AI-DRIVEN 5G AND BEYOND MOBILE NETWORKS: RELIABILITY, MOBILITY INSIGHTS, AND SUSTAINABLE OPERATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SEMİH AKTAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

AUGUST 2025

Approval of the thesis:

AI-DRIVEN 5G AND BEYOND MOBILE NETWORKS: RELIABILITY, MOBILITY INSIGHTS, AND SUSTAINABLE OPERATIONS

submitted by **SEMİH AKTAŞ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering** _____

Assoc. Prof. Dr. Hande Alemdar
Supervisor, **Computer Engineering** _____

Examining Committee Members:

Prof. Dr. İsmail Hakkı Toroslu
Computer Engineering, METU _____

Assoc. Prof. Dr. Hande Alemdar
Computer Engineering, METU _____

Assoc. Prof. Dr. Pelin Angın
Computer Engineering, METU _____

Assist. Prof. Dr. Salih Ergüt
Computer Engineering, Rumeli Üniversitesi _____

Assist. Prof. Dr. Atay Özgövde
Computer Engineering, Boğaziçi Üniversitesi _____

Date:25.08.2025

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Semih Aktaş

Signature :

ABSTRACT

AI-DRIVEN 5G AND BEYOND MOBILE NETWORKS: RELIABILITY, MOBILITY INSIGHTS, AND SUSTAINABLE OPERATIONS

Aktaş, Semih
Ph.D., Department of Computer Engineering
Supervisor: Assoc. Prof. Dr. Hande Alemdar

August 2025, 135 pages

The advancement of 5G and the emerging vision of 6G are transforming mobile networks into intelligent and adaptive infrastructures, presenting significant opportunities alongside considerable technical and operational challenges. In 6G networks, ubiquitous integration of artificial intelligence (AI) is expected across all layers, enabling autonomous decision-making, proactive management, and unprecedented service personalization. This thesis explores how AI can address key challenges in three vital domains of mobile network operations. It investigates AI-driven approaches for predicting radio link failures to enable proactive interventions, minimizing disruptions and ensuring seamless connectivity even under adverse conditions. It further examines how AI can analyze and model human mobility patterns from network data, generating critical insights that support disaster response, urban planning, and public safety, particularly in the wake of large-scale emergencies. Lastly, it focuses on AI-based strategies to optimize energy consumption in mobile networks, developing intelligent methods to reduce power usage without compromising coverage or service quality. Collectively, these contributions underscore AI's transformative role in shaping the future of mobile commu-

nications, positioning next-generation networks to meet escalating demands for performance, resilience, and sustainability, while contributing to societal welfare and environmental responsibility.

Keywords: AI-driven mobile network, 5G and beyond, radio link failure, mobility analysis, disaster management, sustainability, energy saving.

ÖZ

YAPAY ZEKA DESTEKLİ 5G VE ÖTESİ MOBİL ŞEBEKELER: GÜVENİLİRLİK, HAREKETLİLİK İÇGÖRÜLERİ VE SÜRDÜRÜLEBİLİR İŞLETİM

Aktaş, Semih
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Hande Alemdar

Ağustos 2025 , 135 sayfa

5G teknolojisindeki ilerlemeler ve 6G vizyonunun şekillenmesi, mobil şebekeleri akıllı ve uyarlanabilir altyapılara dönüştürmektedir. Bu süreç, yeni fırsatlar sunarken aynı zamanda teknik ve operasyonel zorlukları da beraberinde getirmektedir. 6G şebekelerinde, yapay zekanın (YZ) tüm katmanlara yaygın bir şekilde entegre edilmesi beklenmekte olup, bu sayede otonom karar alma, proaktif yönetim ve benzeri görülmemiş düzeyde hizmet kişiselleştirmesi mümkün olacaktır. Bu tez, YZ'nin mobil şebeke operasyonlarının üç kritik alanındaki zorluklara nasıl çözüm getirebileceğini incelemektedir. Çalışmada, olumsuz koşullar altında dahi kesintisiz bağlantı sağlamak ve aksaklıkları en aza indirmek amacıyla, radyo bağlantı hatalarını öngörmeye yönelik YZ tabanlı yaklaşımlar araştırılmaktadır. Ayrıca, YZ'nin şebeke verilerinden insan hareketliliği kalıplarını modelleyerek, özellikle büyük ölçekli acil durumların ardından afet yönetimi, kentsel planlama ve kamu güvenliğine katkı sağlayacak kritik içgörüler üretebileceği ortaya konulmaktadır. Son olarak, mobil şebekelerde enerji tüketimini optimize etmeye yönelik YZ tabanlı stratejilere odaklanılmakta; kapsama alanı veya hizmet kalitesinden ödün verme-

den enerji kullanımını azaltacak akıllı yöntemler geliştirilmektedir. Bu katkılar bir arada değerlendirildiğinde, YZ'nin mobil iletişimin geleceğini şekillendirmedeki dönüştürücü rolü vurgulanmaktadır. Yeni nesil şebekelerin, artan performans, dayanıklılık ve sürdürülebilirlik taleplerini karşılayacak şekilde konumlandırılmasının yanı sıra, toplumsal refah ve çevresel sorumluluğa da katkı sağlaması hedeflenmektedir.

Anahtar Kelimeler: Yapay zeka destekli mobil şebeke, 5G ve ötesi, radyo bağlantı hatası, hareketlilik analizi, afet yönetimi, sürdürülebilirlik, enerji tasarrufu.

To my wife, and whole family members...

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my thesis advisor, Dr. Hande Alemdar. Before beginning this journey, I believed that the thesis process was primarily technical. However, I gradually came to understand that it is not only about pushing the boundaries of science and exploring new frontiers, but also about enduring the journey itself and learning how to stay on course. I am truly grateful to my supervisor for helping me stay on this path and for supporting me in every possible way throughout the process.

This thesis is supported by Turkcell under the BTK Graduate Scholarship Program and by TÜBİTAK-1515 (Project 5229901 – 6GEN Lab). I also thank my committee members for their insightful feedback, which improved the quality of this work.

I am deeply grateful to Dr. Salih Ergüt for his pivotal mentorship. At a time when I was close to leaving this path, his guidance gave me both the direction and encouragement I needed to continue and complete this thesis.

I am also grateful to all my colleagues, friends, and everyone who has contributed —directly or indirectly— to the development of this work over the years.

My parents' love, patience, and belief in me —together with my brother's and extended family's support— have been a constant source of strength throughout my life. Their quiet strength and steady presence have always carried me through.

Lastly, and most dearly, I would like to express my deepest gratitude and love to my beloved wife, Nihal. I know that she carried the heaviest emotional burden throughout this journey. Every missed holiday, every postponed plan, and every moment we could not share was a silent sacrifice she made so that I could complete this thesis. Without her unwavering support and infinite patience, I would not have made it this far. From now on, I will focus on making up for the time we missed and building new memories together.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xv

LIST OF FIGURES xvii

LIST OF ABBREVIATIONS xxi

CHAPTERS

1 INTRODUCTION 1

1.1 Motivation and Problem Definition 6

1.2 Contributions and Novelties 8

1.3 The Outline of the Thesis 9

2 RADIO LINK FAILURE PREDICTION FOR 5G AND BEYOND USING WEATHER AND NETWORK DATA 11

2.1 Related Work 14

2.2 The RLF Prediction Dataset 17

2.3 Methodology 21

2.3.1 Data Preparation 21

2.3.2 Logistic Regression 22

2.3.3 Support Vector Machine 22
2.3.4 Long Short-Term Memory 22
2.3.5 LSTM+: Long Short-Term Memory with Configuration Parameters Attached 23
2.4 Performance Evaluation 25
2.4.1 Experimental Setup 25
2.4.2 Evaluation Metrics 27
2.4.3 Results 28
2.4.4 Discussion 30
2.5 Conclusions and Future Work 32
3 PREDICTIVE MODELING OF HUMAN MOBILITY FROM MOBILE NETWORK DATA: INSIGHTS FOR DISASTER RESPONSE AND URBAN PLANNING 33
3.1 Background and Related Work 38
3.2 Data Collection and Processing Pipeline 40
3.3 Exploratory Analysis and Visualization of User Mobility 47
3.3.1 Micro Level Mobility Analysis: Grid Based User Mobility 47
3.3.2 Macro Level Mobility Analysis: Inter-city User Mobility 50
3.3.3 Discussion 55
3.4 Predictive Modeling of User Mobility 57
3.4.1 Methodology of Predictive Modeling of User Mobility 58
3.4.1.1 Spatial Grid Definition and Feature Extraction 59
3.4.1.2 Computation of Neighbor-Based Grid Properties 61
3.4.1.3 Grid Similarity Calculation via Cosine Similarity 62
3.4.1.4 Prediction of Target Grid Properties 63

3.4.1.5 Model Evaluation 64
Cosine Similarity Analysis 64
Clustering-Based Structural Comparison 65
Classification Metrics for Structural Assessment 66
Dimensionality Reduction for Visualization 67
3.4.1.6 End-to-End Workflow of Predictive User Mobility Modeling 68
3.4.2 Results 69
3.4.2.1 Cosine Similarity Analysis 69
3.4.2.2 Dimension Reduction with Principal Component Analysis 71
3.4.2.3 Clustering Analysis 72
3.4.2.4 Clustering-Based Classification Metrics 75
3.4.3 Discussion 77
3.5 Conclusion 78
4 AI-ASSISTED SECTOR SHUTDOWN FOR MOBILE NETWORK ENERGY EFFICIENCY: ALGORITHMS AND LIVE NETWORK TRIALS 81
4.1 Related Work 86
4.2 Problem Definition 93
4.2.1 Network Model and Notation 93
4.2.2 Coverage and Service Quality 94
Coverage Matrix. 94
Service Quality. 94
4.2.3 Energy Consumption and Objective 95
4.2.4 Constraints and Final Formulation 95
4.2.5 System Setup 96

4.2.6 Scenario 1: Offline Network Trial 98
4.2.7 Scenario 2: Live Network Trial 99
4.3 Methodology 100
4.3.1 Scenario 1: Offline Network Trial 100
4.3.1.1 Genetic Algorithm Solution (GAS) 100
4.3.1.2 Cost-Aware Heuristic Greedy Shutdown (CAHGS) 103
4.3.2 Scenario 2: Live Network Trial 105
4.3.2.1 Neighbour-Based Expert System Solution (NBESS) 105
4.4 Results and Evaluation 107
4.4.1 Base Station Energy Consumption Model 108
4.4.2 Scenario 1- Offline Network Trial with Real Network Data . . . 109
4.4.3 Scenario 2 - Live Network Trial 112
4.5 Conclusions 116
5 CONCLUSION 119
REFERENCES 123
CURRICULUM VITAE 133

LIST OF TABLES

TABLES

Table 2.1 Summary of the dataset. 19

Table 2.2 The configuration parameters in the rl-kpis. 19

Table 2.3 The summary statistics of the numerical KPIs in rl-kpis. 20

Table 2.4 Comparison of model performances. 28

Table 3.1 IMSI based transaction data. City 1 is the last observed location on Date 1 and City 2 is the last observed city on Date 2. 51

Table 3.2 Aggregated user movement data by city. The table shows the count of users moving from City 1 to City 2 between Date 1 and Date 2. 51

Table 3.3 Comparison of user mobility for Hatay: Percentage of outgoing individuals and their destination cities. 56

Table 3.4 Clustering-based classification performance across different experimental configurations. 76

Table 4.1 Summary of representative energy-saving methods in mobile networks, categorized by granularity and evaluation type. 86

Table 4.2 Variables and functions used in the NBESS algorithm. 107

Table 4.3 104 sites energy consumption table for between 2 AM - 6 AM and network-wide estimation. 112

Table 4.4 Direct effect of the model on energy consumption by presenting 8 sites where our model takes an action. Energy consumptions between 2 AM - 6 AM. 113

LIST OF FIGURES

FIGURES

Figure 1.1 Data-Information-Knowledge-Wisdom pyramid [1]. 4

Figure 2.1 Overview of radio link sites in an operator network 12

Figure 2.2 The architecture of the LSTM model and the proposed model (LSTM+) are compared. The LSTM+ model has the capability of processing time-independent (configuration) inputs and time-dependent (daily) inputs. To process configuration inputs, the new input branch is added and then the two input branches' results are concatenated. k refers to the kernel size, and r refers to the recurrent kernel size for the LSTM layer. 24

Figure 2.3 Illustration of the inputs and model prediction. The daily data refer to the concatenation of numerical radio link KPIs and weather realizations. Configuration parameters refer to frequency band, card type, etc. Historical 10-day data are used to predict the following day. . 26

Figure 2.4 ROC comparison of each model. 29

Figure 2.5 The model performances with different threshold values. 31

Figure 3.1 Overview of the data gathering and processing pipeline, depicting data flow from mobile network interactions to geolocation, data storage, and analysis modules. 42

Figure 3.2 Proposed data analysis pipeline employing a MapReduce-based architecture. The pipeline processes daily snapshots of geolocated network data to detect inter-city user movements, summarizing the process from raw data extraction through mapping, partitioning, and reducing user movements into aggregate migration counts. 44

Figure 3.3 Process of aggregating individual geolocated user events into grid-based spatial representations and summarizing measurement statistics per grid cell. 45

Figure 3.4 Illustration of UE reports to grid-based heatmap conversion. . . 48

Figure 3.5 Comparison of heatmaps for İzmir on October 22, 2022, at 00:00, 06:00, 12:00, and 18:00. 49

Figure 3.6 User movement between cities between 20.01.2023 – 27.01.2023. 52

Figure 3.7 Daily user mobility trends for selected provinces, illustrating temporal variations in user activity and inter-city movements. 53

Figure 3.8 Sankey diagram depicting user flows from Hatay to other cities, comparing the periods before and after the 6 February 2023 earthquake. 55

Figure 3.9 Overview of the predictive modeling framework leveraging inter-city similarities. Test data represents target city, train data represents reference city. 58

Figure 3.10 Visualization of the computation of neighbor-based properties. For a given grid cell g_i, the transition matrix M is used to weight the properties P_j of neighboring grids g_j, producing the neighbor-based property $NP_{g_i}(P)$. 61

Figure 3.11 Schematic overview of cosine similarity evaluation comparing predicted and ground truth feature vectors. 65

Figure 3.12 Workflow of clustering-based classification for assessing structural consistency between predictions and ground truth data. 66

Figure 3.13 Illustration of the overall predictive modeling system pipeline, integrating grid similarity calculation, neighbor-based property prediction, and evaluation against ground truth data. 68

Figure 3.14 Average cosine similarity scores achieved by the predictive modeling framework for different top-k neighbor counts in the reference city. Error bars represent variability across all test grids. 71

Figure 3.15 Cumulative explained variance ratio as a function of the number of principal components. 72

Figure 3.16 Silhouette scores as a function of the number of clusters, comparing results with and without PCA, as well as for different numbers of PCA components. 73

Figure 3.17 Visualization of clustering results based on K-Means applied to PCA-reduced grid features. 74

Figure 4.1 Illustration of the selected region. The blue dots represent the locations of measurement reports created by users. 97

Figure 4.2 Genetic algorithm phases. 101

Figure 4.3 Illustration of a mutation and a crossover (adapted from [2]). . . 103

Figure 4.4 Base station enegy consumption model. 108

Figure 4.5 The GAS results are presented with various RSRP thresholds. . . 110

Figure 4.6 Comparision of the GAS and the CAHGS. 111

Figure 4.7 Energy consumption between 2 AM - 6 AM of 8 sites that our model takes an action. 114

Figure 4.8 Daily unique IMSI count at the work region; blue, red, and green lines represent 10 PM–2 AM, 2 AM–6 AM, and 6 AM–10 AM time slots, respectively. 115

Figure 4.9 The number of missing residents is presented for various thresholds. A missing resident is defined as an IMSI that is normally observed frequently in the region but was not observed on the target date. 116

LIST OF ABBREVIATIONS

ABBREVIATONS	
1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
6G	Sixth Generation
ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
ASMs	Advanced Sleep Modes
AUC	Area Under Curve
BBU	Baseband Unit
BS	Base Station
CDMA	Code Division Multiple Access
CDR	Call Detail Records
CIC	Confident Information Coverage
CSG	Closed Subscriber Groups
CSO	Cell Switch-off
DIKW	Data–Information–Knowledge–Wisdom
DRL	Deep Reinforcement Learning
DTST	Dynamic Transmitter Shutdown Technique
eMBB	enhanced Mobile Broadband

ENI	Experiential Network Intelligence
ESF	Energy Sustainability Framework
ETSI	European Telecommunications Standards Institute
FPR	False-positive Rate
GA	Genetic Algorithm
GAN	Generative Adversarial Networks
GAS	Genetic Algorithm Solution
GHz	Gigahertz
GPS	Global Positioning System
GSMA	Global System for Mobile Communications
IMSI	International Mobile Subscriber Identity
IMT	International Mobile Telecommunications
IoT	Internet of Things
IP	Internet Protocol
ITU	International Telecommunication Union
KPI	Key Performance Indicator
L-BFGS	limited-memory Broyden–Fletcher–Goldfarb– Shanno
LR	Logistic Regression
LSTM	Long Short-term Memory
LTE	Long-Term Evolution
Mbps	Megabits Per Second
MDAS	Management Data Analytics Services
MIMO	multiple input, multiple output
ML	Machine Learning
mMTC	massive Machine-Type Communication
MNO	Mobile Network Operator
NBESS	Neighbour Based Expert System Solution

NGMN	Next Generation Mobile Networks
NOC	Network Operations Center
NWDAF	Network Data Analytics Function
O-RAN	Open Radio Access Network
PCA	Principal Component Analysis
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RBESS	Rule-Based Expert System Solution
RF	Radio Frequency
RIC	Radio Intelligent Controller
RL	Reinforcement learning
RLF	Radio Link Failure
RNN	Recurrent Neural Network
ROC	Receiver Operator Characteristics
RSRP	Reference Signal Received Power
SARSA	State-action-reward-state-action
SHAP	Shapley Additive Explanations
SINR	Signal-to-Interference-plus-Noise Ratio
SMS	Short Message Service
SVM	Support Vector Machines
TPR	True-positive Rate
TRXs	Transceivers
t-SNE	t-Distributed Stochastic Neighbor Embedding
UE	User Equipment
URLLC	Ultra-Reliable Low-Latency Communications
W	Watt
ZSM	Zero Touch Network & Service Management

CHAPTER 1

INTRODUCTION

Mobile networks have undergone a remarkable evolution over the past four decades, progressing through successive generations with each delivering new capabilities and performance improvements. The first generation (1G) systems of the 1980s introduced basic analog voice cellular service, while the second generation (2G) in the 1990s transitioned to digital communication, enabling short message service (SMS) and more secure voice calls through encryption [3]. The 2000s saw the rise of third generation (3G) networks, which were a turning point in offering mobile data services and internet access on phones, albeit at modest data rates (on the order of a few hundred kilobits to a few megabits per second). The advent of fourth generation (4G) in the 2010s marked the era of true mobile broadband: all-IP networks capable of supporting high-speed internet, video streaming, and a new ecosystem of smartphone applications. 4G, also known as Long-Term Evolution (LTE), networks provided dramatically higher throughput (5Mbps-100Mbps) and lower latency, effectively transforming mobile devices into ubiquitous computing and multimedia platforms [3]. Each generational leap not only increased peak data rates and spectral efficiency but also introduced new services – for example, 2G enabled multimedia messaging, 3G facilitated mobile web and video calls, and 4G unlocked high-definition streaming and cloud connectivity.

In the current fifth generation (5G), first deployed around 2020, mobile networks have further expanded their scope and performance. 5G networks are defined by three primary service categories: enhanced Mobile Broadband (eMBB) for extremely high data rate applications, Ultra-Reliable Low-Latency Communications (URLLC) for mission-critical services (such as autonomous driving or remote surgery)

requiring minimal delay and near-perfect reliability, and massive Machine-Type Communications (mMTC) to connect billions of Internet of Things (IoT) devices efficiently. By leveraging new radio technologies (e.g., millimeter-wave frequencies, massive multiple input multiple output [MIMO] antenna arrays) and network architectures (such as network slicing and edge computing), 5G aims to provide an order-of-magnitude improvement in capacity and end-to-end latency over 4G systems [4]. Commercial 5G deployments already offer peak download speeds in the gigabits per second and latency on the order of a few milliseconds for specific applications. These capabilities open the door for advanced use cases like immersive virtual/augmented reality and connected autonomous systems, which were difficult to support with previous technologies.

