticipated eNodeB-wise PRB utilization. The selected optimal deep learning model, BiLSTM, achieves a robust R^2 score of 0.793, notably surpassing the performance of the other deep learning algorithms. Beyond the realm of PRB utilization, the study establishes a Quality of Service (QoS) threshold at 70% – a benchmark rooted in real network experience. This threshold serves as a pivotal trigger for decisions pertaining to soft parameter tuning. Leveraging the projected PRB utilization, the research introduces a pioneering algorithm designed to estimate eNodeB-level soft capacity parameter optimization. This algorithm empowers network operators to address short-term capacity enhancement solutions as well as long-term network expansion, all aimed at maintaining steadfast QoS benchmarks. Situated within the context of network planning, this study not only unravels the intricate dynamics of cellular data traffic but also catalyzes the concept of democratization. By harnessing the capabilities of deep learning, network operators are equipped with potent tools to navigate the intricate landscape of network optimization. Through this research endeavor, strides are made toward an envisioned future where technological advancements seamlessly converge with accessibility, thereby reshaping the contours of mobile network planning.

Keywords: LTE networks, network planning, machine learning in networking, traffic prediction, deep learning, mobile network capacity, physical resource block, resource management.

Acknowledgement

I begin by expressing my profound gratitude to the Almighty Allah, whose unwavering guidance enabled me to successfully complete the writing of this thesis, overcoming various challenges along the way.

I extend my heartfelt appreciation to my mentor, Prof. Md. Golam Rabiul Alam, Ph.D., whose invaluable assistance and guidance significantly contributed to the fruition of this work. His constant support was a pillar of strength, always available whenever I sought advice.

I would also like to express my profound thanks to IEEE Access 2023 and the esteemed judging panel for recognizing the significance of our paper, titled "Deep Learning Based Fusion Model for Multivariate LTE Traffic Forecasting and Optimized Radio Parameter Estimation," and granting it acceptance. The opportunity to contribute to this prestigious publication, as evidenced by the DOI: 10.1109/ACCESS.2023.3242861 has been a significant milestone in my academic journey.

I want to express my gratitude to my family and parents for their unwavering support, encouragement, and prayers. Their belief in me has been instrumental in bringing me to the threshold of graduation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The exponential growth of mobile Internet data traffic in recent decades has presented significant challenges for mobile network operators (MNOs). According to the Ericsson Mobility Report (Jan 2022), global mobile network data traffic is projected to reach nearly 300 exabytes per month by 2027 [54]. As the demand for data continues to surge, MNOs face the crucial task of maintaining quality of service (QoS) benchmarks, optimizing resource utilization, managing mobility, and ensuring high availability of the network system. These challenges predominantly fall within the realm of network planning.
One of the key challenges in mobile network planning is the maintenance of QoS benchmarks. As data traffic surpasses the designed capacity of eNodeBs, the QoS can be significantly affected [31]. Moreover, the rising number of users and their increasing demand for high-speed, high-quality mobile internet further exacerbate the QoS challenge. The Ericsson Mobility Report 2021 reveals that the monthly average internet usage per smartphone is approximately 11.4 GB, a figure projected to quadruple by 2027 [54]. Additionally, the diverse behavior of users in different locations imposes an additional burden on LTE networks and cells. For instance, video traffic currently accounts for nearly 70% of all mobile data traffic [54]. Predictions indicate that video traffic will reach approximately 79% by 2027, resulting in a surge of data transfers over mobile networks. Consequently, it is evident that without proper forecasting and proactive network utilization management, MNOs may struggle to meet user demands promptly, leading to potential QoS degradation.
To address these challenges and enhance mobile network planning, emerging technologies such as deep learning offer promising avenues for improvement. Deep learning techniques have demonstrated remarkable capabilities in various domains, including computer vision, natural language processing, and speech recognition. By leveraging deep learning algorithms, mobile network planning can benefit from advanced predictive analytics, intelligent resource allocation, and optimized decision-making processes. The utilization of deep learning in mobile network planning has the potential to democratize the field, empowering MNOs with efficient and effective tools that enhance their network management capabilities.
In this thesis, we explore the potential of deep learning techniques to empower mobile network planning. We investigate the application of deep learning algorithms in addressing the challenges faced by MNOs, particularly in relation to QoS maintenance, resource utilization optimization, and overall network performance. By analyzing and evaluating various deep learning approaches, we aim to contribute to

the advancement of mobile network planning practices and pave the way for a more democratized and efficient future in the field.

1.1 Research Motivation

The significance of deep learning in mobile network planning is multifaceted. Firstly, deep learning algorithms enable accurate traffic forecasting, allowing MNOs to dimension their networks effectively and allocate resources optimally. By leveraging the power of deep learning models, MNOs can make informed decisions regarding the placement of new cells or sites, minimizing capital expenditure and reducing operational costs [45]. Furthermore, deep learning techniques can capture complex patterns and correlations in sequential data, such as network traffic trends, user behavior, and network performance metrics [51]. This enables MNOs to gain deeper insights into the behavior of their mobile networks and make proactive adjustments to ensure optimal performance and quality of service.

Traffic forecasting is a critical component of network planning and dimensioning [45]. To make cellular network businesses more profitable, investors continuously seek the optimal Capital Expenditure (CAPEX) allocation in the right cells/sites/locations while reducing Operational Expenditure (OPEX). Incorrect traffic forecasting can misguide network dimensioning, resulting in additional CAPEX and OPEX as well as QoS degradation.

In recent years, deep learning-based approaches have been extensively studied to identify patterns in sequential data and classify similar data types together [51]. Various Recurrent Neural Networks (RNN) algorithms have been employed to forecast multiple time series sequential data types. Recognizing the immense potential of deep learning algorithms in predictive analytics, the authors of this thesis focus on building a model to address one of the most critical problems in cellular network dimensioning—traffic forecasting [36]. Modern GPUs are utilized to run complex deep learning algorithms efficiently, incorporating various features in an optimistic runtime.

The advancement lies in the ability to accurately forecast traffic and user demand, enabling the network to promptly manage resource allocation among connected users, ultimately improving the quality of user experience [36]. This research will contribute to a better understanding of mobile network traffic behavior and recommend the expansion triggers for eNodeBs based on deep learning algorithm-based traffic forecasts and utilization correlation charts. By leveraging deep learning techniques, MNOs can achieve enhanced resource optimization, more efficient network planning, and improved QoS.

Additionally, deep learning algorithms provide the ability to intelligently allocate network resources based on predicted traffic patterns. By accurately forecasting future traffic demand, MNOs can dynamically allocate bandwidth, optimize network capacity, and prioritize resources to meet the needs of different user groups and applications [36]. This adaptive resource management enhances the overall quality of user experience by minimizing congestion, reducing latency, and improving network stability. Consequently, deep learning empowers MNOs to deliver a seamless and satisfactory user experience, fostering customer loyalty and satisfaction.

Furthermore, the application of deep learning in mobile network planning con-

tributes to the democratization of the field. Traditional network planning approaches often require specialized expertise and extensive manual effort. In contrast, deep learning models can automate and streamline various planning tasks, reducing human intervention and enabling MNOs of varying sizes and capabilities to benefit from advanced planning capabilities [36]. This democratization leads to more efficient and cost-effective network planning, allowing smaller operators and organizations with limited resources to compete on a level playing field.
In summary, deep learning has the unleashed potential to play a crucial role in empowering mobile network planning. By leveraging deep learning algorithms for traffic forecasting, resource allocation, and network optimization, MNOs can overcome the challenges of maintaining quality of service, efficient resource utilization, and network scalability. The utilization of deep learning techniques not only enhances network performance and user experience but also enables a more democratized and inclusive approach to mobile network planning.

1.2 Research Problem

Understanding the traffic demands in a cellular network poses a significant challenge due to the dense and diverse nature of mobile users, the variety of devices, and the ever-changing user patterns [23]. Additionally, the availability of detailed datasets specific to individual eNodeBs, with valuable features, is limited as most Call Detail Records (CDRs) provide aggregated traffic data without technology segregation or per-protocol categorization [5]. Therefore, there is a pressing need to address these challenges and develop effective solutions for network traffic forecasting and resource allocation in mobile network planning.
A pivotal aspect of mobile network dimensioning hinges on accurate traffic forecasting, which directly influences eNodeB-level utilization. The symbiotic relationship between eNodeB-level traffic and utilization entails that fluctuations in the latter can exert significant impacts on overall network performance and user satisfaction. Uncontrolled spikes in utilization can culminate in resource sharing bottlenecks, triggering declines in user experience and Quality of Service (QoS). Conversely, underutilization of eNodeBs represents an inefficient allocation of network resources. Traditionally, network engineers could only react post-occurrence, leading to delayed responses and customer discontent. It becomes imperative for network planners to discern traffic patterns and estimate radio parameters at the eNodeB level preemptively, thereby enabling proactive measures and minimizing user impact.
In light of the foregoing, two primary research questions arise:

1. How can deep learning algorithms be leveraged to empower mobile network planning, through accurately predicting granular network design components like traffic and utilization, and ensuring the Quality of Service (QoS) benchmarks?

2. Is it possible to devise an innovative algorithm that strategically reduces capacity expansion costs for Mobile Operators through scientific soft radio parameter optimization?

In summary, the ever-changing landscape of cellular network traffic analysis and the limited availability of detailed eNodeB-specific data for academic research underscore the pressing need for viable solutions in traffic forecasting and resource allocation. This research aims to highlight and address these challenges through a systematic and scientific approach, ultimately enabling the identification of issues before they manifest within the network and cause disruptions for customers.

1.3 Research Methodology

This subsection presents the end to end research methodology employed in this study with a visual graphic (in Fig 1.1), outlining the steps and procedures followed to achieve the research objectives. The methodology encompasses the data collection process, the selection and implementation of deep learning models, and the evaluation of the research outcomes.

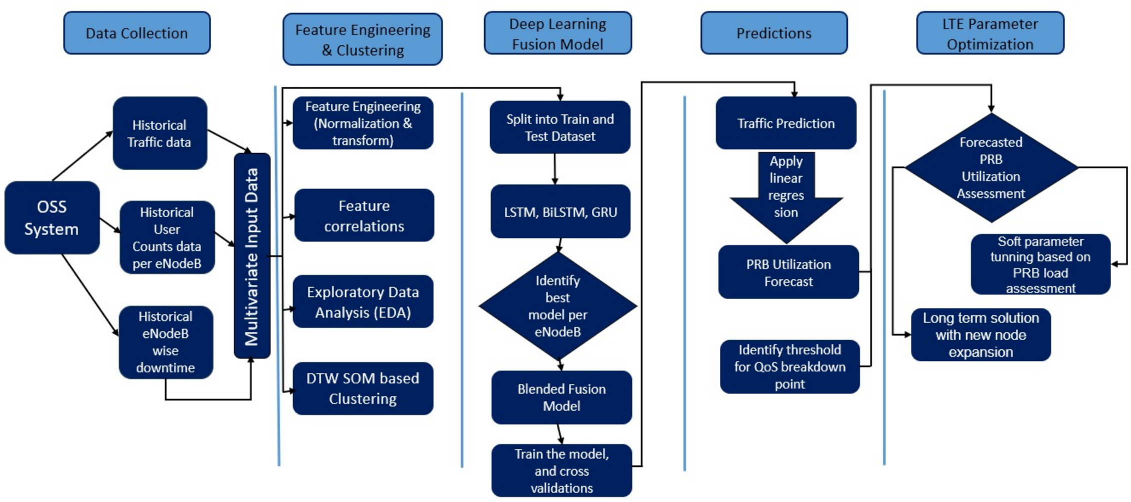

Figure 1.1: High level End to End Research Methodology.

1.3.1 Data Collection

The first step in this research involved collecting relevant data for mobile network planning from MNO. To ensure the availability of diverse and representative data, multiple sources were utilized. These sources included network performance logs, user counts, and network configuration data. The collected data encompassed various parameters such as network traffic, user behavior, and network topology.

1.3.2 Deep Learning Model Selection

The next phase of the research involved selecting suitable deep learning models for mobile network planning tasks. Extensive research and experimentation were conducted to identify the most appropriate models. Considering the complexity and nature of the data, three widely adopted deep learning architectures were chosen: Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Gated Recurrent Unit (GRU). These models have demonstrated strong capabilities in capturing temporal dependencies and have been successfully applied in various time series prediction tasks.

1.3.3 Model Implementation and Training

Once the deep learning models were selected, they were implemented and trained using the collected data. The implementation was performed using a popular deep learning framework, taking advantage of its extensive functionalities and optimization capabilities. The training process involved feeding the models with the historical data and iteratively updating the model parameters to minimize prediction errors. Careful attention was given to hyperparameter tuning and regularization techniques to ensure optimal model performance.

1.3.4 Evaluation Metrics

To assess the performance of the deep learning models, appropriate evaluation metrics were employed. These metrics included but were not limited to Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The selection of evaluation metrics aimed to provide a comprehensive understanding of the models' predictive accuracy and their ability to capture the underlying patterns in the mobile network data.

1.3.5 Experimental Validation

To validate the effectiveness of the proposed deep learning models, extensive experiments were conducted. The experiments involved splitting the collected data into training, validation, and testing sets. The models were trained on the training set, and their performance was evaluated using the validation set. The final assessment of the models' performance was conducted on the testing set, which represented unseen data. Through rigorous experimentation and analysis, the models' strengths, limitations, and potential areas of improvement were identified.

1.3.6 Ethical Considerations

Throughout the research process, the study placed a high emphasis on ethical considerations. The data employed in this investigation underwent anonymization and were managed in adherence to established privacy regulations and industry best practices. The research adhered strictly to ethical guidelines, safeguarding the privacy and confidentiality of individuals. No personal user data or identifiable information was gathered, and all eNodeB data underwent appropriate masking procedures to ensure the preservation of privacy.

1.3.7 Scope and Limitations

Scope

This research sets out to demonstrate the transformative potential of deep learning algorithms in the realm of network planning. The study focuses on a specific mobile network operator (MNO) with a vast subscriber base of over 50 million users across the country. Real network-generated information and the MNO's configuration serve as the foundation for this research.
Multiple machine learning models have been devised to address network planning

objectives. These models are tailored specifically to the characteristics and requirements of the chosen MNO. By leveraging the available data, the proposed models aim to optimize network performance and enhance the overall user experience.

Limitations

While the proposed models demonstrate strong performance under normal operating conditions, it is important to acknowledge certain limitations. Specifically, the models' performance may experience a slight degradation during periods of high user density, such as social gatherings, where the number of users in a specific location surpasses the typical day-to-day usage patterns. The models' effectiveness in such scenarios may be subject to further optimization and refinement.
Additionally, the algorithm for soft parameter tuning primarily focuses on LTE capacity enhancement. As a result, GSM networks may occasionally encounter resource constraints. However, these limitations can be addressed through dedicated optimization techniques that take into account the specific requirements of GSM networks. Furthermore, as the number of eNodeBs increases, it is crucial to enhance computational power to maintain efficient network planning and management.
It is worth noting that this research is conducted within the confines of the selected MNO and its network configuration. The findings and recommendations may not be directly applicable to other MNOs with distinct network architectures or operational environments. However, the methodology and insights gained from this study can serve as a valuable reference for future research endeavors and industry-wide advancements in mobile network planning.

1.4 Research Contributions

The primary objective of this research is to explore the potential of deep learning in revolutionizing mobile network planning. The aim is to make significant advancements in the democratization and optimization of mobile network management, ultimately improving the user experience and meeting the growing demands of the mobile data era.
The research paper makes the following major contributions:

1. **Identification of Optimal Deep Learning Algorithm**: The research presents a pioneering contribution by meticulously evaluating and selecting the most suitable state-of-the-art deep learning algorithm from a range of options including LSTM, BiLSTM, and GRU for Mobile Network design fundamental component predictions. This chosen algorithm is utilized for both forecasting network traffic and predicting granular eNodeB-level utilization through a regression technique. This approach takes into account multivariate inputs, allowing for the modeling and prediction of mobile network data traffic. The accurate prediction of utilization (cell load) stands as a crucial asset for Mobile Network Operators (MNOs), enabling them to make well-informed decisions regarding network expansion while upholding benchmark Quality of Service (QoS) standards.

2. **Categorizing eNodeBs into Precise Clusters**: This research involves categorizing all eNodeBs into distinct clusters to ensure precise data traffic predictions. The accuracy of clustering is attained through a comparative analysis of Euclidean distance and Dynamic Time Warping (DTW) algorithms with the Self Organized Map (SOM) technique. The performance of both Euclidean and DTW algorithms is evaluated for a specific cluster, and based on computational accuracy, the DTW-SOM approach is recommended for further investigation. Clustering facilitates the grouping of similar eNodeBs into common clusters, allowing network planners to efficiently address similar clusters collectively. This approach streamlines the planning process, enhancing efficiency and effectiveness.

3. **Innovative Algorithm for Radio Parameter Estimation to Uphold QoS Benchmarks** This research introduces a groundbreaking algorithm designed to estimate radio parameters, ensuring the maintenance of Quality of Service (QoS) benchmarks. This algorithm empowers network planners to proactively meet customer demands, preventing the breach of QoS thresholds or significant deterioration in customer experience. Additionally, it safeguards Mobile Network Operators (MNOs) from unnecessary capacity-related investments by fine-tuning soft parameters. By leveraging this algorithm, network planners can make informed decisions to optimize network performance and resource allocation, aligning with QoS targets and prudent resource management.

1.5 Thesis Organization

This thesis embarks on a thorough exploration of how deep learning can revolutionize mobile network planning. The following chapters take you on a step-by-step journey, where we blend advanced technology with the intricacies of telecom infrastructure. Chapter 2 discusses about the related works, which delves into previous planning methods and shines a light on how deep learning can transform the field. Moving forward, Chapter 3 provides a clear understanding of the foundation by introducing various deep learning algorithms that predict mobile network trends. As we progress to Chapter 4, we lay out the groundwork—how we collect data, examine it, and design our deep learning systems. Chapter 5 is where we unveil the results, breaking down how different deep learning models perform and highlighting the strengths of best-suited model. In Chapter 6, we take a detour to optimize LTE radio parameter estimation, connecting radio quality with how we use resources. Finally, in Chapter 7, we wrap up our journey by reflecting on what we've learned, acknowledging both progress and limitations, and envisioning how deep learning can democratize mobile network planning.

Chapter 2

Related Works

2.1 Introduction

In this recent era, mobile networks are playing a vital role in driving our society towards digital revolutions. As mobile networks continue to evolve and expand, the challenges associated with network planning become increasingly complex. Network planning is essential for translating customer requirements and quality of services into efficient operational realities. Traditional approaches to mobile network planning often struggle to keep pace with the rapid growth in traffic demand, technological advancements, and changing user requirements. One of the key objectives of this research is to manage the quality of service (QoS) in mobile networks by effectively handling the ever-growing traffic demand. Predicting traffic demand is a crucial component in managing the network ecosystem, as traffic directly correlates with utilization, which in turn affects the performance of cellular networks [61]. This chapter provides various challenges faced in traditional mobile network planning and explores an overview of the state-of-the-art methods in network traffic forecasting and LTE radio parameter estimation, highlighting their relevance to the research objectives.

2.2 Legacy Planning Approaches: Hand tuned manual heuristics

This section critically examines the limitations associated with traditional mobile network planning methodologies. These legacy approaches are typically characterized by their reliance on assumptions and static models, which fail to capture the dynamic nature of network traffic patterns and user behavior. Moreover, they often rely on manual interventions, lacking the scalability necessary to accommodate the growing complexity of modern mobile networks. As a result, these approaches often fall short in terms of timely action, as they lack visibility into future network growth. Instead of proactively addressing potential issues, traditional approaches tend to respond reactively after problems have already occurred, leading to customer dissatisfaction and a decline in quality of service in specific areas.
Furthermore, legacy network planning systems heavily depend on hand-tuned heuristics employed by network planners. However, these heuristics are often inadequate, either due to their inappropriate granularity or the time-consuming nature of their

implementation. Consequently, such legacy approaches fail to provide the necessary level of accuracy and efficiency required for effective network planning.
In light of these specific shortcomings, there is a clear need for more adaptive and data-driven solutions with the help of deep learning algorithms in mobile network planning.

