Addis Ababa University
College of Technology and Built Environment
Telecommunication Network Engineering Graduate Program

Root Cause Analysis of Insufficient Effective Throughput in 4G Core Network using Long Short-Term Memory Networks

By:

Simon Getahun Tirfe

Advisor:

Dr.-Eng Yihenew Wondie

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Telecommunication Network Engineering

June, 2025
Addis Ababa,Ethiopia

Addis Ababa University
College of Technology and Built Environment
School of Electrical and Computer Engineering
Telecommunication Netowrk Engineering Program

Root Cause Analysis of Insufficient Effective Throughput in 4G Core Network using Long Short-Term Memory Networks

by:

Simon Getahun Tirfe

Approval Examiners Name	Signature	Date
1. Dr.-Ing. Dereje Hailemariam	____________	____________
2. Dr. Tsegamlak Terefe	____________	____________

Declaration

I, Simon Getahun, certify that this thesis is my own work carried out under the guidance and supervision of Dr.-Eng.Yihenew Wondie. Any materials, data, or information from other sources have been appropriately credited.

Simon Getahun Tirfe

This thesis has been submitted for examination with my approval as a university advisor.

Advisor Name:	Signature	Date
Dr.-Eng. Yihenew Wondie	______________	____________

Acknowledgments

I would like to begin by sincerely thanking God and St. Mary for blessing me with the strength, wisdom, and inspiration to complete this thesis. I am especially thankful to my advisor, Dr.-Eng. Yihenew Wondie, and my instructor, Dr.-Ing. Dereje Hailemariam, whose support and encouragement have meant a great deal to me. From the very start, they helped shape the direction of this thesis and offered invaluable feedback along the way. Their steady guidance and belief in my work have been a true foundation for completing this thesis, and I remain deeply grateful for their mentorship.

I would like to express my sincere gratitude to Ethio telecom, along with its Service Operation Center (SOC) division and National Network Operation Center (NNOC) department, for their support throughout this study. I am also sincerely grateful to my examiners, Dr.-Ing.Dereje Hailemariam, Dr.Tsegamlak Terefe and Dr.Bethelhem Seifu for their wise counsel and constructive criticism throughout the duration of this thesis.

Finally, and above all, I want to thank my family for their unconditional love and unwavering support. Their encouragement has been my constant source of strength throughout this journey.

Dedication

To

Abstract

In the world of telecommunications today, the 4G Core Network (4G CN) is a critical enabler of higher speed data connectivity for mobile internet users. Unfortunately, many telecom operators continue to struggle with the persistent issue of Effective Throughput (ET), which decreases service quality and user experience. Inspired by these issues, this thesis investigates the causes of ET degradation in various flows within the 4G CN while emphasizing a shift toward higher degree of data-driven additional insight and the need to implement proactive network management strategies.

To tackle the issue, a thesis on a Long Short-term Memory (LSTM) networks method, for time series classification of performance degradation patterns in 4G CN have been proposed. To improve model interpretability and transparency, apply SHapley Additive exPlanations (SHAP) to provide insights into the contribution of feature behaviour on our decisions, to the level of individual features. The methodology includes collecting real-world network performance data, training the LSTM model, and applying SHAP to characterize how each performance metric influenced behaviors contributing to throughput changes.

The resulting LSTM model produces good predictive accuracy with a performance of 92.10% and a ROC AUC score of 97.90%, confirming its effectiveness for the task of identifying anomalies in ET. The SHAP analysis highlights that Client-Side Long RTT, Uplink TCP Out-of-Order Rate, and Network Packet behaviours were important contributors to IET behaviour. As a result, telecom operators can be empowered to dedicate increased effort on high priority performance indicators for targeted optimization.

Key Words: 4G, CN, ET, LSTM, RCA, SHAP.

Contents

List of Figures

List of Tables

List of Abbreviations

2G :2nd Generation

3G :3rd Generation

3GPP :3rd Generation Partnership Project

4G :4th Generation

4G CN :4th Generation Core Network

4G LTE :4th Generation Long Term Evolution

5G :5th Generation

AAA :Authentication, Authorization, and Accounting

AF :Application Function

AI :Artificial Intelligence

APN :Access Point Name

ARIMA :Autoregressive Integrated Moving Average

BiLSTM :Bidirectional LSTM

BPTT :Backward Propagation Through Time

CDMA :Code Division Multiple Access

CEM :Customer Experience Management

CNN :Convolutional Neural Networks

CN :Core Network

CPU :Central Processing Unit

Diameter :Protocol for Authentication, Authorization, and Accounting

DL :Deep Learning

DNS :Domain Name System

E2E :End-to-End Delay

EIR :Equipment Identity Register

ENM :Ericsson Network Manager

EPC :Evolved Packet Core

ET :Effective Throughput

eNodeB :Evolved Node B

ePDG :Evolved Packet Data Gateway

FN :False Negative

FP :False Positives

Gbps :Gigabits per second

GPRS :General Packet Radio Service

GTP :GPRS Tunneling Protocol

GTP-C :GPRS Tunneling Protocol Control

GTP-U :GPRS Tunneling Protocol User

GTPv1 :GPRS Tunneling Protocol version 1

GW :Gateway

Gx :Interface between PGW and PCRF

Gy :Online Charging Interface

HSS :Home Subscriber Server

IET :Insufficient Effective Throughput

IMEI :International Mobile Equipment Identity

IMSI :International Mobile Subscriber Identity

IMS :IP Multimedia Core Network Subsystem

IP :Internet Protocol

IPsec :Internet Protocol Security

KNN :K-Nearest Neighbors

KPI :Key Performance Indicator

LI :Lawful Interception

LSTM :Long Short-Term Memory

LTE :Long Term Evolution

MAP :Mobile Application Part

Mbps :Megabits per second

MME :Mobility Management Entity

NAS :Non-Access Stratum

N4 :Interface between SMF and UPF

N6 :Interface between UPF and external data network

NLP :Natural Language Processing

NMS :Network Management System

NNOC :National Network Operation Center

NT :Network Throughput

OSS :Operations Support System

PCRF :Policy and Charging Rules Function

PDN :Packet Data Networks

PGW :PDN Gateway

PLMN :Public Land Mobile Network

PS :Packet Switched

PSTN :Public Switched Telephone Network

QoS :Quality of Service

RAN :Radio Access Network

RCA :Root Cause Analysis

RMSE :Root Mean Square Error

RNN :Recurrent Neural Networks

ROC AUC :Receiver Operating Characteristic – Area Under the Curve

RTT :Round-Trip Time

Rx :Between PCRF and AF

S1 :Interface between eNodeB and MME/SGW

S1AP :S1 Application Protocol

S1AP :S1 Application Protocol

S1-AP :S1 Application Protocol

S1-MME :Interface between eNodeB and MME (Control Plane)

S1-U :Between eNodeB and PGW

S10 :Interface for MME-to-MME during handovers

S11 :Interface between MME and SGW

S13 :Interface between MME and EIR

S2a :Interface between WLAN and (TWAN) PGW

S2b :Interface between ePDG and PGW

S3 :Interface between MME and SGSN for interworking and mobility

S4 :Interface between SGSN and SGW for user and control planes

S5 :Interface between SGW and PGW (same operator)

S6a :Interface between the MME and HSS

S7 :Interface between PCRF and PGW for policy control

S8 :Interface between SGW and PGW (different operators)

SGi :Interface between PGW and Internet/IMS

SGSN :Serving GPRS Support Node

SGW :Serving Gateway

SHAP :SHapley Additive exPlanations

SIM :Subscriber Identity Module

SMF :Session Management Function

SOC :Service Operation Center

TCP :Transmission Control Protocol

TLS :Transport Layer Security

TN :True Negative

TP :True Positives

TWAN :Trusted Wireless Access Network

UDM :Unified Data Management

UE :User Equipment

UMTS :Universal Mobile Telecommunications System

UPF :User Plane Function

VoIP :Voice over IP

VoLTE :Voice over LTE

Wi-Fi :Wireless Fidelity

WLAN :Wireless Local Area Network

X2 :Interface between eNodeB and eNodeB

Chapter 1: Introduction

This chapter describes the main theme of the thesis, which is developing a Deep learning(DL) based on Root Cause Analysis(RCA) approach to improve Insufficient Effective Throughput (IET) in Fourth Generation Core Network(4G CN). Furthermore, the chapter presents a review of relevant literature and outlines the structure of the thesis. It identifies existing research gaps, defines the scope and limitations of the study, highlights the key contributions, and details the overall organization of the thesis.

1.1 Background

With mobile data usage growing rapidly, the Core Network(CN) is under increasing pressure, leading to heavier signaling traffic and added strain on the system. The need for a flexible and scalable new CN is eminent [10]. The advancement in network complexity, the rapid growth in the number of users, and the increasing demand for data services urged service providers to heavily invest in their network infrastructure.The fact that Long Term Evolution(LTE) is expected to accommodate different types of services such as real time applications, video streaming, Voice over IP(VoIP), web browsing, online gaming ,online transactions and file transfer requires it to be designed carefully taking into consideration different factors , such as data rate, latency, capacity, mobility support, coverage, spectrum flexibility [1].As network complexity rises and user demand for data services escalates, ensuring sufficient effective throughput for customers is essential. And also it is crucial for service providers to understand the differences between data rate, network throughput(NT), and effective throughput(ET).

1. **Data Rate (Bit Rate):**This refers to the maximum theoretical amount of digital data (in bits) that can be transmitted per second over a network connection or device, assuming ideal conditions with no interference or loss, and is typically determined by hardware specifications. For example, a 1 Gigabit per second(Gbps) Ethernet link has a data rate of 1 gigabit per second, meaning it can theoretically transmit 1,000,000,000 bits every second between core network devices like routers or switches.
2. **Network Throughput:** This is the actual data rate at which all types of data including payload, protocol overhead, and retransmissions are successfully transmitted across the network. For example, getting 100 Megabits per second(Mbps) throughput on a 150 Mbps connection due to congestion, packet loss, etc.
3. **Effective Throughput (Goodput):**This is the actual rate at which useful data (payload only, excluding protocol headers, retransmissions, and other overhead) is successfully delivered to the application per second. Fore example, on that 100 Mbps throughput, if 15% is overhead, the effective throughput is 85 Mbps, Or on a 1 Gbps link with 940 Mbps NT, the ET might be around 900 Mbps, representing the true rate at which application data is received by the end user without any protocol overhead.

This thesis is aimed at understanding how insufficient effective throughput(IET) occurs in 4G CN. Sufficient effective throughput for the 4G CN and minimization of the gap between NT and ET will result in effective seamless data delivery and enhanced customer experience.

Using predictive analysis is important in network management since network operators will reduce the down time of the network and complaints from consumers if they are able to predict a failure[2]. Given this desire for prediction, this research has the goal of proposing a Long Short-Term Memory (LSTM) based methodology for RCA of throughput degradation in 4G CN that would facilitate proactive fault detection and mitigation activity.

LSTM networks have demonstrated exceptional performance in capturing temporal dependencies and modeling sequential data, making them highly suitable for time series prediction tasks [9]. In the context of mobile networks, particularly 4G CN infrastructure, throughput performance is influenced by a variety of dynamic and time dependent factors such as packet loss, latency variation, and transmission delays. Traditional diagnostic methods often fall short in uncovering the underlying temporal patterns that lead to throughput degradation. This study leverages LSTM networks to perform RCA of IET in 4G CN, by learning complex temporal correlations from network traffic features. By employing LSTM's ability to retain long-term dependencies in input sequences, the model not only identifies performance bottlenecks but also enables predictive insights that can support proactive network management and optimization.

This research introduces an LSTM based framework for diagnosing the root causes of IET in 4G CN. By leveraging sequential modeling and SHapley Additive exPlanations(SHAP) based feature interpretation, telecom operators can proactively identify key performance bottlenecks such as high Transmission Control Protocol(TCP) retransmission rates and latency patterns. This enables targeted interventions to enhance both throughput and network reliability. The proposed approach provides a practical and data driven solution for optimizing 4G CN operations, improving customer experience, and reducing costly service disruptions ultimately supporting increased user satisfaction and revenue.

1.2 Statement of the Problem

The advancements in the cellular network technologies over the past decade have brought endless services and capabilities to the users. Smartphone users today rely on their phones for work as well as leisure activities, such as online gaming and video streaming. This has increased the load on cellular networks, causing the network traffic to increase rapidly [7]. Despite the continuous improvements in radio access technologies and the deployment of Long Term Evolution(LTE) networks, the 4G CN remains a critical bottleneck that limits the end to end(E2E) performance experienced by users. One of the major challenges faced by network operators is IET within the Core Network(CN), which undermines the benefits of high-speed wireless access. IET occurs when the available capacity in the core network is underutilized due to issues such as packet retransmissions, congestion, latency, and suboptimal resource scheduling. These inefficiencies lead to degraded user experience, particularly during high bandwidth activities like video streaming ,cloud gaming or online payment activities.

The 4G CN of telecom operators faces the challenge of idetfing IET, which negatively impacts service quality and customer satisfaction. Despite the increasing demand for high speed and reliable data connectivity, network performance bottlenecks hinder the network's ability to deliver optimal throughput levels. The root causes of these bottlenecks need to be identified and addressed to enhance network performance and meet the growing demands of customers.

The existing monitoring and troubleshooting techniques used by many telecom operators may not provide a comprehensive understanding of the underlying factors causing IET in their Packet Switched(PS) 4G CN. Traditional methods often struggle to handle the complexity and volume of network performance data, making it difficult to accurately identify and resolve network issues. Therefore, there is a need for an advanced analytical approach which can detect the patterns, anomalies, and possible root causes affecting NT with much more effectiveness. Figure 1.2 shows that the low of ET increases as network volume rises over time, leading to ultimately lower ET at the customer level.

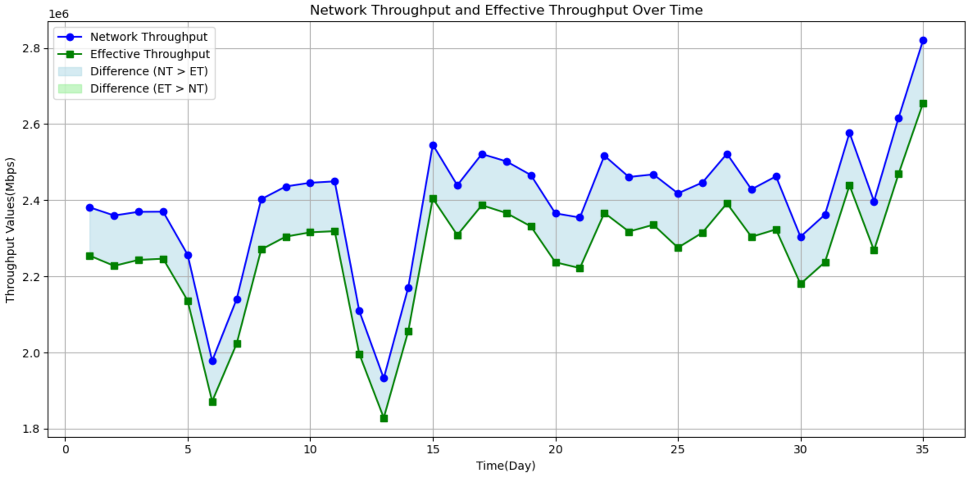

Figure 1.1: NT and ET [51] inMbps

This thesis will address these challenges with a goal in mind. Specifically, this thesis will demonstrate the application of a Deep Learning(DL) algorithm to perform a RCA of IET in the 4G CN of telecommunications operators. Once successful in resolving any throughput sought in the 4G CN, the service would improve, customer satisfaction would

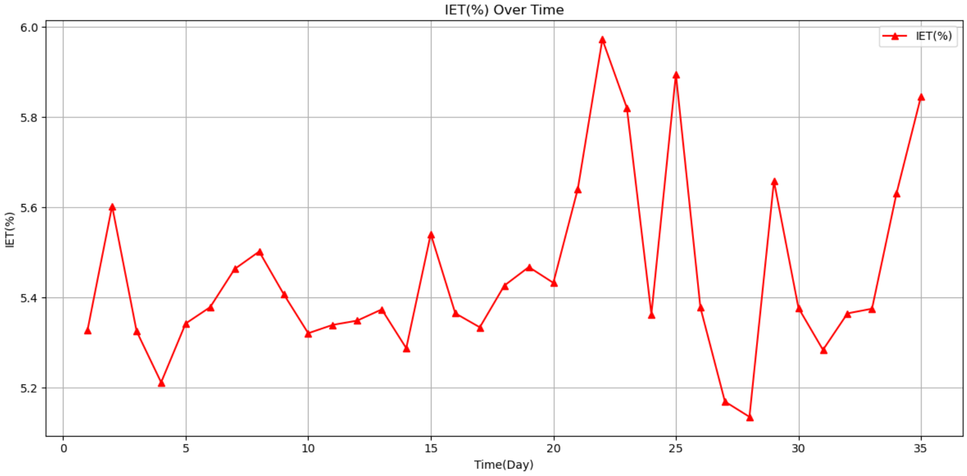

Figure 1.2: The difference level NT and ET (IET) in %

increase, and, ultimately network performance would also improve. Thus, the problem statement of this thesis can be outlined as the need to be able to identify root causes of IET in the 4G CN and establish efficient optimization options using the DL approach for performance improvement on behalf of telecom operators.

Throughput Comparison

Improving the throughput barriers in the 4G CN will improve quality of service, improve customer satisfaction, and improve network efficiency. Therefore, the problem statement for this thesis is to detect the fundamental causes of low effective throughput in the 4G CN and develop optimal methods to resolve these causes using Deep Learning (DL) algorithms and SHapley Additive exPlanations (SHAP) to improve network performance for telecom operators. Table 1.1 highlights the potential cases based on the relationship between ET and NT, and the conditions in each case.

Table 1.1: Performance conditions

Case	Condition	Label	Interpretation	Remarks
1	ET > NT	Sufficient ET	The ET exceeds the NT. Users experience good performance. Network is underutilized with spare capacity.	Mostly a theoretical scenario; rarely occurs in practice.
2	ET < NT	Insufficient ET (IET)	ET is less than the potential of the network. Users receive lower throughput than the network can support.	Common in practice due to issues like congestion, packet loss, or routing.