Looking beyond, research and industry roadmaps are now charting the course for sixth generation (6G) mobile networks, expected to emerge around 2030. While still in the early research phase, the vision for 6G is taking shape through white papers and initial standards discussions. Broadly, 6G is expected to provide unprecedented capacity and speed performance – some predictions cite terabit-per-second data rates and 1,000-fold increases in traffic capacity relative to 5G – along with extreme coverage and reliability [4]. To achieve this, 6G will likely take advantage of terahertz frequencies and advanced modulation, while also incorporating new communication paradigms (such as intelligent reflecting surfaces and holographic MIMO) and tighter integration of satellite/airborne networks for global coverage. Perhaps the most distinguishing aspect of 6G is the emphasis on built-in intelligence and adaptability. There is a growing consensus that 6G networks will artificial intelligence (AI) native, meaning that AI will be deeply embedded into the network's design and operations from the outset [5, 6, 7]. This includes using AI for dynamic resource management, traffic prediction, and autonomous control of network functions. The result would be a network that is not only faster and more ubiquitous, but also more context-aware and self-optimizing. Another key pillar of the 6G vision is sustainability – 6G aims to be significantly more energy-efficient and environmentally friendly in its operation, aligning with global carbon reduction goals. Future networks are expected to intelligently minimize energy consumption (for example, by dynamically powering down unused infrastructure or

using energy-efficient hardware and AI-driven optimizations) while maintaining performance. In summary, the trajectory from 1G to 5G has been one of expanding capabilities (from voice to broadband data) and use cases, and the emerging 6G vision continues this trajectory toward an intelligent, all-encompassing wireless fabric that can support instantaneous, zero-distance connectivity for a fully digital society [4]. The evolution sets the stage for the new challenges and opportunities that come with managing such complex and powerful networks.

In the era of 5G and beyond, data is often touted as one of the most valuable resources for telecom operators. Just as oil was crucial in the industrial age, data has been dubbed "the new oil" in the digital age [8], but its value lies in how it is refined and used. Network data, if properly exploited, can reveal insights into user behavior, network conditions, and emerging problems, enabling a shift from reactive to proactive operations. For example, by analyzing traffic patterns, operators can predict where capacity hotspots will occur and dynamically reroute or provision resources ahead of time. By mining device mobility data, networks can anticipate user movements (e.g., a daily commute route) and pre-allocate resources to cells along that path, thus improving the user's experience with seamless connectivity. Internal network metrics can be used to detect anomalies or degradations (such as an unusual rise in error rates on a base station) and trigger preventive maintenance or self-healing protocols before users even notice an issue. In essence, rich data enables the network to "learn" and adapt to real conditions in ways that static configuration could never achieve.

The exploitation of big data in telecommunications has given rise to the field of data-driven network management and optimization. Traditional approaches to network management relied heavily on human experts and predefined rules; these approaches struggle to cope with the growing complexity and scale of modern networks. By contrast, data-driven methods use techniques from statistics, data mining, and machine learning (ML) to automatically glean insights and make decisions. Studies show that applying big data analytics can significantly improve network Quality of Service (QoS) and user Quality of Experience (QoE) by uncovering patterns invisible to manual analysis [9]. For example, machine learning models can identify subtle correlations between network events (like how certain weather

conditions affect radio link performance, or how traffic surges in one area impact another) and optimize accordingly. Mobile operators are increasingly building data analytics platforms that ingest multi-dimensional data (technical metrics, but also sometimes external data like weather or social events) to drive optimization algorithms. This data-centric innovation also extends to new services, for example, using aggregated mobility data from the network to inform smart city planning or traffic management, while respecting privacy. In summary, the explosive growth of data is a double-edged sword: It stresses the network, but it also empowers the network. The key is to treat data as a strategic asset, investing in analytics capabilities to transform raw data into actionable intelligence to run and improve the network.

From Data to Information to Wisdom: Leveraging Data in Mobile Networks

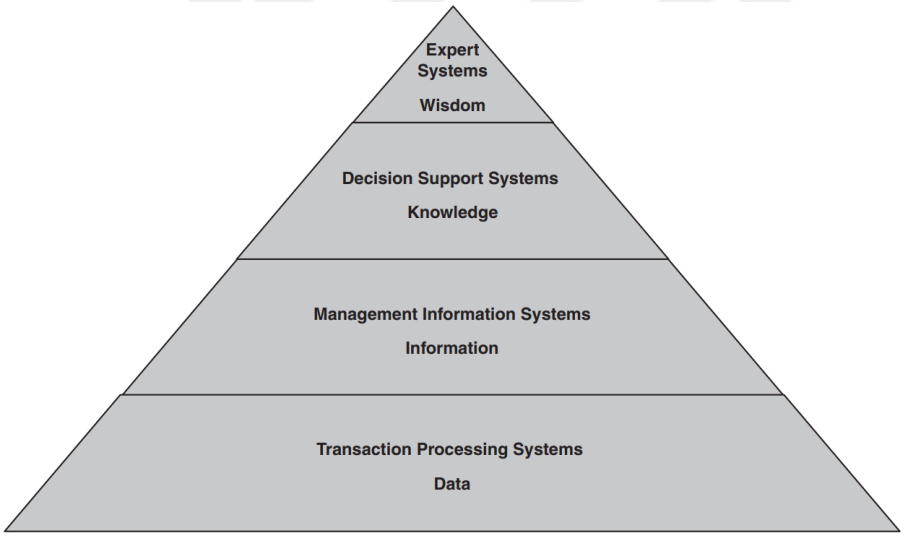


Figure 1.1: Data-Information-Knowledge-Wisdom pyramid [1].

Effectively utilizing data in mobile networks calls for frameworks that guide how raw data can be converted into meaningful insights and decisions. One conceptual model commonly referenced in information science and knowledge management is the Data–Information–Knowledge–Wisdom (DIKW) hierarchy [10]. This model, commonly illustrated as a pyramid, delineates four levels of abstraction in the understanding and use of data. As shown in Figure 1.1, the DIKW pyramid consists of

the following levels:

- **Data:** the base of the pyramid, representing raw facts and observations. In the context of mobile networks, data would be the direct measurements or records – for instance, a stream of timestamped signal strength readings, logs of call drops, user locations, or throughput values on a link. Data in itself has no judgment attached; it is simply the recorded output of network operations or user activities.

- **Information:** the next level, obtained by processing or organizing data to answer basic questions (who, what, when, where). Information adds context and meaning to data. For example, if raw data indicates that a certain cell tower has a 5% call drop rate, turning this into information might involve calculating averages or trends (Is 5% unusually high compared to the norm? Does it occur at particular times or for particular devices?). In networks, information could be a summarized report like "Cell ID 123 experienced a spike in drop rate to 5% between 8–9 PM, which is double its typical value." This begins to flag what the situation is.

- **Knowledge:** the next level where information is further analyzed to understand why something is happening or how to influence it. Knowledge involves recognizing patterns, rules, or models that underlie the information. In our example, knowledge would involve diagnosing why the call drop rate spiked – perhaps by correlating it with other information (e.g., a surge of users at a concert in that area, or a hardware fault, or bad weather interference). Knowledge in telecom could be a model or principle, such as knowing that "an X% increase in traffic load leads to Y% increase in drop rate under current configuration" or that "rain fade on a microwave backhaul link will degrade throughput." It often results from applying expert analysis or machine learning to the information.

- **Wisdom:** the pinnacle of the pyramid, referring to the ability to make sound judgments and decisions based on knowledge. Wisdom entails not just understanding the immediate cause-effect, but also having the insight to decide the best course of action and foresee consequences. In the network ex-

ample, wisdom would be using the knowledge of the cause (say, overload due to a local event) and deciding how to respond – perhaps rerouting traffic, temporarily offloading users to neighboring cells, or scheduling an on-demand coverage solution (like a mobile cell on wheels) for future events. Wisdom in a network context might also involve strategic decisions, like how to design networks to avoid such issues or how to prioritize investments (e.g., reinforcing capacity where recurring events happen). It often requires combining domain experience with the predictive understanding gleaned from knowledge.

In this thesis, the vast amounts of data generated by mobile networks are leveraged through AI techniques, enabling the progression from raw data toward higher levels of the DIKW hierarchy and facilitating the extraction of actionable knowledge and wisdom essential for intelligent and proactive network management. By bridging the gap between data-driven analytical insights and intelligent network control, the approaches developed in this work contribute to the broader vision of creating adaptive, context-aware, and AI-native networks anticipated in the evolution toward 5G and 6G ecosystems.

1.1 Motivation and Problem Definition

The evolution of mobile networks into 5G and the emerging vision of 6G are fundamentally transforming telecommunications infrastructures into intelligent, adaptive ecosystems. These advancements introduce significant opportunities for improving service quality, operational efficiency, and user experience. However, they also bring forth considerable technical and operational challenges driven by the increasing complexity of networks, growing data volumes, and stringent performance requirements such as ultra-low latency, high reliability, and sustainability.

AI and data-driven approaches are poised to play a pivotal role in addressing these challenges. By enabling predictive analytics, autonomous decision-making, and optimization processes, AI methods can unlock new capabilities in proactive network management and resource allocation, ensuring networks can meet future de-

mands while maintaining resilience and efficiency.

This thesis is motivated by three critical and interrelated problem domains within the context of AI-driven mobile networks:

(i) Radio Link Failure (RLF) Prediction. Reliable network connectivity is essential for maintaining uninterrupted service quality. Yet, radio links, particularly in high-frequency bands anticipated in 5G and beyond, are vulnerable to environmental factors such as adverse weather conditions. Traditionally, network management has been reactive, responding to failures only after they occur. However, such reactive measures often result in service interruptions and increased operational costs. This thesis explores AI-driven predictive modeling techniques to forecast radio link failures before their occurrence, enabling proactive interventions and minimizing network disruptions.

(ii) Mobility Analysis and Predictive Modeling. Understanding user mobility patterns is necessary for effective network planning, disaster response, and urban management. However, obtaining accurate insights into population movement remains challenging, particularly during large-scale emergencies where network infrastructures may be disrupted. This thesis investigates AI-based methodologies to analyze and predict human mobility from mobile network data. Using both spatial and temporal features, the proposed approaches aim to generate actionable insights that can inform decision-making authorities. In addition, the results of this research have significant potential for government authorities and public institutions, allowing them to utilize mobile network data to better understand and monitor population movements for urban planning, disaster management, and public safety. The work exemplifies an interdisciplinary approach, integrating domain knowledge about geographic and locational properties with patterns of user mobility to produce insights.

(iii) Energy Saving in Mobile Networks. The environmental impact and economic cost of operating dense mobile network infrastructures are significant concerns, especially as networks become more complex and power-intensive. Energy efficiency, therefore, emerges as a key priority for sustainable network operations. However, achieving meaningful reductions in energy consumption without com-

promising coverage or service quality is a complex problem. This thesis focuses on the development of AI-assisted strategies to optimize energy usage in mobile networks, proposing solutions that intelligently manage network resources to reduce power consumption while ensuring service continuity.

Collectively, these research directions aim to demonstrate how AI and data-driven techniques can contribute to the reliability, adaptability, and sustainability of next-generation mobile networks.

1.2 Contributions and Novelties

This thesis investigates the application of AI-driven methods to enhance reliability, extract mobility insights, and achieve sustainable operations in 5G and beyond mobile networks. The main contributions and novelties of this research can be summarized as follows:

- **AI-Based Radio Link Failure Prediction:** Development of machine learning models, including a novel deep learning architecture, capable of predicting RLFs by jointly processing time-dependent data (weather and radio link key performance indicators [KPIs]) and time-independent configuration parameters. This enables proactive network interventions, reducing service disruptions and operational costs.
- **Human Mobility Analysis and Predictive Modeling:** Proposal of an end-to-end AI-driven framework for analyzing and predicting human mobility patterns from mobile network data. The methodology incorporates grid-based spatial representations, neighbor-based property computation, and inter-city similarity assessments to generate critical insights for disaster response and urban planning.
- **AI-Assisted Energy Saving Techniques:** Design and implementation of intelligent algorithms for optimizing energy consumption in mobile networks. The research introduces both offline and live network trial methodologies, employing solutions such as genetic algorithms and expert systems to re-

duce power usage while preserving coverage and service quality.

By addressing these challenges in an integrated manner, this thesis contributes to the advancement of AI techniques tailored for telecommunications, supporting the vision of reliable, intelligent, and sustainable future mobile networks.

1.3 The Outline of the Thesis

This thesis is organized into three main chapters, along with an introduction and a conclusion, making a total of five chapters. Each main chapter is focusing on a research domain within the context of AI-driven mobile networks.

Chapter 2 investigates the prediction of RLFs in 5G and beyond mobile networks. It introduces ML methodologies, including a novel deep learning architecture, to forecast failures by jointly modeling historical weather data, radio link KPIs, and base stations (BS) configuration parameters. The proposed approaches aim to enable proactive network interventions, thus improving network reliability and reducing operational costs.

Chapter 3 presents an AI-based framework for the analysis and predictive modeling of human mobility patterns using mobile network data. The chapter details methods for transforming network measurements into spatial representations and proposes techniques for modeling mobility, particularly in the context of disaster response and urban planning. The insights generated contribute to more effective decision-making by governmental and public institutions.

Chapter 4 focuses on energy-saving strategies in mobile networks. It explores AI-assisted solutions, including genetic algorithms and expert systems, designed to reduce energy consumption while maintaining network coverage and service quality. The chapter includes both offline and live network trials, demonstrating the practical feasibility and benefits of the proposed methods.

Finally, Chapter 5 concludes the thesis by summarizing the key findings and contributions of this research. It discusses the implications of AI integration into mobile network operations and suggests potential directions for future work in creat-

ing intelligent, sustainable, and resilient communication networks.

CHAPTER 2

RADIO LINK FAILURE PREDICTION FOR 5G AND BEYOND USING WEATHER AND NETWORK DATA

Mobile communications networks evolve in generations. At each generation, improvements in the QoS, efficiency and performance take a big step. Over the past few decades, they have evolved into 4G systems, and currently, we are transitioning to the 5G networks. In this new generation and the ones that will follow (5G and Beyond), uninterrupted service and full connectedness are among the key requirements together with the improvements in other key performance indicators such as speed, data rate, latency, mobility, and energy efficiency. That is why AI will play a central role for the next generations of communication networks in providing improved customer experience on a network serving a diverse set of applications.

Dealing with increased complexity in such networks mandates automation rather than manual operation. To identify potential use cases, analyze gaps, and propose new network architectures, many standard development organizations, including the 3rd Generation Partnership Project (3GPP), Global System for Mobile Communications (GSMA), International Telecommunication Union (ITU), European Telecommunications Standards Institute (ETSI), Next Generation Mobile Networks (NGMN) have initiated dedicated working groups. ITU's focused groups on *Machine Learning for Future Networks including 5G* and *Autonomous Networks*, 3GPP's *Network Data Analytics Function (NWDAF)* specifications, ETSI's *Experiential Network Intelligence (ENI)* and *Zero Touch Network & Service Management (ZSM)* groups are to name a few.

One of the crucial aspects of user experience is the continuity of service without any interruptions. Radio links are deployed in the field to connect base stations to

each other and to the core network in varying types of topologies. 5G and beyond networks will require denser deployments to provide promised high data rates and ultra-low latencies, which can only be realized by even more such wireless radio links. To provide an excellent customer experience, it is essential that radio links are always up and running even in extreme weather conditions.

Cloud, rain, snow, and other weather-related phenomena affect the performance of radio links. This is especially applicable to backhaul links operating at GHz frequencies. It is expected that high-frequency bands will be utilized heavily in next-generation mobile networks [11] and high-frequency bands are more sensitive to weather conditions than traditionally used frequency bands. Therefore, the weather effect on radio links will be more pronounced in the future. The radio link connections are illustrated in Fig. 2.1.

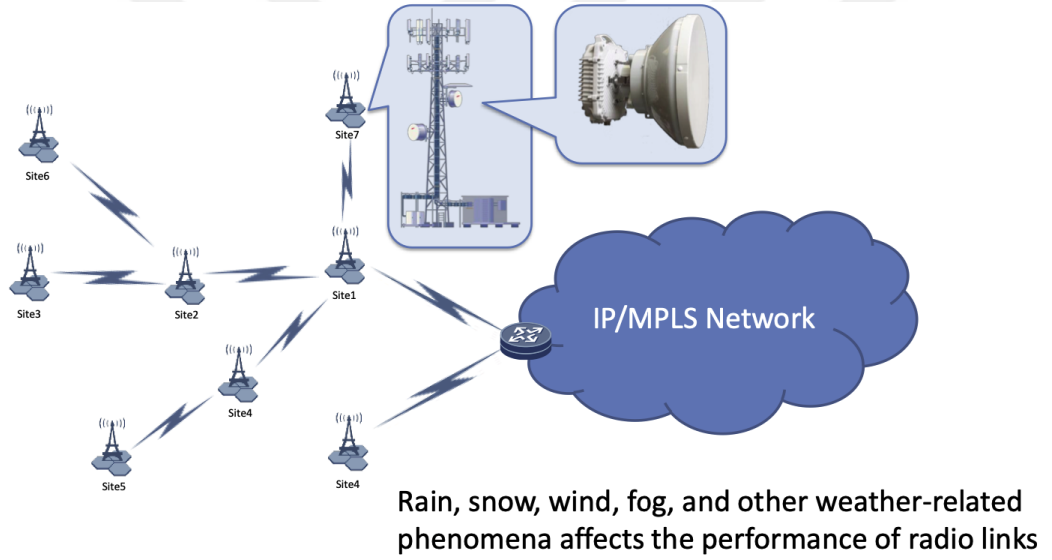


Figure 2.1: Overview of radio link sites in an operator network

Besides the weather conditions, other parameters may also affect the link behavior, and timely detection of such factors is critical for mobile operators. For this reason, companies make large investments in monitoring tools to assist the maintenance, configuration, and optimization of their live networks. Mobile operators make these investments in constant network monitoring so that when a link failure is detected, the configuration could be changed to alleviate the adverse effects in a very short amount of time, and thus, the users will be less affected [12]. How-

ever, since this is a reactive approach, some downtime is unavoidable. It becomes an even bigger problem when link failures occur in multiple places, causing large-scale interruptions in the network. A proactive approach that predicts the radio link failures (RLF) before they occur is more desirable as it provides enough time to take precautionary measures. Assessing risks for link failures is a critical yet not so straightforward task.

In order to address this issue, we propose using ML algorithms to model the complex nature of link failures by using a set of network-related factors together with the weather conditions. To this end, historical data, including weather, KPIs, and configuration parameters are utilized to train a model for predicting link failures. There are several challenges in training such models. To begin with, link failures are infrequent events in the network (typically less than %0.1 percent), and hence we need to train the models using an extremely unbalanced dataset. We use under-sampling to train the models since the class distribution is highly imbalanced. Moreover, weather forecasts that will be used in prediction are not accurately available at the time of prediction.

In this study, we propose using machine learning to predict whether there will be a radio link failure on day $N+1$ given previous N-day data for a specific link. Long Short-term Memory (LSTM) networks are among the novel deep learning architectures that have been shown to perform well on time-series data. We propose a novel LSTM+ architecture to process time-series data (radio link KPI and weather data) and non-times-series data (link configuration parameters) together in the same neural model to improve performance. We validate our approach using k-fold cross-validation on the real-world dataset we introduce in this study. We also compare our LSTM+ model with the traditional approaches such as logistic regression (LR), support vector machines (SVM), and pure LSTM network.

We realize that there is a lack of research on predicting RLFs using network KPIs and configuration parameters and weather data, probably due to the difficulty of accessing live network data from mobile operators. To the best of our knowledge, this study is a pioneer in this area in terms of using this type of data for model creation and making this dataset publicly available to the research community. To

summarize, we make the following contributions:

- We introduce a new real-world dataset collected from a live operator network. The data structure, column definitions, and relations are presented. In addition to that, data collection methods and data statistics are shared. The dataset is not only limited to predicting radio link failures, it can also be used in different studies.
- We define the radio link failure problem. By using previous N-day data, we predict the radio link failure for day $N+1$ with a binary label. We propose using weather conditions data, KPI parameters together with link configuration parameters for the prediction.
- We propose a novel LSTM-based architecture called LSTM+. We use radio link KPIs and weather data in time-series representation and link configuration parameters that do not depend on time together in the same model in LSTM+. We compare our model's performance with other models such as LR, SVM and LSTM. The performances of the models and their effects are interpreted from the operator's point of view.

The rest of this chapter is organized as follows. In Section 2.1, we cite several related studies in the literature. Section 2.2 introduces the RLF prediction dataset we use in this study. In Section 2.3, we describe our methodology for achieving the link failure prediction using ML and we describe our LSTM+ approach. We evaluate our performance on the real-world dataset in Section 2.4. Finally, we conclude and propose future directions in Section 2.5.

2.1 Related Work

Even though AI has transformed many industries, AI applications have been slow to adapt in communication networks considering the vast amount of data these networks offer. Communication service providers are familiar with non-network-related AI use cases, such as churn prediction, customer segmentation, and best offer prediction. With 5G and beyond communication networks, network com-

plexity is increasing to a point where manual configuration and operation of networks is becoming no longer feasible and hence increasingly more deployments of AI-based use cases are expected [13, 14].