2.3 Traffic Demand Analysis and Prediction

In this section, we will explore the challenges and significance of traffic demand analysis and prediction specifically in the context of mobile network planning. The classification of internet traffic into various types has become crucial in today's digital landscape, with different applications requiring specific bandwidth and latency requirements. When considering internet traffic over cellular networks, it can be primarily categorized into five major types: streaming, social networking, browsing, OTT audio and video calling, and online gaming. However, accurately predicting the demand for such traffic poses significant challenges.
Traffic prediction or forecasting is crucial for anticipating the status of cellular networks, identifying user usage patterns, and estimating quality-of-service parameters and resource allocation requirements [45]. One of the major hurdles lies in predicting internet traffic at a spatial granularity suitable for mobile network planning. While mobility is a significant advantage of mobile networks, it is equally essential to forecast traffic patterns with the utmost precision. The ability of network planners to enable precise traffic prediction at a granular level facilitates more effective network design and planning.
The field of mobile network planning recognizes the importance of traffic prediction using spatio-temporal data, which has garnered significant attention in both academic and industry domains. With the advancement of the Internet and location-based technologies, a wealth of spatio-temporal data is collected by government agencies and mobile network operators. For instance, user-centric data, network measurements, and historical traffic patterns can be leveraged to improve the accuracy of traffic prediction models. This data-driven approach empowers network planners to proactively allocate network resources, optimize capacity, and enhance the quality of service.
According to Yuan and Li's survey on traffic prediction (2021), the problem encompasses several aspects within the domain of mobile network planning. Traffic status prediction, for example, involves anticipating the congestion levels of specific network segments in the future, allowing proactive measures to be taken [53]. This prediction can be achieved by estimating traffic speed or travel time, with slower speeds or longer travel times indicating potential congestion areas. Traffic flow prediction aims to forecast the volume of traffic on different network paths, enabling efficient resource allocation and traffic management. Additionally, predicting travel demands within the mobile network context is crucial for capacity planning and optimizing resource utilization.
Several studies have addressed cellular network traffic forecasting using different techniques. For example, Fang et al. [55] presented a city-scale traffic forecasting model based on a cell handover-aware graph neural network. Xu et al. [18] demonstrated the geographical distribution of forecasted traffic in a particular city by analyzing time series data. Similarly, Kirmaz et al. [42] divided the geographic

area into pixels for traffic prediction. However, these studies focused on predicting traffic based on a geographical unit of measurement, which can encompass multiple LTE eNodeBs or cells. In contrast, our research focuses on predicting traffic at the granular eNodeB or cell level, allowing for more localized and accurate predictions. Trinh et al. [23] introduced mobile traffic forecasting using recurrent neural networks (RNNs) at a daily level, while Sun et al. [59] estimated network-level mobile data based on user mobility patterns. Although these studies provide valuable insights into traffic forecasting, our research distinguishes itself by predicting traffic at an hourly level, which provides more detailed information about time series data and can be easily converted to daily-level forecasts [23].

Additionally, L. Lo et al. [56] developed a Thresholded Exponential Smoothing and Recurrent Neural Network (TES-RNN) model for managing network traffic and resources using a hybrid approach of statistical modeling and machine learning. However, this research specifically focused on predicting traffic anomalies at specific times rather than regular hourly or daily traffic patterns. Q. Yu et al. [60] utilized Graph Attention Networks (GATs) and Temporal Convolutional Networks (TCNs) to predict traffic overload considering large amounts of small-scale redundant data. Unlike most related research works, our study focuses on cellular network traffic forecasting at the granular cell level, where each cell represents a different eNodeB. This level of granularity is crucial for mobile network operators' network planning activities. Furthermore, evaluating traffic forecasting at the cell or eNodeB level enables easy conversion to city or province level forecasts by aggregating the traffic of all eNodeBs in a given geographic area. Considering the real-life challenges of network planning, we specifically develop an hourly traffic forecast model, which offers a suitable time horizon for effective planning activities.

In summary, traffic prediction plays a pivotal role in mobile network planning, addressing the challenges associated with accommodating varying traffic demands. By leveraging spatio-temporal data and advanced prediction models, network planners can make informed decisions to meet the dynamic requirements of mobile network users. The key difference between our research and previous works lies in the granular data traffic prediction based on two major factors: hourly time granularity and eNodeB or cell-level granularity. By addressing these aspects, our research aims to provide more accurate and localized traffic forecasts, facilitating efficient network planning and optimization.

2.4 Maintaining QoS Benchmark or Maximizing Throughput

Ensuring a consistent Quality of Service (QoS) benchmark is a significant challenge for Mobile Network Operators (MNOs) in the face of increasing network traffic. With limited resources available for network design, accommodating the growing traffic demands becomes a complex task. In LTE networks, the allocation of physical resource blocks (PRBs) plays a crucial role in resource management.

As more users request data from the network, the QoS and throughput tend to degrade, impacting the overall network experience. However, in this research, we aim to address this challenge and demonstrate how MNOs can maintain the QoS benchmark even with growing traffic.

Our approach involves leveraging advanced machine learning techniques and deep learning algorithms to predict traffic patterns and estimate cell-level utilization. By analyzing historical data, we can anticipate traffic demands with greater accuracy. This enables MNOs to allocate resources more effectively and optimize network performance, ensuring that the QoS benchmark is met even during peak traffic periods.
In addition to traffic prediction, we also focus on estimating radio parameters to further enhance QoS. By applying deep learning algorithms to analyze network data, we can optimize radio resource allocation and improve network efficiency. This enables MNOs to maximize throughput and deliver an enhanced user experience.
Through the utilization of machine learning and deep learning algorithms, our research presents innovative solutions for MNOs to overcome the challenges associated with maintaining a QoS benchmark. By accurately predicting traffic patterns and optimizing radio resource allocation, MNOs can ensure consistent QoS levels, even as network traffic continues to grow.
In conclusion, our research emphasizes the importance of maintaining a QoS benchmark and maximizing throughput in the face of increasing network traffic. By leveraging machine learning and deep learning techniques, MNOs can effectively predict traffic patterns, optimize resource allocation, and deliver an enhanced network experience to their users. Our findings contribute to the development of practical solutions that enable MNOs to meet the growing demands of mobile networks while maintaining high QoS standards.

2.5 Cost-Effective Resource Management and Optimization

The availability of resources, including spectrum, power, and physical infrastructure, poses significant challenges in mobile network planning. All these resources incur a large amount of costs for the network. When the demand increases from the cellular network, operators have to invest more to meet the demand. However, incorrect predictions of demand can result in higher capital expenditure (CapEx) investments as well as increased operating expenditure (OpEx) for the network. In this research, a proposed innovative solution is presented that not only solves the technical problems but also lowers the cost of network investment and operation.

2.5.1 Resource Management Challenges and Cost Implications

1. **Resource Availability and Demand Prediction:** Accurately predicting the future demand for resources is a challenge for mobile networks. Inaccurate predictions can lead to underinvestment or overinvestment in resources, resulting in financial losses for network operators. Careful analysis and management of the availability of spectrum, power, and physical infrastructure are necessary to meet the increasing demands of cellular networks while optimizing costs.

2. **Capital Expenditure (CapEx) Investment:** Network operators need to invest in acquiring additional resources to meet the growing demand. However,

incorrect predictions of resource requirements can lead to higher CapEx investments than necessary. This can result in the wastage of financial resources and negatively impact the overall profitability of the network.

3. **Operating Expenditure (OpEx):** Inefficient resource management can increase the operational costs of the network. Inadequate resource utilization, poor energy efficiency, and ineffective network planning can contribute to higher OpEx, putting additional financial strain on network operators. Therefore, optimizing resource allocation and management is crucial to reducing ongoing operational expenses.

2.5.2 Proposed Solution on LTE Radio Parameter Estimation: Lowering Network Investment and Operation Costs

The proposed research presents an innovative solution that not only addresses technical challenges but also aims to lower the overall cost of network investment and operation. By leveraging advanced techniques such as data analytics, machine learning, and optimization algorithms, the solution focuses on optimizing resource management and allocation to achieve cost-effective network planning.

1. **Accurate Demand Prediction:** The solution incorporates sophisticated demand prediction models that utilize historical data, user behavior analysis, and network traffic patterns to accurately forecast future resource requirements. By improving the accuracy of demand prediction, network operators can make informed decisions regarding resource investments, minimizing the risk of overinvestment or underinvestment.

2. **Optimized Resource Allocation:** The system utilizes advanced optimization algorithms to allocate resources effectively, considering factors such as network capacity, user demand, and cost constraints. By optimizing resource allocation, the aim is to maximize the utilization of available resources while meeting the desired quality of service requirements. This approach helps reduce the need for excessive resource acquisition, resulting in cost savings for network operators.

3. **Energy Efficiency:** The proposed solution emphasizes energy efficiency in resource management. By optimizing power allocation, network operators can reduce energy consumption, leading to lower operational costs. Energy-efficient strategies, such as intelligent sleep mode activation and dynamic resource allocation, are employed to minimize power wastage and improve the overall cost-effectiveness of the network.

By implementing the proposed solution, network operators can achieve cost-effective resource management and optimization, mitigating the risks of excessive CapEx investments and high OpEx. The accurate prediction of resource demand, optimized allocation strategies, and emphasis on energy efficiency contribute to reducing overall network costs while meeting the growing demands of cellular networks. Ultimately, this leads to improved profitability and sustainable operations for network operators. The second part of our research focuses on estimating future network utilization based on predicted traffic and proposing an algorithm for handling expected traffic by estimating LTE radio parameters. While there have been sporadic research efforts on radio capacity analysis at different times, none of the existing studies have specifically addressed the estimation of radio parameters from predicted future traffic.

Han Seung Jang et al. [46] developed a model to estimate the resource block usage rate (RBUR) in order to solve the fixed-length input problem in traditional RNN models. However, their research did not explore how radio parameters could be utilized to estimate RBUR. On the other hand, Mehedi et al. [20] proposed an algorithm for Adaptive Mobility Load Balancing in LTE Small-Cell networks to maintain throughput. However, their approach was reactive and did not provide proactive measures for handling expected traffic based on predicted future patterns. In contrast, our research addresses this gap by investigating the estimation of radio parameters from forecasted traffic. We propose an algorithm that leverages predicted traffic patterns to trigger appropriate radio parameter adjustments. By integrating traffic forecasting and radio parameter estimation, our approach enables proactive and optimized handling of expected traffic demands, leading to improved network performance and quality of service.

By bridging the gap between traffic prediction and radio parameter optimization, our research offers a comprehensive solution for efficient network planning. The proposed algorithm not only anticipates future traffic demands but also provides actionable insights for optimizing the utilization of LTE radio resources. This integration of traffic forecasting and radio parameter estimation contributes to the overall objective of managing network quality of service and addressing the ever-growing traffic demand in mobile networks.

By embracing adaptive planning, operators can respond to dynamic demands and optimize resource allocation. Data analytics provides valuable insights for informed decision-making and proactive optimization. Deep learning algorithms offer automation and intelligent resource allocation.Adopting these strategies enables operators to overcome limitations of traditional approaches and build future-ready networks with superior performance and reliability. In the following chapters, we will discuss the practical implementation of these techniques, showcasing how they modernize network planning activities. In conclusion, this research represents a pivotal convergence of network traffic forecasting and LTE radio parameter estimation, unlocking valuable contributions to the field of mobile network planning. The combination of these techniques offers a pathway for operators to proactively address challenges, efficiently manage resources, and ultimately deliver an unparalleled quality of mobile network experience to their customers.

Chapter 3

Theoretical Background: Deep Learning for Solving Network Planning Problems

3.1 Introduction

Within this chapter, we embark on a journey into the theoretical underpinnings of our groundbreaking approach to network planning, one that harnesses the immense potential of deep learning. Our pursuit leads us to unveil a meticulously crafted cellular network traffic prediction system model, where the deep learning algorithms guide in a new era of precision in traffic projection and proactive resource management. This intricate synergy of Long Short Term Memory (LSTM), Bidirectional LSTMs (BiLSTM), and Gated Recurrent Units (GRU) algorithms bolsters the realm of time series forecasting, enhancing predictive capabilities with unprecedented finesse. The narrative unfurls further as we unravel the intricacies of Self-Organizing Maps (SOM) and Dynamic Time Warping (DTW) based clustering, instrumental in categorizing eNodeBs and orchestrating optimal resource allocation. Beyond theory, this chapter unearths the blueprint for practical implementation, an ode to the transformative prowess of deep learning in reshaping the contours of contemporary network planning.

3.2 LTE Network Architecture

The Long-Term Evolution (LTE) network architecture is a complex system composed of interconnected components and interfaces that work together to provide seamless wireless communication services. Understanding these components and interfaces is essential for grasping the intricate workings of modern cellular networks [10].

3.2.1 Components of LTE Network Architecture

1. **User Equipment (UE):** UE, commonly referred to as mobile devices or terminals, serves as the endpoint for user communication. It encompasses smartphones, tablets, laptops, and various wireless devices used by consumers to access network services [12].

2. **Evolved NodeB (eNodeB):** The eNodeB is the cornerstone of the LTE network. It consists of the base station and radio equipment responsible for transmitting and receiving wireless signals to and from UEs. Each eNodeB covers a specific geographical area known as a cell, which can be further divided into sectors for more efficient coverage.

3. **Evolved Packet Core (EPC):** The EPC is the core network architecture responsible for managing various network functions, including data traffic routing, mobility management, and policy enforcement. It comprises several key components [8] [3]:

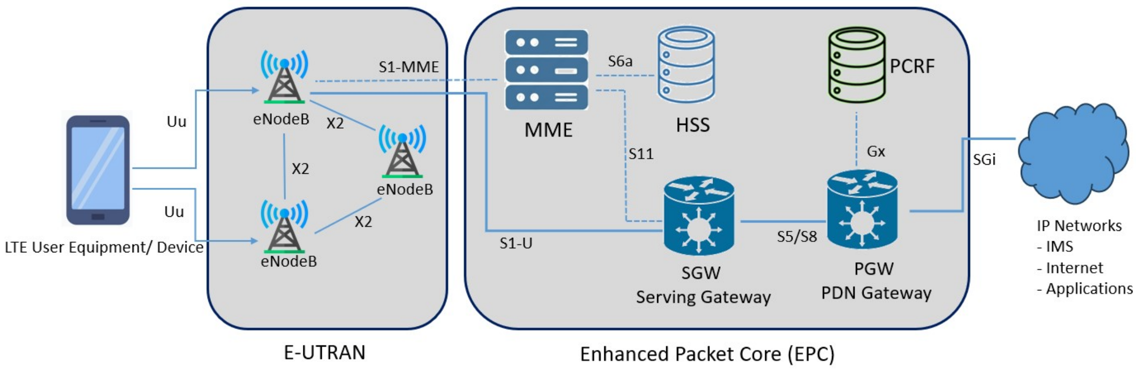

Figure 3.1: LTE Network Architecture.

   (a) **Mobility Management Entity (MME):** The MME handles authentication, UE tracking, handovers, and security functions. It ensures seamless mobility and secure access for UEs.

   (b) **Serving Gateway (SGW):** The SGW manages data traffic routing, packet forwarding, and mobility management within a specific area. It plays a pivotal role in maintaining data connectivity during UE mobility.

   (c) **Packet Data Network Gateway (PGW):** The PGW connects the LTE network to external networks, such as the Internet. It manages IP address allocation, traffic filtering, and charging functions [3].

4. **Home Subscriber Server (HSS):** The HSS stores user-related information, including authentication details, service profiles, and mobility information. It is a central repository for user-related data, facilitating secure and efficient network operations.

5. **Policy and Charging Rules Function (PCRF):** The PCRF manages policy control and charging rules within the LTE network. It ensures that Quality of Service (QoS) requirements are met while controlling and monitoring data usage for accurate billing [3].

3.2.2 Interfaces in LTE Network Architecture

1. **Uu Interface:** The Uu interface is the wireless air interface between the UE and the eNodeB. It enables the transmission of user data, control signals, and

mobility-related information, ensuring seamless communication between UEs and the network.

2. **X2 Interface:** The X2 interface connects neighboring eNodeBs within the same LTE network. It facilitates direct communication between eNodeBs, allowing them to exchange control and data information for optimized handovers, load balancing, and interference coordination.

3. **S1-MME Interface:** The S1-MME interface links the MME in the EPC with the eNodeBs. It handles signaling for functions like initial attach, handovers, and location management, ensuring efficient mobility management and UE session maintenance [4].

4. **S6a Interface:** The S6a interface connects the MME in the EPC with the HSS. It is responsible for authentication, authorization, and the exchange of user-related information, facilitating secure and accurate user management.

5. **S11 Interface:** The S11 interface connects the SGW and PGW within the EPC. It manages session management, traffic routing, and mobility across different ePDNs and external networks, maintaining seamless connectivity.

6. **S5/S8 Interface:** The S5/S8 interface connects the SGW and PGW in the EPC. It is crucial for tunneling user data between gateways, ensuring proper data routing and maintaining quality of service during data transmission. [3]

7. **Gx Interface:** The Gx interface connects the PCRF with the PCEF within the PGW. It facilitates the exchange of policy and charging-related information, enabling accurate enforcement of policies and appropriate charging based on data usage [4] [3].

In conclusion, the LTE network architecture is a sophisticated ecosystem comprising diverse components and interfaces that work in tandem to deliver reliable and efficient wireless communication services. The seamless coordination between these components and interfaces ensures high-speed data transmission, seamless mobility, and optimal network performance for users across various geographical areas.

3.3 Deep Learning Algorithms for Predicting Time Series Data

Predicting time series mobile network data is a multifaceted challenge, often requiring sophisticated tools to decipher intricate patterns within the data. This section delves into the theoretical underpinnings of three such tools: Long Short-Term Memory (LSTM), Bidirectional LSTMs (BiLSTM), and Gated Recurrent Unit (GRU). These Deep Learning techniques offer solutions to the intricacies of temporal data forecasting.

3.3.1 Long Short-Term Memory (LSTM)

In the realm of recurrent neural networks (RNNs), LSTM emerges as a groundbreaking innovation to surmount the limitations of traditional RNNs. The challenge with RNNs lies in their inability to retain contextual information across lengthy sequences, resulting in the vanishing gradient problem [1]. LSTM introduces memory cells and gating mechanisms to address these shortcomings.
LSTM's architecture consists of three gates: forget gate, input gate, and output gate. The forget gate decides what information to retain or discard from the previous state, while the input gate determines the new information to be added to the current state. The output gate regulates the output based on the current state. These gates, orchestrated through intricate calculations, enable LSTM to capture

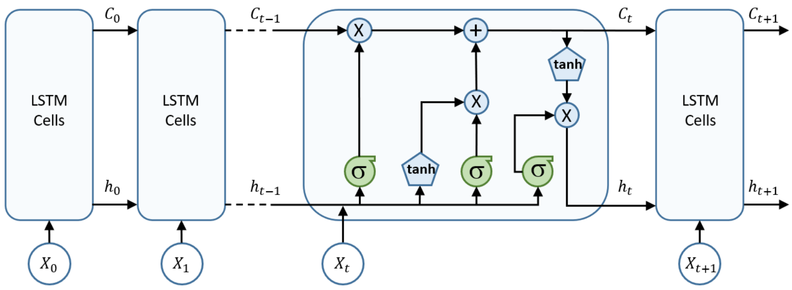

Figure 3.2: LSTM architecture for predicting future traffic.

long-range dependencies and discern patterns that elude traditional RNNs. Equation 3.1 represents the computation of the forget gate, with subsequent equations 3.2 to 3.6 depicting the cascade of operations within LSTM.