1.3 Objectives

1.3.1 General Objective

The main objective of this thesis is to conduct a RCA to look into deeply the factors affecting ET in the 4G CN. This will involve leveraging a DL approach, specifically LSTM networks, in combination with SHAP to enhance throughput and improve overall quality of service.

1.3.2 Specific Objectives

1. Review existing literature on network performance optimization, throughput analysis, and DL techniques, especially in the context of the 4G CN.

2. Collect and preprocess real world network performance data from the 4G CN, focusing on important metrics like TCP Connection Success Rate, Packet Loss Rate, and Latency.

3. Use SHAP for feature selection to identify the most important factors that influence throughput performance.
4. Select the most effective LSTM algorithm type on the preprocessed dataset to perform classification and prediction tasks.
5. Evaluate the performance of the selected LSTM model .

1.4 Litrature Review

The former study [1] indicates that the expansion of 4G LTE networks and increasing service demands require technical network planning, but typically the CN will not get as much attention as the Radio Access Network (RAN). Conventional planning methods utilize estimated traffic values, overlooking traffic generation's highly dynamic and unpredictable nature, which can sometimes lead to underwhelming throughput and unsatisfactory resource allocation. This study aims to close that void by providing a technical planning tool the LTE CN, using realistic traffic parameters, and provides analysis of real traffic data while simulating network operation to support improved accuracy in planning. Findings showed improved bottleneck identification and improved support for decision making, but the tool relied on quality traffic data, might have to be maintained for other networks, and doesn't have a predictive model, which limits wider applicability.

[3], proposed a new scaling mechanism for the Evolved Packet Core(EPC) that utilized both vertical and horizontal scaling, to solve performance issues associated with expanding mobile networks. Their prototype produced a 308% boost in registration per second and

70% less latency, while CPU was less than 90%. Limitations of the study included working in an specific cloud environment and not considering the interaction with other network functions and relying on predetermined thresholds for scaling.

The article [7] explored the potential prediction of network throughput in 4G LTE systems, given the higher demand on mobile data resources. The authors used some machine learning models [**?**] such as Support Vector Regression, K-Nearest Neighbors (KNN), Ridge Regression, Random Forest along with traditional time series models like the Autoregressive Integrated Moving Average (ARIMA) method, and LSTM. The authors were supplied with network and user demand data that had been previously gathered from Kingston, Canada. The authors reported that Random Forest achieved the best results of an R^2 score of 0.78 and a Root Mean Square Error (RMSE) of around 8200, while LSTM performed worse than it should in their opinion due to it being trained on limited data. Additionally, the authors stated that for others in the future, the main work to do would be to collect their own data from network operators with more data and improve the prediction accuracy.

The article [35] looks at the different aspect of predicting trends in network traffic for QoS enhancements in 4G LTE systems. The authors would like to explore the performance of different machine learning methodologies: Support Vector Machines, K-Nearest Neighbors, Random Forest, Bagging, Linear regression in order to predict traffic flow. The authors employed a 1-year dataset for 57 LTE cells, and evaluated each model's performance. In summary, the results suggested: as a rule of thumb, K-Nearest Neighbors produced the best accuracy; followed by Linear Regression and Ridge Regression; Random Forest and Decision Trees performed poorly because of the lack of features in the dataset. The authors add cautionary notes that in many cases that the non-linear models will require extensive hyperparameter tuning, and that results may be limited based on the exact dataset that the authors used.

The paper in [19] looks to classify and diagnose low User Downlink (DL) Average Throughput cases in mobile networks while also considering the added complexity of mobile networks. Their goal is to apply their models done in a supervised machine mode, to predict failure/non-failure states and using the Tree SHAP algorithm. The authors discuss the Boosting classification algorithm where the authors consider applying and then describe their approach to measuring the performance impacts of the counters on throughput performance. The authors note that they were able to succesfully appl their approach to 'cast' the low throughput scenarios from real live networks to misadjustments, radio problems, and network capacity problems. They suggest some method to solve relevant diagnostic problems. Note that the authors highlight some limitations including considerable preprocessing of data prior to analysis, and that implementing in a true real-time solution is a challenge.

Recent work on 4G LTE networks has uncovered difficulties with managing throughput due to static CN planning and the lack of responsiveness with tools that are currently available. Machine learning models, such as KNN and Random Forest, assist with prediction but the LSTM models in the literature perform poorly, mostly due to limitations with training data. The thesis being proposed addresses this by taking advantage of an LSTM model with enough data, to accurately model temporal traffic patterns. With SHAP added for interpretability, the proposed method allows for effective RCA and real time optimization for 4G CN performance in evolving environments.

1.5 Methodology

The summary of the Methoolody can be shown via gragh from the Figure 1.3. By following these steps, telecom operators can effectively identify the root causes of IET in their networks, leading to improved service quality and customer satisfaction. The use of advanced data analysis and LSTM neural network techniques enhances the ability to detect and mitigate performance issues proactively. Furthermore, by incorporating the SHAP method, operators can gain crucial insights into the precise factors influencing the LSTM model's predictions, thereby illuminating the most impactful root causes of IET and fostering greater trust and actionable understanding from the analysis.

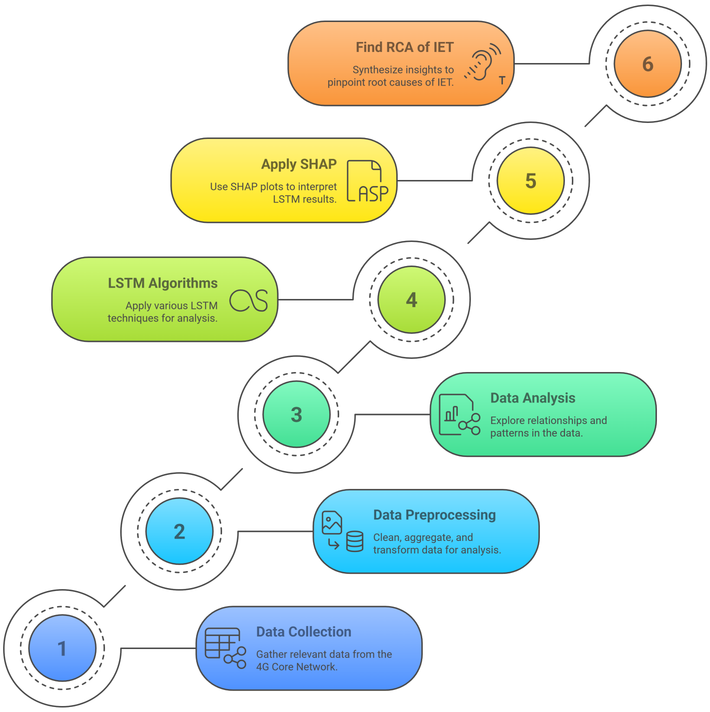

Figure 1.3: Methodology

1.6 Scope and Limitations

1.6.1 Scope

This thesis provides an explainable RCA of IET within 4G CNs. It applies LSTM networks coupled with SHAP to thoroughly examine key CN performance metrics, including throughput, latency, and packet loss ratio. The research utilizes real world data from a telecom operator to analyze practical use cases, aiming to pinpoint performance bottlenecks and optimize ET for enhanced service experience and customer satisfaction.

1.6.2 Limitations

This thesis has some limitations:

- Data Set: This research primarily focuses on 4G CN throughput, and therefore doesn't encompass the performance metrics from the RAN or individual end devices.
- Core Network Bottlenecks: This research exclusively focuses on the throughput within the 4G CN , without investigating other potential bottlenecks or performance constraints that may exist within the 4G CN infrastructure.

1.7 Contributions of the Thesis

This thesis presents significant contributions to any telecommunication company , specifically in the RCA of IET) in the 4G CN, by leveraging LSTM networks integrated with SHAP for model interpretability.:

- As a result of precise detection and elimination of the root causes of network performance problems, the research results will improve the quality of service in the 4G CN. Better ET will make the user experience even better.

- The results and advices obtained from the RCA will help operators to improve thenetwork quality that will lead to an increment of the customersatisfaction. Solving problems that affect Network performance will result in more consistent and faster data connection experience for users.

- The network will be better with the above proposed direct optimization approaches using the conclusions made after the RCA. Telecom operators can optimize resource allocation, solve performance bottleneck problems and transport data more smoothly to achieve better ET.

- By this, the research results would enable operators to predict and debug the problems through which 4G CN throughput is affected. Using the knowledge obtained from the RCA, network operators are able to foresee the performance issues that may arise and perform proactive action in order to keep network performance at high levels.

- The findings and guidelines derived from the research will be helpful for operators in the management and enhancement of 4G CN. The data driven method for network performance bottleneck study and targeted optimization might help other carriers address similar issues.

- The methodology and some findings obtained in this study can be considered as a basis for a similar topic in the future of network optimization. Use of DL such as

LSTM by for RCA is a first of its kind, a precursor for experimenting with more sophistical optimization for 4G CN and further improvements in network efficiency and user quality of experiance.

1.8 Thesis Organization

The thesis is organized in to **Six Chapters** with each chapter looking at a specific part of the research. **Chapter I** describes the introduction, together with the background, problem statement, purpose of the study, significance, asumptions and limitations, and organization. **Chapter II** provided a detailed overview of the 4G CN, including the architecture, components, and operation of the achieving high-speed data services. **Chapter III** examined the LSTM and SHAP methods employed in the research, outlining the different types of LSTM such as Classic, BiLSTM, Stacked, Attention-Based, and Peephole LSTM to predict IET in the 4G CN. In **Chapter IV** the RCA of IET was presented with statistical and graphical analysis tracing through the KPIs associated with throughput and developed a system model. In **Chapter V** the LSTM model outcomes and results was discussed, along with the model performance metrics, feature importance using SHAP, and compared to the previous studies and provided practical insights that could be beneficial for the telecom operators. Finally, **Chapter VI** brings the research to conclusion by summarizing the key findings of the research, explaining how the thesis contributed to stimulatory effects for network optimization and potential future studies.

Chapter 2: 4G Core Network

The 4G CN has the responsibility of managing data traffic and signaling from the UEs to the external networks like the internet.

The 3rd Generation Partnership Project (3GPP) tackled three specific key areas in the design of the CN [53]. The 4G CN is the backbone of the mobile broadband structure, so the 4G CN controls the data exchange and control signaling between mobile devices and the broader internet. It supports seamless user sessions, smooth data exchange, and quality of service among competing application demands. The engineering of the 4G CN supports high throughput, high reliability and high mobility, each of which were, and continue to be, important in maintaining a consistent user experience in LTE.

2.1 Over View of 4G Core Network

The 4G CN is the focal point of the 4G mobile communication network. It manages user access, mobility and connectivity to external services such as the internet. The primary operating capabilities and functionalities of the CN include the following:

2.1.1 Support for High Data Rates and Low Latency

The CN should allow high speed transmission of data with low delay. These characteristics are critical to ensuring high quality services , such as high definition video streaming, online gaming, and real time communications, for the user. Important operational priorities for the CN, to satisfy 4G user performance expectations, are both high throughput (for large files, images, and streaming) and low latency (for real time applications).

2.1.2 Interoperability with Multiple Access Technologies

Besides performance, the 4G CN must be able to interoperate with various radio access technologies. The 4G CN is meant to interoperate with both 3GPP based systems (e.g. LTE, UMTS) and non-3GPP systems (e.g. Wi-Fi, CDMA) so that end users always have access and connectivity to services going from one type of network to another.

2.1.3 Efficient Management of IP Traffic

Since 4G CN are entirely based on Internet Protocol (IP), managing IP traffic effectively is constant and ongoing task that the CN needs to engage in on a daily basis. This is true whether it is routing intellectivity, traffic prioritization, enforcement of quality of service (QoS), or the compliance of secure and reliable data. Effectively managing IP traffic is an important component of network performance, since the outcomes are recorded and tracked as service quality.

2.2 4G Core Network Topology

The 4G CN Topology consists of several key elements that work together to enable the delivery of data services in a 4G CN [53]. These elements include as shown in Figure 2.1 with

—— User plane (User data)

- - - - - Control plane (Signaling messages)

¦ Separation of CN and RAN

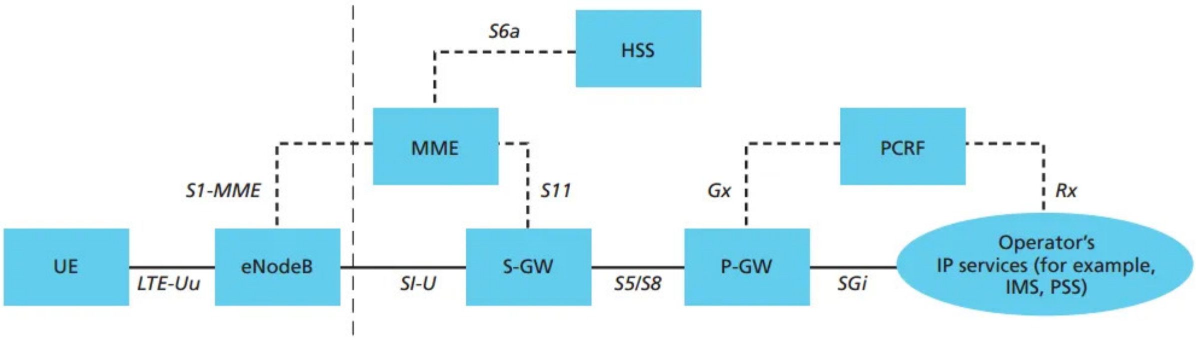

Figure 2.1: 4G CN Topology [51] [53]

In the 4G CN, the architecture is divided into the User Plane and the Control Plane to improve efficiency, scalability, and performance. The User Plane is tasked with carrying actual user data such as web surfing, video streaming, and file downloads between the UE and external PDN via the SGW and PDN PGW. The Control Plane involves the signaling tasks required to set up user data sessions, manage mobility, authenticate users and set up bearers. This involves the MME and HSS. This logical separation allows independent scaling and optimization of user data handling and control signaling, which is critical in supporting high speed, low latency communication in 4G CN.

2.3 4G Core Netowrk Nodes

The cellular network's 4G architecture has a core, its 4G CN, that primarily engages data traffic, and signaling between mobile devices, typically UEs, and the Internet.

The main elements of the 4G CN ([53], [46], [47]) include (MME, SGW, PGW, HSS), etc. -

- **eNodeB:** The eNodeB is the base station of the LTE network, and offers independent communication directly with a UE as long as it communicates with the eNodeB using radio waves. The eNodeB manages radio resource management, mobility management, and data packet decryption/encryption during the user experience. The eNodeB is connected to the CN via the S1 interface.

- **Mobility Management Entity:** The MME is responsible for user authentication and location management similar as a home agent for UE, between it and its current location. HSS when managing authentication and subscriber, data subscribes with MME and S1 to eNodeB.

- **Home Subscriber Server :** The HSS maintains the subscriber data, specifically the authentication data, subscription information and subscriber service profiles. The HSS provides authentication services to MME, and exchanges info with HSS via S6a interface.

- **Serving Gateway (SGW):** The SGW is consider an anchors of the UE session. Additionally, the SGW is responsible for routing, and forwarding packet data traffic between the UE and Internet, and like the MME connects to eNodeB via S1 interface, also it communicates with PGW via the S5 interface.

- **PGW (P-GW):** The PGW connects the CN with external PSTN networks. The PGW must maintain the service policies of the network, which include quality of service and Traffic Shaping, through interfaces S5 for SGW and Gx for external connections.

- **Policy and Charging Rules Function:** The PCRF provides the enforcement of subscriber QoS and billing. It sends the policy decisions to the PGW, and interfaces with PGW using Gx interface.

- **IP Multimedia Subsystem:** The IMS provides multimedia over IP services and is interfaced with CN primarily via the (Rx) interface. In 4G CN the machine, providing VoLTE service is commonly known as the IMS. The IMS architecture provides Voice over IP, Voice and Video over IP, Messaging Service over IP, traditional data services concurrently opposed to original service separately. In VoLTE as example the Benefit is quality voice calls and video concurrently while still data services through design of architecture.

- **AAA Server:** In 4G CN, Authentication, Authorization, and Accounting (AAA) server maintains pivotal control for managing user access to services, providing secure connection, through authentication, AUP authorization, and accounting of mobile subscribers. The AAA server keeps user connected to the network and maintained the user data payments while mobile subscribes used service.

2.4 4G Core Nework Interfaces

The 4G CN architecture consistments of a variety of interefaces and protocols used to communciate and transfter information between the various network elements. Key interfaces in the 4G CN architecture are outlined in Table 2.1 below:

2.5 Protocols in the 4G Core Network

The 4G CN core is a collection of protocols with distinct purposes:

- **Control Plane Protocols**: The protocols are S1-AP, NAS, and GTP-C; they enable signalling to establish sessions, mobility management, and bearer control.
- **User Plane Protocols**: The protocols are GTP-U and IP; they are used to transport user data between UE, SGW, and PGW.
- **Security Protocols**: The protocols are IPsec and TLS; they provide encryption and authentication which secure signalling and user data over an untrusted network.
- **Charging and Policy Control Protocols**: The protocols support Diameter (e.g., Gx, Gy, and Rx) real-time billing policy control, and QoS.
- **Legacy Interworking Protocols**: The protocols are MAP and GTPv1; they provide communication between the 4G EPC, and 2G-3G Networks, allowing handover and backward compatibility for the 4G EPC.

2.6 4G Core Network Operations

The operations of the 4G CN in LTE networks, detailing the steps involved in UE initialization, authentication, and authorization processes [53]. Understanding this workflow is crucial for network engineers and professionals involved in LTE network design and management as shown in Figure 2.3.

The UE initialization procedure starts with the UE's scanning of the available LTE cells and picking the one with the highest signal strength. After finding an appropriate cell, the UE makes a network attachment by sending registration request to the eNodeB.

At the authentication and authorization stage, the request will be forwarded by the eNodeB to the MME. The authentication of the UE will be checked by the MME with the assistance of the HSS to confirm the access rights.

In the session establishment procedure, the MME will establish a session with the SGW that assigns a context to UE. The SGW in this stage, creates a tunnel that associates to connect the UE tunnel to the PGW.