Autonomous systems will be employed widely in operator networks to manage dense networks to meet the demands of a variety of services and deal with the increased complexity in 5G and beyond technologies [15]. As vast amounts of data are generated in the operator infrastructure, more network-focused applications using ML algorithms as well as widely popular deep learning algorithms are being used in operator networks. Especially, LSTM, an artificial recurrent neural network (RNN) architecture used in deep learning, is finding many use cases in traffic forecasting or failure prediction that involve processing time-series and any sequential data produced by the mobile network. Adaptive training-based RNNs are gaining traction as they do not have any underlying assumptions on the historical data, and due to their online nature, efficient algorithms are being studied [16, 17]. Zhang et al. provide a comprehensive survey for the use of deep learning techniques in the next-generation mobile networks [18].

Among the most popular machine learning use cases deployed in mobile networks, network optimization ranks first. Traffic forecasting assists network optimization and thus enables efficient use of network resources. [19], and [20] employ LSTM models for such traffic forecasting problems.

Wei et al. get the benefits of machine learning to address user scheduling and content caching in mobile networks [21]. They train a deep reinforcement learning (DRL) agent to develop a decision mechanism for content serving. Reinforcement learning-based algorithms can also be used to adjust resource assignment for network slices, which are considered vital technologies in enabling running services with varying requirements such as ultra-low latency and high bandwidth [22].

The second use case is energy efficiency. Mohanty and Udgata propose a genetic algorithm to study signal-to-interference-plus-noise ratio (SINR) based adaptive transmission that minimizes the transmission power and required number of scheduling slots [23]. They also discuss their solution's robustness in false data injection attacks and other malicious attacks that may lead ML algorithms to fool.

Another use case is predictive maintenance. Network operations center (NOC) in an operator constantly monitors alarms and performance indicators across the network and takes remedial actions. Predicting performance issues and identifying root causes is crucial for an effective NOC.

A mobile operator network consists of wired and wireless radio links interconnecting network elements. Radio links' performances may be affected by environmental conditions. Changing environmental conditions such as temperature, rain, or humidity affects signal transmission [24]. When the signal cannot be transmitted strong enough to the receiver side, RLF occurs. The effects of weather phenomena are not limited to radio link failures but also link capacity.

Worsened weather conditions cause significant degradation in signal strength [25]. Therefore, robust systems that can handle varying weather conditions should be designed along with methodologies to predict link quality. Javed et al. propose a routing algorithm for adaptive wireless mesh networks to make radio links robust to weather phenomena, where they utilize link quality predictions in their adaptive routing algorithms [26]. According to their predictive method, each link observes their neighbors' Bit Error Rate (BER), and then they record the changes with previous time intervals. Their approach is based on the comparison of the change in time with a constant threshold. Their method is based on statistical, whereas we are using machine learning to analyze time-series data.

The studies in the literature show that machine learning has become widespread in the network. Machine learning algorithms flourish with the availability of data, and it is difficult for researchers to access live network data from operators. To the best of our knowledge, it is the first time that a complete dataset, including long-period weather data and radio link data is shared. In this chapter, real-world telecom operator network data is used while developing and testing our approach. Therefore, the related works are stated to show differences between different datasets. The operators may be inspired by this study to improve their data collection methods and they may use their data to model other problems by using machine learning. The data is publicly available to encourage further research in this area.

2.2 The RLF Prediction Dataset

The data is collected from the mobile network of Turkcell, which is a converged telecommunication and technology services provider, founded and headquartered in Türkiye. Regional radio link performance and configuration data and weather data recorded from weather stations in the same region are collected between the years 2018 to 2020. Due to privacy concerns, certain configuration parameters and performance data are categorized and normalized without loss of information. These fields are generally related to the used equipment name, link IDs, etc. Instead of actual locations of radio link sites and weather stations, pairwise distances between them are provided due to security issues.

The data consist of the following three parts:

1. *Radio link data:* consists of KPI measurements aggregated daily and configuration parameters.
    - **`rl-sites`**: This table contains radio link site identifiers and site-specific parameters such as height and clutter class. There are 1674 unique radio link sites stated in that table. Each radio link sites' deployment environment is categorized under 21 clutter classes.
    - **`rl-kpis`**: This table contains daily aggregated key performance indicators (KPIs) (severally error second, error second, unavailable second, available time, block bit error, capacity), link-specific configuration parameters (card type, modulation, adaptive modulation, frequency band) and radio link failure occurrence. There are 1992986 entities in this table. Each link is uniquely identified with a pair of site-id and mini-link-id, and their daily measurements are presented in this table. There are seven unique card types, 13 unique modulations, and six unique frequency bands.
2. *Weather station data:* consists of hourly weather realizations and five-day weather forecasts for any given day.
    - **`met-stations`**: This table contains weather station identifiers and

site-specific parameters such as height and clutter class information. There are 20 weather stations that are uniquely described by using station no, and they are categorized according to their deployment environment under the eight classes (clutter class).

- **`met-real`**: This table contains hourly historical weather data (realizations) that includes temperature, maximum temperature observed in that time interval, wind direction, wind speed, humidity, precipitation, pressure. There are 479241 entities where each entity is uniquely identified by using station no, observation date, and measurement hour.
- **`met-forecast`**: This table includes the following 5-day weather forecast (rain, snow, hot day, etc.), temperature, humidity, wind speed, wind direction. There are 20323 entities that are uniquely identified by using station no and date. The forecasts include the following five days of the given date in that row. Each prediction column is named with variable and prediction day. Variables are weather, maximum temperature, minimum temperature, maximum humidity, minimum humidity, wind direction, wind speed. These variables are predicted daily for the following five days.

3. *Distance matrix:* consists of pairwise distances between all radio link stations and weather stations.

    - **`distances`**: This table contains pairwise distances between all sites, including radio link sites and weather stations. The distances are scaled while maintaining the distance relationship between stations to ensure privacy. Therefore, the distances specified here should be considered in units. This square matrix consists of 1694 rows, one for each station (radio link station and weather station) that is used in the previous tables.

Table 2.1 summarizes the number of records and features for all the tables in the dataset. Occurrences of radio link failures are given in the **`rl-kpis`** table. It is challenging to predict RLFs since the dataset is highly unbalanced. There are only 1219 radio link failures over 2 million available records within the two-year interval.

Table 2.1: Summary of the dataset.

Table	**Records**	**# of Features**	**Short Description**
rl-sites	1674	3	Radio link site information
rl-kpis	1992986	19	Daily aggregated link KPIs and configurations
met-stations	20	3	Weather station information
met-real	479241	16	Historical hourly weather information.
met-forecast	20323	38	Daily meteorology forecast for following 5 days.
distances	1694	1694	Distances between all stations

When we consider the total number of sites reported in the dataset as being 1674, we can deduce that some sites do not even experience an RLF in the given period.

Table 2.2: The configuration parameters in the rl-kpis.

Columns	**# of unique**	**Categories**
type	2	ENC, NEC
tip	2	FAR, NEAR
card_type	7	Enumerated card types: card_type1, cardtype2...
adaptive_modulation	2	Enable, Disable
freq_band	5	Enumerated frequency bands: f1, f2...
modulation	13	16 QAM, 64 QAM, 128 QAM...

We provide detailed dataset statistics. In the *rl-sites* table, there are three columns: *site_id*, *groundheight*, and *clutter_class*. There are 1674 site-ids that refer to the unique deployment IDs for base stations. Each site may have more than one connection between other sites and these connections are called the *links*. The sites are clustered under the 21 clutter classes according to their deployment environment. In the *rl-kpis* table, there are 19 columns and each entity is uniquely identified by using *datetime*, *site_id*, and *mlid*. In the rl-kpis, link configurations are stored under six columns. In Table 2.2, the configuration parameters are stated. In the rl-kpis, there are eight columns for KPIs. For these columns, distributions of the data are summarized in Table 2.3. According to distribution, we can observe that the link errors (error second, unavailable second, bit block error, etc.) rarely occur in this dataset. Generally, links are running healthy, and it implies the network failures are rarely shown events.

Table 2.3: The summary statistics of the numerical KPIs in rl-kpis.

	severally error second	error second	unavail second	avail time	Block Bit error (bbe)	rxlevmax	scalibility score	capacity
mean	0.10	1.20	1.20	84.20	46.60	-37.50	0.00	336.40
std	3.60	15.70	23.50	21.80	0.19×10^5	6.10	0.00	137.00
min	0.00	0.00	0.00	1.00	0.00	-100.00	0.00	35.00
50%	0.00	0.00	0.00	86.40	0.00	-39.50	0.00	406.00
75%	0.00	0.00	0.00	86.40	0.00	-34.40	0.00	407.00
80%	570.84	600.88	608.80	574.56	12×10^5	3.66	1.68	459.80
max	949.00	991.00	999.00	900.00	180×10^5	0.00	2.80	495.00

The weather station data is presented with three tables: met-stations, met-real, and met-forecast. The *met-stations* table stores the weather stations' deployment environment and their ground heights. There are 20 weather stations and each station is uniquely identified by using *station_no*. There are eight different deployment environments that are given under the *clutter_class* column. Each weather station records hourly observation in the *met-real* table. These observations include temperature, wind direction, wind speed, humidity, precipitation, and pressure. In the *met-real* table, there are 479241 entities and each entity is uniquely identified by using *station_no* and *datetime*.

These forecasts are stored in the *met-forecast* table. There are 20323 entities in this table and each entity is uniquely identified by using *station_no*, *datetime*, and *report_time* (morning, evening). Each forecast includes weather (heavy rain, light rain, snow, windy, clear sky, etc.), temperature, humidity, and wind. The weather data may be used to analyze the weather effects on the radio links by comparing the changes in the radio link KPIs.

The dataset is publicly available on Turkcell's official GitHub page [27].

2.3 Methodology

In this section, we describe our approach for predicting link failures using machine learning. We implement logistic regression, support vector machine, and LSTM algorithms and compare their performance with the LSTM+ model we propose. We also describe the details of our data preparation.

2.3.1 Data Preparation

The network KPI and configuration data that were collected from a live telecom operator network had to be cleaned first as it contained some inconsistencies and errors. These inconsistencies and errors included missing data stemming from hardware and software failures or upgrades, which caused interruptions in the time series data. We performed extensive analyses to determine and correct such inconsistencies in the data. Moreover, data from weather stations also contained some erroneous samples, which may have occurred due to sensor malfunctions or database write operations.

In the preprocessing phase, hourly historical weather data is aggregated with daily statistics and then joined with radio link data, which is already provided on a daily basis.

Due to security issues, the dataset does not include exact global positioning system (GPS) coordinates of weather stations or radio link sites; instead, pairwise distances between all these locations are provided in a distance matrix. As a feature extraction process, each base station's KPI and configuration data is joined with the new input parameters generated by the aggregation of the closest k weather stations' data using min, max, mean and standard deviation.

There are seven input features from radio links and 56 derived features from weather measurements. Therefore, the total number of input features is 63.

2.3.2 Logistic Regression

LR is a kind of nonlinear regression method which is originally designed for binary classification problems, which suits well with our problem of predicting failures.

For a set of input vectors $x_i \in R^d$, the formulation of LR is as follows:

$$P(y = 1|x_{i,1}, ..., x_{i,d}) = \frac{exp(\beta_0 + \beta_1 x_{i,1} + ... + \beta_d x_{i,d})}{1 + exp(\beta_0 + \beta_1 x_{i,1} + ... + \beta_d x_{i,d})} \tag{2.1}$$

where β represents the regression coefficients, x_i refers to the input and $x_{i,j}$ refers to the features. β values are estimated by using the maximum-likelihood [28]. If the $P(y = 1|x_i) > c$, where c is a fixed threshold (generally used as 0.5), then x_i is labelled as $y_i = 1$, and $y_i = -1$ otherwise.

2.3.3 Support Vector Machine

SVM is a kind of supervised machine learning algorithm with the primary objective of finding a decision boundary that separates two classes from each other. A decision boundary is also known as a hyperplane, and the distance between the closest points of each class is known as the margin [29]. These closest points of each class are also known as support vectors. For a set of input vectors $x_i \in R^d$ and their corresponding class labels $y_i \in \{1, -1\}$, the hyperplane is defined by the equation

$$w^T x_i + b = 0, \tag{2.2}$$

where w defines the hyperplane, and b is the offset. SVM algorithm tries to find the hyperplane with the maximum margin, that is the one that maximizes the distance to the closest data points from both classes.

2.3.4 Long Short-Term Memory

Our objective is to predict radio link failures. We hypothesize that the radio links may fail because of the extreme weather conditions, and historical KPIs may hint at the possibility of failure to occur. Since hardware equipment frazzles with time,

the radio links may become more sensitive to weather changes. This motivates us to use time-series measurements to predict radio link failures.

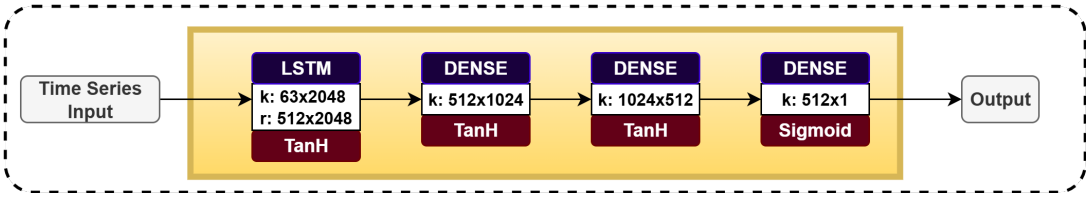
Consequently, the collected measurements represent time-dependent data, and RNN provide promising solutions for time-dependent data processing [30]. More specifically, we use LSTM architecture, which is a kind of RNN architecture and tailored for such data handling. Furthermore, the LSTM architecture overcomes error signal problems such as blowing up or vanishing gradients, which are commonly observable in RNN solutions.

The number of used LSTM cells in a layer is known as a timestamp or look-back parameter. The look-back parameter refers to the number of time sequences data that are used in the LSTM layer. In our case, daily aggregated data is manipulated to be used as time-series data. After the LSTM layer, there are fully connected layers to improve learning capacity.

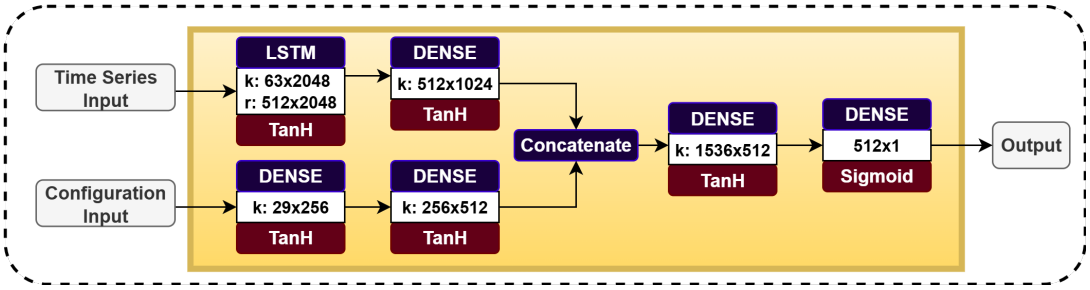
2.3.5 LSTM+: Long Short-Term Memory with Configuration Parameters Attached

The LSTM layer is a neural network layer that has been designed for processing time-dependent data. Each LSTM cell has three major units: the input, the output, and the forget gate. Time-dependent input is jointly processed by a forwarding mechanism between cells. The forget gate is responsible for the forwarding mechanism. The output of a cell depends on the forget gate of the previous cell and the output of the previous cell with the input for that cell. Thanks to the forget gates, the results are produced by preserving the time or sequence dependency. This mechanism works well for time-dependent data but when the dataset contains both time-dependent and time-independent data, a more sophisticated architecture should be proposed.

The dataset we use includes data that change daily (weather, network KPIs) and data that are time-independent, such as the configuration of the base stations/links used. This leads us to design a neural network architecture that has the capability of processing daily data and configuration data at the same time. For that reason, we propose *LSTM+* architecture given in Fig. 2.2b. LSTM+ model processes time-



(a) The LSTM model architecture.



(b) The proposed LSTM+ architecture.

Figure 2.2: The architecture of the LSTM model and the proposed model (LSTM+) are compared. The LSTM+ model has the capability of processing time-independent (configuration) inputs and time-dependent (daily) inputs. To process configuration inputs, the new input branch is added and then the two input branches' results are concatenated. k refers to the kernel size, and r refers to the recurrent kernel size for the LSTM layer.

series input and configuration inputs in two different branches, and then it concatenates both results. The first branch is designed for processing time-dependent data, so we used an LSTM layer for that input branch. The LSTM layer results are processed with a fully connected layer to increase the vector dimension. Since this fully connected layer makes a cartesian product, the decomposition of the data relationship is also increased. The motivation of the second branch is processing time-independent data. Therefore, that branch consists of fully connected layers to process one-hot-vector inputs that describe the configuration. Both branch results are concatenated at the *concatenation* layer. The concatenated results are processed at fully connected layers and the outputs are produced. The difference between the LSTM model and the proposed model (LSTM+) is presented in Fig. 2.2. In this figure, k refers to the kernel size of the layer, and r refers to the recurrent kernel size for an LSTM layer. These are the weight sizes for each layer type.

2.4 Performance Evaluation

The purpose of this study is to predict whether an RLF occurs in given links by using N-day previous data. Since the dataset has a highly imbalanced class distribution, we take an under-sample of the majority class corresponding to no RLF. The under-sampling method is a widely used methodology that helps improve models' learning ability for such imbalanced datasets [31].

Johnson et al. (2019) [32], suggest that the number of available samples for the class is more interesting than the percentage. Therefore, we try to keep as many samples as possible from the majority class while avoiding compromising the performance of the classifier. In our case, the dataset has about 1 million samples, and the class ratio is 1:3300; hence we need to re-balance that. Since we have about 700 RLF samples, we keep all of them, and we reduce the non-RLF sample size to about 14000. Therefore, the sampled subset ratio is 1:20. Thanks to the under-sampling, while the model becomes trainable, the model is prevented from memorizing by making a random selection.

We examine models' success on that sampled data that has about 15000 samples. The dataset is cleaned and prepared for the training phase. To predict RLF, we used LR, SVM, LSTM, and our proposed model (LSTM+) that process has the capability of processing time-series and configuration parameters together. Each model is fed with 10-day data that contains radio link KPI measurements and weather stations observations.

2.4.1 Experimental Setup

In this section, the model details and tuning parameters are given. We used numerical KPIs and weather realization data that are introduced in Section 2.2. For each radio link, these values are concatenated to form the daily data. LR, SVM, and LSTM models use these daily data, whereas the LSTM+ model uses both daily data and configuration parameters. Configuration parameters consist of frequency bands, card types, tips, adaptive modulations, and modulations. These configuration parameters are firstly converted to one-hot-vector, then the LSTM+ model is fed with

daily data, and the one-hot-vector that is formed by using configuration data. The inputs of each model are illustrated in Fig. 2.3. We use 10-day data as a batch when predicting the following day's RLF occurrence, as shown in the figure.

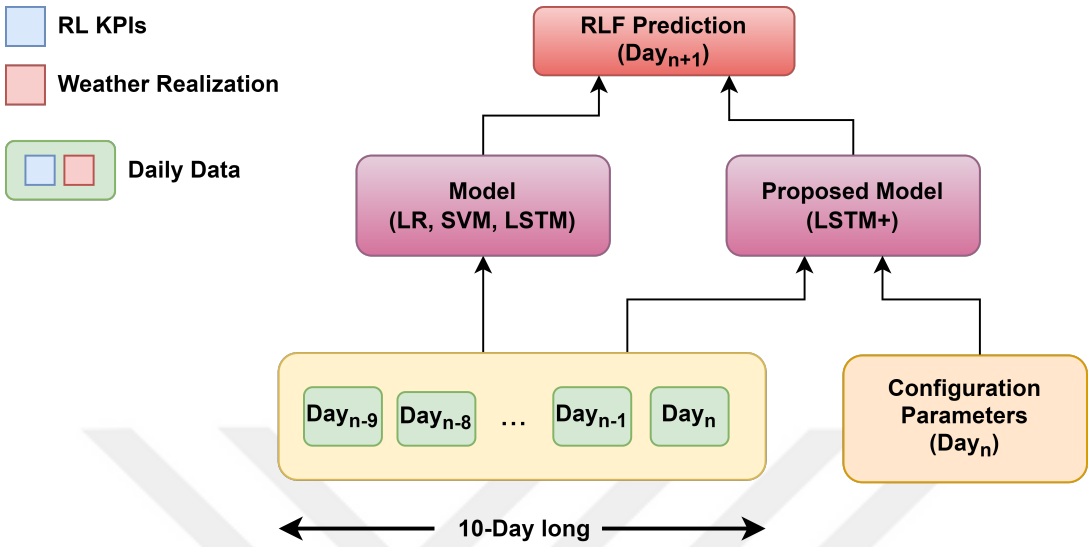


Figure 2.3: Illustration of the inputs and model prediction. The daily data refer to the concatenation of numerical radio link KPIs and weather realizations. Configuration parameters refer to frequency band, card type, etc. Historical 10-day data are used to predict the following day.