Why do we choose LSTM?
We have chosen to utilize Long Short-Term Memory (LSTM) in place of the conventional RNN model, strategically addressing the memory challenge that plagues traditional setups. LSTM stands as a refined iteration of the recurrent neural network (RNN) framework, introducing an architecture that excels in modeling chronological sequences and their intricate long-range dependencies. Unlike conventional RNNs, LSTMs were meticulously crafted to tackle the persistent long-term dependency issue.
The essence of LSTM's architecture lies in its prowess to grapple with the challenges of temporal data. The journey commences with the computation of an output value based on preceding time data. Subsequently, this output value interfaces with input series data, serving as an input to the forget gate [2], [6]. This orchestration captures the essence of LSTM's functioning – an orchestration designed to decode and understand temporal patterns and dependencies inherent in data.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3.1}$$

Here h_{t-1} is the output value of the previous time, as well as x_t , denotes the input value of the present time. f_t is the output gate whose value range is (0,1). The weight of the forget gate is represented as W_f , where b_i is the bias of that forget gate. In addition of that, input to input gate, output value and condition of candidate cell at input gate can also be calculated through output value of previous time and the input value of present time, which can be calculated through the below equations –

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{3.2}$$

$$\tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{3.3}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3.4}$$

$$O_t = \sigma\left(W_0 \cdot [h_{t-1}, x_t] + b_o\right) \tag{3.5}$$

$$h_t = O_t * \tanh\left(C_t\right) \tag{3.6}$$

In the context of equations (3.3),(3.4) and (3.6), C_t signifies the cell state of the candidate cell at time t, with its values ranging from 0 to 1. O_t denotes the output gate, it signifies the input gate, and ht represents the hidden layers within the cell. In this particular context, x_t symbolizes cellular network data traffic. The presence of the network bias finds representation in the b function.
This LSTM model functions as a sequential layer, forming the foundation for the construction of a traffic forecasting model. The architecture of this LSTM has been custom-tailored, drawing inspiration from the insights of references [17] and [13]. The equations (3.1), (3.2), (3.3), and (3.5) pivot on the outcome of a dot product, facilitating information transfer. In cases where the result of the dot product amounts to zero, it signifies a lack of information transfer [1]. Information will transfer, in case of dot product outcome is one.

3.3.2 Bidirectional LSTMs (BiLSTM)

Building upon the foundation of LSTM, Bidirectional LSTMs (BiLSTM) broaden the horizons of temporal modeling. In a conventional LSTM, information flows sequentially from past to future, posing challenges in capturing future context. BiLSTM overcomes this by employing two separate hidden layers – one processing the sequence in its natural order and the other in reverse [32]. This bidirectional processing enables the model to capture past and future context simultaneously, enriching the understanding of intricate temporal relationships [15].

In the context of this research, the Bidirectional Long Short-Term Memory (BiLSTM) model is harnessed to ascertain the optimal prediction methodology. During training, the BiLSTM model capitalizes on input data in a dual-directional manner, embracing information from both forward and backward passes. This process entails two phases: first, analyzing the data from right to left, followed by analyzing it from left to right. This bidirectional mechanism augments the BiLSTM model's precision and performance by addressing and alleviating potential long-term dependencies [15]. Furthermore, this bidirectional framework not only facilitates more comprehensive training but also yields enhanced outcomes in BiLSTM algorithms [58]. As a result of these intrinsic advantages, BiLSTM models often showcase superior performance compared to their traditional LSTM counterparts, a notion that will be extensively discussed in subsequent sections of this thesis.

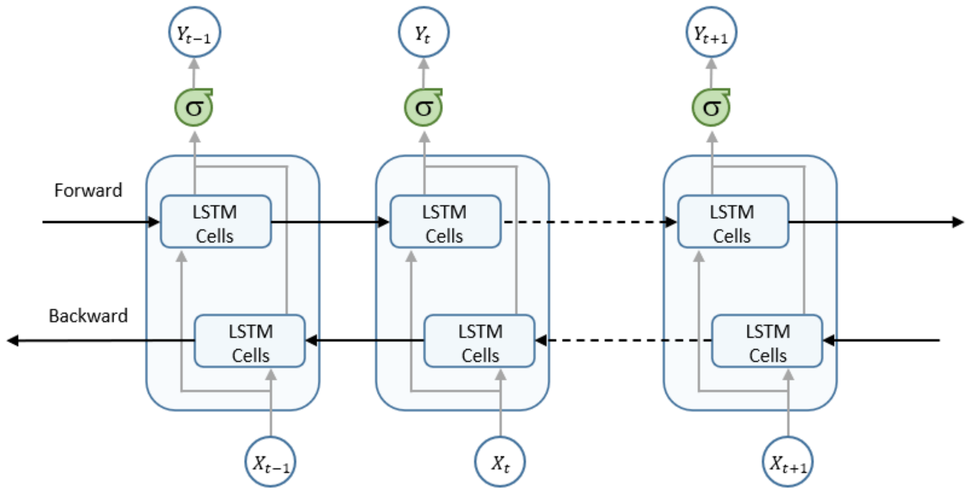

Figure 3.3: BiLSTM architecture for predicting future traffic.

3.3.3 Gated Recurrent Unit (GRU)

Introduced as a sibling to LSTM, the Gated Recurrent Unit (GRU) is a streamlined yet powerful alternative. GRU shares similarities with LSTM, featuring gating mechanisms to control information flow. It condenses LSTM's architecture, combining the forget and input gates into a single update gate and introducing a reset gate. These streamlined components result in simpler training and more efficient computation, making GRU an attractive choice for various sequence modeling tasks [45]. In the context of this research, the process of comparing models involves the

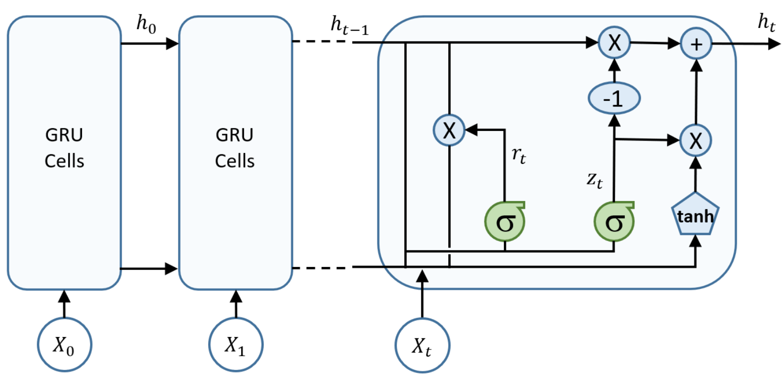

Figure 3.4: GRU architecture for predicting future traffic.

incorporation of a Gated Recurrent Unit (GRU). GRU, introduced by Kyunghyun Cho et al. in 2014, stands as a relatively contemporary member of the RNN family. Although it shares a comparable architecture with LSTM, GRU models exhibit a higher degree of convenience and simplicity in terms of training and implementation. The characteristic architecture of a typical GRU model is showcased in Figure 11, with the design having been adapted from reference [32]. The neural network architecture of GRU carries a distinct advantage in computational efficiency, owed largely to the presence of update and reset gates. These components collectively contribute to the model's ability to retain long-term memory states within the cell [45]. In a manner akin to the LSTM forget gate, the reset gate in the GRU model plays a

pivotal role.
Within the framework of GRU, the hidden state output at time t can be computed through a general expression, as outlined below:

$$h_t = f\left(h_{t-1}, x_t\right) \tag{3.7}$$

In equation (3.7), h_{t-1} is the hidden state status in $t-1$ time and x_t input time series value at t time. For explaining to the GRU NN model as shown in architecture (Fig. 3.4) below equation can be used –

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \tag{3.8}$$

$$z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \tag{3.9}$$

$$\tilde{h_t} = \tanh\left(W_{\tilde{h}}\left[r_t * h_{t-1}, x_t\right]\right) \tag{3.10}$$

$$h_t = \left(1 - z_t\right) * h_{t-1} + z_t * \tilde{h_t} \tag{3.11}$$

$$y_t = \sigma\left(W_o \cdot h_t\right) \tag{3.12}$$

In these equations (3.8),(3.9) and (3.12), Sigmoid function is represented as σ, which output is (0,1). r_t is the updated, which works for determining stored information quantity from one movement to another. Reset gate z_t determines the status of information of the last state, whether the information is kept or erased. The parameter which needs to train are denoted as W_r, W_z, W_h, W_o [36], [15], [25] & [45].
In the quest to forecast time series mobile network data, the arsenal of Deep Learning tools, including LSTM, BiLSTM, and GRU, stands as a testament to the field's advancements. These models transcend the limitations of traditional neural networks, offering mechanisms to capture long-term dependencies, bi-directional context, and streamlined information flow. As the subsequent sections delve into their application and results, the significance of these theoretical foundations becomes abundantly clear in shaping accurate predictions for mobile network planning.

3.3.4 Regression Technique for Utilization Prediction

In the realm of network planning, the journey to accurate utilization prediction is paved by regression techniques. These techniques hold the key to addressing the critical challenge of forecasting continuous values based on inputs [49], [35]. In the ambit of this study, our focus pivoted towards predicting utilization subsequent to deriving forecasted traffic outcomes from our deep learning model. In this context, the prowess of Deep Regression emerges as a beacon, empowered to forecast utilization based on eNodeB-wise traffic predictions. The heart of this endeavor lies within the equation (3.13)

$$\hat{y} = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + b \tag{3.13}$$

Within this equation, w signifies the weight assigned to input traffic x_1 to x_d, and b takes on the mantle of bias or offset.The weight element intricately defines the influence wielded by features within the model [9], [29] & [27]. As we journey through the following sections, the potency of this methodology will be unveiled in its ability to illuminate the path towards optimized utilization prediction.

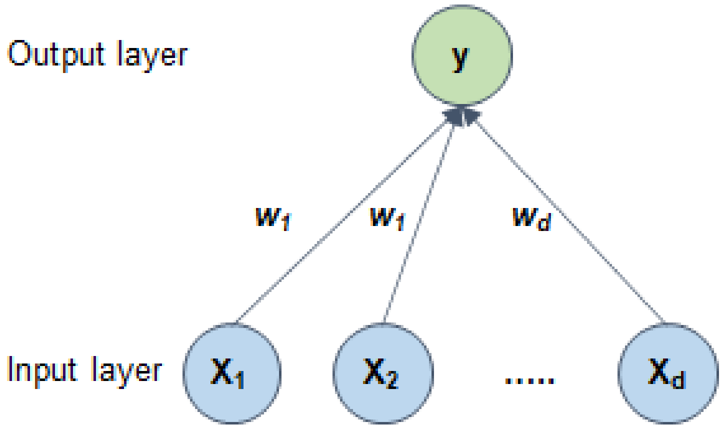

Figure 3.5: Single layer regression with deep neural network.

3.4 Evaluation and Performance Metrics

In the landscape of predictive modeling, the accurate assessment of model performance becomes an essential endeavor. To this end, an arsenal of comprehensive metrics is employed, each with its own distinct purpose in evaluating the efficacy of forecasting. The evaluation criteria comprising Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Squared Correlation (R^2) are the cornerstone of this assessment framework [51]. The formula is describing as below in equations (3.14):

$$\begin{aligned} MSE &= \frac{1}{N}\sum_{k=1}^{n}(y_t - x_t)^2 \\ RMSE &= \sqrt{\frac{1}{N}\sum_{k=1}^{n}(y_t - x_t)^2} \\ MAE &= \frac{1}{N}\sum_{k=1}^{n}|y_t - x_t| \\ R^2 &= 1 - \frac{\sum_{k=1}^{n}(y_t - x_t)^2}{\sum_{k=1}^{n}(\bar{y} - x_t)^2} \end{aligned} \tag{3.14}$$

Let's delve into the intricacies of these metrics and their equations.

Mean Square Error (MSE) serves as a cardinal measure, quantifying the average squared difference between the actual and predicted data points. It enables the capture of the extent to which predictions deviate from the truth across the entire dataset.

Root Mean Square Error (RMSE) acts as an extension of MSE, introducing the element of scale into the equation. By taking the square root of MSE, RMSE offers an understanding of the average magnitude of the prediction errors.

Mean Absolute Error (MAE) adopts an absolute approach, measuring the average magnitude of prediction errors without consideration for their direction. It provides a clearer understanding of the errors' absolute impact.

Squared Correlation (R^2) delves into the realm of correlation between actual and predicted values. It quantifies the proportion of the variance in the dependent variable that can be predicted from the independent variable. Its calculation is expressed as:

These metrics collectively form the cornerstone of performance assessment, guiding the interpretation of forecasting accuracy and directing the course towards refined predictions.

In Chapter 3, a comprehensive theoretical foundation for addressing network planning challenges through deep learning was established. By delving into various techniques, including LSTM, Bidirectional LSTMs, GRU, and Regression, the chapter showcased the versatility of deep learning in predicting time series data. Through rigorous analysis and evaluation using diverse performance metrics, this chapter underscores the potential of these techniques for enhancing network planning solutions.

Chapter 4

Proposed System Model and Methodology

4.1 Introduction

Nestled within this chapter is a meticulous dissection of the structured deep learning paradigm that navigates the complexities of cellular network traffic forecasting. This elucidated journey guides us through a meticulously crafted sequence of steps, illuminating the art and science of modeling within the dynamic realm of mobile networks.
At the beginning of this chapter, we'll briefly discuss the dataset description, including EDA, aggregation of datasets, and feature correlation plot. Then, we formulate the problem scientifically. Next, we present the proposed system model, which aims to predict mobile network traffic demand and optimize resource allocation.
In this chapter, we discuss the methodology employed for data analysis, traffic prediction, and utilization optimization. By leveraging innovative approaches, our methodology ensures proactive decision-making and efficient network planning to meet the dynamic demands of cellular networks. Let's explore these aspects in detail to gain a comprehensive understanding of our research approach.

4.2 Dataset Description

The LTE 4G dataset used in this research was obtained from the Operations Support System (OSS) of a Mobile Network Operator (MNO). The dataset consists of hourly data traffic from the Radio Network for approximately 890 eNodeBs over 351 consecutive days, resulting in a total of 8424 samples. To ensure data privacy, a data masking process was applied, resulting in a dataset of around 6.2 million records.
In addition to the data traffic, various associated features from the eNodeBs were collected, including Utilization, Max_UE, Avg_UE, $Cell\_TP$, $User\_TP$, and traffic. For the analysis, the focus was specifically on Downlink (DL) traffic, as it plays the most significant role in cellular network utilization.
The Cellular Network dataset comprises the following key information and features:

- eNodeB: eNodeB is the Radio network element of the LTE network, which is also known as Evolved Node B.

- Traffic: Traffic means a combination of Uplink (UL) and Downlink (DL) internet Traffic from the Radio network end. The counter formula of traffic is as below:
$$\sum \text{downlink traffic volume for } PDCP$$
$$+\sum \text{uplink traffic volume for } PDCP$$
The unit of traffic is Gigabits here.

- Utilization: Utilization indicates the usage of Physical Resource Block (PRB) in LTE system. The higher number of utilization indicates more usage of LTE resources. Utilization can be formulated in counter level by the below formula:
$$\frac{AvgnumberofusedPRBs}{NumberofavailablePRBs}$$

- Max_UE: Maximum number of Users connected at an instance in a particular node considered as Max_UE.

- Avg_UE: Avg_UE is the average number of connected Users per hour in a particular node

- Cell_TP: Cell_TP means Cell Throughput, which is the sum of all users' throughput in a particular eNodeB or any node for a unit time frame. The counter level formula can be represented as below:
$$\frac{\sum \text{downlink traffic volume for PDCP}}{\sum \text{duration of downlink data transmission in a Node}}$$

- User_TP: A particular user receives an amount of data on average, known as User Throughput or User_TP. In other words, the average number of packets received by the User in a unit time frame. The counter level formula for User_TP as below –
$$\frac{(\sum \text{DL traffic} - \text{DL traffic volume sent in last TTI})}{\text{Data transmit duration except last TTI}}$$

During the data modeling phase for traffic forecasting and utilization prediction, all eNodeBs were classified into different classes based on their time-series behavior. Detailed information regarding the classification procedure will be provided in the later part of this paper.

4.2.1 Exploratory Data Analysis

The process of understanding cellular network data, known as Exploratory Data Analysis (EDA), is akin to piecing together a puzzle with many parts. EDA aids in uncovering concealed patterns and narratives within the data, which is available in various snapshots – think of it as hourly, daily, weekly, or even monthly glimpses into network activity.
The journey begins by amassing this data, akin to assembling puzzle pieces, to

extract its revelations. However, data can sometimes have gaps, much like an incomplete puzzle. Identifying and addressing these gaps is paramount to ensure a comprehensive dataset. To maintain organization, the data is stored in a dedicated repository known as a database, serving as a canvas where the entire picture is visible.

The ultimate objective is to predict network activity accurately. To accomplish this, a foundational understanding of the existing data is indispensable. Enter EDA, likened to the work of a detective seeking crucial clues. EDA facilitates the identification of significant clues hidden within the data. One notable clue pertains to missing data – the gaps previously mentioned. Filling these gaps involves employing certain strategies.

For minor gaps, such as missing data within an hour, the mean of analogous periods on the same day of the month is utilized. This approach resembles the act of deducing the appearance of a missing puzzle piece based on its neighboring pieces. In instances of more substantial gaps, spanning beyond an hour, historical trends are consulted. This strategy is analogous to drawing insights from the past to infer the present.

Thus, in the quest to comprehend cellular network activity, Exploratory Data Analysis emerges as a guiding compass. It aids in locating missing components and uncovering the patterns that imbue the data with vitality, preparing the groundwork for the ensuing predictive endeavors.

Exploratory Data Analysis (EDA) serves as a crucial preliminary step before

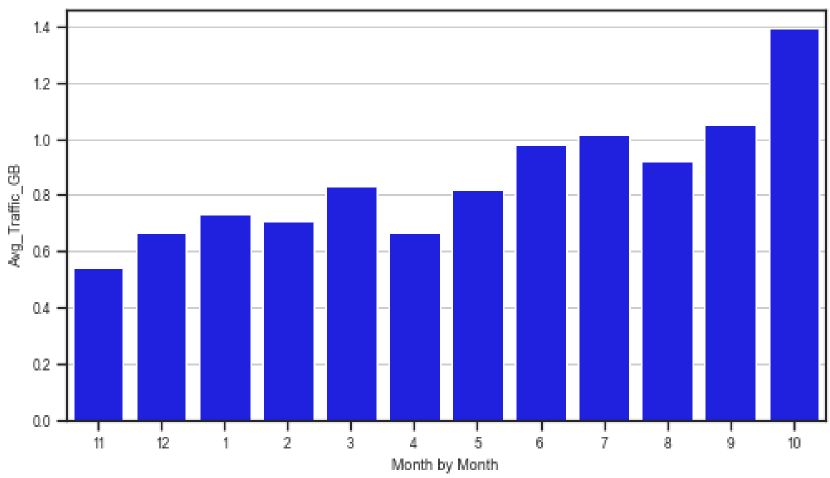

Figure 4.1: Month by Month Average traffic (GB) per eNodeB.

delving into any dataset modeling. In this research, a primary objective is to comprehensively understand the dataset, shedding light on the ebb and flow of traffic over time and pinpointing pivotal hours that contribute significantly to the overall data traffic. The visual cues offered by the four aforementioned figures provide quick insights into the patterns ingrained within the hourly traffic dataset spanning 351 days.