After the session is established, the data transmission will begin. The UE packet will be sent to the eNodeB, which will then route the packets to the SGW. The SGW will then forward the packets to the PGW which will then pass the packets to the external network. The PGW will also enforce policies and manage QoS.

Finally, in mobility management ensuring service continuity in relation to the period of time which the UE is moving between cells, the MME keeps record of the current location of the UE and whenever a handover is necessary the MME will initiate the handover process transitioning the UE to a new eNodeB.

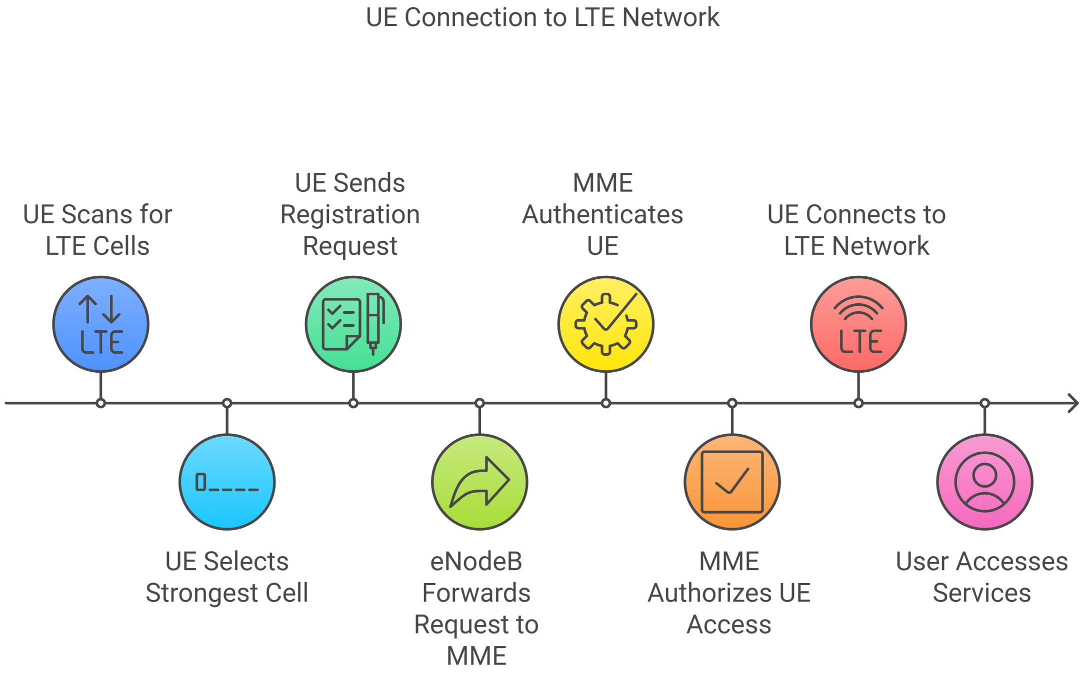

Figure 2.2: 4G CN Operations [53]

2.7 Root Cause Analysis in Core Networks

End-to-end network performance is invaluable to any telecom operator.As shown in Figure 2.3, the RCA process must be applied across different network divisions, such as the GW, CN, RAN, and UE.

2.7.1 Network Troubleshooting

Network troubleshooting is the identification and rectification of issues in a network, it is a vital element of network management and optimization. The aim of network troubleshooting is to get users back up and running as quick as possible while also addressing

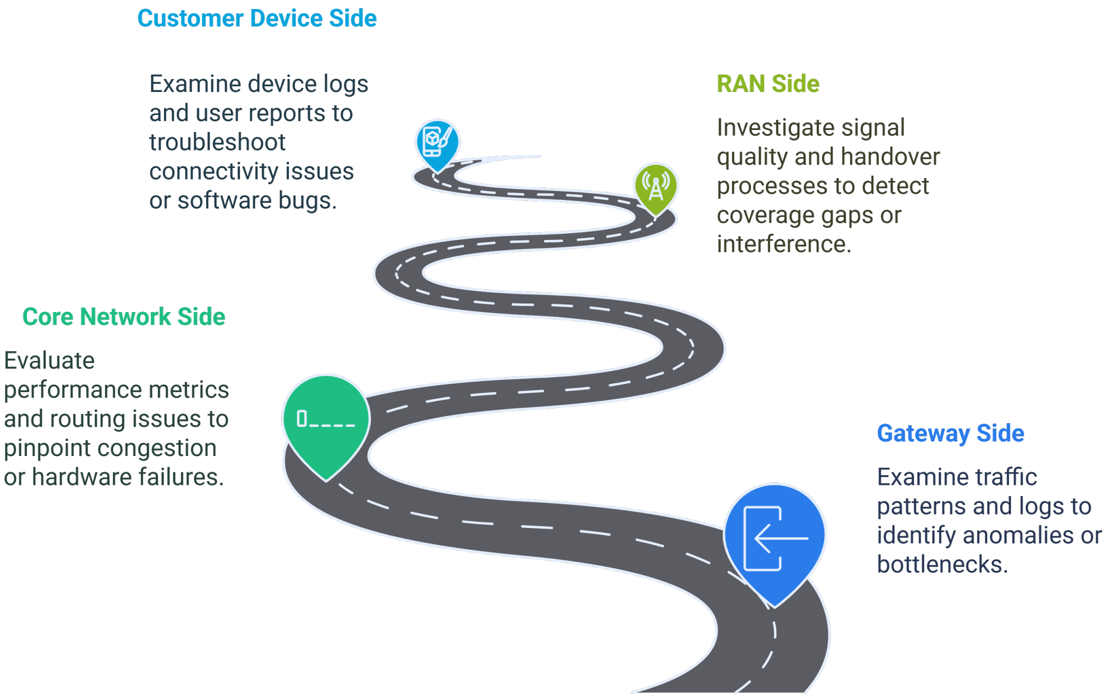

Figure 2.3: Telecommunication RCA catagorization

the underlying issue preventing the user from having a smooth experience [24] .

As far as 4G CN is concerned, troubleshooting in the mobile and wireless environment has become an almost necessitated process because of the intricate networks that are in use, and the multitude of services that they offer users. Troubleshooting involves identifying where the problem is within the components of the network that are supposed to work together, involving signalling protocols and data flow. Issues can take the form of packet loss, latency, and/or IET and pose significant obstacles to ensuring a good user experience.

2.7.1.1 Degradation

Degradation refers to a partial reduction in service quality, where the network remains operational but its performance is noticeably impaired. This may manifest as high latency during bearer establishment for instance, a delayed Create Session Response from the

PGW. Throughput may be reduced by congestion or overload of the core components like the SGW and PGW. In addition, increased packet drops can result from network congestion at the UPF and/or incorrectly configured QoS profiles. There could also be degradation associated with mobility events such as handovers, where the MME is overloaded and unable to allocate resources as quickly as necessary. A tangible example of degradation would be when S1-AP latency suddenly increases and causes a delay in the attach procedure, for example.

2.7.1.2 Outage

Outage is a total failure where part or all the network experiences total loss of service resulting in total unavailability for the user. These types of failures are typically related to critical component failures; as an example, an MME crash could make it impossible for the user to attach to the network. In succession, a complete failure of the PGW, for instance, would break internet connectivity where the user could no longer access external networks. If the HSS is unreachable completely then any attempt to authenticate the user would fail and they would be denied access to the network altogether. One example of outage would be when a region has total loss of 4G data services during a complete PGW service failure.

2.7.1.3 Problem

Problem refers to service failure issues that only affect a subset of users as opposed to the entire user base, or in the case of a network failure would affect the entire network. Generally, these issues are either local in nature or specifically with respect to a particular configuration or set of users. An example would be if there were an incorrect Access Point

Name (APN) profile in the HSS, where this will influence enterprise users specifically and not the general subscriber base. Similarly, a roaming issue might occur for foreign SIM cards if the S6a interface is misrouted or incorrectly configured. A representative case of this category is when a single user is unable to connect to the network due to a corrupted subscription record in the HSS.

2.7.1.4 Failure

Failure refers to the malfunction of a specific network component, rendering it non functional and preventing it from performing its intended role. A failure impedes the immediate function of a particular element in the network. An example of this could be when a corrupted routing table disables the SGW from forwarding anything. Users using the SGW will have service impacted. A failure in the UDM before user profile lookups, means service is impacted for other critical control-plane functions. An additional failure is the UPF stopping packet processing due to a software failure. Fault referes to a component involved in the Network that is no longer entirely functioning as it should, or a major component not functioning at all, which does not always indicate a service disruption has occurred. Fault is also distinct from failure as it can lead to reduced performance, or the inability to provide a service, but not necessarily alluding to a service 'outage'. For example, CPU loading at the MME may be nearing saturation, which begins to introduce latencies, but until that happens, it remains a functioning component of the network. Similarly, an erroneous configuration of a DNS which relates to connectivity in the PGW could block external connectivity from the PGW itself, but nothing to stop it from working internally. A comparable example of a fault would be a misconfigured firewall rule in SMF that would block traffic over one slice of the network but still permitted others to work uninterrupted.

2.7.2 Approaches to Troubleshooting in the Core Network

There are two common methodologies used for troubleshooting the CN [24],[51]:

2.7.2.1 Active Troubleshooting

Active Troubleshooting is the method engineers most commonly utilize. It is a reactive process that requires investigation of the network and service issues once they become known. In the context of the core network, troubleshooting happens mostly with log analysis of control plane components to determine where the network may have failed, such as attach procedure failures in the MME. Engineers may also troubleshoot the issues with diagnostic tools that include trace sessions, packet captures also using tools such as Wireshark, interface level KPIs, etc., to establish the root cause of issues. Active troubleshooting is more responses to alarms that were triggered in the OSS, or Network Management System(NMS), as well as incident reports from the RAN team, or customer reports.

2.7.2.2 Proactive Troubleshooting

Proactive Troubleshooting is the approach that anticipates maintenance on the network, with the purpose of finding and fixing issues before service is affected. In the CN, this practice supports availability and reliability. Proactive troubleshooting supports continuous health and performance monitoring of the system. The main activities for Proactive Troubleshooting include performing routine health checks of highly utilized and critical components in the CN, or monitoring CPU, memory and disk utilization. Proactive troubleshooting also entails, proactive software patching, establishing configuration validation documentation and regularly validating software nightlies and configuration management. Monitoring the critical KPIs attached to the core used by operators attach success rate, bearer set up times, delays in DNS responses these are actions taken to monitor the sign of early warning of degradation on the network. Furthermore, utilizing advanced methods of performance tracking such as predictive analytics and machine learning can allow an operator to distinguish patterns indicative of failing components. Operators can systematically remedy these routines and actions on their network to support performance optimization along with consistent service quality.

2.7.3 Core Network Troubleshooting Phases

To define, and address problems in the core network in an orderly manner we can view troubleshooting in three distinct phases:

From a general fault management perspective, "fault management" of the core network usually starts with **fault detection**, as the process to find all the abnormal behaviors that are stopping users from getting service. These abnormal behaviours might present as a situation where users are receiving attach failures due to an HSS timeouts; or an abrupt drops in the successful bearer establishment rate on the PGW; or latencies on the N6 interface due to an overloaded UPF. The fault detecting processes can include traditional fault detection from OSS/KPI dashboards (for example, ENM from Ericsson, or U2000 from Huawei), real time alerting (Gx session failures are a typical example), or systems for anomaly detection.

After we detect a fault, the next step is to engage in the process of **fault localization**, which is meant to identify the specific network element, or interface, causing the detection of abnormal behaviour. This process is to define if the fault is in the control or user plane, and the process will also include reconstructing all signalling (for example, Diameter time-outs between the MME and the HSS) to find all evidences where failures occurred. The fault localization process will also be able to isolate the fault domain, for example finding if the fault is with the MME, the SGW, some problem with the external DNS, or interconnect? Common tools we might use in this phase will be signalling trace analysis tools and correlation of log files from network nodes, for example; MME trace logs (if we have not used any alarms) and GTP logs from the PGW.

When we have located the fault and have determined the site of fault, we can move into phase of **RCA** to dig to find the true cause of the problem, and not just the symptoms. The RCA phase will start with large amounts of data collection as we will need to collect and review control plane logs (for example S1AP, GTP-C, Diameter), user plane statistics (for example throughput on S5, N4, and N6), node performance metrics (i.e CPU utilization, memory, sessions, etc.), and user impact information for example IMSI traces, etc. The analyst will need to just build a timeline for data for gathered that events correlate to find causation. Taking a hypothetical example, we are also seeing MME did not respond, and we observed attach failures. Cross referencing this with HSS logs showed delayed messages back to the MME which corresponds with observed MME attach failures. This can all be traced back to HSS overload from high signalling.

For the final leg of RCA, we now need to narrow down the actual root cause we found through correlations, or rule logic or even some machine learning. A simple example would be, HSS was oversubscribed from an external system processing aggressive bulk SIM reactivations that resulted in saturation of the CPU. We can then make recommendations and fixes, like throttling the signalling coming into the network from external systems, tuning the MME-HSS retry timers to stop users from seeing false failures, or simply even scaling the number of HSS and migrating to a vendor support cloud native, elastic architecture. Taking proactive action and implementing recommendations promptly will not only allow the offending problem to be dealt with, but will serve to reinforce stability to the network and overall user experience.

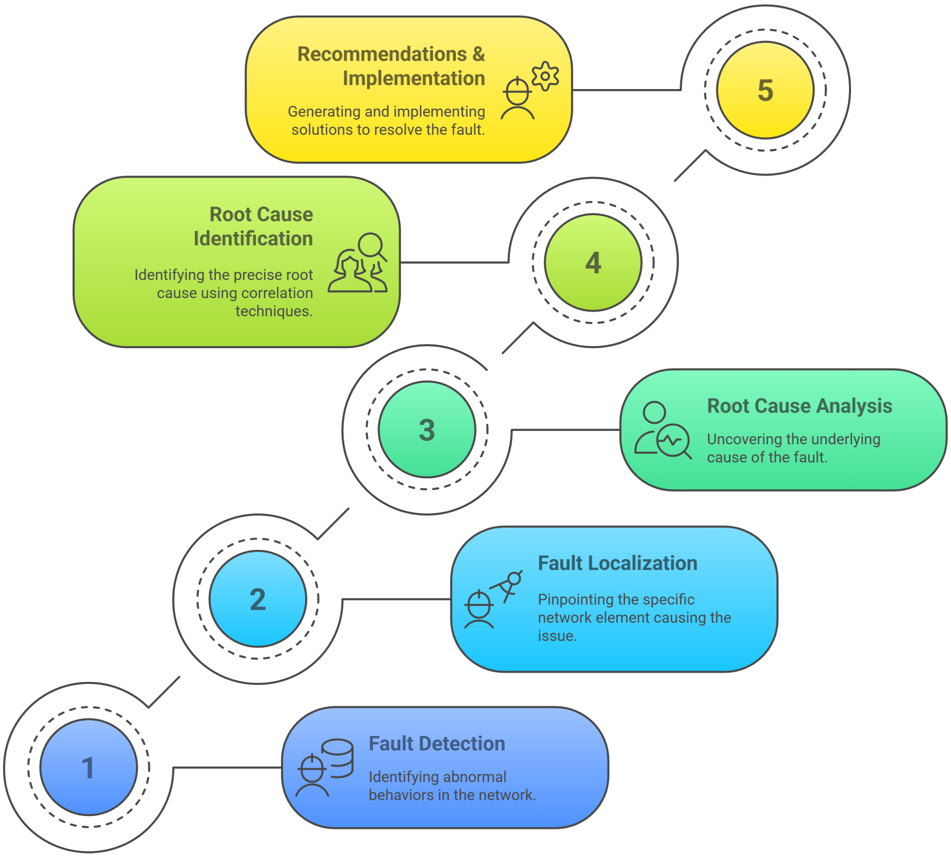

Figure 2.4: Summary of CN Fault Management Process

2.7.4 Methods of Root Cause Analysis

There are a number of methods used to establish the underlying causes of performance issues for RCA and use in CNs. The following methods are often used for RCA:

- **Data Gathering and Analysis**: Real-time data collected from network elements, metrics of performance, and reports rendered by network users; is reviewed to identify patterns and anomalies, and tell a data story.
- **Statistical Analysis**: Using statistical methods to explore relationships between various performance indicators, and what it means to focus on certain issue areas.
- **Machine Learning Methods**: Using machine learning methods and approaches. This can be even more advanced by using LSTM to model the network and its various states based on past network behaviour.
- **Data Visualisation Tools**: Visualisation tools that help to make the data sets to references with more meaning for engineers and to yield a better understanding of ”the what and why of trends and outliers”.
- **SHAP Values**: The role of Shapley Effects is to use SHAP the explain the feature impacts in relationship to performance, as it relates to IET, is also a good practice for establishing a case for the variables contributing to IET.

2.7.5 Benefits of performing RCA on Core Networks

There are various benefits for CN operators in using RCA:

- **Improved Network Performance**: RCA facilitates the operator to improve quality of network performance as the operator can provide better performance of the network by providing reviews of the root cause of those problems to fix and enhance performance reliability; this increase performance yields a better driver of the user experience.

- **Improved Control and Management of the Network**: Using RCA better ensures network control and management for the operator, and predictive capability of anticipated issues in the network prior to them becoming issues. This translates to less downtime, better service availability, and better user satisfaction.

- **Improved Business Decision Making and Planning**. Systemic and sustained RCA pairs well with business intelligence concepts and generates insight to support suggested decision making.

- **Improved Customer Satisfaction and Retention**: This speed of RCA enables the operator to respond and resolve to performance issues quickly in their network and better responds to the improved user experience, enhances customer satisfaction and loyalty, and customer retention.

- **Improved Cost Savings**: The situations that lead to the Root Cause which can be Identification and dealt with easiest, it determines which is more supporting work to manage frequently occurring issues on a net overhead cost, avoiding further attempts at maintenance to fix service performance problems as they remain poor on a service at a overhead support cost or more costly and amount of time.