Each model is trained with the same dataset except the proposed model (LSTM+). For the logistic regression model, the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm is used as a solver with ℓ_2-regularization. The tolerance of stopping criteria is 10^{-4}, and the number of maximum iterations is limited to 10^3. SVM is used with radial basis function (RBF) kernel. It is penalized with ℓ_2-regularization to avoid overfitting. LSTM model is used with four layers. After the input layer, the first layer consists of 10 LSTM cells and each cell has 512 weight parameters. The dropout ratio is 0.1 for this layer to prevent overfitting. The following layers are fully connected layers with weight sizes 1024, 512, and 1, respectively. The first two fully-connected layers' activation function is $tanh$, and the last layer's (output layer) activation function is $sigmoid$. While training the LSTM model, the adaptive moment estimation (ADAM) optimizer with a binary cross-entropy loss function is used. The architecture of the used LSTM model is presented in Fig. 2.2a. The LSTM+ model has two input layers. The first one is connected with the LSTM layer that has the same configuration as the default LSTM

layer. That layer is fed with the daily data. On the other hand, the configuration parameters are directly connected with a fully connected layer that has 256 neurons, and the following layer has 512 neurons. Both layers use $tanh$ as activation function. Initially, daily data and configuration layers are processed in different branches, and then, resulting vectors are concatenated at the concatenation layer. The following layers are fully connected layers with weight sizes 512 and 1, respectively. Except for the output layer, all layers use $tanh$ as activation function. The output layer uses $sigmoid$ as activation function. The overall architecture is presented in Fig. 2.2b.

2.4.2 Evaluation Metrics

For the highly imbalanced datasets, standard evaluation criteria such as classification accuracy are misleading since they are insensitive to skewed domains [33]. Therefore, we used recall, precision and F1-score (harmonic mean of these two metrics), as well as receiver operator characteristics (ROC) and area under curve (AUC) metrics as alternatives to accuracy.

Recall, also known as true-positive rate (TPR), is the ratio of correctly classified positives to all positives. It is formulated as:

$$\text{Recall} = \text{TPR} = \frac{\text{tp}}{\text{tp} + \text{fn}}, \tag{2.3}$$

where *tp* and *fn* refer to true-positive and false-negative counts respectively.

False-positive rate (FPR) is the ratio of misclassified positives to all misclassified ones. The equation of FPR is:

$$\text{FPR} = \frac{\text{fp}}{\text{fp} + \text{tn}}, \tag{2.4}$$

where *tn* stands for true-negative count and *fp* stands for false-positive count.

Precision is the ratio of correctly classified positives to all positive predictions. The formulation of precision is that:

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}. \tag{2.5}$$

The recall helps to observe what percentage of RLF we can detect and precision helps to understand how much we can trust our RLF decisions. Moreover, we

mainly target to maximize the F-score which can be considered as a joint objective that considers precision and recall together. F-score gives more accurate results especially when the problem is imbalanced. F-score is formulated as follows:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{2.6}$$

We also use the ROC graph to visualize the performance. In a ROC curve, perfect classification is located at point $(0,1)$. ROC curve helps us to understand the tradeoff between the true-positive rate and the false-positive rate. The AUC of ROC curve is the probability score that a randomly sampled positive example has a higher probability than a negative one [34]. The area under the ROC curve is in the interval $[0,1]$. An untrained random classifier yields an AUC of 0.5. While we evaluate results, we use 5-fold cross-validation to get more robust and reliable results.

2.4.3 Results

The performance evaluation summary for the models is given in Table 2.4. When we examine the precision scores, we observe that the SVM model gives the best result (0.774). SVM model is followed by LSTM+ (0.652), LSTM (0.587), and LR model (0.344) performs worst. Although SVM model is strong in terms of precision, it gives poor results (0.322) in terms of recall when compared to LSTM based models. This means that it misses many more RLF instances. LSTM+ model's recall (0.626) is higher one from LSTM model (0.619). The precision score of the LR model (0.063) is significantly lower than the other two models. LSTM+ model's precision and recall scores are close to each other and they are high, it has the

Table 2.4: Comparison of model performances.

Model	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**AUC**
LR	0.344	0.063	0.105	**0.995**	0.818
SVM	**0.774**	0.322	0.455	0.963	0.890
LSTM	0.587	0.619	0.601	0.962	0.897
LSTM+	0.652	**0.626**	**0.638**	0.967	**0.931**

highest F1-score (0.626). Then the LSTM model has the second-highest F1-score (0.601). Although the precision score of the SVM model is high, it is ranked third with the F1-score (0.455) because the recall score is substantially lower than the LSTM based models. Since both the precision and recall of the LR model are low, it ranks the last in terms of F1-score (0.105).

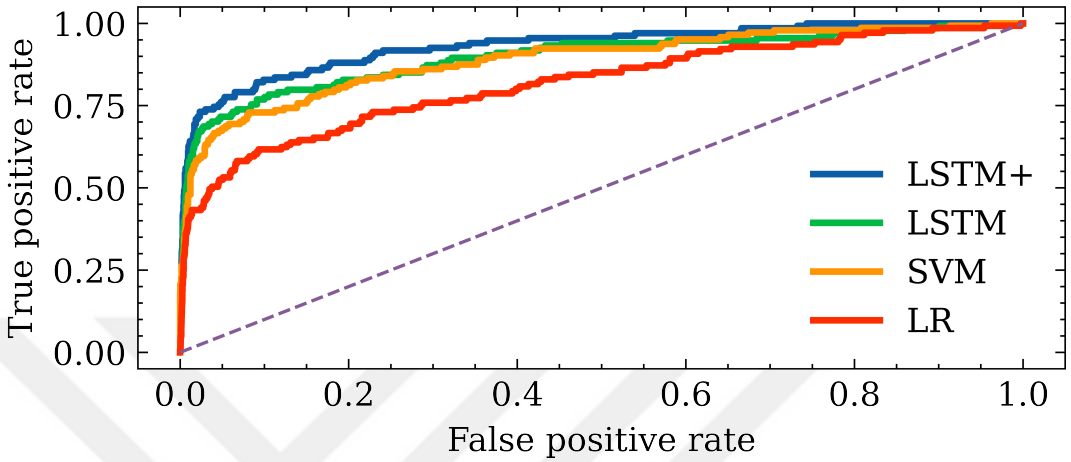


Figure 2.4: ROC comparison of each model.

The ROC curve analysis for each model is presented in Table 2.4 and Fig. 2.4. We observe that the LSTM+, LSTM, and SVM perform similarly to each other. When the false positive rates are low, these models increase their true positive rates while maintaining false-positive rates. After a certain point, TPR and FPR start to rise together for these models. Even if these three models perform very similar to each other, we can observe that LSTM+ model is better than the others. LR model performs poorly. The AUC score is important when performing ROC curve analysis. The ranking of the AUC scores of each model is as follows: LSTM+ (0.931), LSTM (0.897), SVM (0.890), and LR (0.818). LSTM+ model's score is higher than the others.

In binary classification, while labelling according to models' probabilistic results, the following formulation is applied: $f(x) > c$, where x refers to input, $f(x)$ stands for model's probabilistic result within the interval [0, 1], and c refers to the threshold. The previously presented results are based on $c = 0.5$. In Fig. 2.5, the performances of the models at different threshold levels can be compared. This analysis helps to understand models' possible best results with different thresholds. With

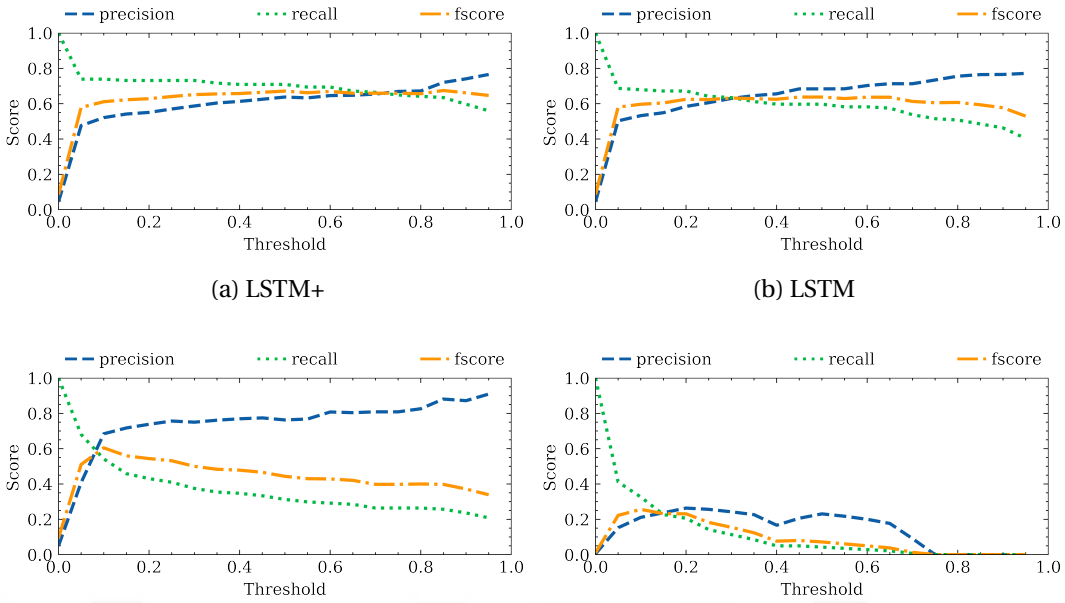
the change in the threshold value, the trade-off between the precision and recall and the F1-score is shown in Fig. 2.5. When the threshold is low, recall is high and precision is low. Because the predictive model labels all instances as positive. On the other hand, when the threshold value is high, the recall is low, whereas the precision is high. Because the model will only positively label samples that it is confident with. However, if the model with a very high threshold value cannot label any results as positive, the precision value may approach to zero as seen in the logistic regression case (Fig. 2.5d).

LSTM+ (Fig. 2.5a) and LSTM (Fig. 2.5b) models perform similar pattern in precision-recall trade-off. However, LSTM+ has better performance. Both models start with high recall and their recall decrease as precision becomes higher. The well-trained and usable models should consider precision and recall together. To understand how a model considers both variables together, we can analyze F1-score. The balance points can be inferred from the F1-score, where LSTM+ model's highest F1-score is 0.680, and for LSTM it is 0.642. SVM model's highest F1-score is 0.607, and LR model's highest F1-score is 0.271. At the highest F1-score point, LSTM+ model's precision and recall scores are 0.743 and 0.627, whereas LSTM model's precision and recall are 0.716 and 0.582, respectively.

2.4.4 Discussion

In this study, we present various model performances on predicting RLF by using historical weather and radio link KPIs and configuration parameters. The impact of RLF is high on a network as it causes service interruption. Therefore, we need to maximize our true-positive throughput or recall to prevent failure before it occurs. When the model predicts an RLF, configurations are updated if needed, and on-site personnel is directed to fix potential problems to prevent failure. It is essential to keep the maintenance cost or operating expenses and workload low by keeping the false positives low.

There is a trade-off between the cost of missing failures and the cost of maintenance. While the LSTM+ model has a high recall value and reduces the cost of missing failures, SVM responds with high precision and reduces maintenance costs.



(a) LSTM+ (b) LSTM

(c) SVM (d) LR

Figure 2.5: The model performances with different threshold values.

Service interruptions from such RLF can have major ramifications such as regulatory fines and reduced customer experience. Therefore, the cost of missing link failures needs to be optimized while keeping the false positive rate low. Among these four classifiers, the LSTM+ model provides the best trade-off in that sense.

The operator may choose the different thresholds to predict link failures. In general, a high threshold increases precision (Fig. 2.5); therefore, operators may increase the threshold to reduce operation costs. When the operator has a limited operation team, they may utilize their service schedule with a high precision score.

The model training time and response time may also be important for operators. LSTM+, LSTM, SVM, and LR are eager learning models; therefore, one-time training is enough to learn target function. After one-time learning, models start to respond to given inputs. In machine learning models, there is a general issue that the trend of data may change over time, which may cause the model to be outdated. The suggested approach is retraining or iterative training for new incoming inputs to catch up with new trends. That model renewal operation can be managed while using the existing model.

2.5 Conclusions and Future Work

With the increasing use of higher frequencies for backhaul data transmission, radio link failure problems may become a bigger problem as they are more sensitive to weather conditions. Proposed RLF detection models can be integrated into the network monitoring systems of mobile operators to reduce operational costs while keeping additional maintenance costs from false negatives at bay.

In this chapter, four different machine learning models are used to predict the radio link failures on data obtained from an operator live network. Each model is fed with daily aggregated radio link KPIs and weather data. The performances of four different models are compared. According to our experimental evaluation, the LSTM+ model that we propose in this study gives the best performance. The LSTM+ model uses configuration parameters, weather, link KPIs together.

There are limitations to the proposed approach. Neural networks are black-box models that capture the complex relationships in the data. Since we used LSTM and LSTM+ models that are fed with different input vectors, we know configuration parameters affect RLF, yet we do not know how and which ones. Therefore, newly emerging "Explainable AI" techniques would be beneficial for gaining further insights.

The dataset used in this study has the potential to trigger several new research directions. One of them could be using auto-regressive models, e.g., Autoregressive Integrated Moving Average model to model the time-variant seasonal effects. Moreover, the dataset we use and share publicly is collected from the live network of an actual telecom operator, and it can be used for other purposes as well. They may conduct several statistical analyses such as card type robustness, frequency robustness, etc. By using generative adversarial networks (GAN), they may try to find better configurations for the bad weather conditions. Researchers may use this data to prepare their simulations more accurately. To conclude, this data gives the opportunity for researchers to use live network data to define new problems or improve their works.

CHAPTER 3

PREDICTIVE MODELING OF HUMAN MOBILITY FROM MOBILE NETWORK DATA: INSIGHTS FOR DISASTER RESPONSE AND URBAN PLANNING

On February 6, 2023, a catastrophic earthquake struck southeastern Türkiye, devastating 11 major cities and impacting nearly one-sixth of the nation's population. In the affected region, home to approximately 12 million mobile subscribers, the telecommunications infrastructure suffered severe damage estimated at over 62 million euros [35, 36]. Network traffic data from Turkcell, the country's leading mobile network operator, revealed dramatic spikes in communication demand, with data and voice traffic surging by 260% and an astonishing 9150% respectively within just the first half hour following the quake. However, widespread base station collapses and power outages severely disrupted mobile connectivity, leaving significant portions of the population in cities such as Adiyaman, Hatay, and Kahramanmaraş without reliable service when it was most critically needed [37].

The operational base stations that survived the initial shocks could function only for limited periods, often less than four hours, due to depleted internal batteries and logistical barriers preventing the delivery of fuel or generator replacements. Innovative solutions, like mounting a civilian base station onto a medium-altitude drone and backhauling via satellite links, were deployed for the first time in Türkiye to restore communications [38]. Yet, even these novel approaches encountered challenges such as radio interference with terrestrial infrastructure, underscoring the complexity of maintaining resilient connectivity during large-scale disasters. Events like earthquakes, floods, and hurricanes starkly reveal that existing mobile networks were not originally designed to withstand extreme disaster scenarios, creating critical gaps in communication precisely when information sharing

becomes a matter of life and death.

Against this backdrop, understanding user mobility emerges as a fundamental pillar not only for designing disaster-resilient mobile networks but also for informing broader governance, policy-making, and legal frameworks aimed at protecting societies during crises. Modeling how populations move in the wake of disasters is essential for orchestrating effective emergency responses, allocating scarce resources, and planning recovery operations. Moreover, insights derived from mobility data can support the development of proactive policies and regulations that enhance urban resilience, ensure equitable aid distribution, and safeguard public safety. Motivated by these urgent challenges, this thesis explores advanced methodologies to transform mobile network data into actionable insights, ultimately aiming to strengthen both technological and institutional capacities for disaster preparedness and response.

With honor, the memory of those who lost their lives in the earthquake that struck southeastern Türkiye on 6 February 2023 is remembered.

Modern mobile networks serve not only as the backbone of communication but also as powerful sensing infrastructures that continuously generate vast volumes of data reflecting the dynamic behaviors of users across space and time [39, 40]. Each interaction between user equipment (UE) and the network produces detailed measurement reports containing information such as cell identities, signal quality metrics, handover events, and in many cases, approximate geolocation. These network-driven data sources offer several key advantages for mobility analysis: they cover a large fraction of the population, remain operational over extensive geographic regions, and capture user behavior at fine temporal resolutions.

The use of mobile network data to analyze and model user mobility represents a promising research direction at the intersection of telecommunications and data science. Unlike traditional methods, such as surveys or manual observations, mobile network data offer the potential to observe population movements without imposing additional burdens on individuals. Although current commercial deployments often provide mobility insights with delays due to privacy constraints, data aggregation processes, and operational limitations, future mobile network

architectures are being designed to facilitate the development of data-driven approaches. In particular, recent standards and architectural developments are paving the way for analytics systems that could provide valuable information for applications such as emergency response and urban management.

In particular, ongoing standardization initiatives within the 3GPP, notably those defined in TS 28.550 [41] for Management Data Analytics Services (MDAS) and TS 23.288 [42] for the Network Data Analytics Function (NWDAF), are establishing standardized interfaces and functional frameworks for the low-latency collection, processing, and dissemination of network intelligence. Such standards are designed to enable real-time or near-real-time analytical capabilities, facilitating timely insights crucial for applications including disaster response and urban mobility monitoring. Currently, the Open Radio Access Network (O-RAN) Alliance is advancing the development of open interfaces and intelligent radio access network (RAN) architectures, where the near-real-time Radio Intelligent Controller (RIC) serves as a key component for rapid data acquisition and analytics-driven decision-making, operating within timescales ranging from subseconds to a few seconds [43]. These architectural and standardization efforts are poised to transform mobile networks into ubiquitous sensing platforms, thereby strengthening the foundations for predictive modeling of human mobility and enhancing the ability to respond swiftly and effectively to both routine urban dynamics and extraordinary events such as natural disasters.

Such evolving architectures and standards promise to transform mobile networks into powerful sensing systems capable of producing timely spatial insights [44]. These advances open pathways for a variety of applications, from urban planning and transportation optimization to public safety and disaster response, where timely understanding of user mobility is crucial [45].

An important and increasingly urgent application area for mobility analysis is disaster management. Natural disasters such as earthquakes, floods, or pandemics can drastically disrupt normal movement patterns, as individuals seek safety, evacuate high-risk areas, or adapt to new restrictions. Understanding how populations move during and after such events is critical for planning emergency responses, al-

locating resources efficiently, and ensuring the resilience of urban infrastructures [46].

In particular in the context of earthquakes, which can cause severe physical damage and sudden large-scale displacements, the ability to model and predict user mobility based on mobile network data becomes invaluable. By analyzing changes in network activity, such as the volume of users connected to particular cells or the frequency of inter-cell transitions, it is possible to detect anomalies, trace evacuation flows, and identify safe or congested zones.

To address these challenges, this chapter brings together two complementary strands of work that together form a unified framework for understanding and predicting user mobility based on mobile network data. The first component focuses on the data collection, processing, and exploratory analysis of large-scale user measurement reports, transforming raw network events into structured representations such as heatmaps and grid-based analyses. These visualizations provide insights into both micro-level urban dynamics and macro-level inter-city movements, serving as an essential basis for understanding normal and disrupted mobility patterns.

Based on these insights, the second component of this research develops predictive models designed to forecast user mobility in various scenarios, including natural disasters. In particular, the chapter proposes a grid-based similarity modeling approach, which uses spatial features, neighboring grid characteristics, and machine learning techniques to predict user movement patterns. This modeling framework aims to translate complex spatio-temporal dynamics into actionable predictions that can support urban planning and emergency response.

This chapter focuses on exploiting mobile network data to develop methodologies for modeling user mobility at both micro- and macro-scales. At the micro-level, mobility patterns within urban areas are studied by aggregating user measurements into grid-based heat maps that capture localized variations in population density and movement. At the macro-level, transitions between cities or regions are analyzed to characterize large-scale evacuation or migration trends. Using machine learning and statistical techniques, these models aim to capture the spatio-

temporal dynamics of user behavior under both normal conditions and disaster scenarios.

The ultimate goal of this research is to establish a data-driven framework capable of:

- Translating raw mobile network measurements into actionable spatial insights through scalable data processing pipelines.
- Modeling urban mobility patterns with high granularity to support city-level analyses and planning.
- Analyzing inter-city mobility to identify broader displacement trends following disruptive events.
- Developing predictive models that can forecast future user mobility patterns under both normal and disaster conditions.
- Integrating mobility models into disaster management systems to enhance situational awareness and inform timely decision-making.

By bridging the fields of mobile network engineering and AI, this work contributes to a deeper understanding of how populations behave spatially and temporally, ultimately aiming to improve urban resilience and public safety in the face of disasters.