To precisely discern data patterns per eNodeB, a fundamental equation, 4.1, comes into play:

$$E(t) = \sum_{i=1}^{351} (\text{Traffic in hours}) / \text{Number of Days} \tag{4.1}$$

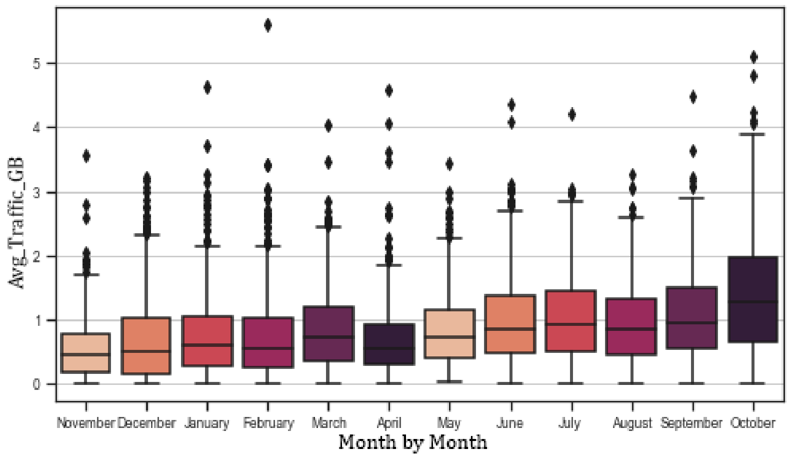

Figure 4.2: Boxplot of Month-by-Month Average traffic (GB).

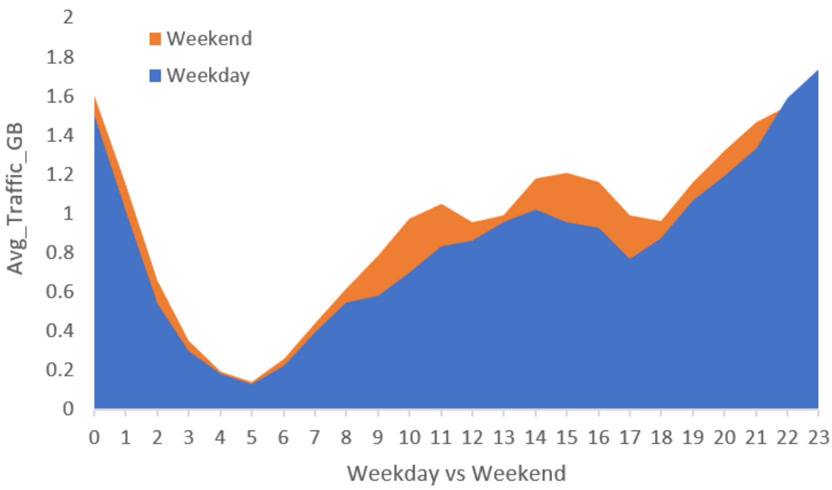

Figure 4.3: Weekday vs. Weekend hourly traffic Pattern.

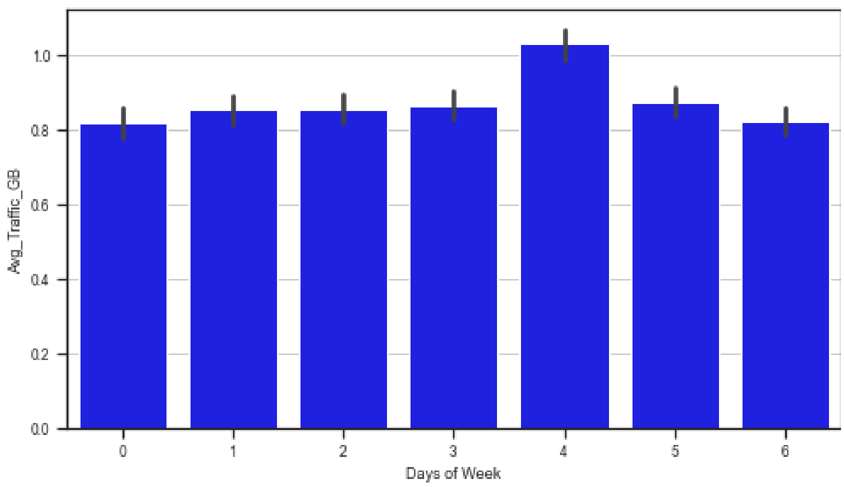

Figure 4.4: Daily Average traffic (GB).

Figure 4.1 vividly elucidates the progressive increase in traffic over time. Further scrutiny of the visualizations brings to light another intriguing observation: the

distinction in hourly traffic between weekdays and weekends, a revelation evident in Figure 4.3.
Shifting our gaze to Figure 4.2, the box plot portraying the monthly average traffic accentuates a consistent uptick in median (Q2) traffic across each passing month. Meanwhile, Figure 4.4 takes us to a distinct day within the week, a day that emerges as an outlier by exhibiting traffic volumes nearly double those of the other days.
In essence, this early stage of pattern identification establishes the groundwork for subsequent modeling endeavors. These insights lay bare the underlying dynamics of network traffic, setting the stage for more informed predictions and meaningful revelations as we navigate through this exploration.

4.2.2 Aggregation of Datasets

The cornerstone of our analytical voyage rests upon a meticulously curated dataset brimming with encrypted eNodeB-wise parameter information, harvested from the bustling urban landscape of South Asia. Let's conceptualize this reservoir of data as $E_t = \{E_{c1t}, E_{c2t}, \ldots, E_{cit}\}$. In this nomenclature, E_t symbolizes the comprehensive collection of all eNodeBs, while E_{cit} encapsulates the rich tapestry of features associated with each individual eNodeB, regardless of the temporal dimension t.
This comprehensive dataset serves as the bedrock for a deeper understanding of network dynamics. To embrace a holistic perspective, we introduce the concept of aggregated eNodeB-wise traffic $A(T)$, meticulously compiled over a defined time frame T. This amalgamated representation manifests as:

$$A(T) = \sum_{r(t) \in R(T)} E(t) \sum_{t \in T} a(t)$$

In this formulation, $A(T)$ resonates as a summation of $E(t)$ across the spectrum of eNodeBs $r(t)$ contained within $R(T)$, coupled with the cumulative influence of features $a(t)$ expressed through the timeframe T. This aggregation affords us a panoramic view of network behaviors and patterns, fueling our journey towards informed forecasting and strategic decision-making.

4.2.3 Feature Correlation Plot

In the pursuit of predicting future eNodeB-wise traffic, this study has meticulously gathered an ensemble of five additional features. These attributes, pivotal in forecasting cellular network traffic and comprehending the overarching LTE network dimensions, form the bedrock of our analysis. Within this context, we delve into the realm of feature correlation plotting, an endeavor facilitated by the equation (4.2) :

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n \sum x^2 - \left(\sum x\right)^2\right]\left[n \sum y^2 - \left(\sum y\right)^2\right]}} \tag{4.2}$$

Enveloped in the graphic representation of correlation, as depicted in Fig. 6, we unveil insights of paramount significance. This visual tapestry reveals a direct correlation between utilization and traffic, an observation that aligns with intuition. In stark contrast, the metric of User Throughput (TP) exhibits a negative correlation with both traffic and utilization. This discovery intimates that heightened traffic manifests as augmented utilization, while also contributing to a decrease in User TP or a potential degradation in the quality of services (QoS). In striving for optimal network design, our paramount objective revolves around maintaining utilization at a level that ensures network integrity. Elevating the discourse, the import of our

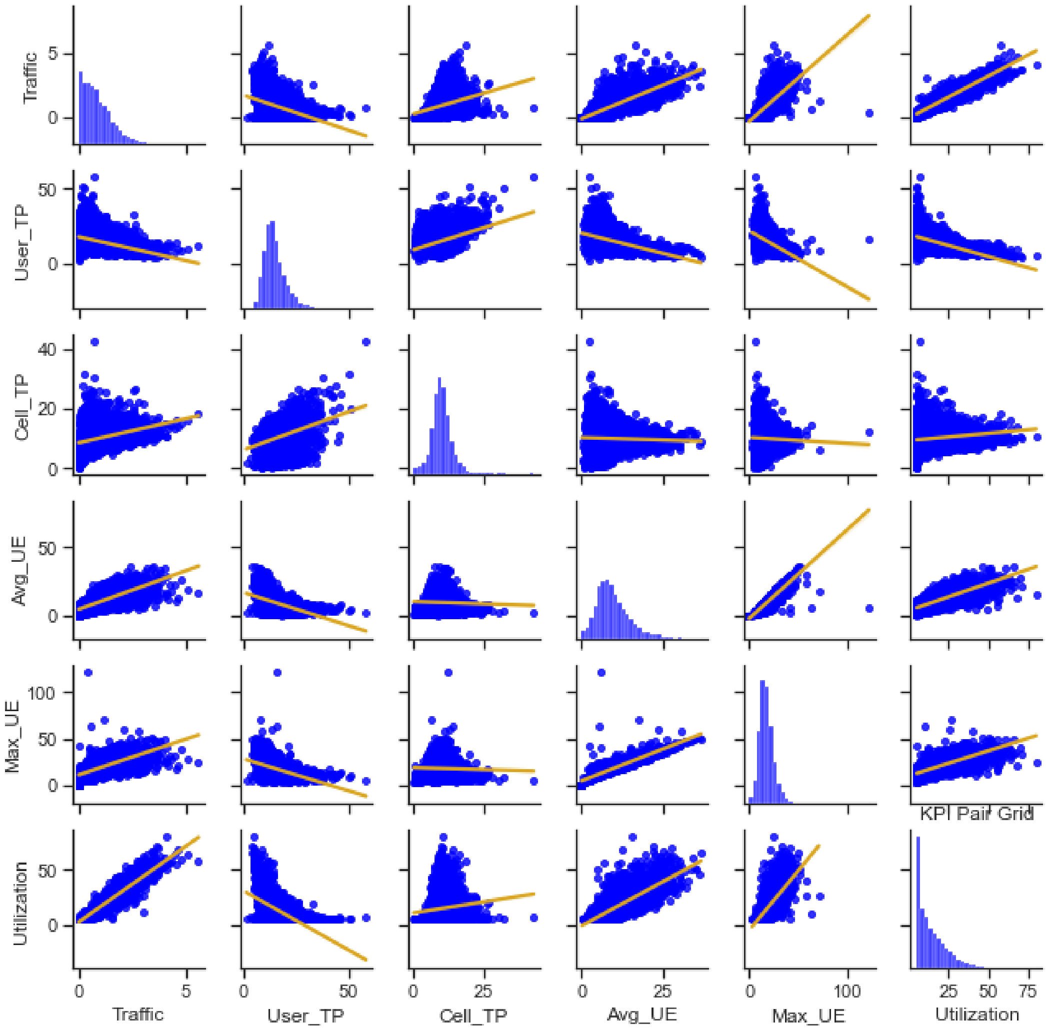

Figure 4.5: Correlation Plot from different features.

findings resonates in network management practices. Elevated utilization flags the advent of burgeoning traffic, thereby triggering the need for network expansion. In this manner, Mobile Network Operators (MNOs) can aptly balance utilization and QoS, steering clear of untenable network conditions. As we traverse deeper into the heart of this research, a subsequent juncture will unveil finely tuned utilization parameters, poised to anchor QoS within the desired threshold.

4.3 Problem Formulation

This research endeavors to tackle these questions by formulating a non-deterministic polynomial *NP*-hard problem that addresses the intricate challenges inherent in mobile network planning. The objective revolves around maximizing user throughput

(Th) for individual eNodeBs, a factor inversely related to Physical Resource Block Utilization (PRBU) and other network variables. The formulation of this problem bridges theoretical research [38], [24], and practical implementation to create a well-rounded approach. The objective function can be defined as follows:

$$\text{Objective function, Max } T_h = \frac{1}{PRBU_t} + C_1 \tag{4.3}$$

In the objective function (4.3), the value of constant C_1 varies based on the configured radio bandwidth of each eNodeB. Notably, the user throughput of a particular eNodeB can vary even with the same PRB utilization due to different configured bandwidths.

Furthermore, predicting future PRB utilization for a cluster of eNodeBs requires considering factors such as predicted traffic volume, average user equipment count, downtime, and other unknown factorsC_2. The equation for PRB utilization in a cluster of eNodeBs can be expressed as follows:

$$median\{PRBU_T\} \times BW \lim_{T\to+\infty} \frac{1}{T} \sum_{t+1}^{t+60} \sum_{e\in E} \quad (Vol_{T,e} + UE_{T,e} - D_{T,e}) + C_2 \tag{4.4}$$

In equation (4.4), the medianPRBUt represents the median PRB utilization for a cluster of eNodeBs, and BW denotes the bandwidth. The equation considers the sum of the predicted traffic volume $\sum_{e\in E} Vol_{T,e}$, average user equipment count (UETe), and downtime(D_T)for each eNodeB in the cluster. It accounts for the factors that affect PRB utilization in a real network scenario.

To simplify the PRB utilization equation for a single eNodeB, we can assume that PRB utilization is directly proportional to traffic growth and the bandwidth of a particular spectrum band. Therefore, the equation for PRB utilization for a single eNodeB can be simplified as follows:

$$PRBU_{t\in T} \times BW = \lim_{T\to+\infty} \frac{1}{T} \left(Vol_{T,e} + \overline{UE}_{T,e} - D_{T,e}\right) + C_2 \tag{4.5}$$

In equation (4.5), T represents the time frame of maximum 60 days hourly future PRB utilization, denoted as $PRBU_{t+60}$, and Vol_{t+60} indicates the predicted traffic volume for the same time frame.

Considering the immense potential of deep learning algorithms for time series data prediction, we have developed prediction models for future traffic volume and PRB utilization by using deep learning algorithms. Our research work extends beyond predicting future traffic and PRB utilization. Based on the assessment of predicted PRB utilization from traffic, we have developed an algorithm for estimating radio network parameters. This algorithm enables the triggering of actions to maintain network QoS benchmarks and optimize resource allocation in mobile network planning.

By addressing the challenges of understanding traffic demands, collecting appropriate datasets, and leveraging deep learning algorithms, our research aims to enhance the efficiency and accuracy of mobile network planning. The proposed models and algorithms provide network planners and engineers with valuable insights into traffic patterns, PRB utilization, and radio parameters, empowering them to make informed decisions and take proactive measures to ensure optimal network performance and user experience.

4.4 Proposed System Model

In this section, we delve into the intricacies of the cellular network traffic prediction system model, elucidating each phase with comprehensive detail. Illustrated in Figure 4.6, our proposed system model is built upon the foundation of deep learning algorithms, signifying a sophisticated approach to achieve accurate predictions.

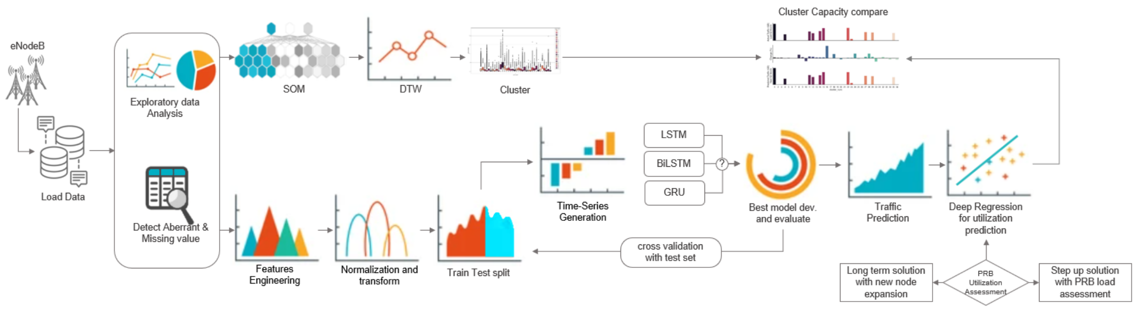

Figure 4.6: Proposed system model for cellular network traffic and PRB utilization prediction driven optimized parameter estimation.

At the outset, our journey commences with meticulous data collection. We meticulously gather eNodeB-wise traffic data and associated parameters through an exhaustive data mining process. This raw dataset finds its abode within a local database, ensuring seamless storage of complex LTE data. Given the intricate amalgamation of underlying features and information, efficient data storage is paramount.

The primary research goal of traffic prediction leads us to the realm of Exploratory Data Analysis (EDA). Through EDA, we not only unveil significant features and patterns embedded within the dataset but also identify any gaps in the data. These gaps, a common occurrence in real-world datasets, require adept handling. We adopt a dual-step strategy: for short, discrete data gaps within a specific hour, we leverage the mean value of that particular day of the month. For more substantial data gaps surpassing an hour, we employ trend predictions extrapolated from preceding data to intelligently populate the missing periods.

With EDA complete and gaps addressed, the dataset takes a central role within our system model. This phase unfolds in two folds. First, we forge clusters of eNodeBs based on their time-series traffic patterns. The prime objective of this clustering lies in grouping eNodeBs with analogous consumption patterns, facilitating subsequent cluster-wise resource utilization comparisons vis-à-vis predicted traffic.

In a research landscape encompassing 890 eNodeBs, clustering stands as a vital instrument for performance visualization. To cater to this need, we adopt the SOM-DTW-based clustering model—a robust mechanism for time series unsupervised data clustering. A comprehensive elucidation of the SOM-DTW model awaits in the Methodology section.

In the subsequent part, we extract critical features from the dataset through feature engineering. Essential features are those that exhibit high correlations with traffic data. As we consider multivariate inputs for the traffic prediction model, these inputs may have different units. To eliminate systematic bias and ensure fair comparisons, data normalization is necessary. In this research, we utilize the min-max method to transform all multivariate inputs to a common range of zero to one. By scaling input data, we effectively reduce bias and enhance the accuracy of the traffic forecasting model. The data transformation equation (Equation 4.6) is used for normalization:

$$z_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \left(New_{\max_x} - New_{\min_x}\right) + New_{\min_x} \tag{4.6}$$

Here, x_{max} and x_{min} represent the maximum and minimum data values, respectively. $Newmin_x$ and $Newmax_x$ represent zero and one, respectively. After normalization and transformation, we split the data into two parts: the test set and the train set. In this research, we maintain a training and test data ratio of 79:21 for 290 days of hourly data from 890 eNodeBs, with the remaining 61 days of hourly data serving as the validation dataset.

4.4.1 Clustering with SOM-DTW

The dataset obtained from traffic and associated parameters revealed a substantial amount of unstructured data. Taming this data requires significant time, resources, and effort. Nonetheless, overcoming the challenges of unsupervised data structuring was imperative to fashion an intelligent traffic forecasting model and an eNodeB utilization model. The efficacy of supervised algorithms is well-recognized in scenarios with properly labeled data. Augmenting the dataset count invariably enhances the precision of the model [34]. In this study, we employ a Self-Organizing Map (SOM) for clustering eNodeBs based on their hourly time series data.

SOM, a variant of unsupervised neural networks, comprises solely two layers [33], an input layer and a mapping layer, which also serves as the output layer. The connection between each neuron in the input and mapping layers is complete in SOM-based clustering. In accordance with its operational principle, each mapping neuron searches for the weight closest to the input vectors during the iterative development of the SOM-based cluster. The optimal matching pair of neurons is referred to as the Best Matching Unit (BMU) [48], [26].

This research explores both Euclidean and Dynamic Time Warping (DTW) matching methodologies for creating SOM clusters. The selection of the appropriate algorithm for cluster formation is the ultimate objective. In the initial stage of the Euclidean matching model, the input comprises d-dimensional Mobile Network time series data (with d input units). Therefore, the input patterns are represented as $x = \{x_i : i = 1, 2, 3, 4, ...d\}$.. The weights connecting the input unit to the computing layer's neurons are denoted by w_{ij}, where j ranges from 1 to N and i ranges from 1 to d.