Table 2.1: 4G CN Interfaces and Functions

Interface Type	**Between**	**Function**
S1-MME	eNodeB ↔ MME	Signaling in the control plane between the eNodeB and MME
S1-U	eNodeB ↔ SGW	Transfer of user plane data between eNodeB and SGW
S3	MME ↔ SGSN	Mobility management and control signaling from LTE to 3G network.
S4	SGSN ↔ SGW	An interface that connects the older GPRS/UMTS standard (for user plane data and signaling) to the EPC.
X2	eNodeB ↔ eNodeB	Handover functions, load balancing, time synchronization.
S5/S8	SGW ↔ PGW	User data transfer, bearer management (S5: same PLMN, S8: roaming and no additional changes).
S7	PGW ↔ PCRF	Charging, Billing and QoS control.
S10	MME ↔ MME	MME to MME handovers and communication during inter-region, inter PLMN, or inter technology mobility conditions.
S11	MME ↔ SGW	Bearer and mobility control for user traffic.
S6a	MME ↔ HSS	User authentication, subscription data and location update.
S13	MME ↔ EIR	Equipment identity check (IMEI) for blacklisting purposes or to reach consensus on validity of equipment.
Gx	PGW ↔ PCRF	Real time charging rules and policy and charging rules.
Rx	PCRF ↔ Application Function (AF)	Service requirements level communicated thus applying QoS and other charging rules.
SGi	PGW ↔ Internet/IMS	Provide an interface to external IP networks such as the internet and IMS.
S2a	TWAN ↔ PGW	Access via trusted non-3GPP networks (ex: Wi-Fi)
S2b	ePDG ↔ PGW	Establish a secure connect ordinary via untrusted non-3GPP networks using secure IPsec - based connectivity.
LI	EPC Nodes ↔ LI Gateway	Access to signaling and data for security and lawful intercept.

Chapter 3: LSTM and SHAP

This chapter focus on the use of LSTM networks and the SHAP method that was applied in this study to examine and classify the degradation of 4G CN. Given that LSTM networks are able to capture complex temporal dependencies in throughput related metrics, they form a logical choice for IET prediction. The SHAP method is used to interpret the predictions made by the LSTM model and reveal how different features contribute to performance degradation in terms of assigning importance scores to the features.

3.1 LSTM

In order to address the problems related to exploding and vanishing gradients that arise using basic RNNs with backpropagation through time (BPTT) training, Sepp Hochreiter and Jürgen Schmidhuber first proposed LSTM networks in 1997. Essentially, LSTM networks can carry information across time as far back as is practically necessary for a model to rely upon that information to predict time series or simulate dynamic systems. This can become problematic for traditional RNNs since they cannot learn or retain long term dependencies of sequential data [25], which is key when handling sequential data.

LSTMs are fantastic at capturing temporal dependencies as a result of their architecture, with memory cells to store information and gates that control their information flow, allow for sequential data to be modelled efficiently, and allow patterns related to IET to be captured.

The key innovation of LSTM networks are the memory cells, which allow data to be remembered over a relatively long time lag, and these memory cells are controlled by three gates comprised of a forget gate, input gate, and output gate, which control the entering, internal, and exiting information entering the memory cell.

As presented in Figure 3.1, the forget gate is used to decide how much of the previous cell state to discard, while the input gate determines how much new information to add to the cell state. The cell state is then updated before it is passed onto the next part of the LSTM. The last gate, as shown in Figure 3.1, is the output gate and it determines what amount of the updated cell state will be output.

The gated architecture of LSTM networks allows them to preserve long term dependencies and manage the vanishing gradient problem, which enables the gradients to propagate through much longer sequences without degrading.

The input data when time step t occurs is the new information that will be processed by the LSTM cell. The information feeds into a number of components of the cell.

The forget gate determines what parts of the previous cell state $t-1$ should be forgotten. The forget gate takes both the current input and the previous cell state input, and outputs values 0-1, where 0 means completely forget and 1 means completely keep.

Next is the input gate which is just like the forget gate and which provides information from the relevant new information, that needed to be added in some part of the cell state. The input gate receives the current input and the previous input and adds it to or updates

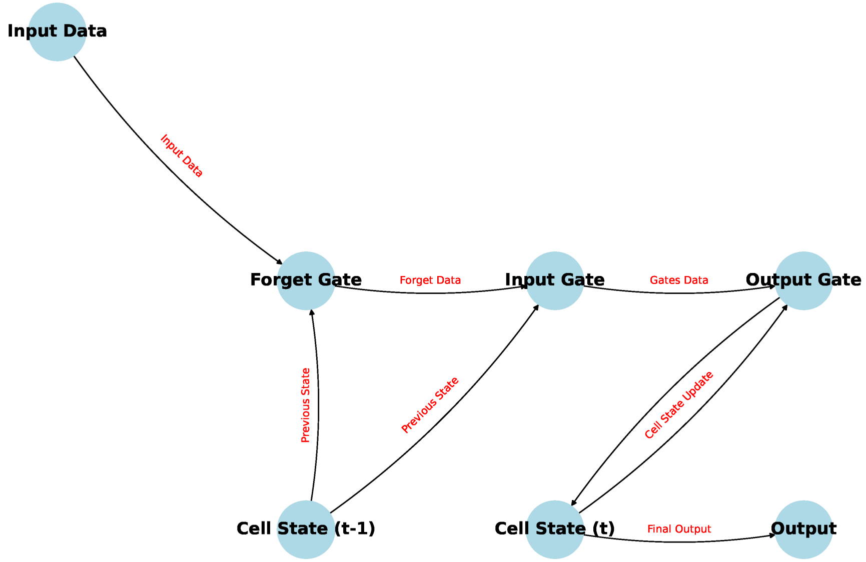

Figure 3.1: LSTM architecture diagram [25]

the cell stat with relevant new information.

The cell state contains memory information and at time $t-1$ it flows through the sequence, it has been updated with the result of the forget and input gate and now at time t provides the current memory of the LSTM. The cell state of the LSTM at time t is given by the equation:

$$\text{Cell State}(t) = (\text{Forget Gate} \times \text{Cell State}(t-1)) + (\text{Input Gate} \times \text{new input values})$$

This new state is the core memory unit that exists across time steps.

The output gate shows what the hidden state (or actual output) of the LSTM should be at time t. The output gate uses the new cell state and output state from the input gate to determine how much of each the hidden state will pass to the next time step, or the model's output layer.

The final output of the LSTM cell at time t is either used as a predictor, or the LSTM cell output is passed as input to the next LSTM cell in the sequence.

Finally, the edges showing the flow of data and control between the components of the LSTM are represented by the edges of the diagram. Labels are attached to each component (often in red) showing what kind of information they will transfer (e.g. “Forget Data”, “Previous State”, or “Final Output”).

Mathematically

The cell state C_t, which acts as a conveyor belt, allows gradients to flow without significant alteration. The cell state is updated as follows:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{3.1}$$

where , f_t, i_t, and $\tilde{C}_t$ are the forget gate, input gate, and candidate cell state, respectively [25].

The output of the LSTM unit, or the hidden state h_t, is then computed as follows:

$$h_t = o_t \cdot \tanh(C_t) \tag{3.2}$$

where o_t is the output gate, which controls how much of the cell state's information should be passed on to the next layer or time step.

This architecture allows for LSTM networks to effectively learn long term dependencies, particularly useful within scenarios in telecommunication, such as throughput analysis in 4G networks. In such situations, network behaviors and performance aspects can change

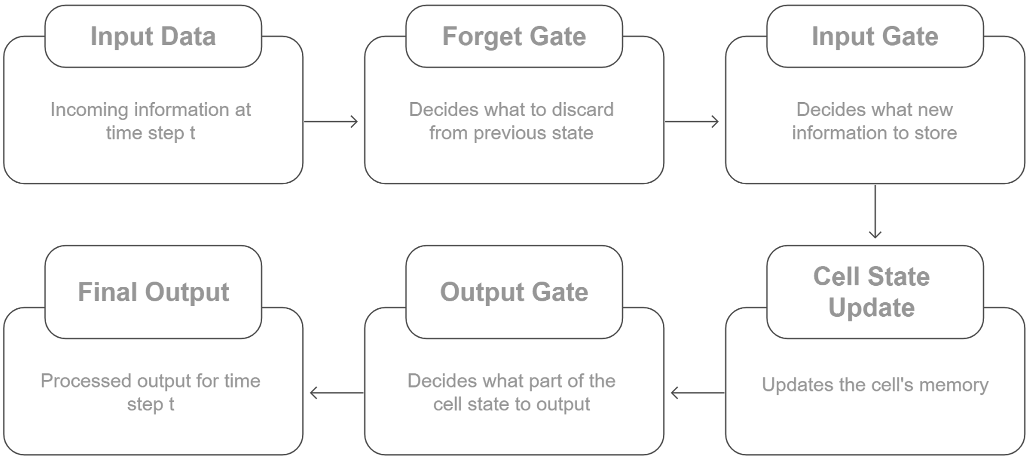

Figure 3.2: Summary of LSTM Cell Information Flow

or develop over long periods of time.

Types of LSTM

3.1.1 Standard LSTM

The Standard LSTM, sometimes known as classic LSTM incorporates the input gate, forget gate, output gate, cell state, and hidden state. It is the basis of learning temporal dependencies in sequential data and it is suitable for problems such as time series forecasting, sequence forecasting and speech recognition, where controlling data flow and remembering relevant data with time is a must.

3.1.2 Stacked LSTM

A Stacked LSTM is more advanced than the standard LSTM model as it incorporates several LSTM layers on top of each other. This arrangement allows the network to learn a more abstract and complex representation of the input. Therefore, Stacked LSTM s are suitable for deep sequence modeling tasks such as advanced time series analysis or video understanding.

3.1.3 Attention Based LSTM

The Attention Based LSTM model includes an attention mechanism with context vector, which emphasizes selective areas of the input sequence. Attention Based LSTM yield remarkable performance increases in tasks such as machine translation, text summarization, or image captioning, where some inputs are of more importance and produce more semantic meaning than others.

3.1.4 Bidirectional LSTM

Bidirectional LSTMs(BiLSTM) processes the input sequence by using forward to backward and backward to forward processing. A BiLSTM allows for full context processing with a double path system. Bidirectional LSTMs are useful in processing temporal data in applications such as natural language processing (NLP), where the word meaning can depend on both the preceding words as well as the succeeding words.

3.1.5 Peephole LSTM

Peephole LSTMs incorporates a connection from the cell state to the input gate, forget gate, and output gate thus allowing the gates immediate access to the current cell state. This architecture allows for improved detailed timing and long term dependencies to be captured which enhance the our ability to a good job in real time systems, financial forecasting, and situations where detailed memory control is highly influential. Where

LSTM Variant	**Equations**
Standard LSTM	Forget Gate: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$ Input Gate: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ Cell State Update: $C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$ Output Gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ Hidden State: $h_t = o_t \cdot \tanh(C_t)$
BiLSTM	Forward LSTM: $\overrightarrow{h_t} = \text{LSTM}(x_t, \overrightarrow{h_{t-1}})$ Backward LSTM: $\overleftarrow{h_t} = \text{LSTM}(x_t, \overleftarrow{h_{t+1}})$ Final Output: $h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$
Stacked LSTM	Layer Output: $h_t^{(l)} = \text{LSTM}^{(l)}(h_t^{(l-1)}, h_{t-1}^{(l)})$
Peephole LSTM	Peephole Forget Gate: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + V_f \cdot C_{t-1} + b_f)$ Peephole Input Gate: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + V_i \cdot C_{t-1} + b_i)$ Peephole Output Gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + V_o \cdot C_t + b_o)$
Attention-Based LSTM	Attention Weights: $\alpha_t = \text{softmax}(e_t)$ Context Vector: $c_t = \sum_{t'} \alpha_{t'} h_{t'}$ Final Output: $\tilde{h}_t = \tanh(W_c[c_t; h_t])$

Table 3.1: Equations for different LSTM Variants [25]

- x_t: Input at time step t

- h_t: Hidden state at time step t

- C_t: Cell state at time step t

- f_t: Forget gate output at time step t

- i_t: Input gate output at time step t

- o_t: Output gate output at time step t

- $\tilde{C}_t$: Candidate cell state at time step t

- σ: Sigmoid activation function

- tanh: Hyperbolic tangent activation function

- W_f, W_i, W_o: Weight matrices for forget, input, and output gates

- b_f, b_i, b_o: Bias terms for forget, input, and output gates

- $\overrightarrow{h_t}$: Forward direction hidden state in BiLSTM

- $\overleftarrow{h_t}$: Backward direction hidden state in BiLSTM

- $h_t^{(l)}$: Hidden state at layer l at time step t in Stacked LSTM

- V_f, V_i, V_o: Peephole connections' weight matrices

- α_t: Attention weight at time step t

- c_t: Context vector at time step t

- e_t: Alignment score for the attention mechanism

All variations of LSTM architecture have advantages by allowing users to select the best model for their specific purpose. Classic LSTMs are especially good at maintaining long term dependencies, and adoption of a bidirectional LSTM can help users build more context. With continued advancement of DL, there will likely be further developments of LSTM models which will provide an increased array of tools for managing challenging sequential data problems.

3.2 SHapley Additive exPlanations

SHAP is a method designed to interpret the outputs of complex machine learning models by attributing predictions to individual features [24]. SHAP analysis is rooted in Shapley values, a concept from collaborative game theory [26]. Rooted in cooperative game theory, SHAP employs the Shapley value to fairly distribute contributions among features, allowing for a clearer understanding of how they influence a model's predictions. This additive property means that any prediction can be expressed as the sum of the average model output and the contributions of each feature, making it intuitive to visualize and comprehend the impact of different factors.

: (SHAP) is helpful for explaining predictions made from a data sample by allowing each feature's contribution to the predictions made by the algorithm to be measured and explained as the Shapley value from coalitional game theory used to assign the predictions to the feature values [44].

When using SHAP in conjunction with : (CNN) models, we have a useful method of examining the : (KPI) and counters [24] as presented in the LTE framework in the thesis, and will help in identifying the contribution of features that affect accessibility and integrity of the communication networks. This will give network operators information about which features are the most impactful in the event of incidents that effect performing degradation. SHAP allows improved insights into the importance of a feature to a model and increases transparency of the function of a model to enable targeted intervention to mitigate degradation and improve network performance and customer experience ..

SHAP is a tool to interpret complex machine learning models by providing an allocation of the predictions from the model to their features. SHAP is based on cooperative game theory Shapley value, which is an equitable distribution of the contributions each feature makes to the prediction from a model. The additive property of SHAP means any prediction can be written as the average model output plus a sum of all the individual contributions feature values to the prediction as a way to help visualize how all the difference contributors have an affect.

In terms of 4G CN throughput, SHAP measures the contribution of each of the features to the algorithm's predictions. it applies concepts from coalitional game theory (Shapley values) to obtain Shapley values, which show how each prediction is allocated to each of the features. This adds transparency to the model and gives network operators the knowledge to recognize the features that contribute to (or add to) the network performance and accessibility they desire.

SHAP was applied as described in this thesis, to LSTM networks to analyze KPIs and features. The integration will enable users to analyze the network behavior due to specific features. By determining feature importance, SHAP can facilitate intervention to fix a performance deterioration, increase network management, and improve user experience. The general formula for calculating the SHAP value of a feature is taken from [24]:

$$\phi_i = \sum_{S \subseteq F} \frac{|S|! \cdot (|F| - |S| - 1)!}{|F|!} \left(f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \right) \tag{3.6}$$

where:

- ϕ_i is the Shap value of feature i.

- $\frac{|S|!(|F|-|S|-1)!}{|F|!}$ is the number of ways to choose set S from the set of all features F, excluding feature i.

- $f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)$ is the difference between the model's prediction when feature i is included in the set S and the model's prediction when feature i is excluded from the set S.

- S is a subset of features in F that does not include feature i.

- $|F|!$ is the total number of ways to choose a subset of features from F, used to normalize the SHAP values.

3.2.1 Global Interpretability

Global Interpretability is the understanding of a model's behavior over many predictions. SHAP enables this by taking the average of the SHAP values of many individual predictions, providing a view on how features impact the model on average. The semi-aggregated results of the last section show you an unbiased view of feature importance while allowing you to see the full use of features, so you have a representation of your best effort to explain your model. And you get some insight into what variables are most consistent for the model, or rather, which variables the models rely on most overall.

3.2.2 Local Interpretability

Local Interpretability is the ability to explain specific model predictions. SHAP is capable of local explainability because it assigns SHAP values for every model feature, which represents how much a feature provided to a specific prediction. This is achieved by assessing a model's output change when a feature is added to different feature-subset inputs. The final value of the contribution a feature has for a specific prediction is computed by averaging the differenced-output contribution from all possible combinations of every feature, excluding the feature. Each feature-subset combination's amount of contribution to output is weighed according to the amount of features in the subset.

Consider the prediction for a sample i, $f(i)$. The SHAP values for all features j of sample i, $\phi_{i,j}$, satisfy the following formula [24]:

$$f(i) = \phi_0 + \sum_{\text{features}} \phi_{i,j} \tag{3.3}$$

where,

Model Prediction $f(i)$: This represents the predicted output for a specific sample i. It could be a classification or regression result from a machine learning model.

SHAP Values $\phi_{i,j}$: The SHAP value for feature j of sample i quantifies the contribution of that feature to the prediction $f(i)$. It indicates how much the presence of feature j increases or decreases the prediction compared to the average prediction.

ϕ_0: This term is the base value or average prediction across all samples. It represents the model's expected output when no features are considered. It acts as a starting point for the prediction.

Summation $\sum_{\textbf{features}} \phi_{i,j}$: This summation adds up the contributions of all of the relevant features j for sample i. By summing the individual SHAP values for all of the features, we can see that collectively they model the prediction for this sample.

The equation shows that the prediction for sample i is the average prediction and all of the contributions from all the features. The contributions are added in a way that satisfies an additive function so that we can explicitly detail contributions of each feature to the prediction, and thereby make the predictions easier to understand. Additionally, with SHAP values, we can also guarantee that attributions are 'fair' and consistent, informed by properties from cooperative game theory.

3.2.3 SHAP plots

The main types of SHAP plots are listed in the Table 3.2 [52].