The rest of this chapter is organized as follows. Section 3.1 reviews the background and related work on mobility analysis, and mobile network data usage. Section 3.2 describes the data collection process and the processing pipeline used to transform raw mobile network measurements into structured datasets suitable for analysis. Section 3.3 presents exploratory analyses and visualizations that reveal insights into user mobility patterns across different regions and scales. Based on these insights, Section 3.4 details the predictive modeling framework developed to estimate user mobility under disaster conditions. Finally, Section 3.5 summarizes the key findings and outlines directions for future research.

3.1 Background and Related Work

In the domain of human mobility research, a substantial shift has occurred over the past two decades, driven largely by technological advancements and the growing availability of large-scale digital traces of population movements. Early studies in mobility modeling predominantly relied on simulation-based frameworks and theoretical constructs, which offered flexibility in designing hypothetical evacuation scenarios and required relatively modest data inputs. These models typically incorporated parameters such as pedestrian speed, spatial layout, and evacuation behavior under controlled assumptions. However, scholars have increasingly recognized that purely theoretical models can fall short of accurately capturing real-world human behavior, particularly during high-stress events such as natural disasters, where behavioral dynamics diverge significantly from normal conditions. Studies like those by Shiwakoti et al. (2008) have emphasized the necessity of complementing theoretical simulations with empirical datasets to validate model performance and enable meaningful comparative analyses [47].

With the advent of ubiquitous mobile communication, researchers have turned to mobile network data as a promising source for capturing fine-grained human mobility information. Mobile network signaling records, including handover events, cell-level associations, and measurement reports, provide an indirect yet ubiquitous sensing mechanism capable of covering entire populations with high spatial and temporal fidelity [48, 49]. Studies leveraging Call Detail Records (CDRs) and mobile phone data have shown significant potential in uncovering urban dynamics, commuting patterns, and responses to extraordinary events, despite inherent challenges in spatial precision [50, 51, 52]. While CDRs typically capture user activity only during communication events, more advanced data sources such as passive measurement reports offer continuous monitoring, enabling analyses at sub-hourly granularity [53].

Recent work has explored multiple analytical frameworks for modeling mobility using mobile network data. One line of research focuses on purely descriptive analyses, employing visualizations such as heatmaps and spatiotemporal density plots to identify hotspots, periodic patterns, and deviations [46, 54]. For example,

Yabe et al. [55] analyzed the changes in mobility patterns following the COVID-19 pandemic, highlighting the potential of mobile network data to capture collective behavioral adaptations. However, descriptive approaches, while valuable, are limited in their ability to predict future movement patterns, especially in unprecedented scenarios.

To address the limitations of generalizing mobility prediction models across diverse urban contexts, recent research has explored advanced data-driven approaches that utilize large-scale spatio-temporal data and sophisticated modeling techniques. One study proposes a deep residual neural network architecture that effectively models citywide crowd flows by capturing spatial correlations and temporal patterns such as closeness, periodicity, and long-term trends, while also incorporating external factors like weather, thereby achieving high accuracy within specific urban environments [56]. Meanwhile, another study demonstrates how mobile phone data can be transformed into detailed origin-destination matrices and integrated into transportation models, offering valuable insights into urban travel demand and traffic assignment processes [57]. Despite these significant advancements, both works underline that transferring such models across different cities remains challenging due to variations in urban morphology, socio-economic conditions, and behavioral patterns, underscoring the ongoing need for approaches that enhance model adaptability and generalization.

A particularly relevant trend is the integration of transfer learning techniques to mitigate the limitations posed by data scarcity in crisis situations. Studies have investigated transferring knowledge from cities with abundant data to similar but data-poor regions, leveraging urban structural similarities to inform predictions [58, 59]. These strategies are particularly pertinent for disaster scenarios, where immediate local data may be unavailable, but rapid situational awareness remains critical. Recent research has advanced this paradigm by introducing methods that identify structurally similar urban regions across cities and employ deep learning to adapt spatio-temporal patterns from data-rich environments to new contexts, achieving notable improvements in predictive accuracy even under limited data conditions. Such techniques not only facilitate rapid deployment of analytical tools during emergencies but also contribute to more resilient and adaptable

urban mobility models, ensuring that insights derived from one city can effectively inform planning and response efforts in others.

Despite these advances, significant gaps remain in the field. Many existing models assume stable mobility patterns and struggle to adapt to abrupt changes induced by crises. Furthermore, while structural urban features provide a powerful lens for understanding mobility, there remains a need for systematic frameworks capable of quantitatively linking urban morphology to dynamic human behavior across diverse contexts.

This chapter positions itself at the intersection of these research directions. It proposes a novel framework that combines grid-based spatial representations, machine learning-driven similarity modeling, and transfer learning to predict user mobility under normal and disaster conditions. Using neighbor-based properties and structural urban characteristics, the proposed approach aims to enhance the robustness and transferability of mobility predictions. This framework seeks not only to model intra-city dynamics at high spatial resolution but also to capture inter-city migration flows, ultimately contributing to improved urban resilience and informed decision-making in the face of emergencies.

3.2 Data Collection and Processing Pipeline

Modeling human mobility based on mobile network data requires a systematic and scalable approach for collecting, processing, and structuring vast volumes of user activity records. Mobile networks continuously produce detailed measurement reports that capture diverse aspects of user interactions, including unique device identifiers, timestamps, radio signal quality indicators, and approximate spatial references [60]. These reports form the primary data source for reconstructing both individual mobility trajectories and aggregate patterns of population movement across urban and regional scales. However, managing this data presents significant challenges due to its massive scale, rapid generation rate, and the necessity for precise spatial interpretation.

The architecture designed to address these challenges is depicted in Figure 3.1.

The pipeline begins with the interactions between mobile devices and the surrounding radio network infrastructure, which comprises base stations and other network elements. Each user event—such as initiating a call, maintaining a data session, or executing periodic network updates—triggers the creation of measurement reports that are transmitted to the core network for further processing.

The machine learning-assisted geolocator takes a critical step in transforming these network events into meaningful spatial information. Instead of relying solely on the identity of the serving cell - which offers only a coarse approximation of the position of the user - this geolocation module employs advanced machine learning algorithms to estimate precise geographic coordinates. It leverages signal strength measurements from the serving cell and multiple neighboring cells, alongside knowledge of the physical network layout, to model the relationship between observed radio conditions and geographic positions. This approach significantly improves spatial resolution, particularly in complex urban environments where overlapping cell coverage and multipath effects can complicate location estimation.

Once the geolocation step is complete, the enriched data, now comprising timestamped device identifiers and their estimated coordinates, is transmitted to a centralized data warehouse. This data warehouse is designed to manage the immense scale and high velocity of mobile network data, as typical operators may generate billions of records each day. To maintain operational efficiency, periodic data retention and removal policies are employed, ensuring that storage capacity remains available for continuous data input while retaining the most relevant datasets for analysis.

Complementing the storage infrastructure, the pipeline incorporates a data analysis module and a visualization module, both of which operate directly on the data stored within the warehouse. The data analysis module employs statistical methods and machine learning algorithms to detect patterns, quantify user mobility metrics, and model behavior under various conditions, including normal operations and disaster scenarios. The visualization module translates these analytical results into intuitive graphical representations—such as heatmaps, temporal trends, and spatial flow diagrams—enabling clear communication of insights to

stakeholders and supporting effective decision-making processes.

Altogether, this integrated pipeline converts raw measurement reports into structured, high-resolution representations of human mobility, forming the essential foundation for the exploratory analyses and predictive modeling approaches presented in the subsequent sections of this chapter.

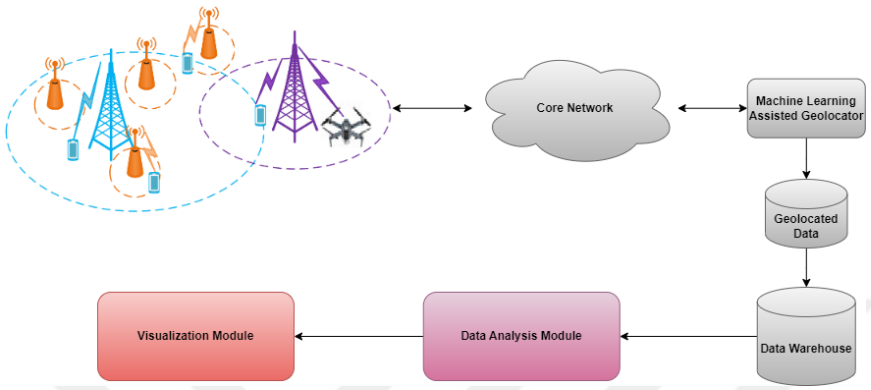


Figure 3.1: Overview of the data gathering and processing pipeline, depicting data flow from mobile network interactions to geolocation, data storage, and analysis modules.

Modern mobile networks generate enormous volumes of data as a byproduct of user activity. In typical operational environments, networks serving approximately 30 to 50 million subscribers can produce between 8 and 10 billion user measurement reports each day. Managing and extracting meaningful insights from such massive datasets poses significant challenges in terms of storage, computational efficiency, and analytical scalability. To address these challenges, this thesis proposes a dedicated analysis pipeline designed around the principles of big data processing, which relies on distributed computing paradigms such as the MapReduce framework [61] to enable the systematic transformation of raw network data into actionable mobility insights.

The architecture of the proposed analysis pipeline is illustrated in Figure 3.2. The primary problem addressed by this pipeline is to determine the location of a user at time $T + n$ based on their last observed position at time T. By linking user records across different days, the pipeline reconstructs individual user trajectories, allowing the identification of where each user has moved over the specified interval. From this user-level movement information, it becomes possible to compute ag-

gregated flows between different locations, enabling quantitative analysis of how populations relocate over time. The pipeline operates as follows:

- **Map Phase:** In this stage, daily data snapshots for two distinct days—denoted as T and $T + n$—are extracted from the data warehouse. For each user identifier (represented by IMSI), the pipeline retrieves the last observed location on day T and on day $T + n$. This step transforms raw measurement records into a structured representation, associating each user with a pair of locations: a source city and a target city.
- **Partition Phase:** Following the mapping of user movements, the data is partitioned based on the combination of source and target cities. This partitioning enables subsequent aggregation to be performed independently for each unique movement corridor, thereby improving computational efficiency and facilitating parallel processing.
- **Reduce Phase:** For each partition, the pipeline aggregates the number of users who have moved between the corresponding source and target cities. This counting operation yields a matrix of user flows, quantifying the volume of inter-city mobility across the network. The results provide actionable information on regional migration trends, evacuation behaviors in the context of disasters, or routine population flows under normal conditions.

By employing the proposed approach, it becomes feasible to process large-scale datasets and extract valuable insights from user location information. This methodology enables the transformation of raw geospatial data into structured representations that capture meaningful patterns of human movement. In particular, it proves highly effective for modeling user transitions between cities, allowing for the identification of macro-level mobility flows and regional connectivity dynamics. Such analyses can reveal critical insights into population distribution, commuting behaviors, and changes in movement patterns under both normal and disruptive conditions, thereby providing a robust analytical foundation for various applications, including urban planning, transportation management, and emergency response strategies.

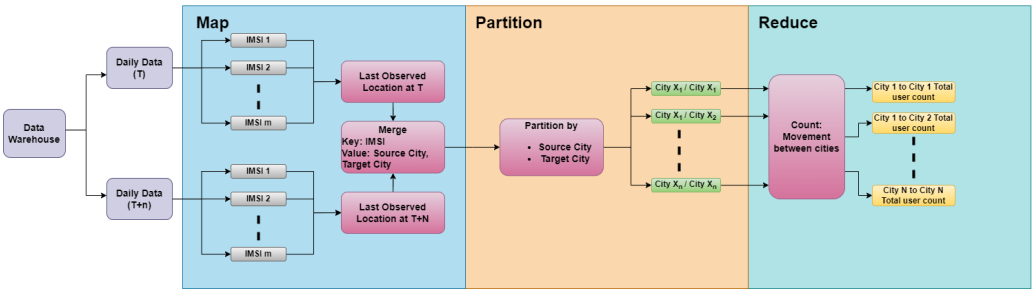


Figure 3.2: Proposed data analysis pipeline employing a MapReduce-based architecture. The pipeline processes daily snapshots of geolocated network data to detect inter-city user movements, summarizing the process from raw data extraction through mapping, partitioning, and reducing user movements into aggregate migration counts.

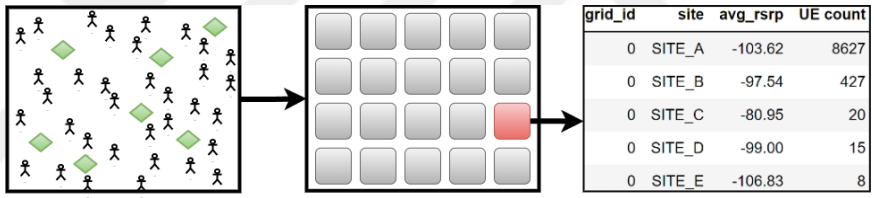
In general, the proposed data analysis pipeline bridges the gap between raw, voluminous network data and high-level mobility insights, supporting urban planning initiatives and disaster response strategies through scalable and systematic data processing.

Beyond the analysis of macro-level movement patterns, the proposed pipeline also enables micro-level investigations that provide a fine-grained understanding of user mobility dynamics within urban environments. A grid-based approach is employed to transform scattered, pointwise location measurements into structured spatial representations suitable for detailed spatial analysis. This methodology leverages the H3 indexing system, which partitions geographic space into hexagonal grid cells of various resolutions, allowing for flexible spatial granularity depending on the analytical requirements [62].

By assigning individual user measurement records to grid cells, the process effectively reduces the complexity and volume of the raw data while preserving essential spatial detail. This enables the computation of meaningful spatial metrics, such as hourly user densities, average signal strength indicators, and localized user distributions across urban landscapes. Moreover, a key enhancement of this approach lies in the integration of connected cell information. For each grid cell, the pipeline associates dominant serving and neighboring cell sites based on observed radio signal measurements, providing insights into how specific parts of the network infrastructure serve different geographic regions. This association facilitates

an enriched spatial analysis, enabling the identification of high-density zones, understanding service areas of network elements, and tracing user movements across grid cells over time.

This grid-based transformation process is illustrated in Figure 3.3. As depicted, individual user events—initially represented as scattered geospatial points—are mapped into a structured grid system, where each cell aggregates all user events occurring within its boundaries. The grid representation simplifies the management of vast volumes of pointwise data and supports the calculation of spatial statistics. The right side of the figure demonstrates how aggregated data for each grid cell is summarized in tabular form, including attributes such as grid identifiers, associated cell sites, average radio signal quality indicators (e.g., reference signal received power [RSRP]), and user counts. This structured format provides the analytical basis for micro-level insights into user density, network performance, and user-network interactions within specific urban areas.



grid_id	site	avg_rsrp	UE count
0	SITE_A	-103.62	8627
0	SITE_B	-97.54	427
0	SITE_C	-80.95	20
0	SITE_D	-99.00	15
0	SITE_E	-106.83	8

Figure 3.3: Process of aggregating individual geolocated user events into grid-based spatial representations and summarizing measurement statistics per grid cell.

The outputs of this micro-level analysis serve as valuable inputs for visualization techniques such as heatmaps, which vividly depict spatial variations in user density and network coverage across the city. These insights are crucial for diverse applications, including urban planning, transportation optimization, network management, and disaster response strategies. By bridging the gap between raw geolocation data and interpretable spatial patterns, the grid-based approach complements macro-level analyses and contributes to a comprehensive understanding of human mobility within modern urban environments.

Taken together, the methods and pipeline presented in this chapter constitute a comprehensive framework for modeling human mobility using mobile network

data. The primary focus has been on developing macro-level analysis techniques capable of detecting inter-city user transitions and large-scale movement flows, as well as micro-level analysis methods that enable grid-based investigations of user density and localized mobility within urban environments. Central to both approaches are robust data preparation processes that transform vast volumes of raw measurement reports into structured, interpretable datasets suitable for further analysis and visualization. These processes facilitate the generation of heatmaps, spatial flow diagrams, and statistical models that support various analytical tasks. Importantly, the pipeline is designed with the flexibility to operate not only under normal conditions but also to analyze user behavior during and after disruptive events such as natural disasters. Using post-disaster user reports, the framework developed in this chapter provides a powerful tool for assessing situations and supporting decisions, ultimately enhancing capabilities in urban planning, transportation management, and disaster response strategies.

The key contributions of this work can be summarized as follows:

- **Big data processing framework:** Design and implementation of a scalable data pipeline capable of handling massive volumes of data from mobile networks employing distributed computing principles, such as the MapReduce paradigm, to ensure efficient processing and analysis in large datasets.

- **Macro-level mobility analysis:** The development of analytical techniques to detect and quantify inter-city user transitions, enabling the examination of large-scale mobility flows, regional connectivity, and migration trends, as well as providing insights into evacuation behaviors during disruptive events.

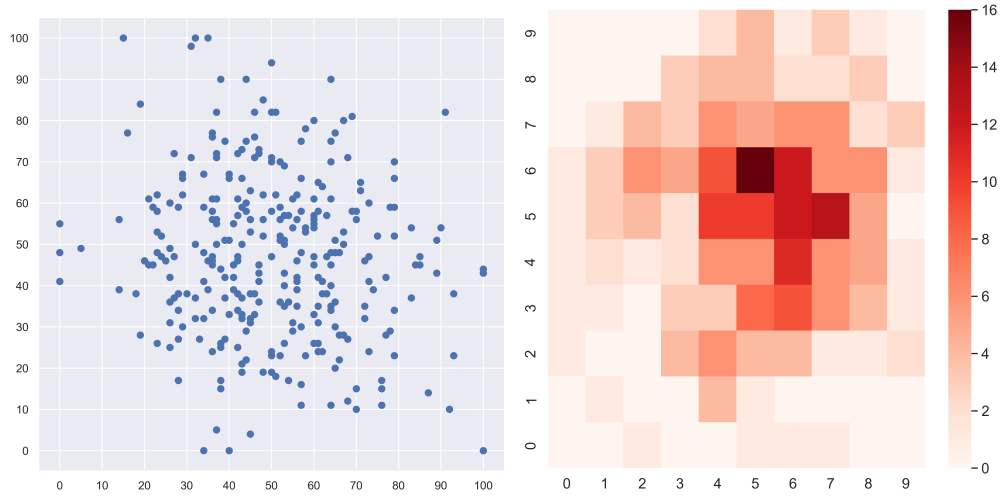
- **Micro-level mobility analysis:** The establishment of a grid-based approach using the H3 indexing system to transform scattered user location measurements into structured spatial representations, facilitating detailed analyses of intra-city mobility, user density patterns, and localized network interactions. This approach also integrates connected cell information, linking mobility data to network infrastructure and enabling enriched spatial insights.

3.3 Exploratory Analysis and Visualization of User Mobility

Understanding human mobility requires not only the processing of vast volumes of network data but also the ability to transform this data into interpretable insights through visualization techniques. This section focuses on presenting the exploratory analyses that translate raw measurement reports into meaningful spatial and temporal mobility patterns. Two distinct analytical perspectives are employed: micro-level analysis, which investigates intra-city dynamics using grid-based methodologies to capture localized fluctuations in user density and movement; and macro-level analysis, which examines inter-city mobility flows to reveal broader regional trends and transitions. Together, these analyses provide a comprehensive view of mobility behavior, enabling the identification of both routine patterns and anomalies that may emerge under exceptional circumstances such as natural disasters. The following subsections detail the findings and visualizations derived from each analytical level.

3.3.1 Micro Level Mobility Analysis: Grid Based User Mobility

A fundamental aspect of understanding urban mobility dynamics involves examining how user presence and movement vary across fine-grained spatial and temporal scales. In this micro-level analysis, grid-based heatmaps are employed to visualize and interpret localized fluctuations in user density within urban environments. Raw user measurement reports—containing anonymized device identifiers, timestamps, and geospatial coordinates—are initially recorded as scattered data points, which complicates direct analysis and interpretation. To address this challenge, the data is transformed using the H3 spatial indexing system, which partitions geographic space into hexagonal grid cells of uniform size. Each measurement is assigned to its corresponding grid cell, thereby aggregating the scattered point data into structured spatial units. This process reduces data complexity and enables the computation of meaningful spatial statistics, such as hourly user densities and average signal quality metrics. The transformation from scattered measurement reports to a grid-based heatmap is illustrated in Figure 3.4.