Consider N as the total number of neurons. The Euclidean distance (ED) between the input vector x and the weight vector w_j for each neuron j can be computed as follows:

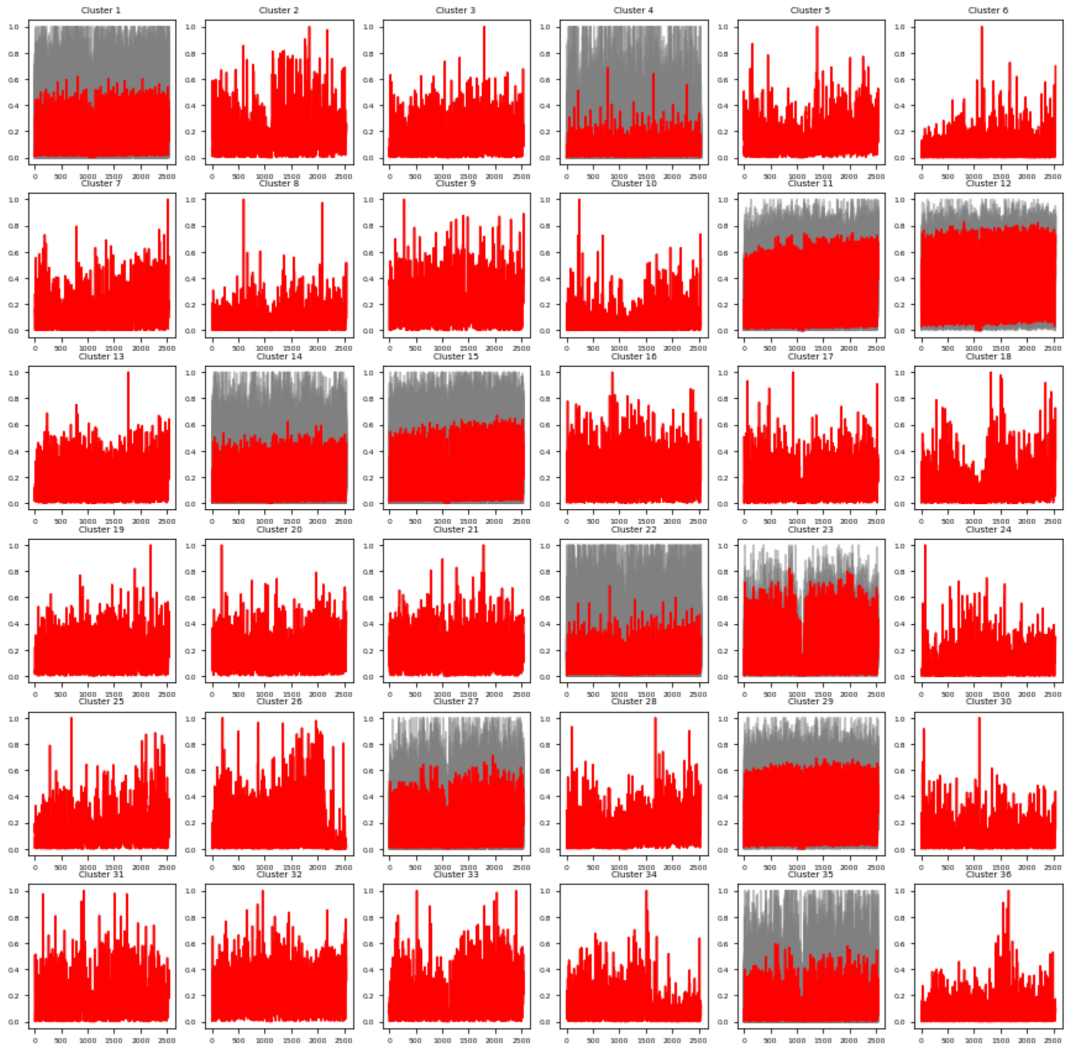

Figure 4.7: DTW Matching based 36 SOM cluster according to eNodeB hourly time series data pattern of 351 days.

$$ED_j\left(x\right) = \sqrt{\sum_{i=1}^{d}\left(x_i - w_{ij}\right) \wedge 2} \tag{4.7}$$

Utilizing equation (4.7), the Euclidean distance matching results in the division of 890 eNodeBs into 36 clusters. Equation (4.7) is analogous to the Pythagorean theorem in Cartesian coordinates, representing the distance between two points. However, this method has limitations; while combining eNodeBs with similar traffic loads into clusters, the SOM-Euclidean matching considers only the endpoints. This might not accurately depict the most representative cluster trend-line for all eNodeBs [48], [28].

To surmount the shortcomings of Euclidean Matching-based SOM clustering, an alternative approach is proposed: Dynamic Time Warping (DTW)-driven SOM clustering. Unlike Euclidean distance, DTW-based clustering isn't confined to endpoints; it involves a comprehensive evaluation. Within DTW, the neuron corresponding to the Best Matching Unit (BMU) is sought using a minimal DTW sample from eNodeB-wise traffic data (as depicted in Fig. 4.7). Moreover, the DTW-based distance method incorporates a distance decay kernel function [50].

$$\mathrm{W}_{dt^{New}} = \mathrm{W}_{dt^{previous}} + \Theta \bullet \mathrm{K}_{rs} \bullet (x - \mathrm{W}_{dt^{previous}}) \tag{4.8}$$

The DTW matching algorithm updates the model using learning rates, in contrast

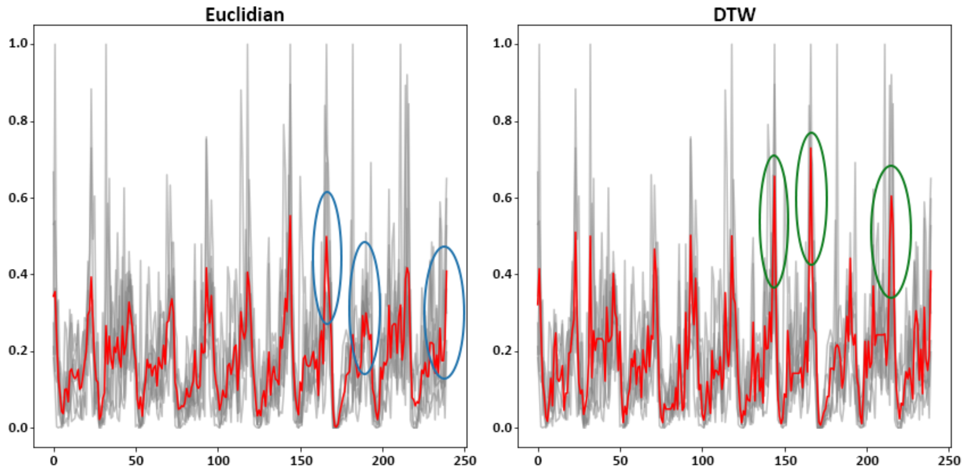

Figure 4.8: DTW-based cluster represented traffic trend line able to capture each spike or pattern of long traffic trend (circular green marked), whereas Euclidean based clustered trend line missed those details.

to Euclidean matching which merely computes distances between two points. Consequently, DTW-SOM affords greater accuracy in generating cluster representative trend-lines, as illustrated in Fig. 4.8.

4.4.2 Deep Learning Algorithms for Traffic and Utilization Prediction

In this section, we embark on an exploration of the intricate landscape of multivariate deep learning algorithms—a realm that stands as the cornerstone of our endeavor to achieve unparalleled accuracy in the prediction of cellular network traffic time series. Our journey traverses the integration of recurrent neural networks (RNNs). The fusion of these methodologies not only amplifies the predictive capabilities of our models but also unveils new vistas in predictive accuracy. With an overarching commitment to precision, we have scientifically crafted an architecture that resonates with the network-specific complexities of cellular data traffic prediction.

Architecture for Multivariate Time Series Prediction

Cellular network traffic is influenced by various factors beyond the immediate historical trend. Numerous variables can impact the data traffic volume of a specific eNodeB. For instance, traffic might experience a significant drop if an eNodeB experiences extended downtime. Moreover, factors like local social and religious events can lead to increased gatherings under specific eNodeBs, resulting in heightened traffic. Acknowledging these complex network dimensioning challenges, we embarked on a path of predicting time-series traffic through a Multivariate input-based approach

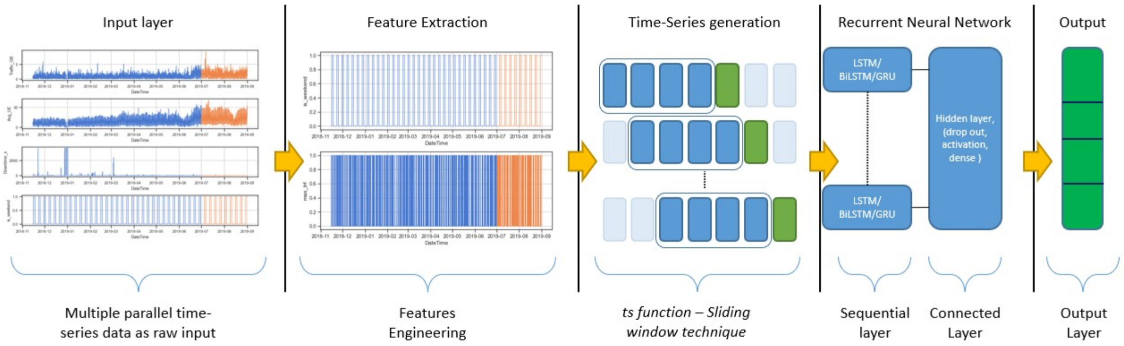

Figure 4.9: Architecture of the proposed multivariate Deep Neural Network for multiple parallel time-series prediction.

[17].

Taking into account these diverse influencing elements for traffic forecasting, let us denote the relevant variables as $x_{1,t}, x_{2,t}, \ldots, x_{k,t}$. At time t, the traffic $T_{1,t}$ can be represented as follows (equation (4.9)):

$$\mathrm{T}_{1,t} = f1\left(\mathrm{x}_{1,t}, \mathrm{x}_{2,t}, \cdots, \mathrm{x}_{k,t}, \mathrm{x}_{1,t-1}, \mathrm{x}_{2,t-1}, \cdots, \mathrm{x}_{k,t-1} \cdots\right) \tag{4.9}$$

Upon forecasting the traffic $T_{1,t}$, the subsequent time $t+1$ traffic becomes reliant on the entirety of the preceding variables. Given this notion, equation (10) can be extended for predicting the $t+k$ time traffic $T_{k,t+k}$:

$$\mathrm{T}_{k,t+k} = \mathrm{f}_k\left(\mathrm{x}_{1,t}, \mathrm{x}_{2,t}, \cdots, \mathrm{x}_{k,t}, \mathrm{x}_{1,t-1}, \mathrm{x}_{2,t-1}, \cdots, \mathrm{x}_{k,t-1} \cdots\right) \tag{4.10}$$

Aligned with the principles of multivariate time series analysis, where variables interdependently rely on their past values and other associated features [17], the goal of multivariate time series prediction, akin to univariate cases, is to anticipate data trends. However, the inclusion of additional correlated parameters allows the multivariate approach to yield more precise outcomes. This research incorporates these principles in its approach.

Algorithm 1: Time-Series generation with sliding window technique

Data:
A: array of traffic and feature
p: number of days in past as sliding window
f: number of total features
Result: return array of X and target Y

initialization;
$x, y \leftarrow 0$;
for $i \leftarrow p$ **to** $length(A)$ **do**
    $append(A[i-p:i, 0:f])toX$
    $append(A[i:i+1, 0])toY$
end
return X, Y

Illustrated in Figure 4.9 is the architecture of the deep neural network model. The process initiates with the collection of raw time-series input data, from which salient features are extracted. Subsequently, we employ the sliding window technique algorithm to generate time series from the feature-extracted data. The sliding window technique operates on historical time series data up to $N - 1$ points [14]. Following feature extraction, the ts function generates time series data (as depicted in Figure 4.9), a process further elucidated in Algorithm 1.

Algorithm 1 employs a comprehensive approach, integrating multivariate input data encompassing historical time-series traffic, downtime, and user counts denoted as X, along with the output data (traffic at $t+1$) denoted as Y. The dataset is partitioned, assigning 79% for training purposes, while reserving the remaining 21% for test data. The algorithm harnesses the sliding window technique, orchestrating a sequential shift of both X and Y across windows, orchestrated in correspondence to array $A[i..]$. This procedural operation is effectively illustrated in the third segment of Figure 4.9.

Optimal Model Selection for Predictions

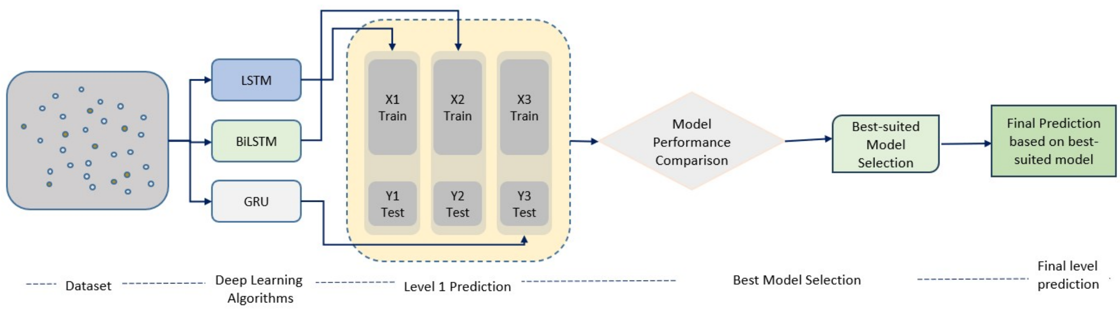

Figure 4.10: Best-suited Model Selection Approach.

The next crucial phase in our system model development centers on training the prediction model. Here, we determine the most suitable model for training eNodeB-level data from a range of deep learning algorithms: LSTM (Long Short-Term Memory), BiLSTM (Bidirectional LSTM), and GRU (Gated Recurrent Unit). These algorithms are well-regarded for their prowess in time-series prediction and are synergistically combined to enhance accuracy. If you're interested in a deeper dive into these algorithms, you can refer to the theoretical background section.

Our study introduces a method for selecting the best-fitting model from various Recurrent Neural Networks (RNNs) to construct a robust predictive model for data traffic forecasting. LSTM, BiLSTM, and GRU lay the foundation for our multivariate Deep Neural networks, designed to make simultaneous parallel time-series predictions, as illustrated in Figure 4.9. This selection strategy stems from a meticulous analysis of the RNN model's performance, specifically evaluated for individual eNodeBs. Ultimately, we pinpoint a single model that exhibits optimal performance across the majority of eNodeBs. The objective here is to bolster predictive precision, simplify the model, and enhance its overall practicality.

Our research journey culminates in the prediction phase. Equipped with a trained model, we embark on predicting traffic patterns through deep regression. These predicted traffic values serve as the cornerstone for utilization forecasts, facilitating a comprehensive comparison of overall cluster utilization. This pivotal step marks the heart of our research—an endeavor where we successfully forecast mobile network traffic demand, estimate resource utilization, and engage in enlightening comparisons across clustered eNodeBs. The approach of identifying the best-fitting model among deep learning algorithms and employing innovative data analysis techniques underscores the creation of an efficient cellular network traffic prediction system model. Through this multifaceted approach, our model is primed to provide accuracy and valuable insights into cellular network traffic forecasting.

4.4.3 Network Optimization Decisions from Predictions

Having predicted the crucial network components for network design, our subsequent task is to scientifically develop heuristics for network optimization. As previously discussed in an earlier chapter, network demand is continually on the rise, and Mobile Network Operators (MNOs) strive to optimize their networks in a manner that minimizes investment costs.

Various popular solutions exist for network capacity expansion scenarios, including cell splitting solutions, dynamic power allocation solutions, additional spectrum resource allocation solutions, or the deployment of new eNodeBs to cater to the increased demand [7] [22] [11] .

These solutions can broadly be categorized into two sections. Leveraging the predicted PRB utilization assessment, we can classify network optimization techniques into either "step-up" or "soft parameter tuning" approaches. As a long-term, sustainable solution, we propose the deployment of new eNodeBs to maintain QoS benchmarks [61]. Also, we need to derive a threshold point for PRB utilization; at that point, quality of service breakdown starts or throughput degrades rapidly. Further details about this optimized LTE radio parameter estimation are briefly discussed in Chapter 6.

In summary, Chapter 4 has unravelled the intricacies of methodology and the underlying system model, shedding light on the core mechanisms that drive our pursuit of accurate cellular network traffic prediction. The journey commenced with data aggregation, a critical step that laid the groundwork for subsequent analysis. Through feature correlation plotting, we unearthed insightful relationships between variables, setting the stage for informed decision-making.

The SOM-DTW clustering technique emerged as a powerful tool for unravelling patterns within unstructured traffic data. This chapter seamlessly transitioned into the realm of multivariate deep learning algorithms, where the architecture for time series prediction was meticulously elucidated. Choosing the best-suited model strategy introduced within Recurrent Neural Networks (RNNs) promises to elevate our predictive models to new heights.

The system model, offered a comprehensive overview of the stages that lead to cellular network traffic prediction and optimization decisions. From data collection and exploratory data analysis to feature engineering, training deep learning models, forecasting traffic and utilization, and optimization decisions, each phase was expounded upon in detail.

This chapter, an embodiment of structured methodology and a coherent system model, not only equips us with the tools to navigate cellular network data complexities but also paves the path for accurate predictions and enhanced resource utilization within the network. The fusion of innovation, deep learning, and rigorous analysis underscores our mission to unlock insights into the dynamic world of cellular network traffic prediction.

Chapter 5

Experimental Outcomes and Performance Evaluation

5.1 Introduction

This chapter takes us on a journey to evaluate how well our deep learning models perform and to uncover the outcomes of our experiments for predicting mobile network traffic data. We're exploring this territory because predicting mobile network traffic and understanding its utilization is becoming increasingly vital as well as a necessary step for network planning and optimization. This chapter is like a treasure hunt, where we'll follow the path from our research methods to the actual results we've achieved.

In this chapter, we'll carefully examine how our models performed and the selection process of the best-suited model. We'll use numbers, graphs, and comparisons to understand how well our predictions match the real data. As we continue, we'll dive into the details of our analysis. We'll talk about how our different models did and how they worked together. We'll present the actual numbers and predictions side by side to see how close they are. Through this process, we'll uncover how successful our methods are in real-world situations. So, let's embark on this journey and explore the insights awaiting us in the following sections.