Table 3.2: SHAP Visualization Types

Type	**Interpretability**	**Interpretation**
Summary Plot	Global	Shows importance and effect of all features across the dataset; the summary plots can show two different types of style: (1) Dot plot (beeswarm), and (2) Bar chart.
Force Plot	Local	Show contribution of features to a particular prediction; This is an Interactive Visualization Visualization that shows push and pull towards model output, and what influences are happening in local space.
Dependence Plot	Feature effect	displays SHAP values for how changes in value of a single feature effects SHAP values (with option to color by another feature); good for identifying interaction effects.
Decision Plot	Local and global	show how features, cumulatively contribute to predictions (usually through multiple samples); this one modal represents an ideal format for models with sequential decision making.
Heatmap Plot	Overview for many samples	Show values for SHAP values in a heatmap, for features and instances - this is helpful for seeing patterns quick.

Chapter 4: Root Cause Analysis Modeling

This chapter presents a RCA model that utilizes LSTM networks and SHAP values to analyze the root causes of low IET in 4G CN. At the beginning of this chapter, relevant network performance metrics throughput, latency and packet loss are reviewed before introducing the proposed RCA framework which utilizes LSTM model representations to capture complex temporal dependencies in the network data, and includes SHAP for conducting feature importance analysis to facilitate accurate results. This chapter then covers the methods used to set up the experiments, including gathering data, preprocessing the data, creating the LSTM model, and finishes with a narrative of the evaluated metrics that were used to determine model performance; namely, accuracy metrics, the recall metric group, and the overall effectiveness of the knowledge framework pertaining to diagnosing IET issues.

4.1 The Designed RCA Architecture

In this section we introduce an RCA technique to address the IET difficulty in the 4G CN. The approach combines LSTM networks and SHAP.

The method involves gathering KPI data from the 4G CN. Then, the raw dataset is preprocessed so that can begin analysis. Preprocessing involves combing the data to remove any irrelevant instances, formatting it for modeling and normalizing the data to prepare it for training.

After preprocessing, separate the data into training, validation, and test sets. It is important to separate these datasets to create a fair evaluation of the LSTM model are going to train. The LSTM model is expected to learn temporal dependencies in the data sequences. Then, trained and tuned the model to predict when IET degradation would happen, as summarized in Figure 4.1 and in the Table 4.1.

After the model makes predictions, SHAP is used to explain those predictions. This leads to potential identification of the most important sources of IET issues by communicating which metrics were dominant in impacting throughput performance. This explanation added value by elucidating the moat predictive model's predictions, then applying interventions that potentially improve aspects of the network's overall performance.

By better illuminating the causation path of IET the methodology enables additional proactive management and optimization of the 4G CN.

4.2 Experimental Design

4.2.1 Data Collection and Preprocessing

The dataset for this study is in the file *TH.csv*", and has roughly 30,000 rows of data per hour. The study has a total of 27 features, which is what we found in the initial exploration phase. The features were collected from reputable sources and saved in a

Step	**Description**
1.4G CN KPI Data	The first stage of the model begins by extracting live, real time KPIs from the 4G CN.
2.Preprocess Data	The KPI data are now cleaned and normalized, which includes: handling any missing data, selecting or engineering types of features, based on their importance to the model, normalizing (i.e. scaling) each variable.
3.Split Data	The data will be split into: The Training Set that will be used to perform training of the model, The Validation Set that will be used for hyperparameter tuning, and finally The Test Set that will ultimately evaluate the trained model.
4.Training & Parameter Tuning	Now the LSTM model will be trained on the training set, and tuned on the validation set.
5.Test	The test set will be used to evaluate the model and how it performed on the unseen data.
6.Model Evaluation	Output metrics such as accuracy and confusion matrix will be calculated.
7.LSTM Classifier	The best LSTM model will be chosen on classifier for IET status (e.g tolerable or intolerable IET).
8.SHAP Explainer	SHAP was used as an predictor explainer to provide and interpret the final LSTM model output, rank order complete with rankings of their importance to the predictions and identify the KPI(s) that were most important to the final prediction, in relation to IET status.
9.IET RCA	The SHAP data will be used to provide understandable RCA, help identify the root-cause(s) to IET, assist with network optimization.
10.New Data Feed	When new data arrives, preprocessing can occur and forwarded into the model; along with predictions that can be interpretable with SHAP which can be valuable in RCA assessments and continued observation of IET.

Table 4.1: System Model Steps

structured format to allow better workflows and processing. Initial exploration of the data allowed us to understand the layout of the data along with the types of features and inconsistencies.

As part of the data preprocessing stage, the handled missing values through deletion or imputation approaches. The conducted feature selection to ensure that can retained the most relevant inputs for the model. Normalization methods were implemented to bring features in similar ranges across variables and normalize variables was particularly relevant

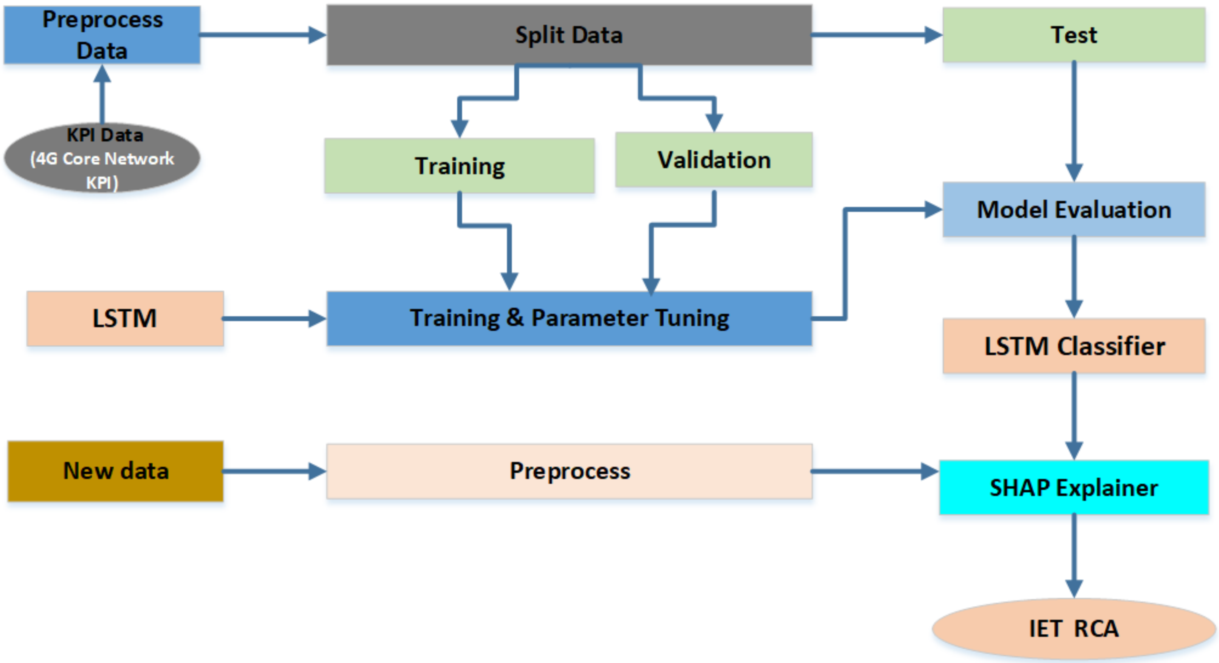

Figure 4.1: System model

to networks. Categorical features were converted to numeric features, and the dataset was divided into a training data set and testing data set to accurately evaluate performance. Finally, general data preprocessing methodology was not always different, but would also accordingly use the features of outlier detection and the many data transformation techniques which would assist with the accuracy and the model's generalization.

4.2.2 Software

Python

Python, as well as the Anaconda environment, are very popular programming environments in DL because of the ease of use, readability, and library support [48]. Since Python programming is mainly appropriate for building and deploying deep learning models, it has strong implications to RCA in 4G CNs. Also, Python is fully compatible with many of the popular deep learning frameworks like TensorFlow, Keras, and Pytorch allowing for rapid design, training, and deployment of deep neural networks.

4.3 Model Evaluation

For performance evaluation of the LSTM models Standard LSTM, Stacked LSTM, and Bidirectional LSTM(BiLSTM) in classifying user throughput into either not tolerable (Class 0) or tolerable (Class 1), number of model performance metrics can be applied. The model performance metrics allow determination of how effective the model is in discriminating between the two classes.

4.3.1 Confusion Matrix

	Predicted Tolerable (Class 1)	**Predicted Not Tolerable (Class 0)**
Actual Tolerable (Class 1)	TP	FN
Actual Not Tolerable (Class 0)	FP	TN

Table 4.2: Confusion Matrix for Tolerable and Not Tolerable Classes

- True Positive (TP): The number of instances correctly predicted as Tolerable (Class 1).
- False Positive (FP): The number of instances incorrectly predicted as Tolerable (Class 1) when they are actually Not Tolerable (Class 0).
- False Negative (FN): The number of instances incorrectly predicted as Not Tolerable (Class 0) when they are actually Tolerable (Class 1).

- True Negative (TN): The number of instances correctly predicted as Not Tolerable (Class 0).

Interpretation

- A high TP indicates effective identification of Tolerable instances.
- A low FP suggests that the model is not misclassifying many Not Tolerable instances as Tolerable.
- A low FN shows that the model is effective in identifying Tolerable instances, minimizing the risk of overlooking degraded performance.
- High TN indicates that the model accurately recognizes Not Tolerable instances.

4.3.2 Feature Correlation

It is necessary to select the fetures that impact the IET and that contribute to lead IET the 4G CN using SHAP correlation coefficient as shwon in the figure (4.2).

The visualization demonstrates feature selection by ranking network performance metrics according to their absolute mean SHAP values, which quantify each feature's average impact on model predictions. The horizontal bar chart clearly ranks features in decreasing order of importance from top (most important) to bottom (least important), with **Network Packets** ranked as the most important predictor of IET performance (longest bar with a $\sim$ 1.6 SHAP value) and **Effective Throughput** as the least relevant (shortest bar near 0). In addition, the bar chart allows for data driven feature selection in terms of what IET classifying metrics would contribute significantly to the classification of IET performance:

- **High priority features (top 1/3):** Specific Network metrics such as Packet Volume, TCP Retransmissions, and RTT Long Rates
- **Medium priority features (middle 1/3):** Packet Loss Rates, Connection Delays, and Out of Order Rates
- **Low priority features (bottom 1/3):** Effective throughput metrics and Connection Success Rates

Selection Strategy: Model features above the 0.8 SHAP value thresholds (top 10) should be favoured for model optimising, whereas features below 0.4 (bottom 7) may be removed to reduce complexity with little loss of model accuracy. The colour gradient will further reinforce the rankings visually by hue - warm colours indicate features with the highest impact. Selected features presented in the diagrams consists of Network Packets, Downlink TCP Retransmission Rate, Client Side RTT Long Rate, Server Side Uplink TCP Packet Loss Rate, Delay Between First TCP and First Get, Uplink TCP Out-of-Order Rate, and Server Side RTT Long Rate..

4.4 Model selection

Based on the same features, target variable, and IET threshold level, the selection process among the three LSTM models considering their prediction or classification capabilities. Based on the evaluation metrics, the Standard LSTM model is the most suitable

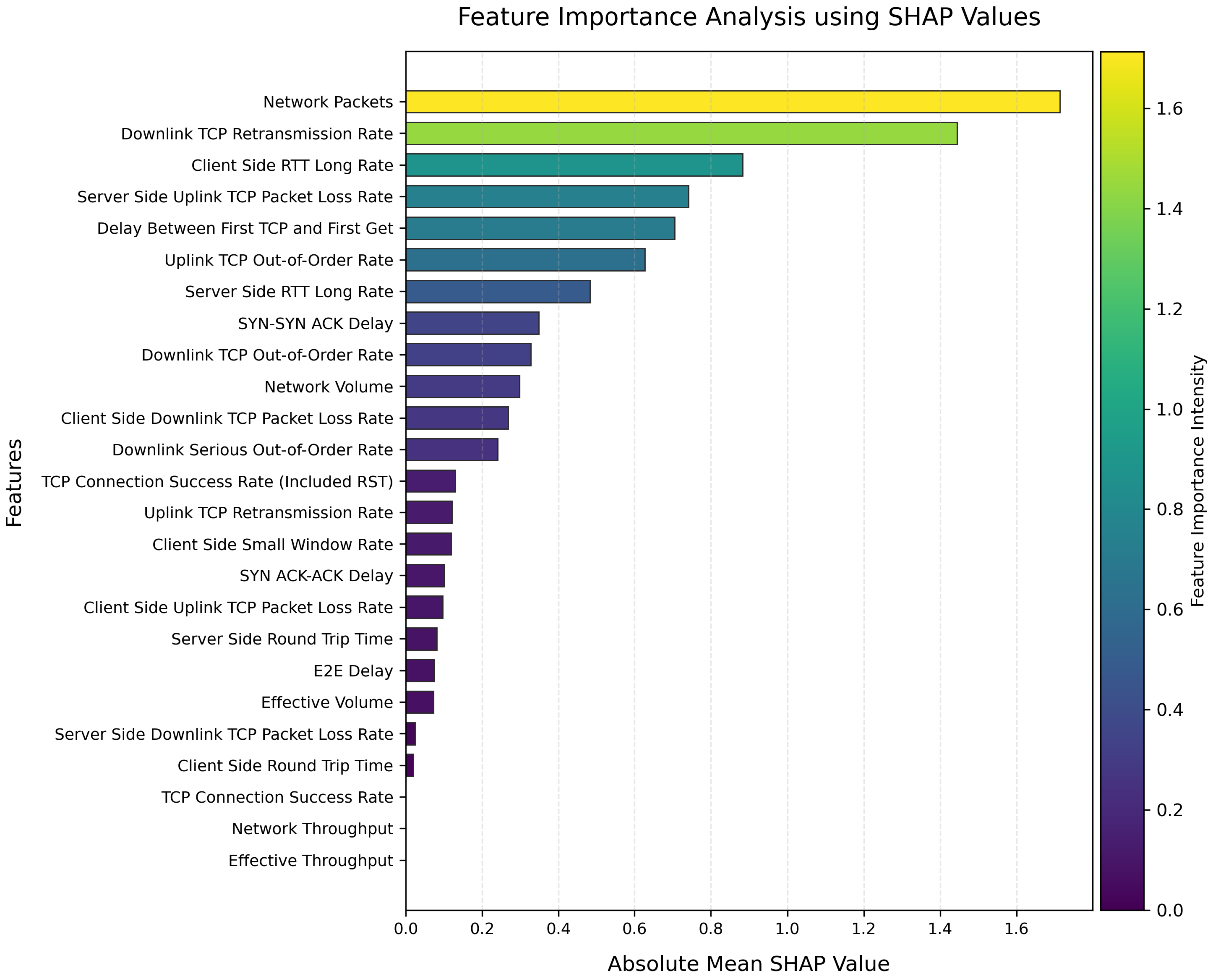

Figure 4.2: Correlations with target varibale (IET)

choice among the three. The Standard LSTM had the highest overall accuracy (95.87%), precision (96.86%), and F1 score (97.74%) when considering performance. This indicates good overall classification capability. The BiLSTM performed better with respect to recall (98.82%) and ROC AUC (97.72%). The differences are minimal however, and do not overcome the balanced overall performance of the Standard LSTM in the main performance metrics. While Stacked LSTM was a close performance model, nothing performed better overall in any of the metrics listed. So, the model chosen for its overall consistent and well rounded prediction and classification ability is the Standard LSTM.

Chapter 5: Results and Discussions

In this chapter, the evolution and model selection discussed in previous chapters to evaluate the performance of a specific algorithm the standard (classic) LSTM. The analysis is conducted using data sampled at a rate of 30,000 samples per hour from the Customer Experience Management(CEM) tool of the CN.

5.1 Data Visualization of target Varibale

The dataset does not contain any missing values that necessitate deletion; however, to enhance the analysis of network performance, a new column will be added. This column will represent "Insufficient Effective Throughput", along with its percentage, referred to as "IET" or "Throughput Efficiency". This additional measure shall provide the dataset with a more robust means to illustrate throughput efficiency levels, allowing for a more thorough analysis of performance, and identifying potential optimizations. Currently, ET and NT are the only fields included in the dataset, without considerations for the efficiency of throughput, or throughput inefficiencies in the name of network and effective throughput. The "IET" values range from approximately 6.11% to 8.02%, with as shown Figure 5.1 :

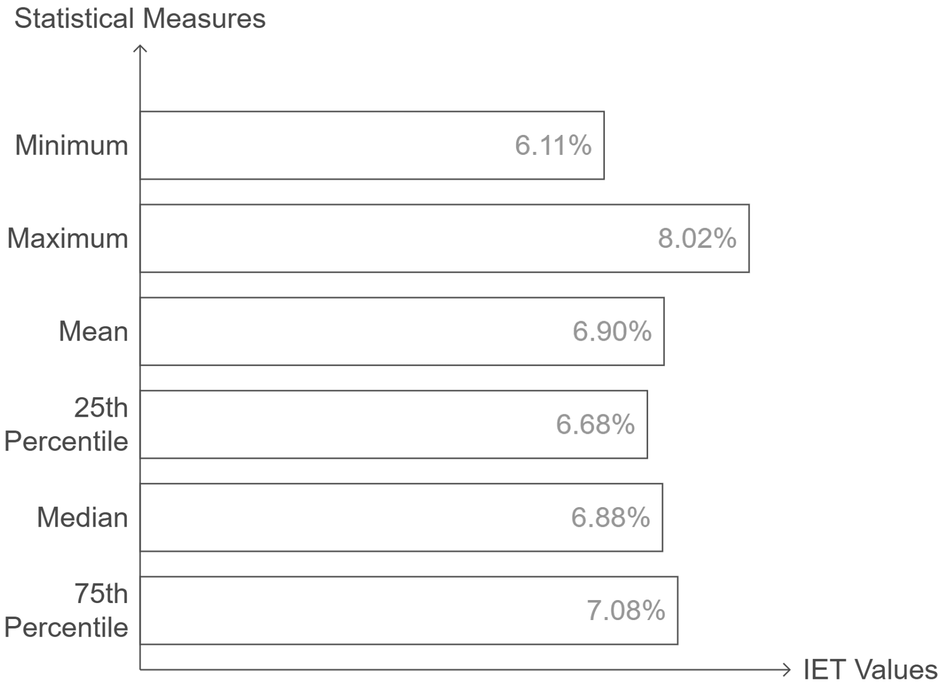

Figure 5.1: Statistical Measures of IET

5.1.1 Classification of IET

- Class 1 (Tolerable): IET ≤ 6.88:-This category includes instances where the effective throughput is less than or equal to 6.88%. In this case, there are 15,127 instances classified as tolerable, meaning that the network is performing adequately within this range.