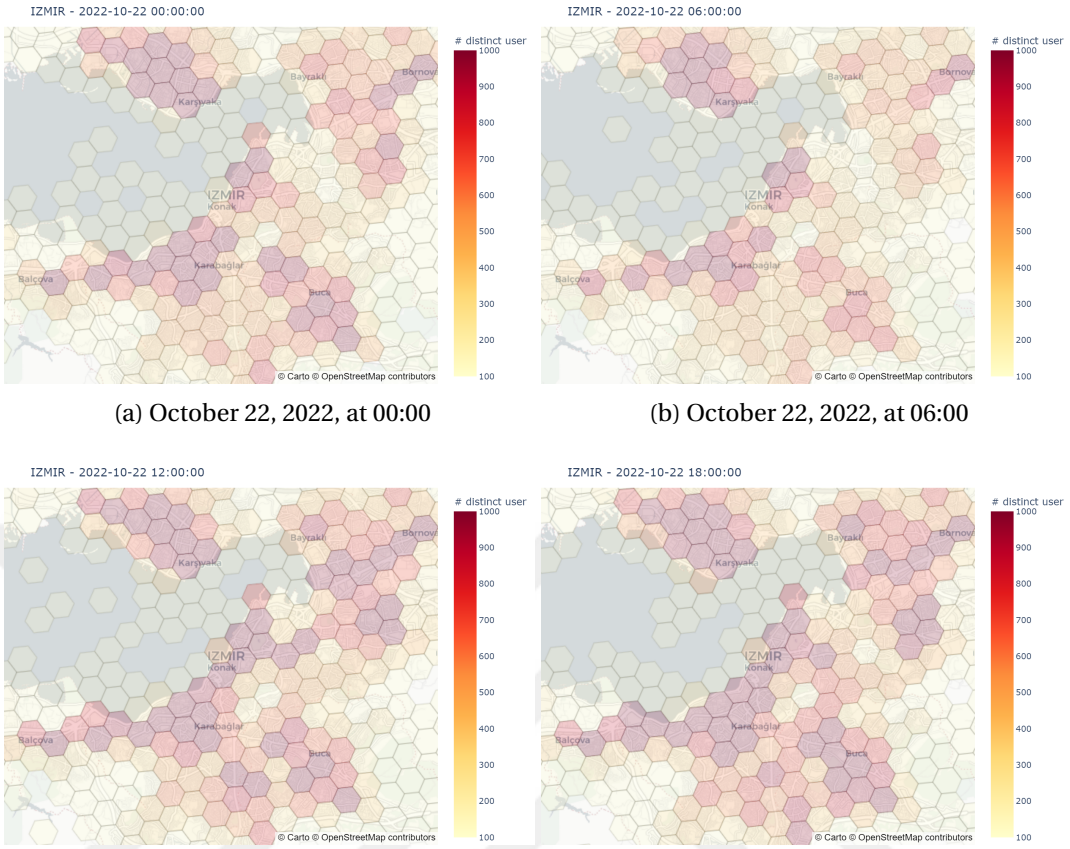
(a) UE measurement reports distribution. (b) Heatmap of the number of unique UE.

Figure 3.4: Illustration of UE reports to grid-based heatmap conversion.

Heatmaps generated from these aggregated datasets serve as a powerful tool for visualizing variations in user activity across different urban areas. By mapping user counts or signal metrics to color gradients, heatmaps make it possible to identify zones of high and low activity, urban hotspots, and regions of sparse usage. This visualization technique is particularly valuable not only under normal conditions but also for monitoring changes during extraordinary events such as natural disasters. Comparative analysis of heatmaps captured at different times can reveal temporal anomalies, shifts in population density, and patterns of evacuation or congregation.

To demonstrate the practical application of this approach, the İzmir metropolitan region was chosen as a case study. As a major urban center in western Türkiye, İzmir exhibits complex mobility patterns shaped by its dense urban core and diverse socio-economic activities. Figure 3.5 displays heatmaps for İzmir on October 22, 2022, captured at four different times of the day. The visualizations indicate consistently higher user densities in central districts, with noticeable variations during early morning hours. These fluctuations correspond to typical urban phenomena such as commuting, shifts between residential and business areas, and evening leisure activities. Such analyses illustrate the value of heatmaps in differentiating between routine urban dynamics and abnormal deviations that may

signal emergent events.



(a) October 22, 2022, at 00:00

(b) October 22, 2022, at 06:00

(c) October 22, 2022, at 12:00

(d) October 22, 2022, at 18:00

Figure 3.5: Comparison of heatmaps for İzmir on October 22, 2022, at 00:00, 06:00, 12:00, and 18:00.

The insights gained from these heatmap analyses extend beyond simple visual depiction. They enable the identification of persistently high-activity areas, detection of temporary surges or reductions in population density, and the monitoring of evolving urban mobility trends over time. In disaster management contexts, such analyses are crucial for rapid situational awareness and decision-making. By comparing pre-disaster and post-disaster heatmaps, it becomes possible to pinpoint zones experiencing significant shifts in human activity, revealing potential evacuation routes, temporary gathering points, and areas requiring immediate response. Integrating grid-based heatmaps into the analysis pipeline thus plays a vital role in transforming raw measurement data into actionable knowledge for both routine urban operations and emergency scenarios.

In summary, micro-level mobility analysis using grid-based heatmaps offers advantages for interpreting urban dynamics. By visualizing user counts distributed across spatial grids, heatmaps provide a detailed understanding of how people move and concentrate within a city throughout the day. Such insights make it possible to identify popular locations and periods of peak activity, informing decisions on urban planning and infrastructure development. Additionally, understanding localized mobility patterns enables more targeted investments by aligning services and resources with actual user needs and demands. From a network management perspective, these analyses also contribute to enhancing energy efficiency, as operators can optimize base station utilization by considering variations in user density and mobility over time. Overall, micro-level analysis serves as a powerful tool for bridging raw network data with actionable strategies in both commercial and public-sector applications.

3.3.2 Macro Level Mobility Analysis: Inter-city User Mobility

While micro-level analyses provide vital insights into localized dynamics within urban areas, understanding human mobility at a national or regional scale necessitates a broader perspective. Macro-level mobility analysis focuses on capturing inter-city user movements, revealing how populations flow between major urban centers and regions. Such insights are crucial for numerous applications, including transportation planning, infrastructure development, epidemiological modeling, and, critically, the management of natural disasters and emergency response operations.

The methodological foundation for macro-level analysis in this work is based on the MapReduce-inspired data processing architecture illustrated in Figure 3.2. This scalable pipeline enables the efficient processing of massive mobile network datasets by distributing computation across multiple nodes. Each day, a snapshot of user locations—determined through geolocation processes—is recorded in the data warehouse. The pipeline systematically extracts each user's last observed location for two different days, forming the basis for tracking user transitions between cities.

This process results in a transactional dataset in which each record captures a

unique user's movement between two cities over a specified time interval. For instance, as shown in Table 3.1, a user identified by International Mobile Subscriber Identity (IMSI) ID 1 was last recorded in Istanbul on 1 January 2023 and subsequently in Ankara on 8 January 2023. This granular dataset captures millions of such transitions across the entire mobile network subscriber base, offering an unprecedented view of human mobility dynamics.

Table 3.1: IMSI based transaction data. City 1 is the last observed location on Date 1 and City 2 is the last observed city on Date 2.

IMSI ID	Date 1	City 1	Date 2	City 2
1	01.01.2023	Istanbul	08.01.2023	Ankara
2	01.01.2023	Izmir	08.01.2023	Antalya
3	01.01.2023	Ankara	08.01.2023	Bursa
4	01.01.2023	Antalya	08.01.2023	Istanbul
5	01.01.2023	Bursa	08.01.2023	Izmir

From this user-level data, aggregated transition tables are constructed to quantify flows between city pairs. Table 3.2 exemplifies this aggregation, presenting counts of users moving from City 1 to City 2 between two dates. These aggregated figures form the core analytical dataset for macro-level mobility analysis, enabling the identification of dominant corridors of movement, migration trends, and intercity connectivity.

Table 3.2: Aggregated user movement data by city. The table shows the count of users moving from City 1 to City 2 between Date 1 and Date 2.

Date 1	City 1	Date 2	City 2	Count
01.01.2023	Istanbul	08.01.2023	Ankara	c_1
01.01.2023	Izmir	08.01.2023	Antalya	c_2
01.01.2023	Ankara	08.01.2023	Bursa	c_3
01.01.2023	Antalya	08.01.2023	Istanbul	c_4
01.01.2023	Bursa	08.01.2023	Izmir	c_5

Such aggregated transition tables serve as the foundation for powerful visualization techniques. One particularly insightful tool is the heatmap, which provides

an overview of the magnitude and directionality of inter-city flows over a given period. Figure 3.6 presents a heatmap visualizing user movements between cities across Türkiye between 20 January and 27 January 2023. The horizontal and vertical axes represent origin and destination cities, respectively, while color intensity reflects the proportion of users making each transition.

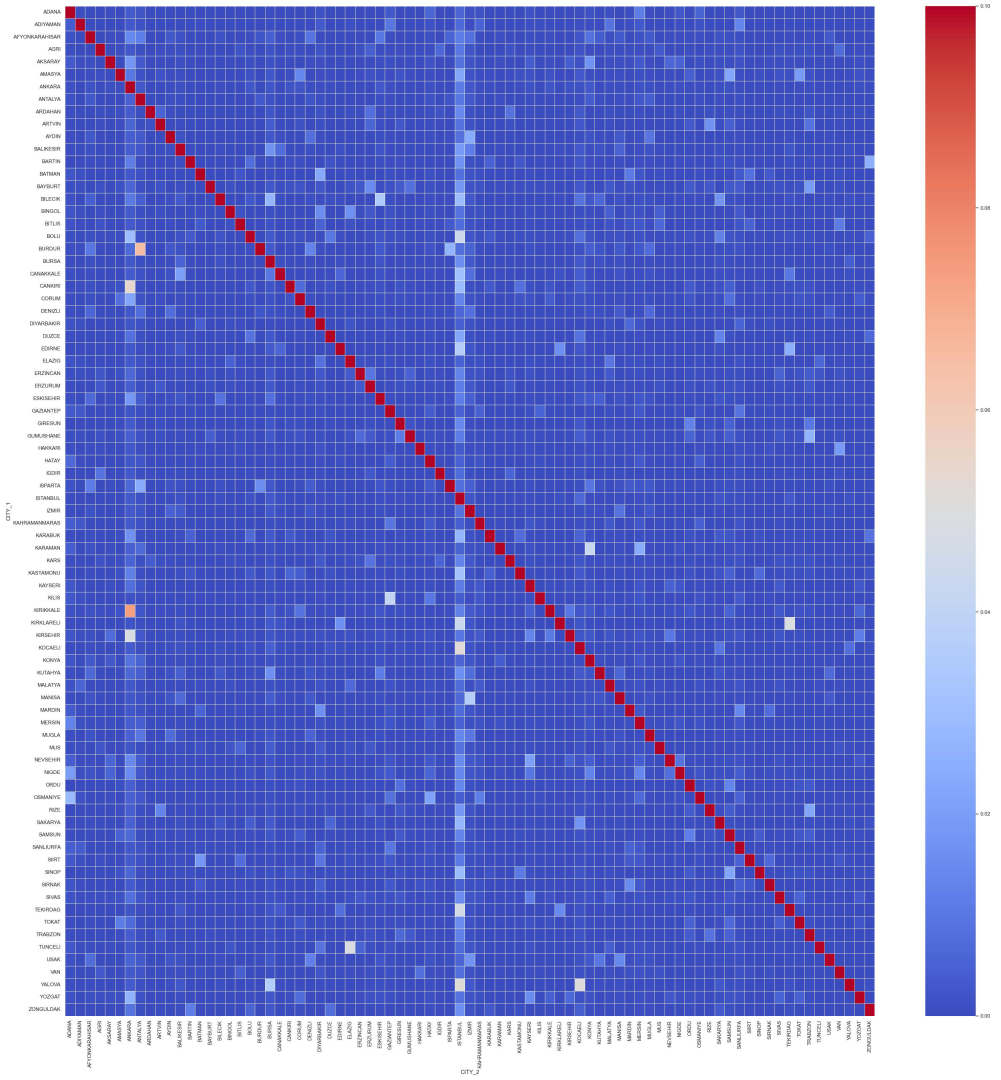
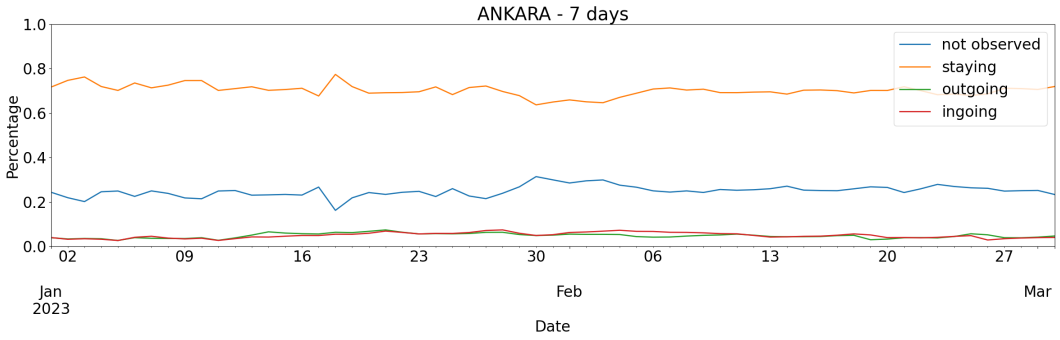


Figure 3.6: User movement between cities between 20.01.2023 – 27.01.2023.

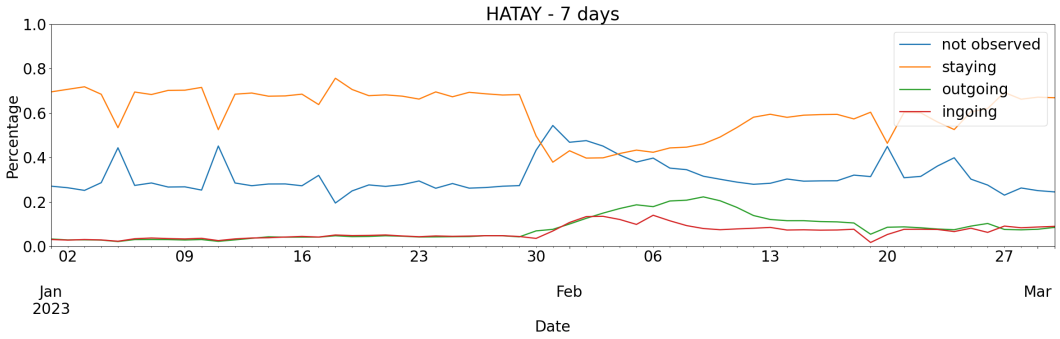
Careful examination of Figure 3.6 reveals notable patterns. Major urban hubs such as Istanbul, Ankara, and Izmir serve as significant attractors and dispersal centers, exhibiting both strong incoming and outgoing flows. The diagonal dominance in the heatmap indicates a large proportion of users remaining within their cities, consistent with routine residential stability. However, off-diagonal elements high-

light the corridors of significant inter-city mobility. For example, substantial flows are visible between Istanbul and Ankara, reflecting the dense socio-economic ties and high transportation connectivity between these two metropolises.

Beyond static transitions, it is critical to analyze how inter-city mobility patterns evolve over time, particularly under disruptive circumstances such as natural disasters. To this end, time-series visualizations such as line charts provide a dynamic perspective on population movement. Figure 3.7 presents line charts for Ankara and Hatay, illustrating daily trends in four key mobility categories: *not observed*, *staying*, *outgoing*, and *ingoing* users.



(a) User mobility trends in Ankara.



(b) User mobility trends in Hatay.

Figure 3.7: Daily user mobility trends for selected provinces, illustrating temporal variations in user activity and inter-city movements.

These categories are defined as follows:

- **Not Observed:** Users not appearing in network records on a given day, potentially indicating either physical absence from the area or reduced device

usage.

- **Staying:** Users consistently located in the same city across consecutive days, reflecting stable residential patterns.
- **Outgoing:** Users moving from the city of interest to another location, indicative of departures or evacuations.
- **Ingoing:** Users arriving into the city from other locations, reflecting inbound mobility or returns.

The charts in Figure 3.7 reveal significant differences between cities. For Ankara, the relatively stable trend across all categories suggests steady urban dynamics and limited influence from extraordinary events during the observed period. In contrast, Hatay's trends show a pronounced spike in the outgoing category immediately following the earthquake of 6 February 2023, coupled with a corresponding drop in the staying population. Such shifts signal a large-scale evacuation and population displacement, underscoring how macro-level analyses can rapidly detect crisis-induced mobility changes.

In addition to time-series plots, Sankey diagrams provide an effective tool for visualizing the magnitude and direction of population flows between multiple destinations [63]. Figure 3.8 depicts flows of users departing from Hatay before and after the February 2023 earthquake. The width of each link represents the volume of users moving between city pairs.

Before the disaster, significant flows from Hatay were directed toward cities such as Mersin, Adana, and Istanbul, reflecting established socio-economic and transportation links. Post-earthquake, the Sankey diagram reveals substantial changes in these patterns. Movements toward Mersin and Antalya increased dramatically, suggesting large-scale displacement and relocation as residents sought refuge in neighboring cities perceived as safer or better resourced. Conversely, flows toward cities closer to the earthquake's epicenter decreased, highlighting disrupted connectivity and potential infrastructure damage.

These observations are quantitatively reinforced by Table 3.3, which details the

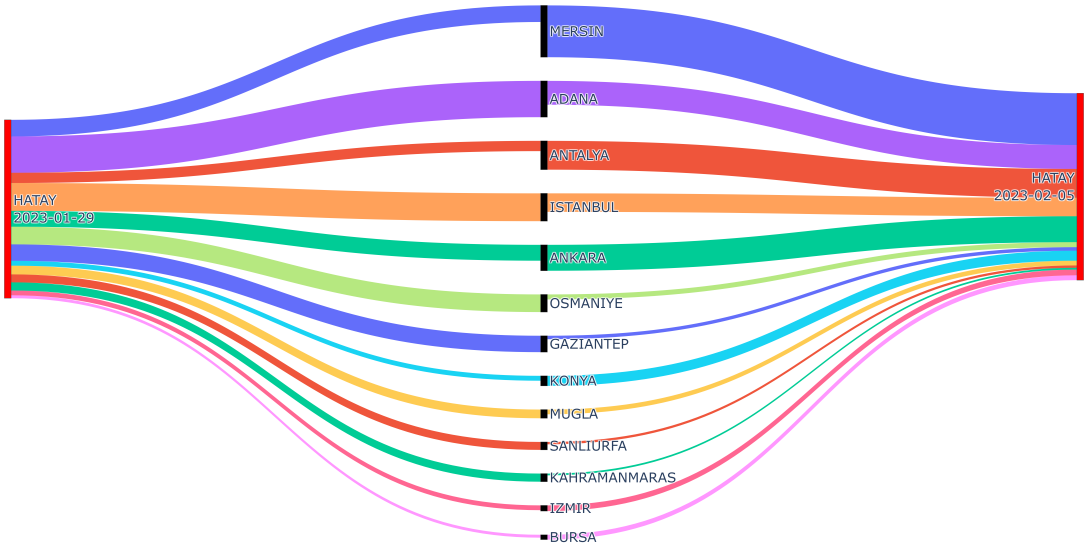


Figure 3.8: Sankey diagram depicting user flows from Hatay to other cities, comparing the periods before and after the 6 February 2023 earthquake.

proportion of outgoing users from Hatay to various cities before and after the earthquake. For example, migration from Hatay to Mersin surged from 6.9% to 21.6%, reflecting Mersin's critical role as a receiving hub during the crisis.

Together, these advanced analytics and visualization techniques form a robust framework for understanding inter-city mobility under both normal and crisis conditions. They transform raw network data into actionable knowledge, enabling informed decision-making for urban planning, disaster response, and socio-economic policy. By bridging the gap between massive digital traces and high-level strategic insights, macro-level mobility analysis contributes a crucial perspective to modern data-driven governance and crisis management.

3.3.3 Discussion

The analyses presented in this section highlight the potential of mobile network data for modeling and interpreting human mobility across both micro and macro scales. Through a robust data processing pipeline, raw measurement reports have been transformed into structured datasets that enable meaningful analyses and visualization.

Table 3.3: Comparison of user mobility for Hatay: Percentage of outgoing individuals and their destination cities.

Cities	**1 week Before Disaster (29.01.2023 to 05.02.2023)**	**After the Disaster (05.02.2023 to 07.02.2023)**
MERSIN	0.069	**0.216**
ANTALYA	0.043	**0.120**
ANKARA	0.066	**0.108**
ADANA	**0.152**	0.100
ISTANBUL	**0.116**	0.077
KONYA	0.020	**0.042**
IZMIR	0.020	**0.024**
OSMANIYE	**0.074**	0.021
BURSA	0.012	**0.020**
MUGLA	**0.036**	0.020
GAZIANTEP	**0.069**	0.014
SANLIURFA	**0.033**	0.010
KAHRAMANMARAS	**0.034**	0.007

At the micro level, grid-based heatmaps have proven effective for visualizing urban dynamics, revealing localized patterns of user density and movement. By comparing consecutive heatmaps that represent snapshots at given times, crowd mobility can be inferred, allowing the detection of temporal shifts in population distribution and identifying emerging hotspots or areas of declining activity. Techniques like the H3 spatial indexing system allow for efficient aggregation of point data, enabling the identification of urban hotspots, commuting patterns, and anomalies, as demonstrated in the İzmir case study. Such insights are valuable for urban planning, network management, and rapid assessment during emergencies.