5.2 Experimental Setup

The experimental setup for this best-suited model selection has been successfully configured using TensorFlow, scikit-learn, and several standard Python libraries such as pandas, seaborn, among others. This setup is tailored for compatibility with Windows 10 as the operating system. The hardware specifications include an AMD Ryzen 5 3600 processor, 32GB of RAM, and an RTX 3070 GPU. The model's parameter configuration, which ultimately yields the best evaluation score, is as follows:

- For each model, the number of epochs: 100
- Batch size: 128

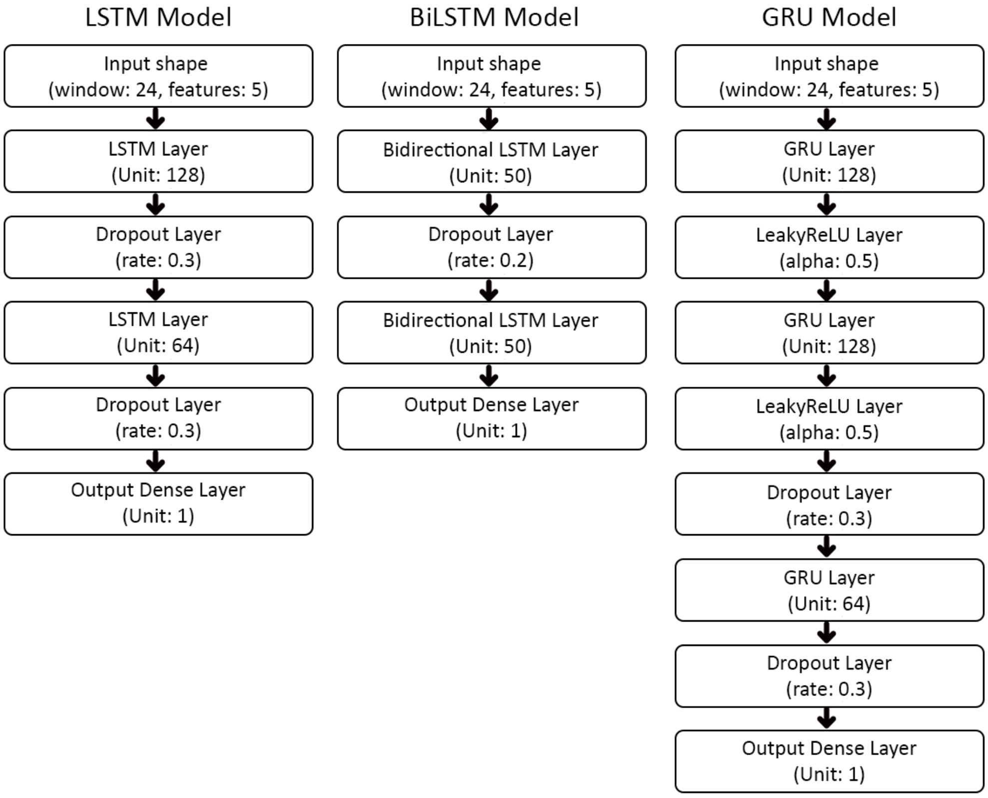

Figure 5.1: Sequential model for LSTM, BiLSTM, and GRU.

- Adam optimizer employed with a learning rate of 0.001

To ensure the training process is optimized and efficient, specific callback mechanisms have been integrated:

- EarlyStopping: This halts training when a monitored metric no longer shows improvement.
- ModelCheckpoint: This facilitates the preservation of Keras models or their weights at regular intervals.
- ReduceLROnPlateau: This dynamically reduces the learning rate when a metric's improvement reaches a plateau.

Fig 5.1 illustrates the proposed architecture of the multivariate LSTM, BiLSTM, and GRU models designed for predicting multiple time-series instances. These models are configured with five features and 24-time steps for the prediction process. The entire ensemble of 890 nodes across various sites undergoes training through this tailored model and proceeds to generate predictions for the subsequent 62-day period.

5.3 Experimental Outcomes

The experimental results of forecasting traffic using distinct deep learning algorithms are showcased in Fig. 5.2 and 5.3. These figures present sample outcomes for two different eNodeBs, highlighting the predictive performance of each algorithm.

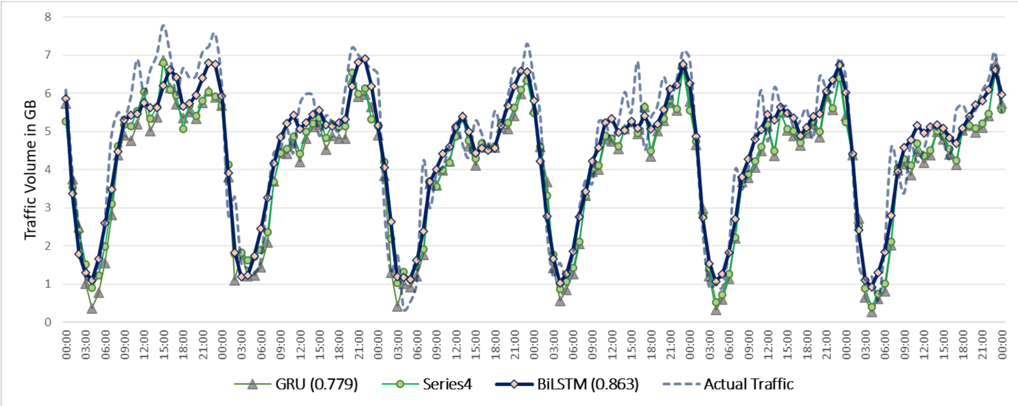

Figure 5.2: Actual vs. Predicted Traffic based on different Model.

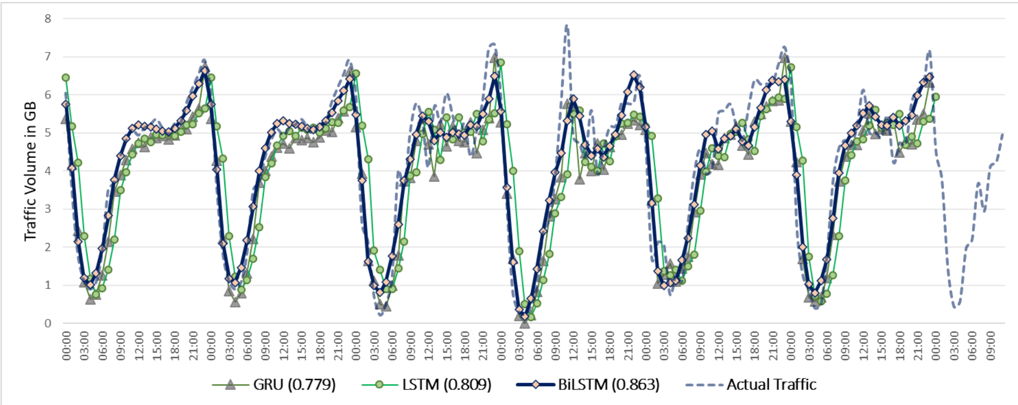

Figure 5.3: Actual vs. Predicted Traffic based on different Model.

This analysis is conducted across all 890 eNodeBs in our dataset, allowing us to comprehensively compare the prediction outcomes. While the deep learning algorithms exhibit similar performance overall, we remain vigilant about even minor spikes due to their significance in precision-driven design. As a result, we delve into a more scientific approach for evaluating performance in the subsequent section.

5.4 Performance of Deep Learning Models

Leveraging the prosed system architecture and trained model, we extend our capabilities to predict the subsequent 62 days' traffic and anticipate the utilization of each eNodeB across 36 distinct clusters within the total 890 eNodeBs. This process unfolds in two stages. Initially, we forecast traffic for each individual eNodeB using the three considered deep learning algorithms: LSTM, BiLSTM, and GRU. Subsequently, we scrutinize the compatibility of each deep learning model with specific eNodeBs, aiming to discern which models exhibit superior fitting. This endeavor culminates in the identification of the most suitable models for the majority of eNodeBs, enabling us to ascertain the collective performance of all three models.
Evaluating the model's overall accuracy Evaluation Criteria are combined and shown

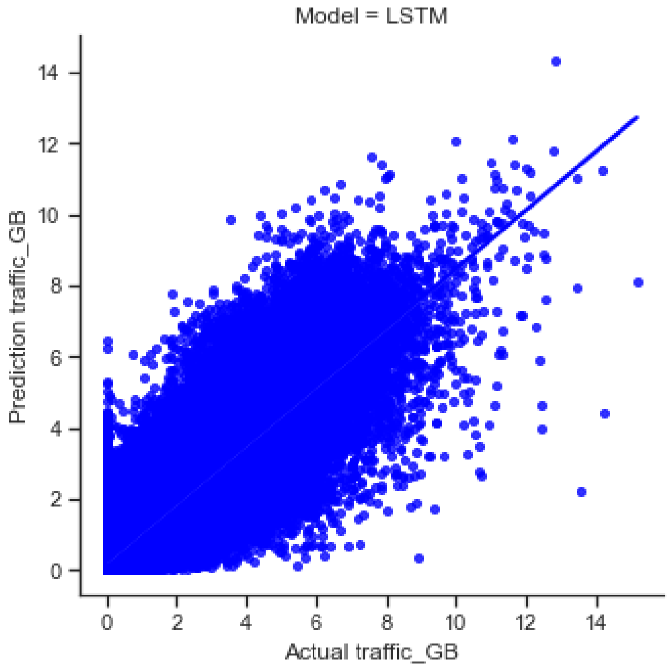

Figure 5.4: Regression plots of the LSTM models at the training phase.

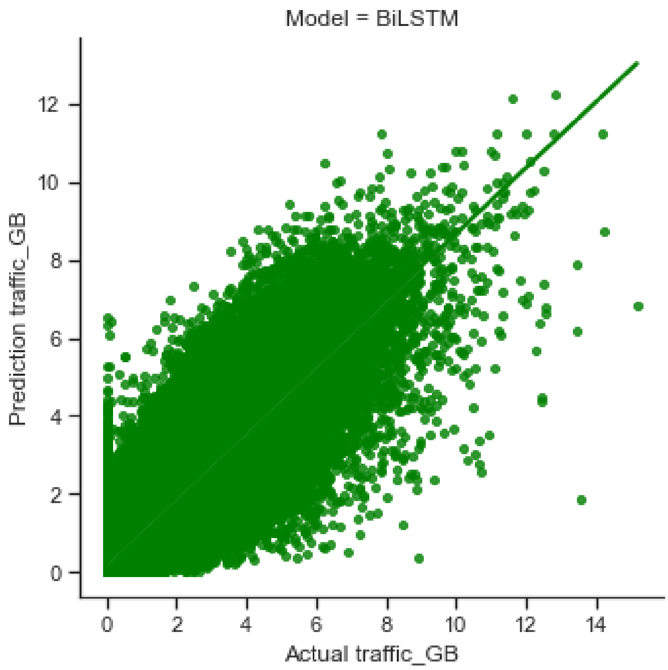

Figure 5.5: Regression plots of the BiLSTM models at the training phase.

Model	MSE	MAE	RMSE	R^2
LSTM	0.4478	0.4355	0.6692	0.7635
GRU	0.4461	0.4346	0.6679	0.7644
BiLSTM	0.3922	0.4158	0.6262	0.7929

Table 5.1: Performance of the model in the testing phase.

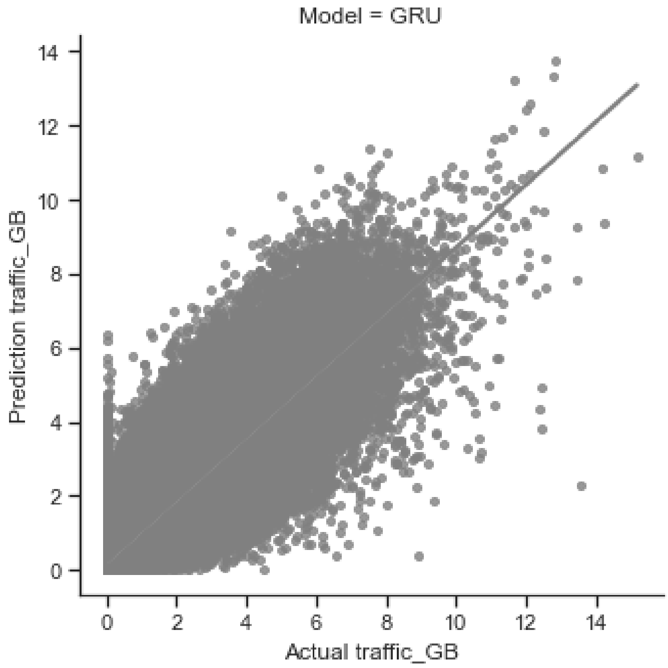

Figure 5.6: Regression plots of the GRU models at the training phase.

in Fig. 5.4 , 5.5 , 5.6. This plot is used to find the predicted and actual values relationship. Also, Table 1 presents the testing results of the proposed models.

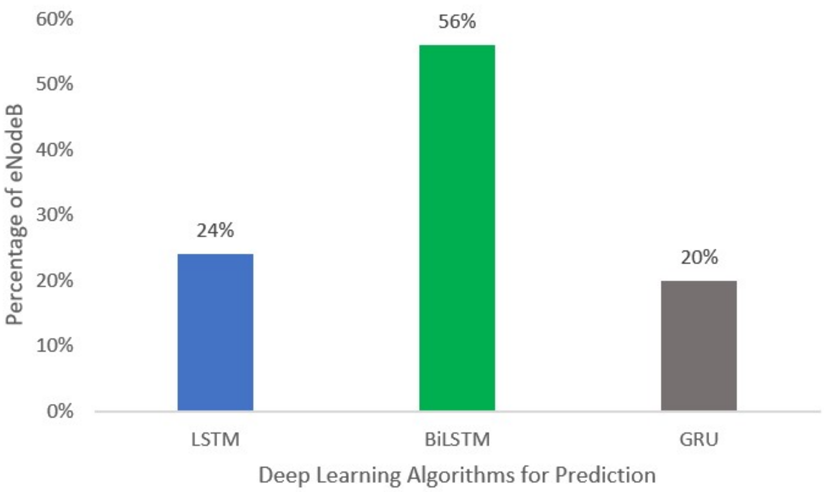

Figure 5.7: Percentage of eNodeB wise Best-suited Model.

Analyzing the data collected from each node, we find that among the 890 nodes, BiLSTM delivers promising results in 65% of cases, followed by LSTM with 24%, and GRU with 20%. Fig. 5.7 and Table 5.1 clearly illustrate that the BiLSTM model emerges as the dominant performer. BiLSTMs impressive R^2 value of 0.7929 outperforms other comparable models, positioning it as the **Best-suited** model to work further.

5.5 Best-suited Model Driven Outcome

Armed with this intricately designed system architecture and our devised best-suited model (which is BiLSTM), we embarked on the voyage of forecasting the forthcoming two months' traffic and predicting the utilization patterns of specific eNodeBs within the designated clusters. In the ensuing Table 5.2, a detailed panorama emerges:

- Summation of Vol_{Act}: Pre-actual Total traffic (GB) over the past 2 months.
- Summation of Vol_{Pred}: Post-predicted Total traffic (GB) for the ensuing 2 months
- Count PRB Utilization actual U_{Act}: The tally of sample utilizations exceeding 70% in the pre-timeframe.
- Count PRB Utilization predict U_{Pred}: The count of sample utilizations surpassing 70% in the post-timeframe.
- The % column illustrates the percentage change between the pre and post traffic.

Table 5.2 comprehensively presents the traffic details for the past and forthcoming 2 months across the 36 clusters. Moreover, we applied regression techniques to anticipate the count of instances with high utilization within each cluster (details refer to Chapter 3, subsection 3.5.4). This information is also included in the table. Notably, the table 5.2 offers valuable insights into the occurrence of sample utilization surpassing the 70% threshold during the specified timeframes. In the subsequent chapter, we dedicate a section to delve into the rationale behind setting the utilization threshold at 70% within this specific LTE network configuration.

5.6 Discussion of Performance Evaluation

In this chapter, we have provided an overview of the approach we employed to select appropriate deep learning models tailored for each individual eNodeB, enabling accurate traffic prediction. Upon identifying the optimal deep learning model for eNodeB granular prediction, we compared diverse deep learning algorithms to sort out the best-suited model for comprehensive training.
Subsequently, we harnessed the trained model to forecast traffic volume (Vol) and then predict LTE channel Physical Resource Block (PRB) Utilization (U) for distinct eNodeB clusters through the application of the linear regression technique. PRB utilization plays a pivotal role in maintaining Quality of Services (QoS) by ensuring adequate throughput benchmarks [61]. Our best-suited BiLSTM model and regression technique demonstrated the ability to accurately capture the count of highly utilized samples with minimal discrepancies (refer to Table 3).
Having forecasted both traffic and utilization, we delved deeper into addressing cases of overutilization by employing network optimization techniques. These techniques involved estimating radio parameters with a focus on minimizing Capital Expenditure (CapEx) investments for Mobile Network Operators, a discussion further elaborated upon in Chapter 6.

Cluster	Vol_{Act}	Vol_{Pred}	%	U_{Act}	U_{Pred}
1	234193	242947	3.74	1015	1697
2	2215	2276	2.77	86	14
3	1092	1126	3.17	0	0
4	88575	86776	-2.03	64	197
5	1565	1599	2.2	0	0
6	464	461	-0.62	0	0
7	1219	1218	-0.04	0	0
8	719	852	18.51	0	0
9	1067	1123	5.25	0	0
10	424	355	-16.31	0	0
11	128599	140779	9.47	8202	8846
12	103936	111979	7.74	11570	14646
13	3485	3630	4.18	2	0
14	133977	138778	3.58	1470	2053
15	174525	182734	4.7	2704	3105
16	1085	1791	65.13	0	0
17	1703	1699	-0.23	0	0
18	1068	1301	21.85	0	0
19	1343	1326	-1.23	0	0
20	2461	2308	-6.22	9	9
21	1574	1579	0.33	0	0
22	178059	182311	2.39	329	756
23	23017	23647	2.74	1193	547
24	321	400	24.44	0	0
25	919	896	-2.56	0	0
26	853	847	-0.63	1	0
27	118642	128411	8.23	3146	4626
28	1739	2164	24.42	0	0
29	113749	122865	8.01	5747	5883
30	1189	963	-19.04	2	0
31	1496	1520	1.61	0	0
32	3473	3697	6.45	105	150
33	2015	2060	2.23	8	0
34	815	758	-6.95	0	3
35	106649	108547	1.78	703	657
36	521	473	-9.25	0	0

Table 5.2: Best-suited Model (BiLSTM) and regression driven outcome.

Chapter 6

Optimized LTE Radio Parameter Estimation

6.1 Introduction

As we embark on a deeper journey through the intricate landscape of optimizing LTE networks, this chapter casts a resolute spotlight on the paramount importance of estimating LTE radio parameters. These parameters, often hidden behind the scenes, play an instrumental role in elevating the efficiency and quality of mobile network services. Think of them as the architects behind the scenes, crafting the blueprint for a seamless and robust network experience.

In the realm of Long-Term Evolution (LTE) networks, the configuration of radio parameters holds a profound influence on network reliability and performance. This influence ripples across the entire network, affecting everything from call clarity to data speeds. In essence, these parameters act as the guardians of Quality of Service (QoS), ensuring that your mobile experience is nothing short of exceptional.

Yet, the path to network optimization is riddled with challenges. The delicate equilibrium between delivering impeccable QoS and keeping operational costs in check is a puzzle that Mobile Network Operators (MNOs) constantly grapple with. Enter the art of estimating LTE radio parameters – a strategic tool that empowers MNOs to make informed decisions. By understanding the intricacies of these parameters, MNOs can walk the fine line between network excellence and prudent resource allocation.

Within the chapters ahead, we'll journey through the foundational aspects of LTE radio parameters, diving into the heart of their functionalities. We'll unveil a groundbreaking algorithm designed to estimate QoS parameters, an algorithm that promises to be a game-changer for the industry. But this isn't just about innovation – it's about practicality. We'll illustrate how this algorithm can strategically reduce the investment demands for network planning. The ultimate goal? To lay out a strategic roadmap that allows MNOs to strike the perfect harmony between optimal network performance and astute cost-effectiveness. So, let's navigate this path together and uncover the blueprint for a future where seamless connectivity and intelligent resource utilization coexist harmoniously.

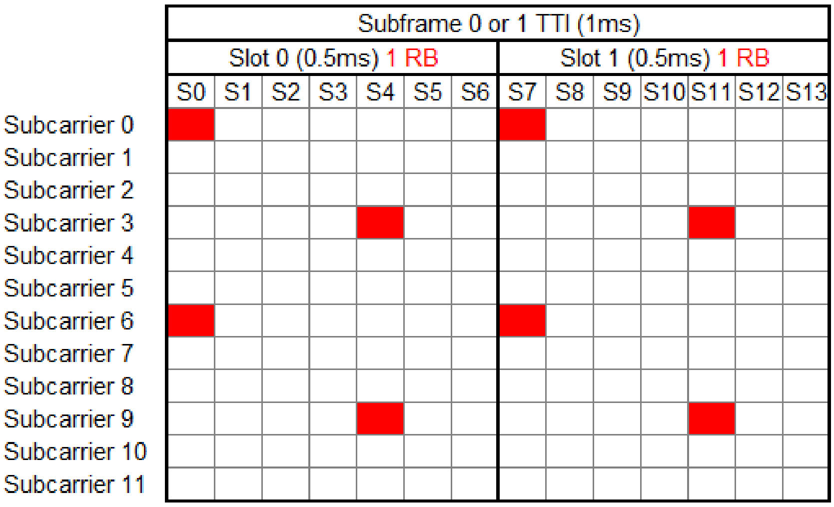

Figure 6.1: Physical Resource Block the LTE air interface.