- Class 0 (Not Tolerable): IET ≥ 6.88:-This category encompasses instances where the effective throughput is greater than or equal to 6.88%. Here, there are 14,873 instances classified as not tolerable, indicating that the network's performance is insufficient and may lead to degraded user experience as shwon in Figure 5.2.

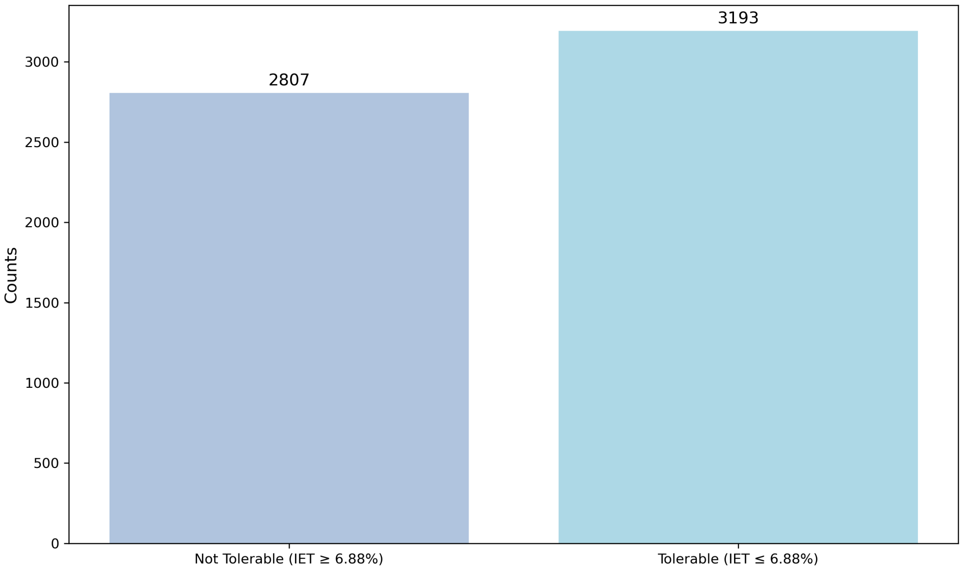

Figure 5.2: Threshold for IET catagories

5.1.2 Distribution Plot of IET

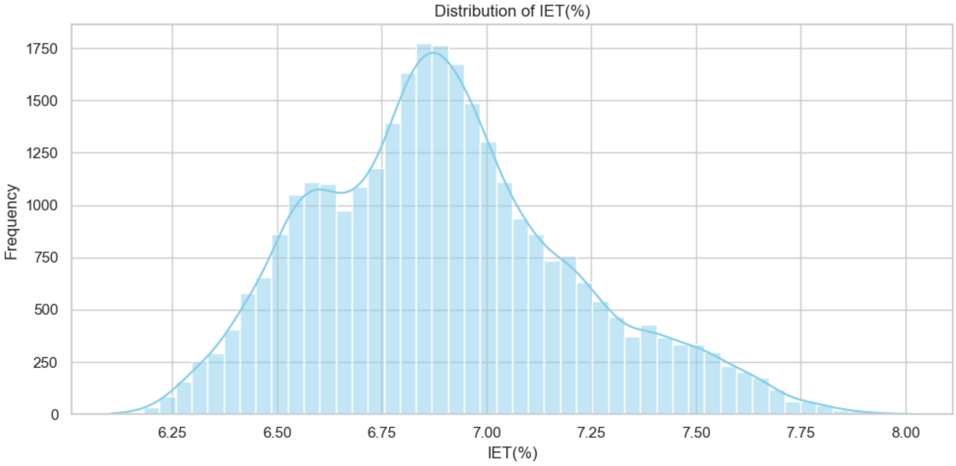

Figure 5.3: Distribution of IET

IET follows a somewhat normal distribution, slightly skewed to the right with peak around 6–8%.

5.1.3 Boxplot for Outlier Detection

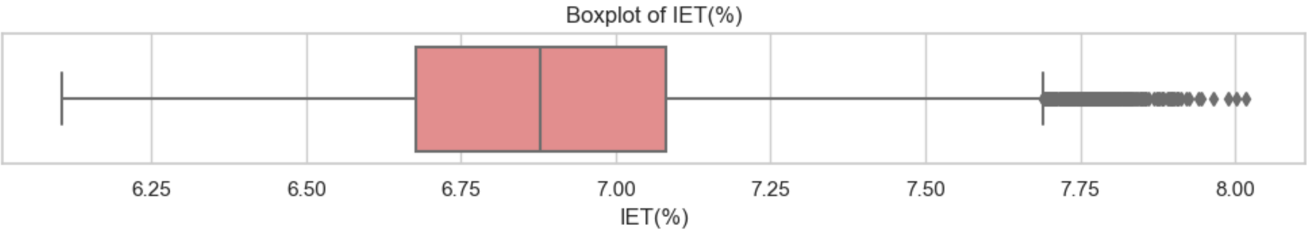

Figure 5.4: Boxplot of IET

Shows the presence of mild to moderate outliers on the higher end of IET with the central tendency is around 6–7%.

5.1.4 Time Series Trend

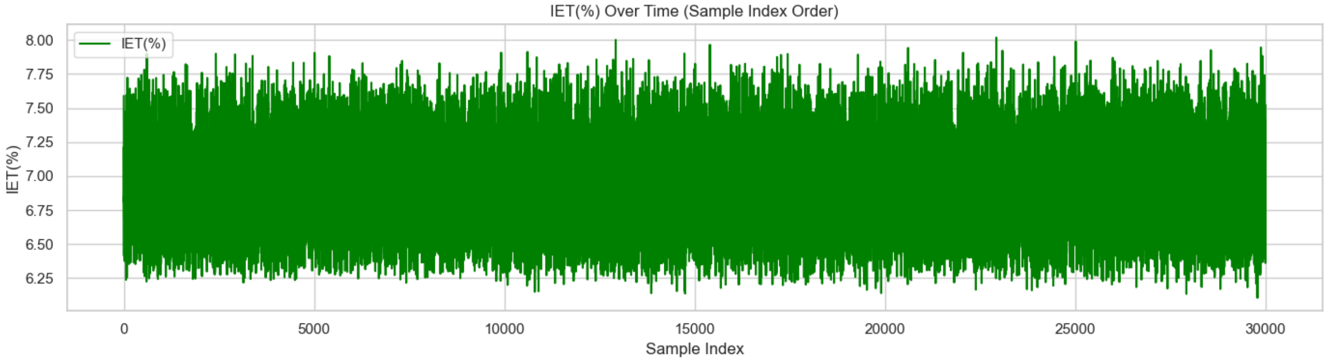

Figure 5.5: IET Over Time

Fluctuations in IET across samples suggest variability, but no obvious trend and useful for time series modeling like LSTM.

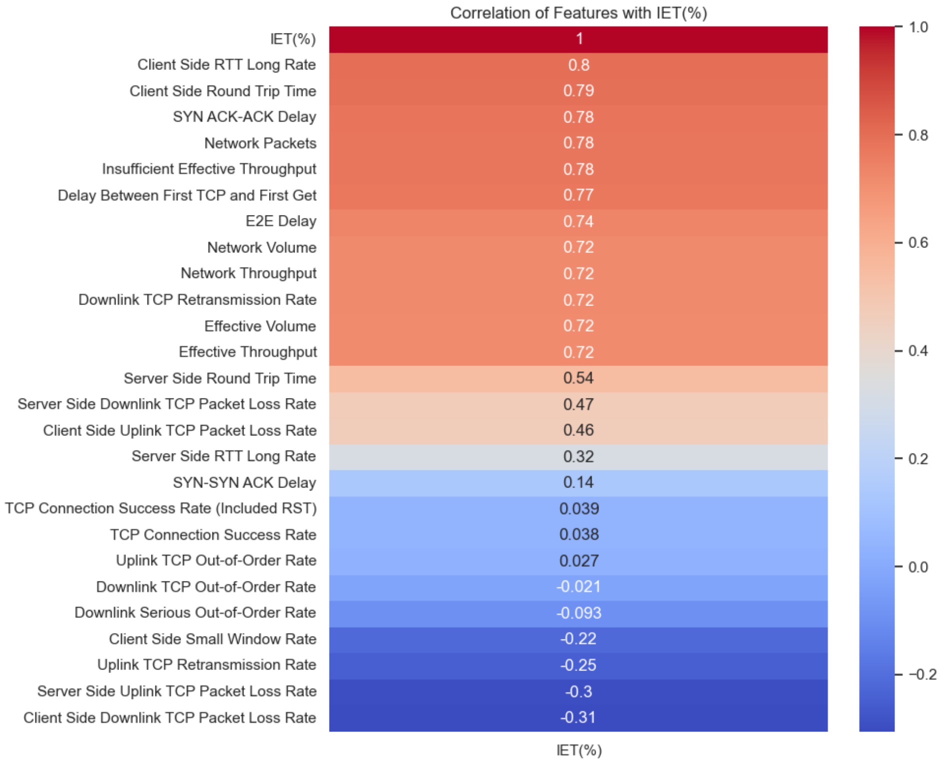

Figure 5.6: Correlation of Features with IET

5.1.5 Correlation Heatmap with IET

Identifies features most correlated with IET. Helps select relevant inputs for the LSTM model.

5.1.6 Determining Model Inputs and Outputs

To establish important features for input into the LSTM model, a dataset is first created using pandas and NumPy, simulating various network metrics. After constructing the DataFrame, correlation coefficients for each feature with respect to the target variable, *"Insufficient Effective Throughput Level (%)"*, are computed. These coefficients are then converted into a Series for easier manipulation.

Two feature thresholds of correlation are created, knowing specifically, there will be features above 0.5 and features below -0.5 that are important features in the data for filtering using python programming. As a filtering method to show which, of the many important indicators, correlate strongly with throughput, which will then become input features to the : (LSTM) model. To conclude, the condensed features resulting from the process are printed to verify that the model is only trained on important features to enhance performance and predictive capabilities.

The features chosen for the LSTM model have these metrics where IET has a high correlation with, based on Figures 4.2 and 5.6, and SHAP correlation is more trustworthy than the heatmap correlation. Based on those Figure 4.2 salient features (7 selected features) are listed below:

- **Network Packets:** Number of packets can show the load on the network, and may help in predicting throughput or latency.

- **Downlink TCP Retransmission Rate:** This feature illustrates amount of retransmitted packets which can indicate network reliability problems.

- **Client Side RTT Long Rate:** This feature speaks to the frequency of long RTT which could suggest issues in network performance.

- **Server Side Uplink TCP Packet Loss Rate:** This will indicate reliability of the data from the server to the client.

- **Delay Between First TCP and First Get:** The time can provide useful temporal insight into how responsive the network is.

- **Uplink TCP Out-of-order Rate:** This feature tracks out-of-order packets which can impact how effective a TCP connection is.
- **Server Side RTT Long Rate:** This features provides an idea of the overall latency experienced by the server and was a useful input into performance prediction.

Key Observations

From the SHAP analysis, we see that, of all the features evaluated, Network Packets is the most important feature because it has the highest mean SHAP value and has a large impact on the target variable, IET. There are a few features that strongly influence the model's predictions as well. Specifically, the Downlink TCP retransmission rate, and the Client Side RTT Long Rate are two key features that contributed to the disabled throughput output, there is some indication these features are important in terms of diagnosing throughput issues. Similarly, the Delay Between First TCP and First Get and the Uplink TCP retransmission rate are also relevant features that influence model predictions about network behavior. The Downlink TCP Rate further contributes substantially to the prediction process. In contrast, certain features show relatively lower impact. These include the Client Side Out-Of-Order Rate, Server Side Downlink TCP Retransmission Rate, and TCP Connection Success Rate , which have minor influence on the target variable.

5.1.7 Developing LSTM Model

- The LSTM model designed for this thesis employs a deep sequential architecture tailored for binary classification of IET:

  - Class 1 indicates tolerable conditions when IET $< 6.88\%$.
  - Class 0 signifies non-tolerable conditions when IET $> 6.88\%$.

- Selected features: Network Packets, Downlink TCP Retransmission Rate, Client Side RTT Long Rate, Server Side Uplink TCP Packet Loss Rate, Delay Between First TCP and First Get, Uplink TCP Out-of-Order Rate and Server Side RTT Long Rat ,Features were normalized using MinMaxScaler to ensure numerical stability during training.

- Data split: - Data was split into development, validation and test (80,10,10).

- Model structure/design: - The model had 4 lstm layers, with the first distal layer comprising 128 units and getting increasingly closer to the output layer comprising of 16 units. There were dropout layers (0.3 and 0.2) to mitigate overfitting (0 / .0 etc). The last dense layer had output binary type using a sigmoid activition.

- Model compile: - The model was compiled with Adam optimizer, binary cross entropy for the loss, and trained as for 250 epochs in total.

- After training: - The prediction from the test set was thresholded at .5 to make binary classes and then a confusion matrix was used to determine the predictions accuracy.

- Evaluation metric(s): - Metrics utilized for demonstrating the models strong separation of tolerable and non-tolerable (IET) conditions for Precision; Recall; F1 score; ROC AUC.

- Visualisation: - A custom style confusion matrix was developed with Seaborn to demonstrate a visual application for the model's classification capability to express available possible options in practice.

Compilation

The loss function is binary cross-entropy, the optimizer is Adam, and accuracy is chosen as the evaluation metric in the compilation process of the model.

Table 5.1: LSTM Model Architecture and Trainable Parameters

Layer	**Output Shape**	**Trainable Parameters**
Input Layer	(n, 1, 7)	0
LSTM Layer 1 (128 units)	(n, 1, 128)	69,120
Dropout Layer (0.3)	(n, 1, 128)	0
LSTM Layer 2 (64 units)	(n, 1, 64)	49,408
Dropout Layer (0.3)	(n, 1, 64)	0
LSTM Layer 3 (32 units)	(n, 1, 32)	12,416
Dropout Layer (0.3)	(n, 1, 32)	0
LSTM Layer 4 (16 units)	(n, 16)	3,136
Dropout Layer (0.2)	(n, 16)	0
Dense Layer (sigmoid)	(n, 1)	17

LSTM Layer Parameters

For each LSTM layer, the parameters are calculated using:

$$\text{Params}_{\text{LSTM}} = 4 \times (input\_dim + units + 1) \times units \tag{5.1}$$

Dense Layer Parameters

The Dense layer has:

$$\text{Params}_{\text{Dense}} = (input\_dim \times units) + units \tag{5.2}$$

Based on the above training parameter the following output have been found .

Metric	**Value**
Test Loss	18.73%
Test Accuracy	92.10%
Precision	92.46%
Recall	92.41%
F1 Score	92.44%
ROC AUC	97.90%

Table 5.2: Model Performance Metrics

If the above in the Table 5.2 is not satisfacory different techniques can be held for improving the model performance.

5.1.8 Develop LSTM Model Performance

1. **Change the Threshold:**

It is possible that the currently used threshold of 0.5 is not the best threshold for maximizing accuracy. Consider testing lower thresholds as a lower threshold may also improve the model performance when using accuracy as a metric.

2. **Change Class Weights:**

If balancing the dataset for class imbalance use class weights when training the model, as this can improve accuracy by preventing the model from under-learning the smaller classes. Class weights will allow for a small degree of a more balanced partial learning process so accuracy should only improve.

3. **Decrease Parameters in the Model:**

As models tend to become over-parameterized in relation to the input features, if so, restricting architecture parameters may lead to better generalization which should lead to better accuracy for validation and test.

4. **Feature Engineering / Feature Selection:**

Assessing the feature importance will also give some indication of what features are having a meaningful contribution to the model. Limiting noise and insignificant features obviously should allow the model to achieve maximum performance and should also limit overfitting.

5. **More Data, or Data Augmentation:**

Usually increasing the training set size helps a model generalize. If dataset is small enough perhaps consider ad data augmentation/techniques such as synthetic oversampling to enrich the training set.

6. **Adjust the Learning Rate:**

Changing the learning rate for optimizer should more effectively help converge the model while training it, but find that lowering the learning rate can improve the model's reach an optimized solution. Because the Standard or Classic LSTM model, produced very useful results as shown in Table 5.2 there is no significant outstanding reason to seek further improvement by boosting model performance.

5.1.9 Confusion Matrix

Each fold produces a confusion matrix, and its results are displayed in Figure 5.7. Then all confusion matrices are averaged, across the folds.

Changing the training parameters (number of epochs, batch size, validation split, and verbosity) could affect the resulting confusion matrix's results substantially. For example, by using more epochs giving the model more time to more effectively learn from the training data increasing the possibility that it can generalize from the training dataset. In general this would mean a higher true positive rate, and lower false positive rates which is essentially better classification overall.

Conversely, using very large batch sizes which may drive the training faster, by speeding up the training process, can reduce the model's ability to generalize, producing more possible misclassifications. More validation splits would mean the model would be evaluated by an even larger quantity of unseen data training data and by looking at the larger measure of the confusion matrix, a better measure can be made of the overall performance. Finally, a more verbose approach to training could give a learner a chance to keep track of the learning process, potentially adjusting based on biases or overfitting, which would be evident in the per-class confusion matrix.

The confusion matrix gives a detailed view of the model's performance when classifying instances into two categories:"Tolerable" (Class 1) or "Intolerable" (Class 0) for the application layer, as also visible in Figure 5.7.

- **TP**: The model correctly predicted 1438 instances as "Tolerable" (Class 1).