At the macro level, the MapReduce-based pipeline facilitates the extraction of inter-city user movements from large-scale datasets. Visualization tools such as heatmaps, line charts, and Sankey diagrams provide complementary perspectives on regional mobility flows, capturing both routine movements and disruptions, such as those observed following the February 2023 earthquake. These visualizations make com-

plex data accessible for decision-makers and support effective planning and crisis response.

Overall, this section demonstrates how a data-driven approach leveraging mobile network data can transform vast digital traces into actionable insights. The combination of scalable processing and advanced visualizations offers valuable tools for applications ranging from urban infrastructure planning to disaster management, contributing to a deeper understanding of human mobility patterns in both normal and crisis conditions.

3.4 Predictive Modeling of User Mobility

Modeling human mobility has traditionally focused on describing and visualizing observed patterns derived from large-scale network datasets. Although such descriptive analyses provide valuable information, they are inherently limited in their ability to anticipate future behaviors, particularly under rapidly changing conditions. The increasing frequency of disruptive events such as natural disasters underscores the importance of moving beyond observation towards predictive modeling.

Under normal circumstances, human mobility follows relatively stable and predictable patterns shaped by factors such as urban infrastructure, economic activity, and daily routines. However, during crises, these patterns are often dramatically disrupted, leading to large-scale shifts in population movements. Such abrupt changes pose significant challenges for urban management, disaster response, and resource allocation.

The primary motivation of this study is to bridge this gap by developing a predictive framework capable of forecasting user mobility not only during routine conditions but also in the aftermath of disasters. Specifically, our objective is to investigate whether knowledge derived from one city - capturing both its spatial characteristics and user mobility dynamics - can be transferred to predict mobility behaviors in another city experiencing similar circumstances. This capability has potential for proactive decision making, allowing authorities to prepare for and re-

spond effectively to emerging situations based on data-driven insights.

3.4.1 Methodology of Predictive Modeling of User Mobility

The predictive modeling framework proposed in this study aims to estimate human mobility patterns in a target city by leveraging structural and functional characteristics learned from another city. This approach is particularly relevant under scenarios such as natural disasters, where timely, local mobility data may be unavailable or insufficient for modeling abrupt changes in population behavior.

The problem is formally defined as follows: given a target city T and a reference city R, the objective is to predict user movement behavior in T by identifying structurally and functionally similar regions across the two cities. Figure 3.9 illustrates this conceptual framework.

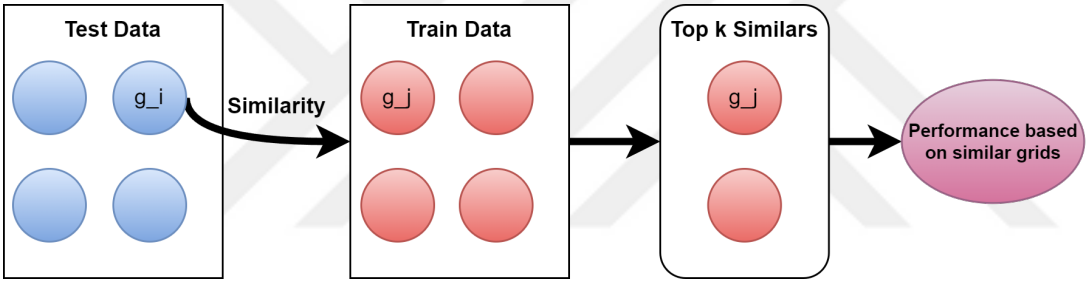


Figure 3.9: Overview of the predictive modeling framework leveraging inter-city similarities. Test data represents target city, train data represents reference city.

The proposed methodology for predictive modeling of user mobility comprises a structured sequence of analytical steps. These steps are summarized below:

1. **Spatial Grid Definition and Feature Extraction:** Partitioning the urban space into uniform grids and extracting detailed urban features for each cell (Section 3.4.1.1).

2. **Computation of Neighbor-Based Grid Properties:** Incorporating spatial interactions by calculating weighted neighborhood feature profiles based on observed user transitions (Section 3.4.1.2).

3. **Grid Similarity Calculation via Cosine Similarity:** Quantifying structural

similarity between grids across cities to identify functionally analogous regions (Section 3.4.1.3).

4. **Prediction of Target Grid Properties:** Estimating properties in the target city using weighted aggregation of properties from similar grids in the reference city (Section 3.4.1.4).

Finally, in Section 3.4.1.5, the evaluation methodologies employed to assess the predictive performance and structural consistency of the proposed framework will be presented, providing quantitative and qualitative information on its effectiveness in different urban contexts.

3.4.1.1 Spatial Grid Definition and Feature Extraction

Both the target city T and the reference city R are spatially partitioned into uniform square grids of size 1 km^2. This discretization enables a detailed representation of spatial heterogeneity and urban structure.

Let P denote the set of all spatial and urban features considered in the analysis:

$$P = \{p_1, p_2, \ldots, p_{21}\}. \tag{3.1}$$

For each grid cell g, the corresponding feature vector is defined as:

$$P_g = \{p_1^{(g)}, p_2^{(g)}, \ldots, p_{21}^{(g)}\}, \tag{3.2}$$

where $p_k^{(g)}$ denotes the value of the k-th feature measured for grid g. The features in P capture a wide range of urban and infrastructural characteristics relevant to human mobility modeling. They are described as follows:

- **Urban vs. Rural Land Proportion** (p_1): The ratio of urbanized land to total grid area, indicating the intensity of urban development.

- **Presence of Industrial Zones** (p_2): The proportion of the grid occupied by industrial facilities such as factories, manufacturing plants, or logistics centers.

- **Road Types** (p_3 **to** p_6): Four separate features quantifying the spatial extent of different road categories in the grid, namely highways, primary roads, secondary roads, and local roads.

- **Energy Facilities** (p_7): The fraction of the grid containing energy-related infrastructure, including power plants, substations, or renewable energy sites.

- **Military Installations** (p_8): The percentage of the grid covered by military zones or restricted defense areas.

- **Airports** (p_9): An indicator of whether the grid contains civilian or military airport infrastructure.

- **Agricultural Lands** (p_{10}): The proportion of the grid utilized for agricultural purposes, such as crop cultivation or orchards.

- **Forested Regions** (p_{11}): The percentage of the grid covered by forested or wooded areas, contributing to green space within the urban environment.

- **Educational Institutions** (p_{12}): The spatial footprint of schools, universities, and other educational facilities within the grid.

- **Religious Buildings** (p_{13}): The presence and spatial coverage of places of worship, including mosques, churches, temples, and other religious institutions.

- **Residential Zones** (p_{14}): The proportion of the grid allocated to residential housing and dwellings, reflecting population density and settlement patterns.

- **Open Spaces** (p_{15}): The fraction of the grid designated as parks, recreational areas, or public green spaces.

- **Government Buildings** (p_{16}): The proportion of the grid occupied by administrative or governmental institutions and facilities.

- **Healthcare Facilities** (p_{17}): The spatial extent of hospitals, clinics, and other healthcare services located within the grid.
- **Social Facilities** (p_{18}): The presence of community centers, social service buildings, and cultural institutions in the grid.
- **Commercial and Mixed-Use Zones** (p_{19} **to** p_{21}): Three features capturing the presence and coverage of commercial areas, retail districts, and mixed-use developments that integrate residential, commercial, and recreational functions.

Together, these features construct a high-dimensional representation of each grid's physical, infrastructural, and socio-economic profile, enabling nuanced modeling of urban environments and human mobility patterns.

3.4.1.2 Computation of Neighbor-Based Grid Properties

To capture spatial interdependencies and the contextual influence of urban surroundings, neighbor-based properties (NP) are computed for each grid cell. This step ensures that the feature profile of any given grid reflects not only its intrinsic characteristics but also the attributes of neighboring regions, which is essential for modeling of human mobility patterns, as illustrated in Figure 3.10.

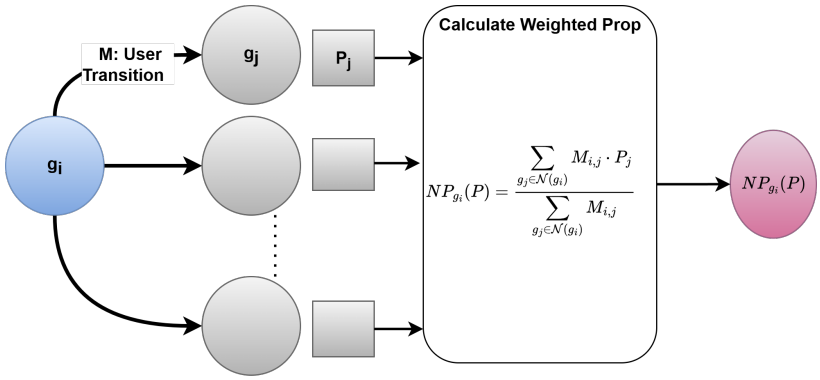


Figure 3.10: Visualization of the computation of neighbor-based properties. For a given grid cell g_i, the transition matrix M is used to weight the properties P_j of neighboring grids g_j, producing the neighbor-based property $NP_{g_i}(P)$.

Let M denote the transition matrix of size $|G| \times |G|$, where $|G|$ is the total number

of grid cells in the city under analysis. Each element $M_{i,j}$ quantifies the number of user transitions observed from grid g_i to grid g_j during a specified observation period. Thus, M encodes the intensity and directionality of spatial interactions within the urban environment.

The neighborhood $\mathcal{N}(g_i)$ of a grid cell g_i is formally defined as:

$$\mathcal{N}(g_i) = \{g_j \in G \mid M_{i,j} > 0\}. \tag{3.3}$$

That is, $\mathcal{N}(g_i)$ comprises all grid cells g_j that have received at least one transition from g_i during the observation window, capturing the set of locations directly connected to g_i through human movement.

For any feature vector P, the neighbor-based property for grid g_i, denoted $NP_{g_i}(P)$, is computed as the weighted average of the corresponding feature values in its neighborhood:

$$NP_{g_i}(P) = \frac{\sum_{g_j \in \mathcal{N}(g_i)} M_{i,j} \cdot P_j}{\sum_{g_j \in \mathcal{N}(g_i)} M_{i,j}}. \tag{3.4}$$

Equation (3.4) ensures that the influence of each neighboring grid g_j on g_i is proportional to the volume of user transitions between them. This computation produces a smoothed estimate of feature vector P for grid g_i, effectively incorporating information from the surrounding spatial context and reducing the impact of data sparsity. Such context-aware representations enhance the reliability of predictive models, particularly in scenarios characterized by sudden shifts in mobility patterns, including those that occur during natural disasters.

3.4.1.3 Grid Similarity Calculation via Cosine Similarity

To facilitate the transfer of knowledge between cities, it is essential to quantify the structural similarity between grid cells in the target city T and those in the reference city R. In this study, cosine similarity is adopted as a metric for measuring the

degree of alignment between the feature profiles of different grids.

Let $g_i \in T$ and $g_j \in R$ denote grids from the target and reference cities, respectively, each represented by their corresponding feature vectors. The cosine similarity score $S(g_i, g_j)$ is computed as:

$$S(g_i, g_j) = \frac{\sum_{k=1}^{21} p_k^{(i)} \cdot p_k^{(j)}}{\sqrt{\sum_{k=1}^{21} (p_k^{(i)})^2} \times \sqrt{\sum_{k=1}^{21} (p_k^{(j)})^2}}, \tag{3.5}$$

where $p_k^{(i)}$ and $p_k^{(j)}$ denote the value of the k-th feature for grids g_i and g_j, respectively. A similarity score $S(g_i, g_j)$ close to 1 indicates a high degree of structural and functional resemblance, suggesting that the grids exhibit similar urban characteristics and land use patterns. This metric serves as the foundation for identifying functionally analogous regions and enables the effective transfer of spatial insights from the reference city to support predictive modeling in the target city.

3.4.1.4 Prediction of Target Grid Properties

In the final stage of the predictive modeling framework, the goal is to estimate the neighbor-based properties for each grid cell g_i in the target city T by leveraging information from structurally similar regions in the reference city R.

For each grid $g_i \in T$, the k most similar grids in the reference city are selected based on their cosine similarity scores. Let $\mathscr{S}(g_i)$ denote the set of these top-k grids:

$$\mathscr{S}(g_i) = \left\{ g_j^{(1)}, g_j^{(2)}, \ldots, g_j^{(k)} \right\}, \tag{3.6}$$

where the elements are ordered such that:

$$S(g_i, g_j^{(1)}) \geq S(g_i, g_j^{(2)}) \geq \cdots \geq S(g_i, g_j^{(k)}).$$

The predicted neighbor-based property $\hat{P}_i$ for grid g_i is computed as a weighted

average of the neighbor-based properties of the grids in $\mathscr{S}(g_i)$, with weights determined by their similarity scores:

$$\hat{p}_i = \frac{\sum_{g_j \in \mathscr{S}(g_i)} S(g_i, g_j) \cdot NP_{g_j}(P)}{\sum_{g_j \in \mathscr{S}(g_i)} S(g_i, g_j)}. \tag{3.7}$$

Equation (3.7) ensures that the prediction for each grid cell in the target city is informed by patterns learned from structurally comparable regions in the reference city. This approach exemplifies a generic and systematic framework, wherein the spatial and functional similarity between grids is explicitly quantified and leveraged for predictive purposes. By identifying grids with analogous characteristics and aggregating their neighbor-based properties through similarity-weighted calculations, the methodology enables informed estimation of spatial attributes in areas where direct observations may be sparse or unavailable. This framework thus provides a basis for transferring spatial knowledge across different urban contexts and for supporting data-driven modeling and analysis of human mobility patterns.

3.4.1.5 Model Evaluation

This part presents the methodology employed to evaluate the predictive modeling framework introduced in this study. The objective of this evaluation is twofold: first, to quantitatively measure the alignment between predicted and observed spatial patterns; and second, to qualitatively assess the structural consistency of predictions relative to actual urban dynamics. To this end, two complementary strategies are integrated into the evaluation process: vector similarity analysis and clustering-based structural comparison.

The subsequent parts detail the specific procedures and metrics used in this comprehensive evaluation framework.

Cosine Similarity Analysis The first analytical approach involves assessing the similarity between the feature representations of pairs of grid cells g_i and g_j in the

target city T. For each grid cell, let $\mathbf{P}_i$ and $\mathbf{P}_j$ denote the feature vectors characterizing the spatial properties of g_i and g_j, respectively. The cosine similarity metric is employed to quantify the degree of alignment between these two vectors:

$$\text{CosineSimilarity}(g_i, g_j) = \frac{\mathbf{P}_i \cdot \mathbf{P}_j}{\|\mathbf{P}_i\| \cdot \|\mathbf{P}_j\|}, \tag{3.8}$$

where the numerator represents the dot product of the two vectors, and the denominator scales this result by the product of their Euclidean norms. Values of $\text{CosineSimilarity}(g_i, g_j)$ approaching 1 indicate high similarity, suggesting that the two grid cells exhibit comparable spatial characteristics. Although defined here for generic pairs of feature vectors, this metric can also be applied to evaluate the similarity between predicted feature representations and ground truth data. A schematic overview of this evaluation process is provided in Figure 3.11.

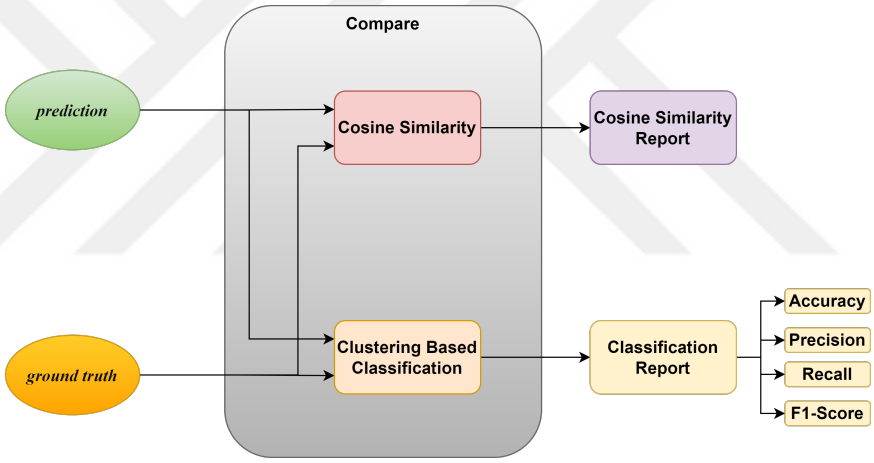


Figure 3.11: Schematic overview of cosine similarity evaluation comparing predicted and ground truth feature vectors.

Clustering-Based Structural Comparison To complement vector-based similarity analysis, a qualitative structural comparison is conducted through clustering-based classification. The rationale behind this approach is to examine whether the predictions preserve the same spatial segmentation patterns that exist in the ground truth data. This procedure unfolds in several steps:

1. A clustering algorithm—such as K-Means—is trained on the ground truth

feature vectors derived from the target city.

2. The trained clustering model is then employed to assign cluster labels to:
   - the ground truth feature vector $\mathbf{P}_i$, representing the spatial properties of grid cell g_i in the target city,
   - and the predicted feature vector $\widehat{\mathbf{P}}_i$, representing the estimated spatial properties for the same grid cell g_i.
3. The predicted cluster labels are compared to the ground truth labels, enabling an evaluation of how accurately the model captures the underlying spatial structure.

This methodology is visualized in Figure 3.12, which illustrates how both predicted and actual feature sets are subjected to identical clustering analyses, thereby enabling rigorous label-based comparisons.

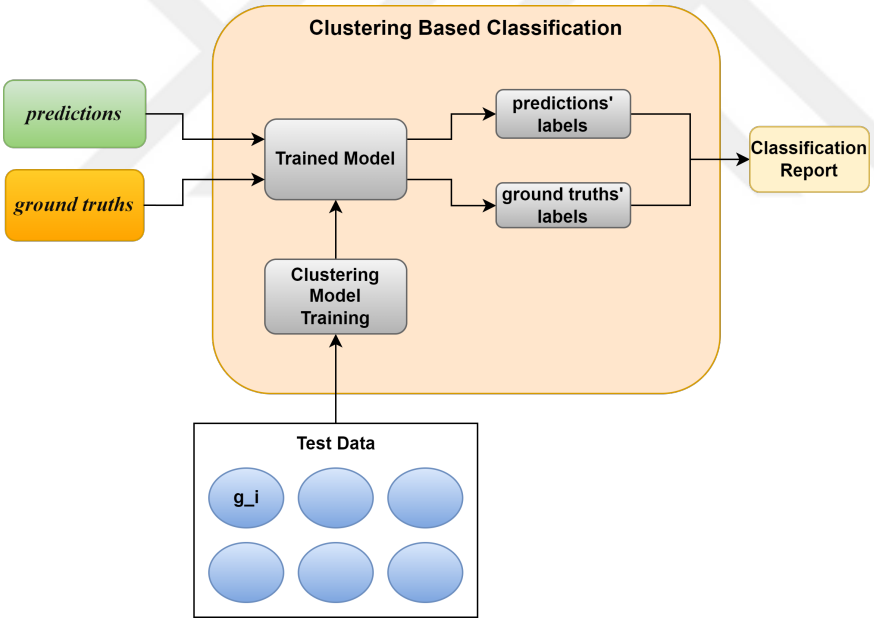


Figure 3.12: Workflow of clustering-based classification for assessing structural consistency between predictions and ground truth data.

Classification Metrics for Structural Assessment The agreement between predicted and true cluster labels is quantified using several well-established classification metrics:

- **Accuracy**, which measures the proportion of correctly assigned cluster labels:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}.$$

- **Precision**, which evaluates the proportion of grids assigned to a cluster that truly belong to it:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}.$$

- **Recall**, which quantifies the proportion of actual cluster members that are correctly identified:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}.$$

- **F1-Score**, which balances precision and recall through their harmonic mean:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

These metrics collectively provide a robust statistical basis for evaluating the structural fidelity of the predictions relative to the true spatial organization of the city.

Dimensionality Reduction for Visualization Given the high dimensionality of the feature space, Principal Component Analysis (PCA) is employed as a dimensionality reduction technique. PCA facilitates visualization of both predicted and ground truth feature vectors in two or three dimensions, enabling direct visual comparison of cluster separability and the preservation of urban structural patterns.

In summary, the dual strategy of cosine similarity analysis and clustering-based structural comparison establishes a rigorous, multi-perspective evaluation framework for the predictive modeling system. This approach ensures that the model's outputs are assessed not merely in terms of numerical similarity, but also in relation to their ability to replicate the spatial patterns that characterize real-world urban environments.