6.2 LTE Air Interface

4G cellular networks embrace the Long-Term Evolution (LTE) wireless communication standard. Within LTE, the wireless channel facilitating data transmission and reception between user equipment (UE) and the LTE base station (eNodeB) is referred to as the air interface. LTE strategically segments available frequency and temporal resources into smaller units termed Resource Blocks (RBs) and Physical Resource Blocks (PRBs) to proficiently manage the finite radio spectrum.

Resource Block known as The fundamental radio resource unit for LTE. Within a specified bandwidth, it represents a certain portion of the frequency spectrum. Each RB in the LTE standard has 12 subcarriers and covers a frequency range of 180 kHz[43], [30]. An RB's time duration varies depending on the type of specified subframe, but it normally corresponds to one slot, which lasts for 0.5 ms.

- 1 RB = 12(Sub-carriers) x 7 (Symbols) = 84 Resource Elements. (For Normal CP: 7 symbols)
- 1 RB = 12(Sub-carriers) x 6 (Symbols) = 72 Resource Elements. (For Extended CP: 6 symbols)

Physical Resource Blocks (PRB) are allotted collectively in A particular group of RBs that are in both the time and frequency domains. A PRB is, in other words, a two-dimensional unit for resource allocation. The PRB occupies a frequency range of 180 kHz and is made up of 12 sub-carriers that follow one another throughout a period of time (0.5 ms).
The eNodeB assigns PRBs to specific UEs for downlink data transmission, allowing them to transmit data. This allocation mechanism extends to uplinks as well, where PRBs are assigned to UEs for data transmission from the UE to the base station. The LTE scheduler dynamically allocates PRBs to UEs based on their quality of service specifications and prevailing channel conditions.
By segmenting the available spectrum into PRBs, LTE efficiently manages radio resource distribution, ensuring users receive sufficient bandwidth to fulfill their communication requirements. This division empowers the system to adapt data rates

according to channel conditions for each UE, facilitated through the utilization of PRBs for adaptive modulation and coding techniques.
In essence, LTE Resource Blocks and Physical Resource Blocks are essential components of the air interface, fostering efficient and adaptable radio resource utilization. This symbiotic relationship provides customers with high-speed data connectivity and a diverse array of services.

6.3 LTE Radio QoS Parameters

LTE networks deploy Quality of Service (QoS) mechanisms to ensure various types of traffic receive appropriate levels of service based on their distinctive demands. These QoS mechanisms enable network operators to effectively allocate resources and prioritize services in alignment with their value and performance criteria. The following key LTE QoS parameters shed light on their significance and interpretation:
QoS Class Identifier (QCI): Assigned to specific traffic types or services, the QCI is a numeric value ranging from 1 to 9. It encapsulates the priority level and associated QoS features. Different services such as telephony, video streaming, and best-effort data are allocated distinct QCIs. Parameters like packet latency, packet loss rate, and data rate are determined by predefined specifications for each QCI. A higher QCI value signifies elevated priority and enhanced QoS [44], [47].
Bit Rate: The highest achievable data rate for a specific QoS flow is represented in bits per second (bps) and linked to various QCIs. For instance, speech services may require lower bit rates compared to streaming high-definition video.
Block Error Rate (BLER): Calculated as the ratio of incorrect blocks to total blocks, BLER relies on the cyclic redundancy check (CRC) method to detect errors in the transport block. If the calculation yields undesired results, the receiver initiates a Hybrid Automatic Repeat Request (HARQ) Negative Acknowledgment (NACK) for retransmission. Ensuring a standard BLER objective of 10% guarantees 90% effective transmission at the receiver's end, upholding service quality [5]. If the BLER target isn't met, additional retransmissions may be needed, resulting in increased radio resource consumption. Optimal resource scheduling strategies are crucial to maintain desired QoS benchmarks of maximizing throughput, ensuring user equity, and minimizing Block Error Ratio [37].
Resource Allocation: Based on QoS requisites, LTE dynamically allocates radio resources such as PRBs (Physical Resource Blocks). Critical services receive the resources necessary for consistent performance when flows with higher priorities or QCI require additional resources. The scheduler continually adapts PRB allocation to accommodate evolving network load and environmental variables. Resources are allocated preferentially to flows with higher priorities or stringent QoS requirements, ensuring essential services like voice calls and real-time video streaming receive the necessary resources to maintain their quality.
PRB Utilization: This metric gauges the efficiency of assigned PRBs in transmitting data, control signals, and other communication elements within the network. PRB utilization showcases the effective use of radio spectrum to fulfill communication needs of user equipment (UE) and optimize network performance. Low PRB utilization may indicate underutilized or wasted resources, while high utilization reflects the substantial use of allocated resources. PRB utilization is influenced by

factors such as traffic load, service types, QoS requirements, channel conditions, scheduling algorithms, and network congestion.

6.4 Identifying QoS Breakdown Point from PRB Utilization

The primary aim of this subsection is to pinpoint the threshold point at which the breakdown of quality of service or user throughput User_TP occurs within the PRB utilization graph. This facet is of utmost importance in diverse networks. By identifying the threshold point, network engineers can make informed decisions regarding the adoption of either immediate step-up measures or longer-term solutions (Fig. 6.2). To conduct this analysis, we gathered PRB Utilization and User Throughput

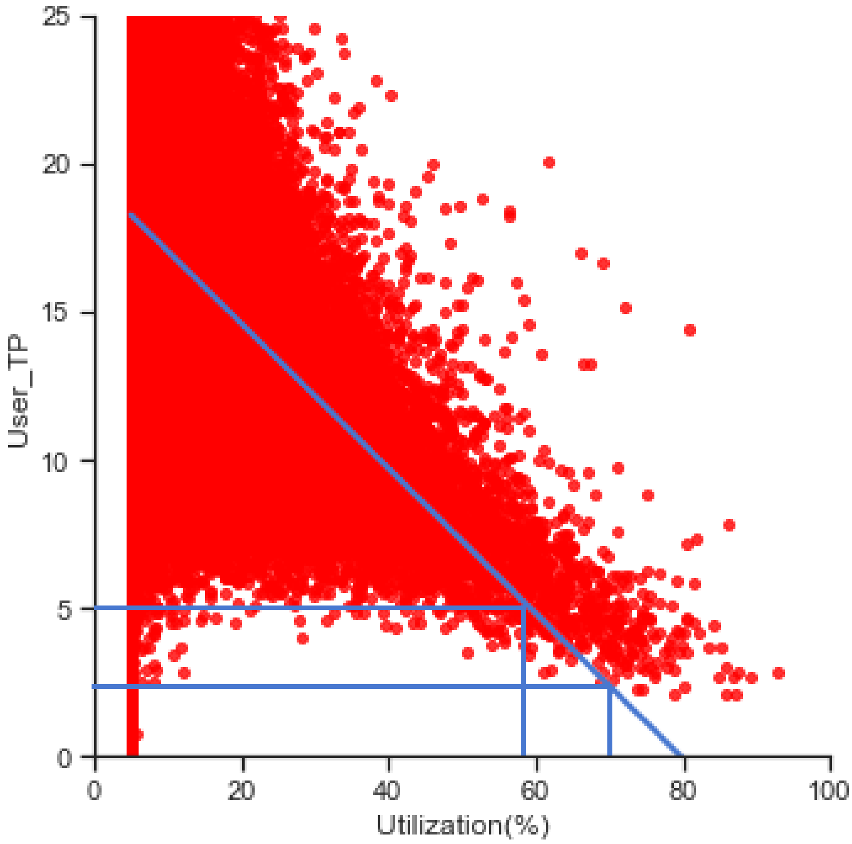

Figure 6.2: User Throughput (User_TP) vs. PRB Utilization Graph.

(User_TP) data from 890 eNodeBs over a span of 120 days, collected at hourly intervals from a live operator network. Our findings revealed that when PRB Utilization reaches or surpasses 70%, user throughput experiences a 50% degradation (from 5 Mbps to 2.5 Mbps). This level of degradation corresponds to a significant impact on user experiences. In such instances, it becomes imperative for responsible engineers to promptly initiate step-up solutions, aiming to reduce PRB utilization and enhance user satisfaction. Similarly, as PRB utilization exceeds 80%, our analysis indicates a nearly negligible user throughput, effectively representing a lack of Quality of Service (QoS). It's important to note that this threshold point may vary based

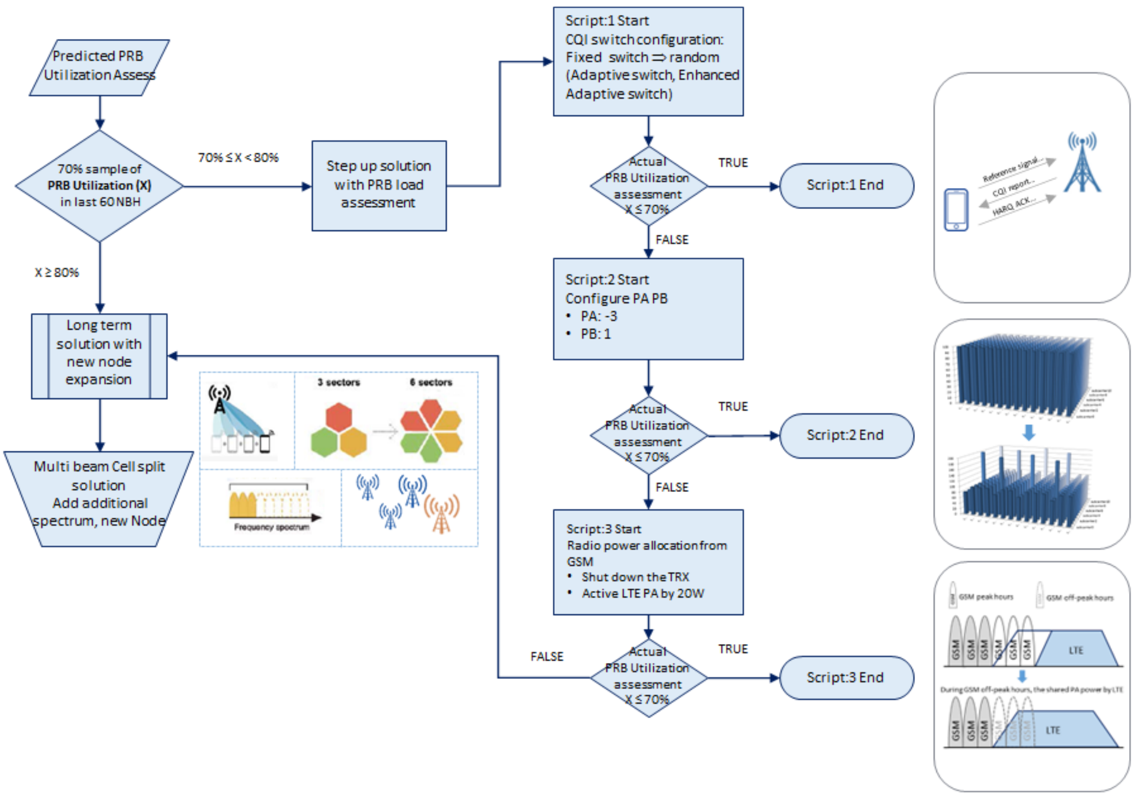

Figure 6.3: Algorithm for predicted PRB Utilization based QoS Parameter Estimation, (a) CQI switch adjustment, (b) Resource block power allocation, (c) Radio power allocation for LTE from GSM.

on specific network characteristics. For reference, our analysis was conducted on a live LTE network configured at 1800 Mhz, with a per eNodeB frequency bandwidth of 10 Mhz.

6.5 Proposed Algorithm for QoS Parameter Estimation

We have put forth an algorithm for radio parameter estimation grounded in predicted PRB utilization. The algorithm initiates with an initial step where the capacity requirement of each eNodeB (Fig. 18) is estimated based on an evaluation of 60 days of projected PRB utilization. If 70% of the sample of 60 days' eNodeB busy hours (NBH) equals or surpasses 80% PRB utilization (X), it is recommended to implement hard capacity expansion through solutions like Multibeam Cell Split Solution [16], New Spectrum addition [57], and the planning and deployment of new nodes [57], [24]. However, the addition of new nodes incurs higher capital expenses, making it a final option for maintaining QoS [52]. In cases where PRB utilization (X) falls within the range of $70\% \leq X < 80\%$,, three soft step-up solutions are triggered based on the estimated parameters to mitigate PRB utilization and uphold QoS standards.

The assessment of actual PRB utilization takes place during the execution of the soft step-up solutions. The rationale behind the selection of 70% and 80% as thresholds for activating action points for Optimized QoS parameter estimation is discussed later in this paper. STEP UP SOLUTION 1: ADJUSTMENT OF CQI SWITCH: Downlink CQI adjustment, interactively compensates for inaccurate CQIs reported by UEs, optimizes MCS (Modulation and Coding Scheme) selection, and increases throughput [39]. If the network has moderate or heavy loads in the downlink, the downlink user-perceived rate will increase by 1% to 3% after CQI adjustment Fig. 6.3(a) [14].

SCRIPT 1 (CQI SWITCH ADJUSTMENT)
MOD CELLALGOSWITCH: LocalCellId=0,
CqiAdjAlgoSwitch = DlVarIBLERtargetSwitch-1
CqiAdjAlgoSwitch = DlEnVarIblerTargetSwitch-1;

`DlVarIBLERtargetSwitch:`In Adaptive configuration, The downlink target initial block error rate (IBLER) is adaptively adjusted from a fixed configuration based on the Transport Block Size (TBS) to improve spectral efficiency. The higher the block size, the higher the throughput. In this scenario, the eNodeB adjusts the target IBLER to 10% for UEs with large packet services at non-edge locations and 30% for UEs with small packet services at edge locations.

`DlEnVarIblerTargetSwitch:`In Enhanced Adaptive configuration, the downlink target IBLER is adaptively adjusted from a fixed configuration based on TBS and as well as CQI fluctuation. In this scenario, the eNodeB adjusts the target IBLER to 5% and 10% for slightly fluctuated CQI values. With heavily fluctuated CQI values, the target IBLER is 10% and 30% [55].

STEP UP SOLUTION 2: P_A AND P_B POWER ALLOCATION:
4G LTE RS RE Power (RSRP power) boosting depends on ρ_B/ρ_A parameters Fig. 6.3(b). this value is determined by many parameters, such as the max power of the RRU channel (P_{max}), the number of RBs of the cell Nrb, P_A, P_B etc. Below describes the impact.

- The higher P_A Implies, the lower RS power (ERS), the smaller cell radius and the higher throughput can get.
- With changing P_A, P_B need to change as well to make full usage of power.
- The max value of ERS is determined by (P_{max}) and P_A.
- To configure RS power, first determine P_A, then RS and, P_B is determined following the Table 6.1 (for 2T cells).

The following definition from 3GPP. 36.213 protocol [21]:

- $TypeA$: the PDSCH OFDM symbol without RS
- $TypeB$: the PDSCH OFDM symbol with RS
- EA: the power of one element in $TypeA$, in W
- EB: the power of one element in $TypeB$, in W

P_A	*ERS*	ρ_B/ρ_A	P_B
0	*EA*	5/4	0
-3	*2*EA*	1	1
-1.77	*1.5*EA*	3/4	2
-6	*4*EA*	1/2	3

Table 6.1: The cell-specific parameter for P_B.

- ERS: the power of $referenceSignal$, in W
- RS: $referenceSignal = 10log\,(ERS * 1000)$, in dbm
- P_{max}: the max power of RRU channel, in W
- Nrb: the number of RBs in the cell
- $\rho_A = EA/ERS$
- $\rho_B = EB/ERS$
- $P_A = 10log\,(\rho_A) = 10log\,(EA/ERS)$
- $\rho_B/\rho_A = EB/EA$

	ρ_B/ρ_A	
P_A	One Antenna Port	Two and Four Antenna Ports
0	1	5/4
1	4/5	1
2	3/5	3/4
3	2/5	1/2

Table 6.2: The cell-specific ratio ρ_B/ρ_A.

Script 2 (P_A and P_B Power Allocation):

MOD CELLDLPCPDSCHPA: LocalCellId=0, PaPcOff=-3 dB; MOD PDSCHCFG: LocalCellId=0, Pb=1 `PaPcOff`: Indicates the PA to be used when PA adjustment for PDSCH power control is disabled, DL ICIC is disabled, and the even power distribution is used for the PDSCH [40].
`Pb`: Indicates the Energy Per Resource Element (EPRE) scaling factor index on the PDSCH. The value of this parameter and the antenna port control this scaling factor.
After executing script 2, PRB utilization (X) is expected to reduce by 70%. But if it does not happen next script will be executed.

Step Up Solution 3: Dynamic Radio Power Allocation from GSM:

As per definition, Dynamic Cell Power Off is a BSC feature [40] that enables power dynamically off or on the TRXs (in GSM Cell) based on the traffic demand of the co-coverage cell within a certain time frame. When LTE load is high, some radio power allocates from existing GSM PA to LTE PA through GSM TRX shutdown Fig. 6.3(c). In this way, the LTE network can be boosted up to 20W PA power.

Dynamic Power Allocation:
SET GCELLDYNTURNOFF: IDTYPE=BYID, CELLID=0,
TURNOFFENABLE=ENABLE,
SAMECVGCELLIDTYPE=BYID,
SAMECVGCELLID=1,
TURNOFFCELLSTRTIME= [Time PRB Util>70%],
TURNOFFCELLSTPTIME= [Time PRB Util<70%] `TURNOFFENABLE`: to enable the Dynamic Cell Power
`SAMECVGCELLIDTYPE`: Index type of a co-coverage cell. If the coverage area of a cell is under the coverage area of another cell, the cell can be disabled during off-peak hours. In this situation, another cell is considered as the co-coverage cell of the cell.
`TURNOFFCELLSTRTIME`: Start time for dynamically disabling a cell.
`TURNOFFCELLSTPTIME`: End time for dynamically disabling a cell.
After executing script 3, PRB utilization still persists above 70% then we have no other option rather implement a long-term solution with node expansion.

Algorithm 2: Decision and QoS Parameter implementation with PRB utilization assessment

---

Data:
s : *sample count in NBH*
l : *last n sample*
prb : *prb utilization*
Function QoS-Decision-Estimation(s, prb):
  if $\max_{prb} \in s_{70\%}$ & $l_{60} > 80\%$ **then return** $decision : NewNodeExpansion$;
  else
  $Script_1 \leftarrow CqiAdjAlgoSwitch$
  $Script_2 \leftarrow PowerAllocation$
  $Script_3 \leftarrow DynamicCellPower$
  for *each*, $i \in Script_i$ **do**
    $U \leftarrow \arg\max_{u \in prb} \text{Implement}(u, Script_i)$
    $decision \leftarrow$ QoS-Decision-Estimation(s,U)
  end
  return $decision$;

6.6 Minimizing Investment through the Proposed Algorithm

Imagine saving money while optimizing network planning – a challenging feat for Mobile Network Operators (MNOs). MNOs face the tightrope act of delivering great service without overspending. The solution? Our groundbreaking algorithm – a financial compass. By shrewdly distributing resources, MNOs can supercharge network capabilities without straining their budgets.
At its core, this algorithm merges advanced math with historical data to foresee future network needs. It predicts potential bottlenecks, enabling MNOs to make well-

informed choices proactively. No rushed decisions, no costly equipment splurges.
But its versatility goes beyond. This algorithm is a Swiss army knife for networks. It tweaks operations without adding extra parts, much like a mechanic fine-tuning a car. It optimizes data flow, power consumption, and resource sharing, enhancing network efficiency.
What's truly remarkable is the timing of these changes – seamless and non-disruptive. It's like a master chef enhancing a dish without changing its essence. By embracing this algorithm, MNOs discover that sweet spot where networks thrive, costs stay in check, and users stay happy.
In a nutshell, this algorithm goes beyond complex math. It employs historical data to foresee the future, introduces intelligent tweaks, and identifies the golden balance where networks flourish without financial strain. This innovation proves that clever thinking can heighten network performance without emptying pockets. As we move forward, this algorithm sets the stage for networks that are streamlined, user-friendly, and budget-conscious all at once.