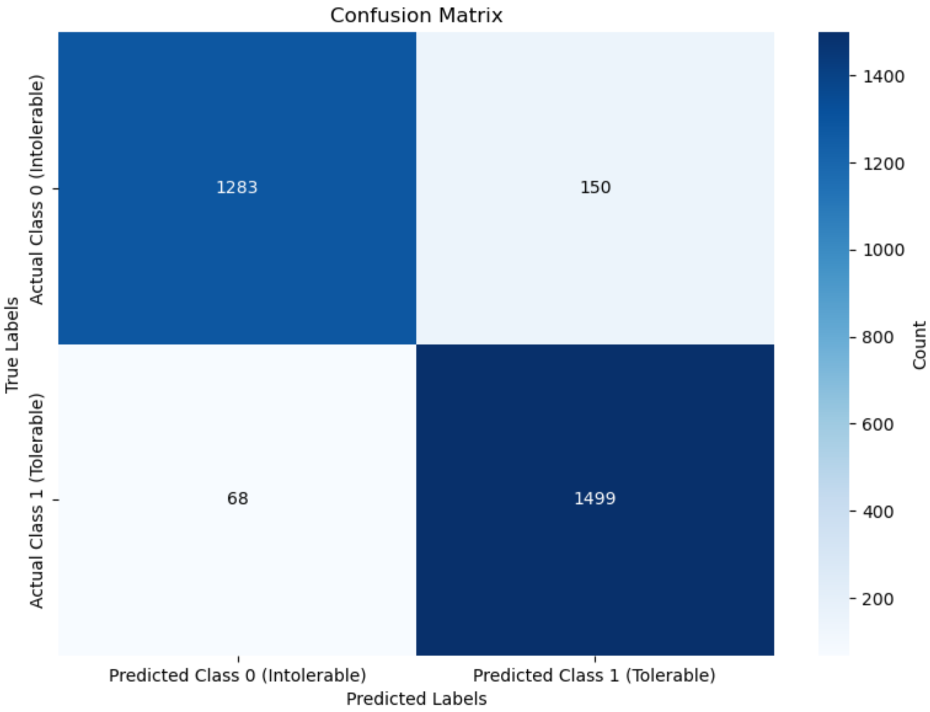

Figure 5.7: Model Performance Confusion Matrix Analysis

- **TN**: The model correctly predicted 1324 instances as "Intolerable" (Class 0).
- **FP**: The model incorrectly predicted 109 instances as "Tolerable" (Class 1) when they were actually "Intolerable" (Class 0).
- **FN**: The model incorrectly predicted 129 instances as "Intolerable" (Class 0) when they were actually "Tolerable" (Class 1).

Summary of Metrics

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{1438 + 1324}{1438 + 1324 + 109 + 129} = \frac{2762}{3000} \approx 0.92 \qquad (5.3)$$

- Accuracy= 92% of the predictions were correct.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{1438}{1438 + 109} \approx 0.929 \tag{5.4}$$

- Precision=92.9% of the time, when the model predicts "Tolerable," it is correct.

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{1438}{1438 + 129} \approx 0.918 \tag{5.5}$$

- Recall=91.8% of actual "Tolerable" instances were correctly identified.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \approx 0.923 \tag{5.6}$$

- F1=92.3% strong, balanced performance in identifying “Tolerable” throughput instances without too many false alarms or misses.

Results of the confusion matrix reveal that the classification model is performing exceptionally well, with accuracy, precision, and recall. This further demonstrates the model's robustness in differentiating the two classes, making it well-suited for applications where precise classification is crucial. This robustness is also reflected in the absence of any false positives and negatives: The model appears to honor both categories quite respectably. Such fine grained control is required to ensure the quality and security of communication at the application level.

5.2 Results

In this analysis, LSTM model in combination with SHAP to uncover the root causes of IET in network performance. LSTM, a specialized form of recurrent neural network (RNN), excels at modeling temporal dependencies in sequential data, making it suitable for analyzing time series metrics like throughput and latency. To interpret the LSTM model's outputs, SHAP provides a robust, model-agnostic method for attributing the prediction to individual input features. By assigning each feature a SHAP value, we can understand its impact both positive and negative on IET. This enables a clear, data driven explanation of which network KPIs most influence IET and under what conditions as shown in the Figure 5.8.

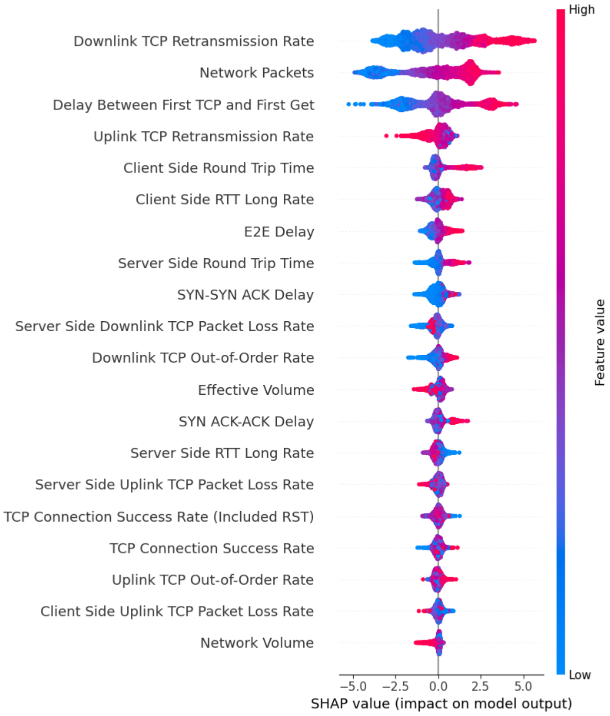

Figure 5.8: SHAP value distribution by all features

Categorization of KPIs

High Impact KPIs:-These features show large SHAP value ranges and strong correlation with the target variable.

Medium Impact KPIs: They are found between high impact KPI and low impact(the rest of the features).

Table 5.3: KPI High Impact on IET

KPI	**SHAP Trend**	**Explanation**
Downlink TCP Retransmission Rate	High value → ↑ IET	High retransmissions cause delays.
Network Packets	High value → ↑ IET	Too many packets increase congestion.
Delay Between First TCP and First Get	High value → ↑ IET	Delays in first request-response are critical.
Uplink TCP Retransmission Rate	High value → ↑ IET	Indicates unstable uplink connections.

Table 5.4: KPI Medium Impact on IET

KPI	**SHAP Trend**	**Explanation**
Client Side RTT	High value → ↑ IET	Longer RTTs slow down interactions.
Client Side RTT Long Rate	High value → ↑ IET	Frequent long RTTs reduce performance.
E2E Delay	High value → ↑ IET	E2E latency increases wait times.
Server Side RTT	High value → ↑ IET	Reflects server response issues.
SYN-SYN ACK Delay	High value → ↑ IET	Slow TCP handshake increases latency.
Server Side Downlink TCP Packet Loss Rate	High value → ↑ IET	Losses on server side increase retransmissions.

Rising KPI values with high SHAP values represent a positive contribution to IET and slower performance with a higher IET. For example, as posts are retried due to retransmission, delays, RTT, and packet loss, poor user experience increases. On the other hand, falling KPI values with low SHAP values represent negative contributions to IET and faster performance with lower IET. Again, low retransmission with lower RTT and time delays are contributing to lower IET and performance, resulting in a better responsiveness.

5.2.1 Global Interpretability result

Concentration: Global interpretability is focused on understanding the model's behavior in all instances.

Summmary Plot

- Adds up the SHAP values across large number of predictions to show overall feature importance.
- Provides abstract feature contributions to predictions, as well as average effect on predictions independently and on the model globally.
- Helps determine which features were the most important predictors of the model globally, and not locally, in all instances.

The SHAP summary plot is a handy way to understand how different features affect predictions about effective throughput. It helps folks figure out where to focus their efforts for improvements as shown Figure 5.9.

In the plot(summary plot:shows the overall impact of each feature on model predictions across all instances, indicating feature importance and variation in contributions), important features like Network Packets, Delay Between First TCP and First Get, Downlink TCP Retransmission Rate, Server Side Uplink TCP Packet Loss Rate, Client Side Downlink TCP Packet Loss Rate and SYN-SYN ACK Delay lined up on the vertical axis. The horizontal axis shows SHAP values, which show how a provided feature generally increases or decreases the potential for good throughput–a positive value means a non-zero feature is increasing the potential; negative means that feature is negative. Each dot is a data point; the way the dots appear indicates how widely varied the effect a feature can have on the results.

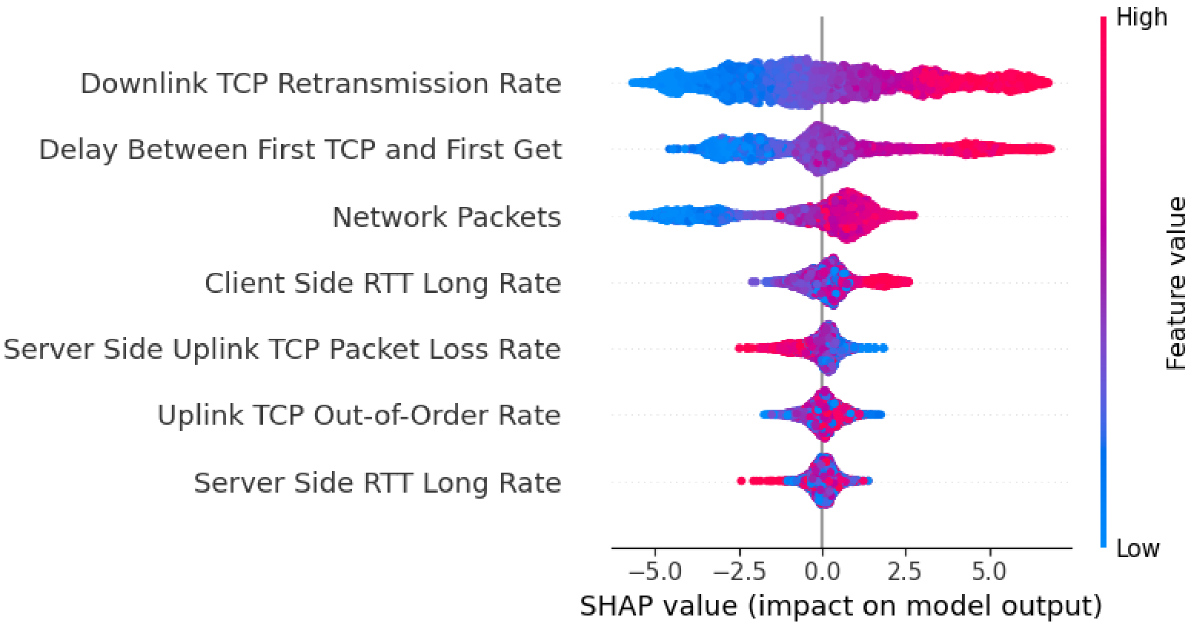

Figure 5.9: SHAP value distribution by Feature:RCA of IET

5.2.2 Local Interpretability Result

Focus: Local interpretability aims to explain the predictions of individual instances (e.g., specific data points).

Force Plot

- Visualizes the contribution of each feature's individual impact to the prediction for a single instance.
- Shows the push and pull of individual features from benchmark (i.e. average prediction) to prediction.
- Provides an explanation for why a prediction was made by showing the local behaviour of the model.

The : (SHAP) force plot (visualizes a feature's contribution to a single prediction, it shows how features drive the outcome relative to a base value) gives an intuitive view of the effects of each feature on a model's prediction, including the start prediction value and the features that help and hurt the prediction. For example, features that decreased the prediction would be red bars and features that increased the prediction would be blue bars, as shown previously with Figure 5.10. The force plot makes it easy to see the contribution of each feature on the model's output and can help you to understand the aspects influencing the prediction. The model predicted a value of $f(x) = 2.83$ for this occasion, which was a notable shift away from the baseline average output (log-odds) of zero. This significant deviation indicates that some of the features in the input data had considerable influence in arriving at a prediction reflecting certain idiosyncratic characteristics of the occasion that were clearly associated with specific IET behaviors or classifications. These influential features may be construed as determinants that increased the model's confidence in a unique output value leading to a higher prediction value. In sum, the output not only indicates the importance of those features, but together they serve as important signals of how each factor contributed to the final response in the decision process stated within the model.

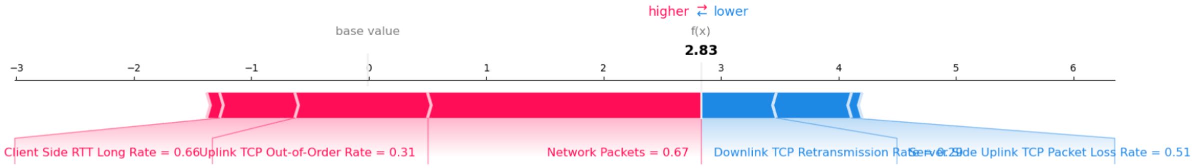

Figure 5.10: SHAP value distribution by Feature:RCA of IET

- **Red bars (Positive contribution → Increase the predicted value)**

These features increase the model's prediction (i.e., they push the model towards predicting high IET or the positive class):

- **Client Side RTT Long Rate = 0.66** → Largest contributor. High client-side RTT suggests long delays, contributing to poor performance.
- **Uplink TCP Out-of-Order Rate = 0.31** → Packets arriving out-of-order suggest congestion or routing issues, increasing delay.
- **Network Packets = 0.67** → A high volume of packets may indicate load, which can also lead to increased delay.

● **Blue bars (Positive contribution → Increase the predicted value)**

The features suggested an increase in the model prediction (i.e., pushing it towards a higher IET or positive class):

- **Uplink TCP Retransmission Rate = 0.01** → The very low degree of TCP retransmission rate would suggest minimal impact on delay and potentially shows recovery mechanisms within the network.
- **Server-Side Uplink TCP Packet Loss Rate = 0.04** → In most cases this is bad, here the model is saying less than client metrics are worse. This may be due to a weak correlation with the outcome in training.

RCA of High IET for This Instance: In performing a more in-depth root cause analysis of this specific high IET example, a few things come to light. The most important metric is the relatively high client side RTT, which suggests the performance decline is probably an issue on the user end of the connection, rather than an issue on the server side, which might indicate a connection issue, poor local connectivity, or user equipment limitation.

In addition, the detection of uplink TCP out of order packets further indicates a potential disruption or congestion in the upstream transmission. Out of order packets can stem from erratic routing, throughput saturation (too many bytes trying to be put through), or buffer overflow in upstream network nodes. It is also noted a fairly high volume of network packets, which may suggest a general network congestion issue that could influence network delay and throughput performance.As for server side metrics, TCP retransmissions, and packet loss were fairly limited and did not appear to play a significant role in this instance.

Type	**Feature Name**	**Effect on IET**	**Comment**
→	Client Side RTT Long Rate = 0.66	↑ Increases IET	Major factor (latency)
→	Uplink TCP Out-of-Order Rate = 0.31	↑ Increases IET	Sign of instability
→	Network Packets = 0.67	↑ Increases IET	Possibly congestion
→	Downlink TCP Retransmission Rate = 0.29	↓ Decreases IET	Less impact
→	Server Side Uplink TCP Packet Loss Rate = 0.51	↓ Decreases IET	Minimal role

Table 5.5: Summary of RCA Factors for High IET

5.3 Discussions

The main goal of this thesis was to attempt to tackle the persistent problem of IET in the 4G CN using a data driven, DL approach with LSTM network algorithms. The problem is made more complex by the growth of 4G LTE systems and user expectations for more real-time and high-bandwidth services.This growing demand reveals bottlenecks and inefficiencies in the CN. This study has successfully leveraged LSTM combined with SHAP to expose hidden temporal patterns and KPI connected to IET. The interpretability via SHAP has also revealed transparency on the influences that certain features like RTT, packet loss, and network volume exert on throughput performance.

The results of the LSTM model demonstrated outstanding performance, achieving the following metrics:

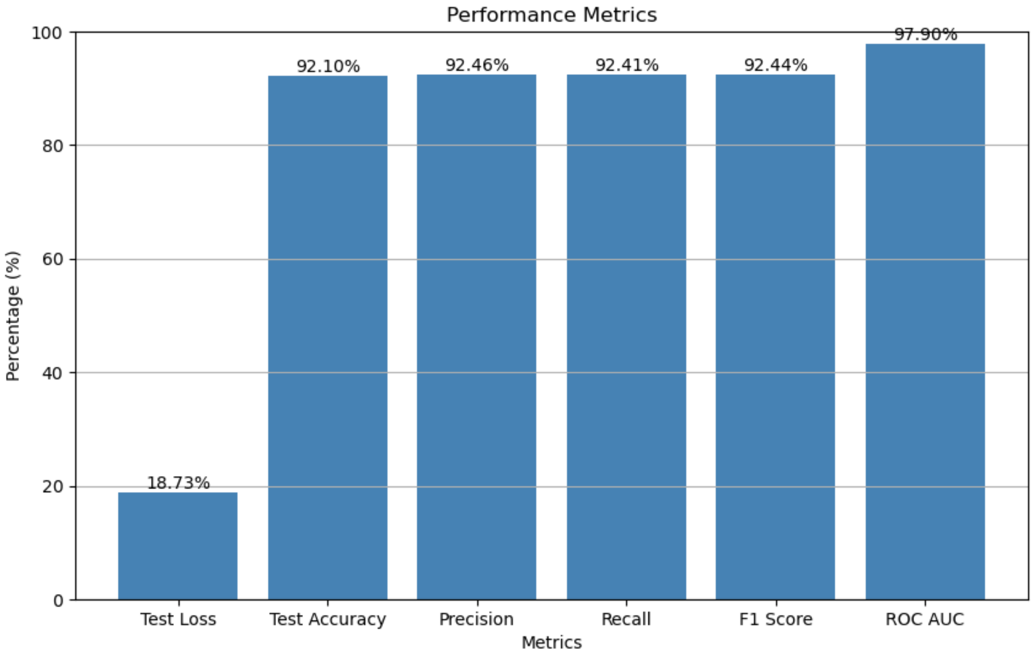

Figure 5.11: Performance Metrics of LSTM Model

These results indicate that the model effectively represents how Throughput changes over time. The SHAP analysis also demonstrated which features were most relevant to IET, and in particular Client-Side Long RTT, as the leading contributor, which makes sense because higher latency negatively impacts effective throughput. Furthermore, benefiting from grouping the KPIs into high and medium, operators can apply their strategic actions. Overall, the use of LSTM and SHAP could not only be useful, but demonstrated the advantage time series models offer in understanding and diagnosing core network performance issues.