6.7 A Path to Democratization

The idea of democratization become popular after it's well explanations and importance by Thomas Friedman (1999) [41]. in the early ages, cutting edge technologies like high speed internet or space technology typically developed for military [41], after that it speeded to civilians, the process is known as democratization process. However, today's scenario has changed in many cases, as example, tech giant companies like Microsoft, google, IBM, meta has mission to democratize their developed technologies by open sources, API, train up peoples by seminars, workshops and many other formal and informal ways [19].

The one major objective of this research of empowering mobile network planning through deep learning is to democratize scientific process we developed for mobile network planning among the relevant parties who become beneficial with this. In this research we have shown end to end process for traffic and utilization prediction for LTE network, not only that, how to handle those predicted traffic with our innovative algorithm of radio parameter estimations. In this process we have shown how can we select the best suited deep learning algorithm for training and performance evaluation of those models.

This research thesis is the major steppingstone of democratizing scientific process to empower mobile network planning and make the network management system more efficient. In future, based on the further interest of relevant parties, we will organize workshop, seminars on this approach we are proposing for mobile network planning as a part of democratization.

Chapter 7

Discussion and Conclusion

In this concluding chapter, we engage in a comprehensive discussion of the findings and contributions presented in this thesis. Our journey has led us through the intricate terrain of optimizing mobile network planning using the transformative power of deep learning. We explore the limitations that our proposed model encounters and delve into the future prospects that lie ahead in this realm of research. Let's reflect on how we are empowering mobile network planning through deep learning and why this path is being hailed as a route to democratization.

7.1 Discussion and Limitations: Progress and Navigating Challenges

In the realm of mobile network planning, our research acts as a guiding light, harnessing the transformative power of deep learning. As we delve into our findings, achievements, and challenges, we uncover a tapestry woven with advancements and opportunities. This section reflects on our journey while envisioning a more inclusive network planning future through the fusion of cutting-edge technologies.

7.1.1 Advancements and Transformations

In our quest for network optimization, our proposed model emerges as a potent agent of change. It operates smoothly under normal conditions, showcasing its adaptability and effectiveness. This adaptability is especially evident during busy social events, where it maintains solid performance even under high user demand.
A significant stride lies in our model's emphasis on boosting LTE capacity. This strategic pivot addresses the growing appetite for data-driven services, strengthening the network's foundation. However, this focus on LTE capacity could introduce challenges related to resource allocation for GSM networks. The coordination of these distinct networks underscores the need for a delicate balance

7.1.2 Recognizing Limits and Overcoming Challenges

To truly understand our research, we must acknowledge its limitations. While our model excels under normal circumstances, it might experience a slight dip in performance during times of intense user activity. Yet, these challenges motivate us to recalibrate and optimize the model's parameters, ensuring it adapts consistently to

changing conditions.
A crucial consideration revolves around the balance between LTE and GSM networks. Our commitment to enhancing LTE capacity raises the question of potential resource constraints for GSM networks. Addressing this challenge involves careful resource distribution and comprehensive optimization strategies, fostering a harmonious relationship between these network technologies.
As we peer into the future of network expansion, the increasing number of eNodeBs demands heightened computational capabilities. Managing this growing network landscape efficiently necessitates robust processing, highlighting the importance of optimal computational resource allocation.

7.1.3 Empowering Network Planning and Democratization

Central to our discussion is the idea of empowerment and democratization. Through deep learning, we unveil a new avenue for network planning, driven by data insights and predictive capabilities. This approach opens the door for more people to access sophisticated technologies, equipping network operators with tools to navigate complexities with precision.
Our research journey transcends limitations and challenges, echoing the spirit of progress. Every constraint becomes a catalyst for innovation, urging us to refine strategies and develop transformative solutions. As we stand at the intersection of mobile network planning's evolution, we envision a landscape where AI and predictive analytics redefine network optimization.

7.2 Conclusion and Future Research

In conclusion, this thesis embarks on an innovative journey to empower mobile network planning through the integration of deep learning. The utilization of three distinct deep learning algorithms has led to exceptional granular-level cellular network traffic prediction. A pivotal step involves comparing various deep learning algorithms and selecting the most fitting one that addresses the NP-hard optimization challenge of maximizing user throughput.
Our chosen best-suited model, BiLSTM, demonstrates an outstanding R^2 score of 0.739, surpassing other model's (for example LSTM and GRU) performance. Also, Leveraging a DTW-based self-organized map (SOM), we have streamlined the clustering of distinct eNodeB time series data. Moreover, by anchoring our analysis in the reference LTE network radio configuration, we've identified the QoS breakdown threshold point, coinciding with the PRB Utilization graph at 70%.
However, the achievements don't stop there. The introduction of a robust parameter estimation algorithm triggers dynamic capacity enhancement solutions two months ahead. This, combined with optimized radio power allocation based on forecasted LTE network traffic and PRB utilization, leads to a new era of network capacity optimization.
We dub this endeavor as a path to democratization in mobile network planning. Why? Because our approach bridges the gap between traditional network planning intricacies and a more accessible, data-driven decision-making process. By harnessing the power of deep learning, we are enabling network engineers to not

only plan and execute soft parameters more effectively but also democratize the access to cutting-edge technology for optimizing network performance. This approach significantly reduces the time lag associated with capacity expansion, enhancing customer experiences and empowering Mobile Network Operators (MNOs) to allocate resources with astute precision.

Looking ahead, our future research trajectory points to addressing the prediction of traffic peaks during social events, specifically in localized eNodeB serving areas, through innovative techniques like Restricted Boltzmann Machines (RBM) with Conditional Random Fields (CRFs). Furthermore, we aspire to achieve seamless dynamic resource allocation in complex heterogeneous networks, encompassing GSM, LTE, 5G, and beyond, based on forecasted traffic patterns and customer demand. This holistic approach aims to propel the mobile network planning landscape into a new era of efficiency and accessibility, fostering a democratic network planning ecosystem.

Bibliography

[1] I. Toyoda, F. Nuno, Y. Shimizu, and M. Umehira, "Proposal of 5/25-ghz dual band ofdm-based wireless lan for high-capacity broadband communications," in *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, IEEE, vol. 3, 2005, pp. 2104–2108.

[2] S. Adhikari, "Critical analysis of multi-antenna systems in the lte downlink," in *2009 IEEE International Conference on Internet Multimedia Services Architecture and Applications (IMSAA)*, IEEE, 2009, pp. 1–6.

[3] E. Dahlman, S. Parkvall, and J. Sköld, "Chapter 8 - radio-interface architecture," in *4G LTE/LTE-Advanced for Mobile Broadband*, E. Dahlman, S. Parkvall, and J. Sköld, Eds., Oxford: Academic Press, 2011, pp. 109–126, ISBN: 978-0-12-385489-6. DOI: https://doi.org/10.1016/B978-0-12-385489-6.00008-4. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123854896000084.

[4] *Lte; architecture enhancements for non-3gpp accesses (3gpp ts 23.402 version 10.4.0 release 10);* 3GPP – The Mobile Broadband Standard, 2011.

[5] V. D. Blondel, M. Esch, C. Chan, *et al.*, "Data for development: The d4d challenge on mobile phone data," *arXiv preprint arXiv:1210.0137*, 2012.

[6] M. Caretti, M. Crozzoli, G. Dell'Aera, and A. Orlando, "Cell splitting based on active antennas: Performance assessment for lte system," in *WAMICON 2012 IEEE Wireless & Microwave Technology Conference*, IEEE, 2012, pp. 1–5.

[7] H. Ishii, Y. Kishiyama, and H. Takahashi, "A novel architecture for lte-b: C-plane/u-plane split and phantom cell concept," in *2012 IEEE Globecom Workshops*, IEEE, 2012, pp. 624–630.

[8] A. Kumar, J. Sengupta, and Y.-f. Liu, "3gpp lte: The future of mobile broadband," *Wireless Personal Communications*, vol. 62, pp. 671–686, 2012.

[9] L. Yang, Z. Ouyang, and Y. Shi, "A modified clustering method based on self-organizing maps and its applications," *Procedia Computer Science*, vol. 9, pp. 1371–1379, 2012.

[10] J. Cao, M. Ma, H. Li, Y. Zhang, and Z. Luo, "A survey on security aspects for lte and lte-a networks," *IEEE communications surveys & tutorials*, vol. 16, no. 1, pp. 283–302, 2013.

[11] H. Lee, S. Vahid, and K. Moessner, "A survey of radio resource management for spectrum aggregation in lte-advanced," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 745–760, 2013.

[12] S. B. H. Said, M. R. Sama, K. Guillouard, *et al.*, "New control plane in 3gpp lte/epc architecture for on-demand connectivity service," in *2013 IEEE 2nd international conference on cloud networking (CloudNet)*, IEEE, 2013, pp. 205–209.

[13] J. Xiao, R. Q. Hu, Y. Qian, L. Gong, and B. Wang, "Expanding lte network spectrum with cognitive radios: From concept to implementation," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 12–19, 2013.

[14] H. Al-Zayadi, O. Lavriv, M. Klymash, and A.-S. Mushtaq, "Increase throughput by expectation channel quality indicator," in *2014 First International Scientific-Practical Conference Problems of Infocommunications Science and Technology*, IEEE, 2014, pp. 120–121.

[15] N. Morozs, T. Clarke, and D. Grace, "Intelligent secondary lte spectrum sharing in high capacity cognitive cellular systems," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, IEEE, 2015, pp. 1–2.

[16] M. B. Shahab, M. A. Wahla, and M. T. Mushtaq, "Downlink resource scheduling technique for maximized throughput with improved fairness and reduced bler in lte," in *2015 38th International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2015, pp. 163–167.

[17] M.-C. Nguyen, H. Nguyen, D.-H. Nguyen, E. Georgeaux, P. Mege, and L. Martinod, "Adaptive physical resource block design for enhancing voice capacity over lte network in pmr context," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, IEEE, 2016, pp. 1–5.

[18] F. Xu, Y. Lin, J. Huang, *et al.*, "Big data driven mobile traffic understanding and forecasting: A time series approach," *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 796–805, 2016. DOI: 10.1109/TSC.2016.2599878.

[19] C. Garvey, "A framework for evaluating barriers to the democratization of artificial intelligence," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.

[20] M. M. Hasan, S. Kwon, and J.-H. Na, "Adaptive mobility load balancing algorithm for lte small-cell networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2205–2217, 2018. DOI: 10.1109/TWC.2018.2789902.

[21] *Lte; evolved universal terrestrial radio access (e-utra); physical layer procedures (3gpp ts 36.213 version 14.6.0 release 14)*, 3GPP – The Mobile Broadband Standard, 2018.

[22] J. A. Sue, P. Brand, J. Brendel, R. Hasholzner, J. Falk, and J. Teich, "A predictive dynamic power management for lte-advanced mobile devices," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2018, pp. 1–6.

[23] H. D. Trinh, L. Giupponi, and P. Dini, "Mobile traffic prediction from raw data using lstm networks," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, IEEE, 2018, pp. 1827–1832.

[24] D. Babicz, A. Tihanyi, M. Koller, C. Rekeczky, and A. Horváth, "Simulation of an analogue circuit solving np-hard optimization problems," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, 2019, pp. 1–5.

[25] D. Chmieliauskas and D. Guršnys, "Lte cell traffic grow and congestion forecasting," in *2019 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, IEEE, 2019, pp. 1–5.

[26] J. Kim and N. Moon, "Bilstm model based on multivariate time series data in multiple field for forecasting trading area," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–10, 2019.

[27] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "A comprehensive analysis of deep regression," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 9, pp. 2065–2081, 2019.

[28] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "A comprehensive analysis of deep regression," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 9, pp. 2065–2081, 2019.

[29] H. R. Medeiros, F. D. de Oliveira, H. F. Bassani, and A. F. Araujo, "Dynamic topology and relevance learning som-based algorithm for image clustering tasks," *Computer Vision and Image Understanding*, vol. 179, pp. 19–30, 2019.

[30] H. R. Medeiros, F. D. de Oliveira, H. F. Bassani, and A. F. Araujo, "Dynamic topology and relevance learning som-based algorithm for image clustering tasks," *Computer Vision and Image Understanding*, vol. 179, pp. 19–30, 2019.

[31] A. N.D. and R. A., "Avoiding queue overflow and reducing queuing delay at enodeb in lte networks using congestion feedback mechanism," *Computer Communications*, vol. 146, pp. 131–143, 2019, ISSN: 0140-3664. DOI: https://doi.org/10.1016/j.comcom.2019.07.015.

[32] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of lstm and bilstm in forecasting time series," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 3285–3292. DOI: 10.1109/BigData47090.2019.9005997.

[33] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of lstm and bilstm in forecasting time series," in *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, 2019, pp. 3285–3292.

[34] N. Tavakoli, "Modeling genome data using bidirectional lstm," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, IEEE, vol. 2, 2019, pp. 183–188.

[35] J. Yuan, H. Wang, C. Lin, D. Liu, and D. Yu, "A novel gru-rnn network model for dynamic path planning of mobile robot," *IEEE Access*, vol. 7, pp. 15 140–15 151, 2019.

[36] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications surveys & tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[37] P. H. Braga, H. R. Medeiros, and H. F. Bassani, "Deep categorization with semi-supervised self-organizing maps," in *2020 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2020, pp. 1–7.

[38] L. Feng, W. Li, Y. Lin, L. Zhu, S. Guo, and Z. Zhen, "Joint computation offloading and urllc resource allocation for collaborative mec assisted cellular-v2x networks," *IEEE Access*, vol. 8, pp. 24 914–24 926, 2020.

[39] *Huawei eran channel state management feature parameter (feature id: Lbfd-00101501 cqi adjustment)*, Huawei, Shenzhen, China., 2020.

[40] *Huawei power control feature parameter (feature id: Lbfd- 002016 dynamic downlink power allocation)*, Huawei, Shenzhen, China., 2020.

[41] T. J. Karlovitz, "The democratization of technology–and its limitation," in *Managing Customer Experiences in an Omnichannel World: Melody of Online and Offline Environments in the Customer Journey*, Emerald Publishing Limited, 2020, pp. 13–25.

[42] A. Kirmaz, D. S. Michalopoulos, I. Balan, and W. Gerstacker, "Mobile network traffic forecasting using artificial neural networks," in *2020 28th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, 2020, pp. 1–7. DOI: 10.1109/MASCOTS50786.2020.9285949.

[43] H. Liu and Z. Long, "An improved deep learning model for predicting stock market price time series," *Digital Signal Processing*, vol. 102, p. 102 741, 2020.

[44] K. B. Anne Delaporte. "The state of mobile internet connectivity 2021." (Sep. 2021), [Online]. Available: https://www.gsma.com/r/wp-content/uploads/2021/09/The-State-of-Mobile-Internet-Connectivity-Report-2021.pdf.

[45] J. L. Bejarano-Luque, M. Toril, M. Fernandez-Navarro, C. Gijon, and S. Luna-Ramirez, "A deep-learning model for estimating the impact of social events on traffic demand on a cell basis," *IEEE Access*, vol. 9, pp. 71 673–71 686, 2021.

[46] H. S. Jang, H. Lee, H. Kwon, and S. Park, "Deep learning-based prediction of resource block usage rate for spectrum saturation diagnosis," *IEEE Access*, vol. 9, pp. 59 703–59 714, 2021. DOI: 10.1109/ACCESS.2021.3073670.

[47] K. Li, K. Sward, H. Deng, *et al.*, "Using dynamic time warping self-organizing maps to characterize diurnal patterns in environmental exposures," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.

[48] N. Li, L. Hu, Z.-L. Deng, T. Su, and J.-W. Liu, "Research on gru neural network satellite traffic prediction based on transfer learning," *Wireless Personal Communications*, vol. 118, no. 1, pp. 815–827, 2021.

[49] A. Masood, T.-V. Nguyen, and S. Cho, "Deep regression model for videos popularity prediction in mobile edge caching networks," in *2021 International Conference on Information Networking (ICOIN)*, IEEE, 2021, pp. 291–294.

[50] D. Rügamer, C. Kolb, C. Fritz, *et al.*, "Deepregression: A flexible neural network framework for semi-structured deep distributional regression," *arXiv preprint arXiv:2104.02705*, 2021.

[51] W. Shen, H. Zhang, S. Guo, and C. Zhang, "Time-wise attention aided convolutional neural network for data-driven cellular traffic prediction," *IEEE Wireless Communications Letters*, vol. 10, no. 8, pp. 1747–1751, 2021.

[52] C. S. Wickramasinghe, K. Amarasinghe, D. L. Marino, C. Rieger, and M. Manic, "Explainable unsupervised machine learning for cyber-physical systems," *IEEE Access*, vol. 9, pp. 131 824–131 843, 2021.

[53] H. Yuan and G. Li, "A survey of traffic prediction: From spatio-temporal data to intelligent transportation," *Data Science and Engineering*, vol. 6, pp. 63–85, 2021.

[54] "Ericsson mobility report." (Jan. 2022), [Online]. Available: https://www.ericsson.com/49d3a0/assets/local/reports-papers/mobility-report/documents/2022/ericsson-mobility-report-june-2022.pdf.

[55] Y. Fang, S. Ergüt, and P. Patras, "Sdgnet: A handover-aware spatiotemporal graph neural network for mobile traffic forecasting," *IEEE Communications Letters*, vol. 26, no. 3, pp. 582–586, 2022. DOI: 10.1109/LCOMM.2022.3141238.

[56] L. Lo Schiavo, M. Fiore, M. Gramaglia, A. Banchs, and X. Costa-Perez, "Forecasting for network management with joint statistical modelling and machine learning," in *2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2022, pp. 60–69. DOI: 10.1109/WoWMoM54355.2022.00028.

[57] A. Mostafa, M. A. Elattar, and T. Ismail, "Downlink throughput prediction in lte cellular networks using time series forecasting," in *2022 International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications (CoBCom)*, 2022, pp. 1–4. DOI: 10.1109/CoBCom55489.2022.9880654.

[58] M. Pirani, P. Thakkar, P. Jivrani, M. H. Bohara, and D. Garg, "A comparative analysis of arima, gru, lstm and bilstm on financial time series forecasting," in *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2022, pp. 1–6. DOI: 10.1109/ICDCECE53908.2022.9793213.

[59] F. Sun, P. Wang, J. Zhao, *et al.*, "Mobile data traffic prediction by exploiting time-evolving user mobility patterns," *IEEE Transactions on Mobile Computing*, vol. 21, no. 12, pp. 4456–4470, 2022. DOI: 10.1109/TMC.2021.3079117.

[60] Q. Yu, H. Wang, T. Li, *et al.*, "Network traffic overload prediction with temporal graph attention convolutional networks," in *2022 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2022, pp. 885–890. DOI: 10.1109/ICCWorkshops53468.2022.9814643.

[61] S. T. Nabi, M. R. Islam, M. G. R. Alam, *et al.*, "Deep learning based fusion model for multivariate lte traffic forecasting and optimized radio parameter estimation," *IEEE Access*, vol. 11, pp. 14 533–14 549, 2023. DOI: 10.1109/ACCESS.2023.3242861.