From an operational standpoint, incorporating explainable AI approaches into network analysis now allows for emerging directions in proactive network management. Operators can now understand not only what is the problem, but why it is a problem, and more precise actions. This is particularly valuable in networks that have shortcomings in diagnosing dropped, intermittent, or latent failures, where strategies to precisely adjust active IP functions is critical. The testbed methodology and approach could be expanded from 4G CNs to developing 5G networks and beyond where dynamic traffic patterns and QoS impose severe proactive fault diagnosis requirements. In conclusion, this research presents a technically competent and operationally relevant outcome to a long-term issue in the optimization of telecommunication networks.

IET is how we define the gap between NT and ET, and then separate these between tolerable and intolerable IET. The gap can be constituted by high latency, packet losses, and network congestion, which can not allow the network to be able to deliver the maximum data rate in an ideal event. To manage IET, you would need to conduct a RCA, where do you go through the process of identifying what the specific contributors to the gap are. If you use SHAP, it gives you an understanding of how the features which influence throughput. You can then start to have targeted actions to mitigate IET, such as changing routing protocols or a target to improve QoS techniques to mitigate delay. As the gap reduces there lies the probability to experience an adequate effective throughput directly correlating to customer satisfaction. Once these strategies are continuously monitored and adjusted to minimize the gap, the effective throughput will continue to improve effectiveness and thus performance is compensated with the user's experience!

Chapter 6: Conclusions and Future Work

6.1 Conclusions

This thesis examined the issue of IET in the 4G CN, a persistent challenge that affects both service quality and user satisfaction. Using real traffic data from Ethio Telecom's network, an LSTM model was applied to detect and classify throughput-related performance issues. The model was able to categorize tolerable and intolerable (IET) examples accurately, with 92.10% accuracy and an ROC AUC score of 97.90%. These findings demonstrate that LSTM models can identify temporal patterns in network data that can significantly be missed by conventional analysis.

The thesis is also used SHAP values to examine single features and their weightings in improving model predictions. And the process of interpretability of the LSTM model revealed important results that formed the basis of understanding the factors affecting effective throughput was poor. Identified KPIs such as, Downlink TCP Retransmission Rate, Network Packets, and the time delay between the initiating TCP and get request, were all factors that significantly contributed to below target resource throughput, but the Client-Side RTT Long Rate was the feature that dominated IET measurements. These findings offer a path for telecom operators to take more targeted actions to address latency and retransmission issues.

In summary, the thesis presents a practical and explainable approach to addressing throughput degradation in the 4G CN using LSTM and SHAP. The integration of DL with interpretability improves both prediction accuracy and root cause analysis, making it a useful tool for real-time network optimization. The approach developed here offers value for telecom operators seeking to enhance 4G CN performance based on actual traffic behavior. And ,overall, the evidence points to client-side and uplink path conditions as the primary causes of the ineffective throughput observed.

6.2 Future Work

The experimentation carried out in this dissertation could be transferred to newer mobile network technologies such as 5G. The application of these results in such a context may assist in revealing new performance issues and could help drive the development of improved network management that uses real time processing and adaptive learning.

For further improving the use of LSTM networks in evolving network environments, a potential avenue for future work resides in integrating the LSTM networks with other DL techniques, such as reinforcement learning. Another potential area of development could involve the use of non traditional KPIs that may offer a more comprehensive picture of the service quality from the users perspective, especially where standard measures are inaccurate.

Further extending this work may involve collaborating with telecom operators to trial the approaches proposed herein in an operational environment. Given the evolving nature of mobile network user demands, this kind of project could validate these methods in practice, as well as improve and refine these methods for the changing circumstances.

Bibliography

Literatures

[1] *"S. Sadchikova and M. Abdujapparova. Traffic and throughput for 4G packet core networks.* 2021 International Conference on Information Science and Communications Technologies , Tashkent, Uzbekistan, 2021". doi: `https://doi.org/10.1109/ICISCT52966.2021.9670029`.

[2] *"M. Choi, T. Kim, J. P. Lee, and S. Koh. An Empirical Study on Root Cause Analysis and Prediction of Network Failure Using Deep Learning.* 2021 International Conference on Information and Communication Technology Convergence , Jeju Island, Korea, Republic of, 2021". doi: `https://doi.org/10.1109/ICTC52510.2021.9621097`.

[3] *"C. H. T. Arteaga, F. B. Anacona, K. T. T. Ortega, and O. M. C. Rendon. 11A Scaling Mechanism for an Evolved Packet Core Based on Network Functions Virtualization.* IEEE Transactions on Network and Service Management, June 2020". doi: `https://doi.org/10.1109/TNSM.2020.2978494`.

[4] *"J. Prados-Garzon, A. Laghrissi, M. Bagaa, T. Taleb, and J. M. Lopez-Soler. A Complete LTE Mathematical Framework for the Network Slice Planning of the EPC.* IEEE Transactions on Mobile Computing, 1 Jan. 2020". doi: `https://doi.org/10.1109/TMC.2019.2892612`.

[5] *"S. Vitale. A CNN-Based Pansharpening Method with Perceptual Loss."* 2019 IEEE International Geoscience and Remote Sensing Symposium , Yokohama, Japan, 2019, pp. 3105–3108. doi: `https://doi.org/10.1109/IGARSS.2019.8898627`.

[6] *"L. Bai, H. Abe, and C. Lee. RNN-based Approach to TCP Throughput Prediction.* 2020 Eighth International Symposium on Computing and Networking Workshops (CANDARW), Yamaguchi, Japan, 2020" doi: `https://doi.org/10.1109/CANDARW51189.2020.00082`.

[7] *"H. Elsherbiny, H. M. Abbas, H. Abou-zeid, H. S. Hassanein, and A. Noureldin. 4G LTE Network Throughput Modelling and Prediction.* GLOBECOM 2020 - 2020 IEEE Global Communications Conference, Taipei, Taiwan, 2020", pp. 1–6. doi: `https://doi.org/10.1109/GLOBECOM42002.2020.9322410`.

[8] *“D. P. Quang, H. N. Tai, S. Le Van, Q. Le Hong, H. Le Van, and A. T. Tuan. Advanced Mechanisms for Guaranteeing the High Performance of Virtualized 4G Core Networks.”* 13th International Conference on Information and Communication Technology Convergence , Jeju Island, Korea, 2022. doi: `https://doi.org/10.1109/ICTC55196.2022.9955600`.

[9] “I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, *Applications of Long Short-Term Memory Networks in Polymeric Sciences: A Review,* Artificial Intelligence Technology Scientific and Education Center, Bauman Moscow State Technical University, Moscow, Russia, 2023”.

[10] *“E. B. Hamza and S. Kimura. A Scalable SDN-EPC Architecture Based on OpenFlow-Enabled Switches to Support Inter-domain Handover.* In IEEE 2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)”, doi: `https://doi.org/10.1109/imis.2016.62`.

[11] *“B. Wu, L. Ge, J. Zeng, X. Zheng, Y. Tang, and X. Su. An Innovative EPC Architecture based on Not Only Stack for Beyond 5G Mobile Networks.”* Adv. Sci. Technol. Eng. Syst. J. 4(2), 319-332 (2019); doi: `https://doi.org/10.25046/aj040241`.

[12] *“K. K. Jha, Nishant, A. K. Jangid, and D. Das. Method for Improvement in Throughput and Network Resource Utilization in 5G.* In IEEE International Conference on Electronics, Computing and Communication Technologies , Bangalore, India, 2019”; doi: `https://doi.org/10.1109/CONECCT47791.2019.9012922`.

[13] *“M. Moradi A., S. A. Sadrossadat, and V. Derhami. Long Short-Term Memory Neural Networks for Modeling Nonlinear Electronic Components.* In IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 11, no. 5, pp. 840-846, May 2021”; doi: `https://doi.org/10.1109/TCPMT.2021.3071351`.

[14] *“A. M. Miyim and A. Wakili. Performance Evaluation of LTE Networks.* In 15th International Conference on Electronics, Computer and Computation , Abuja, Nigeria, 2019”; doi: `https://doi.org/10.1109/ICECCO48375.2019.9043271`.

[15] *“B. -H. Lee, E. K. Dewi, and M. F. Wajdi. Using Load Balancing Mechanism to Reduce Overload in LTE/EPC Defined Network.* In 7th International Conference on Smart Computing and Communications (ICSCC), Sarawak, Malaysia, 2019”; doi: `https://doi.org/10.1109/ICSCC.2019.8843622`.

[16] *“C. Kim, V. B. Mendiratta, and M. Thottan. Unsupervised Anomaly Detection and Root Cause Analysis in Mobile Networks.”* In International Conference on Communications Systems and Network’s (COMSNETS), Bengaluru, India, 2020; doi: `https://doi.org/10.1109/COMSNETS48256.2020.9027328`.

[17] *“D. K. Kiluiba, W. V. Kambale, J. C. Mudilu Kafunda, A. M. Nzung, M. Hamed, and K. Kyamakya. Enhancing Downlink Throughput Prediction in 4G LTE Networks*

Using Ensemble Learning Techniques." In 2024 International Conference on Applied Mathematics and Computer Science, Venice, Italy, 2024; doi: `https://doi.org/10.1109/ICAMCS62774.2024.00036`.

[18] *"M. Alazab, M. Broadbent, and M. Hobbs. Core Network Performance Measurement of 4G-LTE Network."* In Proceedings of the 10th International Conference on Cloud Computing, Data Science and Engineering (Confluence), pp. 1-6, 2020.

[19] *"M. Cilínio, D. Duarte, P. Vieira, M. P. Queluz, and A. Rodrigues. Root Cause Analysis of Low Throughput Situations Using Boosting Algorithms and the TreeShap Analysis."* In Proceedings of the 2022 IEEE 95th Vehicular Technology Conference: VTC2022-Spring, Helsinki, Finland, 2022.

[20] *"Y. Zhang, L. Niu, and G. Zhang. Research on Optimization of 4G Core Network Architecture Based on SDN."* In 2nd International Conference on Artificial Intelligence, Automation and Control Technologies, 2020.

[21] *"Y. Chen, H. Chen, and X. Zhang. Fault Diagnosis of 4G Core Network Based on Random Forest Algorithm."* In IEEE 5th Information Technology and Mechatronics Engineering Conference, 2021.

[22] *"A. Saleem, M. Raeiszadeh, A. Ebrahimzadeh, R. H. Glitho, J. Eker, and R. A. F. Mini. A Deep Learning Approach for Root Cause Analysis in Real-Time IIoT Edge Network."* In NOMS 2023 - 2023 IEEE/IFIP Network Operations and Management Symposium, Miami, FL, USA, 2023. doi: `https://doi.org/10.1109/NOMS56928.2023.10154218`.

[23] *"Y. A. Matebie and Dr.–Eng. Yihenew Wondie. Root Cause Analysis of Unsatisfactory User Throughput in UMTS for Bahir Dar City."* Master's Thesis, Addis Ababa Institute of Technology, Addis Ababa, Ethiopia, October 2021; URI: `http://etd.aau.edu.et/handle/123456789/29885`.

[24] *"F. Tazeb and Dr.–Ing. Dereje Hailemariam. Machine Learning for Improved Root Cause Analysis of LTE Network Accessibility and Integrity Degradation."* Master's Thesis, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia, October 4, 2023; URI: `http://etd.aau.edu.et/handle/123456789/1029`.

[25] Y. Zhao, L. Lin, and A. Schlarb, *"Long Short-Term Memory Networks for Automated Identification the Stationary Phase in Tribological Experiments"*, 2024. DOI: 10.20944/preprints202409.2269.v1

[26] A. V. Ponce Bobadilla, V. Schmitt, C. Maier, S. Mensing, and S. Stodtmann, *"Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development"*, 2024. DOI: 10.1111/cts.70056

[27] *"R. Farhat, Y. Mourali, H. Ezzedine, and M. Jemni. An Overview of Machine Learning Technologies and Their Use in E-learning."* In 2020 IEEE (pp. 1-6), Date of Conference: 06-08 February 2020; doi: `https://doi.org/10.1109/OCTA49274.2020.9151758`.

[28] *"O. De Jesus, J. M. Horn, and M. T. Hagan. Analysis of Recurrent Network Training and Suggestions for Improvements."* In IJCNN'01 - International Joint Conference on Neural Networks. Proceedings , Washington, DC, USA, 15-19 July 2001; doi: `https://doi.org/10.1109/IJCNN.2001.938785`.

[29] *"K. Rusek, J. Suárez-Varela, P. Almasan, P. Barlet-Ros, and A. Cabellos-Aparicio. RouteNet: Leveraging Graph Neural Networks for Network Modeling and Optimization in SDN.* In IEEE Journal on Selected Areas in Communications," Oct. 2020; doi: `https://doi.org/10.1109/JSAC.2020.3000405`.

[30] *"A. Aggarwal, D. Jain, A. Gupta, and P. Garg. Analysis and Prediction of Churn and Retention Rate of Customers in Telecom Industry Using Logistic Regression.* In 2024 International Conference on Emerging Innovations and Advanced Computing , Sonipat, India, 2024"; doi: `https://doi.org/10.1109/INNOCOMP63224.2024.00124`.

[31] *"M. I. Habibie and N. Nurda. Performance Analysis and Classification Using Naive Bayes and Logistic Regression on Big Data."* In 2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS), Surakarta, Indonesia, 2022, pp. 48-52; doi: `https://doi.org/10.1109/APICS56469.2022.9918793`.

[32] *"G. Shiwei. An Improved KNN Based Decision Algorithm for Vertical Handover in Heterogeneous Wireless Networks.* In 2021 40th Chinese Control Conference (CCC), Shanghai, China, 2021"; doi: `https://doi.org/10.23919/CCC52363.2021.9550412`.

[33] *"S. Henna. A Throughput Analysis of TCP Variants in Mobile Wireless Networks.* In Proceedings of the 2009 Third International Conference on Next Generation Mobile Applications, Services and Technologies , 2009"; doi: `https://doi.org/10.1109/ngmast.2009.71`.

[34] *"N. Alzibdeh, M. T. Alrashdan, and A. Almabhouh. Bandwidth Utilization with Network Traffic Analysis.* In Proceedings of the 2023 3rd International Conference on Mobile Networks and Wireless Communications (pp. 1-5), 2023", Tumkur, India; doi: `https://doi.org/10.1109/ICMNWC60182.2023.10435712`.

[35] *"B. Panjavarnam, P. Poornima, and K. Subathra. Network Traffic Prediction in 4G-LTE Using Machine Learning Techniques.* In 2024 International Conference on Communication, Computing and Internet of Things, Chennai, India, 2024." doi: `https://doi.org/10.1109/IC3IoT60841.2024.10550364`.

[36] *"S. Li, Y.-Q. Deng, Z.-L. Zhu, H.-L. Hua, and Z.-Z. Tao. A Comprehensive Review on Radiomics and Deep Learning for Nasopharyngeal Carcinoma Imaging.* In Diagnostics", doi: `https://doi.org/10.3390/diagnostics11091523`.

[37] *"S. M. Kala, V. Sathya, K. Dahiya, T. Higashino, and H. Yamaguchi. Identification and Analysis of a Unique Cell Selection Phenomenon in Public Unlicensed Cellular Networks Through Machine Learning."* , 2022. doi: `https://doi.org/10.1109/ACCESS.2022.3199409`.

[38] *"L. Jančička, J. Koumar, D. Soukup, and T. Čejka. Analysis of Statistical Distribution Changes of Input Features in Network Traffic Classification Domain."* in Proceedings of the 2024 IEEE Network Operations and Management Symposium (NOMS), Seoul, Korea, 2024, pp. 1-4. doi: `https://doi.org/10.1109/NOMS59830.2024.10575630`.

[39] *"M. N. Reza, S. F. Kabir, N. Jahan, and M. Islam. Evaluation of Machine Learning Algorithms using Feature Selection Methods for Network Intrusion Detection Systems.* in Proceedings of the 2021 5th International Conference on Electrical Information and Communication Technology, Khulna, Bangladesh, 2021". doi: `https://doi.org/10.1109/EICT50815.2021.9401715`.

[40] *"I. Sülo, Ş. R. Keskin, G. Doğan, and T. Brown. Energy Efficient Smart Buildings: LSTM Neural Networks for Time Series Prediction.* 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications , 2019". doi: `https://doi.org/10.1109/Deep-ML.2019.00012`.

[41] *"S. Leone, M. Koroma, and I. Sawaneh. A 4G LTE Evolved Packet Core Planning and Deployment in Freetown."* In *Asian Journal of Interdisciplinary Research*, 2020. doi: `https://doi.org/10.34256/ajir20231`.

[42] *"A. Issa, N. Kandil and N. Hakem. Deep Learning-based Anomaly Detection for 5G Core Mobility Management.* 2023 Fifth International Conference on Advances in Computational Tools for Engineering Applications, Zouk Mosbeh, Lebanon, 2023". doi: `https://doi.org/10.1109/ACTEA58025.2023.10194147`.

Websites

[43] "Core Network Dimensioning." Available at: ITU.

[44] "End-to-End Delay." Available at: ScienceDirect.

[45] "Online Library".Available at: Wiley Online Library.

[46] "Medium.Com" Available at: Medium.

[47] "4G Mobile Core." Available at: *Mobile Packet Core*.

[48] “Machine Learning with the Help of Python Complete Guide”. Available at: Techprofree.

[49] “LTE Network Architecture”. Available at: Huawei Support.

[50] “Overview of Congestion Avoidance and Congestion Management.” Available at: Huawei Support.

[51] “Core Network Elements and 4G PS Traffic”. Available at: Ethio telecom.

[52] “SHAP,” available at: SHAP Documentation.

Books

[53] “A. R. Mishra, *Fundamentals of Network Planning and Optimisation 2G/3G/4G: Evolution to 5G*, John Wiley and Sons Ltd, 2018”. Available at: Google Books.

[54] “Advances in Knowledge Discovery and Data Mining: 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017”. Available at: Google Books.

[55] “J. Brownlee, *LSTM With Python: Develop Sequence Prediction Models with Deep Learning*, 2017”. Available at: Google Books.

[56] “Computer Organization and Architecture: A Formula Handbook,” (N.B. Singh), available at: Google Books.