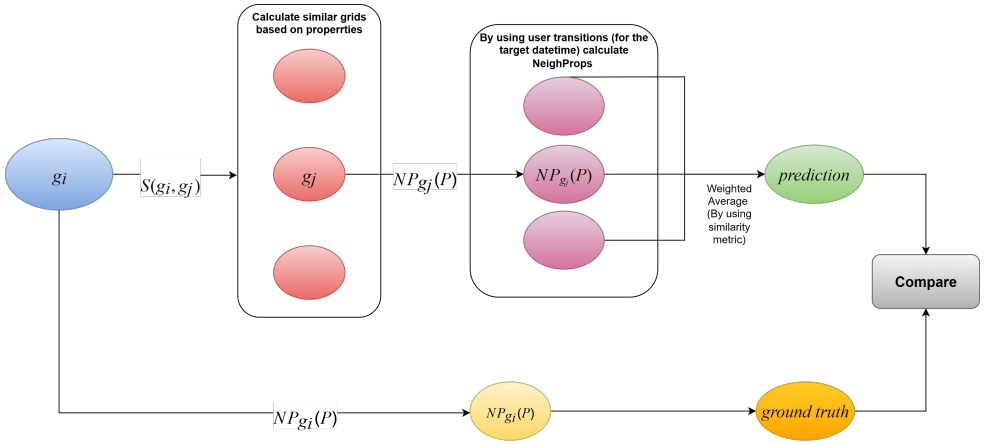


Figure 3.13: Illustration of the overall predictive modeling system pipeline, integrating grid similarity calculation, neighbor-based property prediction, and evaluation against ground truth data.

3.4.1.6 End-to-End Workflow of Predictive User Mobility Modeling

Figure 3.13 illustrates the complete pipeline of the proposed predictive modeling framework. The process begins with a selected grid cell g_i in the target city T. To identify structurally similar regions, the cosine similarity $S(g_i, g_j)$ is computed between g_i and all other grid cells g_j in the reference city R, based on their feature vectors. This yields a set of grids in R that are most similar to g_i in terms of spatial characteristics.

For each similar grid g_j, the neighbor-based property vector $NP_{g_j}(P)$ is computed using observed user transitions in the reference city. These neighbor-based properties are then aggregated to estimate the corresponding neighbor-based property vector for g_i, denoted $\widehat{NP_{g_i}(P)}$. This aggregation utilizes a weighted average, where the weights are proportional to the similarity scores $S(g_i, g_j)$.

In parallel, the ground truth neighbor-based property vector $NP_{g_i}(P)$ is computed directly from transition data available in the target city. Finally, the predicted neighbor-based property vector $\widehat{NP_{g_i}(P)}$ and the ground truth vector $NP_{g_i}(P)$ are compared to evaluate the accuracy of the predictive framework. This comparison employs measures such as cosine similarity to quantify the alignment between predicted and actual spatial dynamics.

3.4.2 Results

This section presents the experimental results obtained from applying the proposed predictive modeling framework to real-world urban mobility data. The analysis is based on datasets from two Turkish cities: Adana and Malatya. Adana serves as the reference city, while Malatya functions as the test city to evaluate the transferability and prediction performance of the proposed method.

The grid-based datasets for both cities were constructed at a resolution of 1 km^2. Specifically, Adana and Malatya comprise 458 and 214 grids, respectively. Each grid cell is characterized by a 21-dimensional feature vector encoding diverse urban and infrastructural attributes, as detailed in Section 3.4.1.1.

Several key analyses have been performed to assess the predictive model's accuracy and structural consistency:

1. Evaluation of cosine similarity between predicted and ground truth neighbor-based properties, as a quantitative measure of prediction alignment.
2. Dimensionality reduction through PCA to explore the data structure and support subsequent clustering.
3. Clustering analysis to identify meaningful groupings in the data and to compare predicted and true spatial patterns.
4. Clustering-based classification metrics (accuracy, precision, recall, and F1-score) to quantitatively evaluate the agreement between predicted and actual cluster assignments.

3.4.2.1 Cosine Similarity Analysis

A central component of evaluating the proposed predictive modeling framework is the assessment of how closely the predicted grid-level property vectors align with the actual spatial characteristics observed in the test city. This alignment is quantified via the cosine similarity metric, previously defined in Equation (3.8). Higher values of cosine similarity indicate a stronger correspondence between predicted

and ground truth feature vectors, reflecting greater accuracy in capturing the urban spatial structure.

Figure 3.14 presents the average cosine similarity results obtained for different values of k, where k denotes the number of top similar reference grids used in forming the predictions for each test grid. The figure reveals several important trends. First, there is a clear improvement in the average cosine similarity as k increases from 3 to 20. When only $k = 3$ neighbors are considered, the average similarity remains relatively lower, indicating that relying on a small number of reference grids may not sufficiently capture the complex spatial patterns present in the test city.

As k increases to 5, 10, and 20, the mean cosine similarity progressively rises, surpassing values of 0.85 for $k = 10$ and approaching 0.9 for $k = 20$. This suggests that aggregating information from a broader set of similar grids in the reference city enhances the stability and representativeness of the predictions. The error bars in Figure 3.14 further indicate that higher k values reduce variability across grid predictions, underscoring improved consistency in the modeling outcomes.

However, it is also observed that the magnitude of improvement diminishes as k becomes larger. While increasing k from 3 to 5 or 5 to 10 yields notable gains, the incremental benefit from moving to $k = 20$ is more modest. This plateau effect suggests that there is a balance between incorporating sufficient spatial context and avoiding excessive smoothing that may obscure local distinctions in urban structure.

Overall, these findings demonstrate that the predictive model effectively leverages structural similarities across cities to estimate grid-level features with high fidelity, and that tuning the parameter k is crucial for optimizing performance. The results highlight the capacity of the framework to transfer spatial knowledge between distinct urban contexts, supporting robust predictions even in regions where direct measurement data may be limited.

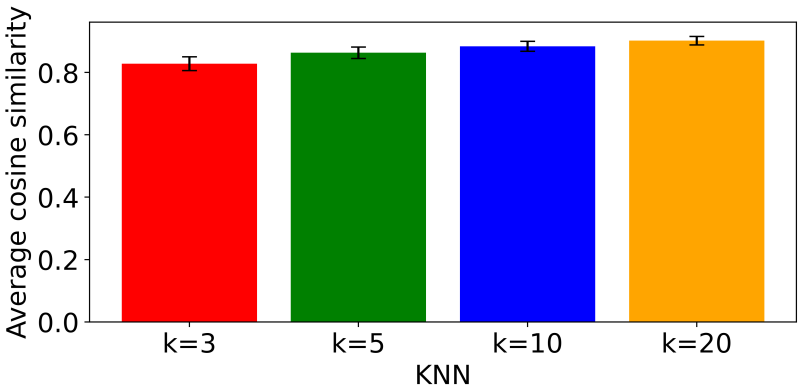


Figure 3.14: Average cosine similarity scores achieved by the predictive modeling framework for different top-k neighbor counts in the reference city. Error bars represent variability across all test grids.

3.4.2.2 Dimension Reduction with Principal Component Analysis

Given the high dimensionality inherent in the grid-based urban feature space, PCA is employed as a method for dimensionality reduction. The primary objective of applying PCA within this framework is to identify a minimal set of orthogonal components that capture the majority of the variance present in the original 21-dimensional feature vectors $\mathbf{p}_i$. This reduction facilitates more efficient computational processing while retaining the essential structural information needed for downstream analytical tasks.

Figure 3.15 illustrates the cumulative explained variance ratio as a function of the number of principal components retained. As depicted, the first two principal components alone account for approximately 88% of the total variance in the feature space. Incorporating four principal components increases the cumulative explained variance to beyond 95%, demonstrating that a significant proportion of the spatial and functional variability among grids can be effectively captured using a relatively low-dimensional representation.

This observation has critical implications for subsequent analytical stages. Reducing the feature space dimensionality to two or four principal components minimizes redundancy and mitigates the risk of overfitting in machine learning models such as K-means clustering. Moreover, lower-dimensional representations facilitate the visual exploration of the spatial organization of grids, enabling the detec-

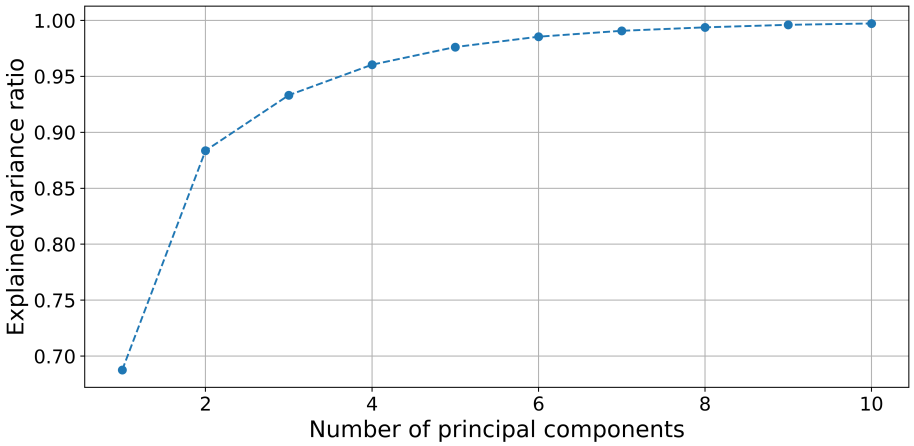


Figure 3.15: Cumulative explained variance ratio as a function of the number of principal components.

tion of meaningful clusters and structural patterns that might otherwise remain obscured in the high-dimensional feature space.

In this study, the derived principal components are subsequently used as inputs for clustering analyses (see Section 3.4.2.3), where they serve as a compact yet informative feature set to characterize the spatial similarities and differences among grids in the target city. By balancing dimensionality reduction with information preservation, PCA thus acts as a crucial preprocessing step, enhancing both the interpretability and computational efficiency of the predictive modeling framework.

3.4.2.3 Clustering Analysis

Clustering analysis serves as a crucial tool for assessing the structural coherence between predicted and observed urban mobility patterns. Given the high dimensionality of the grid feature space, it is first necessary to reduce the feature set to a more tractable size to facilitate meaningful clustering results.

To this end, PCA is applied prior to clustering. The principal purpose of PCA in this context is twofold: (i) to mitigate the curse of dimensionality by projecting the original feature vectors into a lower-dimensional subspace that retains the majority of the data's variance, and (ii) to improve the performance and stability of subsequent clustering algorithms. The number of principal components retained is

determined based on the cumulative explained variance ratio, ensuring that the reduced representation remains sufficiently descriptive of the urban grid structures while removing redundant information.

Once dimensionality reduction is completed, K-Means clustering is performed on both the ground truth and predicted neighbor-based property vectors. The aim is to examine whether the spatial segmentation induced by the model's predictions aligns with the inherent patterns present in the actual data. Figure 3.16 presents the silhouette scores obtained for various combinations of the number of clusters and the number of PCA components retained.

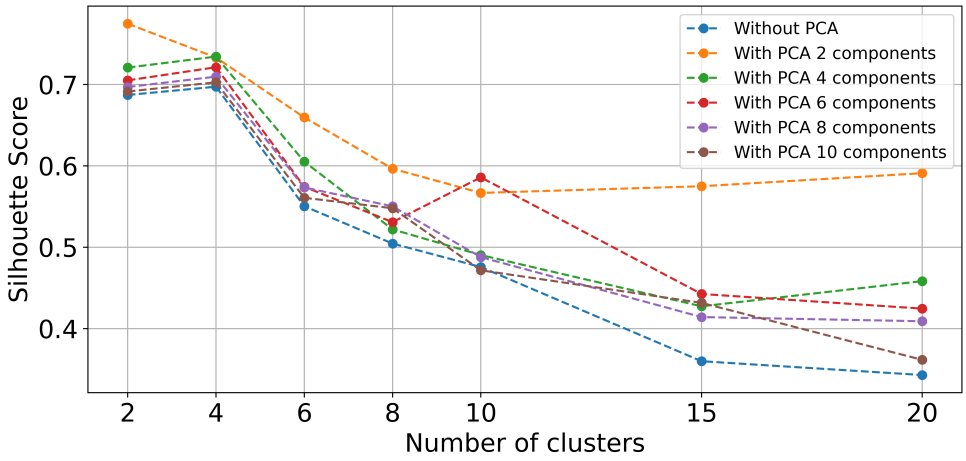


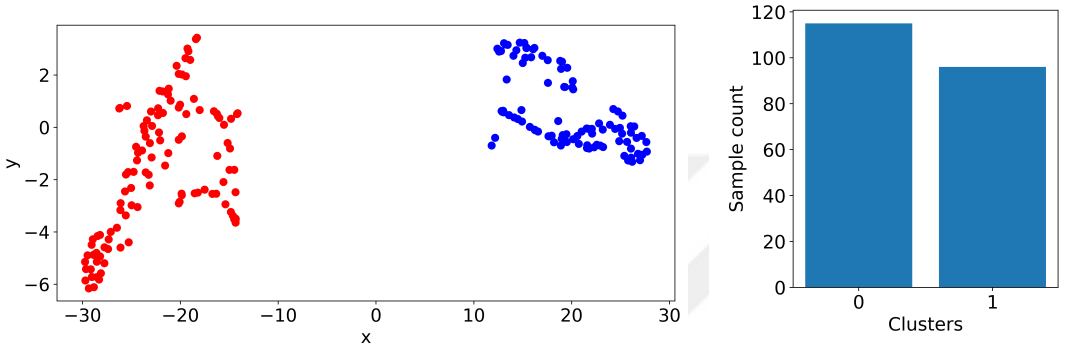
Figure 3.16: Silhouette scores as a function of the number of clusters, comparing results with and without PCA, as well as for different numbers of PCA components.

From Figure 3.16, it is evident that clustering quality, as measured by the silhouette score, generally declines as the number of clusters increases. This behavior reflects the greater difficulty in cleanly partitioning data into a large number of well-separated clusters, especially in urban environments where spatial features may overlap. Notably, the use of PCA substantially improves silhouette scores across all cluster sizes, with particularly high scores observed when retaining only two or four principal components. This finding suggests that the essential spatial patterns governing urban grid properties are effectively captured in a low-dimensional space.

Moreover, the highest silhouette scores are consistently achieved for configurations involving two or four clusters, indicating that the urban grids naturally organize into a small number of distinct functional groups. This observation aligns

with the expectation that urban areas exhibit relatively coarse-grained segmentation, with zones such as residential, commercial, industrial, and open spaces forming prominent clusters.

Further qualitative insights are obtained by visualizing the clustering outcomes. Figure 3.17 illustrates the two-dimensional embedding of the clustered data, obtained via t-Distributed Stochastic Neighbor Embedding (t-SNE), using two clusters and four principal components as inputs to the clustering process.



(a) t-SNE projection of clustered grid features using two clusters and four PCA components. The spatial separation between clusters demonstrates the distinctiveness of urban regions captured by the model.

(b) Distribution of data points across the two identified clusters, indicating the relative size and balance of each cluster.

Figure 3.17: Visualization of clustering results based on K-Means applied to PCA-reduced grid features.

In Figure 3.17a, a clear spatial separation between clusters is evident, suggesting that the reduced feature space successfully preserves the key distinctions among urban grid types. Furthermore, Figure 3.17b shows that the two clusters are relatively balanced in terms of data point counts, confirming that the segmentation achieved by the clustering algorithm is both meaningful and statistically robust.

Overall, the clustering analysis demonstrates that the predictive modeling framework not only generates feature vectors that are quantitatively similar to ground-truth representations, but also successfully reproduces the underlying structural organization of urban spaces. The combination of PCA and K-Means emerges as a methodological pipeline to validate the spatial consistency and fidelity of mobility predictions in high-dimensional urban data.

3.4.2.4 Clustering-Based Classification Metrics

The evaluation of the proposed predictive modeling framework further extends to a quantitative comparison between the spatial structures inferred from the predicted feature vectors and those derived from the ground-truth data. This comparison is operationalized through a clustering-based classification procedure, where cluster memberships assigned to predicted and actual data are compared to assess structural agreement.

Table 3.4 summarizes the results of this analysis across different experimental configurations, varying the number of top-k similar grids considered, the number of clusters chosen for segmentation, and the dimensionality of the feature space following PCA reduction.

Several notable trends can be observed from the results:

- **Impact of Cluster Count.** Across all configurations, the use of four clusters consistently yields substantially higher classification metrics than two clusters. For instance, with top-$k = 20$ neighbours and four clusters, the framework achieves an accuracy of 92.9%, reflecting a strong ability to replicate the structural segmentation of the urban environment.
- **Effect of Top-k Neighbours.** Increasing the number of top-k similar grids generally leads to improvements in all metrics. This suggests that leveraging a broader contextual pool of reference grids enhances the robustness of the predictions, capturing more nuanced urban variations.
- **Role of PCA Dimensionality.** Variations in the number of PCA components introduce modest differences in performance. However, the results show that both low-dimensional representations (e.g., 2 or 4 components) and higher-dimensional configurations (up to 10 components) yield comparably strong results, indicating that the essential spatial structures are effectively preserved even under aggressive dimensionality reduction.
- **Balanced Performance Across Metrics.** Not only do the models achieve high accuracy, but they also maintain strong precision and recall values, leading

to consistently high F1-scores. This indicates that the predictive framework is capable of reliably distinguishing among urban grid types without significant biases towards particular classes.

Table 3.4: Clustering-based classification performance across different experimental configurations.

Top-k Neighbours	**# of Clusters**	**PCA # of Components**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
5	2	2	0.767	0.879	0.767	0.810
5	2	4	0.773	0.882	0.773	0.815
5	2	10	0.776	0.883	0.776	0.817
5	4	2	**0.911**	**0.950**	**0.911**	**0.928**
5	4	4	0.895	0.946	0.895	0.917
5	4	10	0.890	0.944	0.890	0.913
10	2	2	0.759	0.873	0.759	0.802
10	2	4	0.769	0.877	0.769	0.810
10	2	10	0.773	0.877	0.773	0.813
10	4	2	0.919	0.953	0.919	0.935
10	4	4	0.900	0.947	0.900	0.921
10	4	10	0.897	0.945	0.897	0.918
20	2	2	0.801	0.866	0.801	0.828
20	2	4	0.811	0.870	0.811	0.836
20	2	10	0.814	0.871	0.814	0.838
20	4	2	**0.929**	**0.957**	**0.929**	**0.941**
20	4	4	0.910	0.951	0.910	0.928
20	4	10	0.907	0.948	0.907	0.925

Taken together, the results in Table 3.4 provide compelling evidence for the efficacy of the proposed predictive approach. The framework demonstrates a robust capacity to reconstruct spatial patterns consistent with ground truth observations, validating its potential utility in urban planning, spatial analysis, and mobility management applications.

3.4.3 Discussion

The findings presented in this chapter demonstrate that the proposed predictive modeling framework can effectively estimate urban mobility patterns in a target city by leveraging structural similarities with a reference city. Through rigorous quantitative and qualitative evaluations—including cosine similarity analysis, dimensionality reduction via PCA, clustering assessments, and classification metrics—the model has shown a robust capacity to replicate both fine-grained spatial properties and broader urban structures.

The results underscore several important insights. Firstly, increasing the number of top-k similar grids incorporated into the prediction generally improves model performance, suggesting that drawing from a broader pool of structurally analogous regions helps capture the inherent complexity of urban environments. Secondly, dimensionality reduction through PCA proved essential for stabilizing clustering results and facilitating the detection of coherent spatial patterns, while preserving critical variance in the data.

Despite these promising outcomes, certain limitations should be acknowledged. The reliance on structural and functional features assumes that analogous urban forms will yield comparable mobility dynamics, an assumption that may not fully hold under unique socio-economic, cultural, or disaster-specific conditions. Moreover, while the framework demonstrates strong performance under the tested configurations, its generalizability to cities with drastically different urban morphologies or mobility behaviors remains an open question requiring further empirical validation.

In conclusion, this study establishes a systematic, transferable approach for predicting urban mobility using cross-city knowledge. The integration of similarity-based modeling, dimensionality reduction, and clustering-based evaluation provides a robust analytical toolkit for applications ranging from urban planning to emergency response. Future research could explore the inclusion of temporal dynamics, multi-modal mobility data, and adaptive learning mechanisms to further enhance predictive accuracy and adaptability in diverse urban contexts.

3.5 Conclusion

This chapter has presented a comprehensive exploration of modeling and predicting human mobility through the lens of mobile network data, bridging descriptive analyses with predictive frameworks. Mobile networks generate vast volumes of user-level measurement reports, which, when systematically processed and analyzed, yield unprecedented insights into both routine urban dynamics and mobility disruptions induced by extraordinary events.

A central contribution of this work lies in demonstrating that mobile network data can serve as a powerful proxy for observing and forecasting human movement, especially valuable in the context of disaster response. The proposed methodologies—including scalable big data processing pipelines, grid-based spatial aggregation, and predictive modeling driven by inter-city similarity measures—enable the transformation of raw, high-frequency measurements into actionable intelligence for urban planning, transportation management, and emergency preparedness.

Crucially, this chapter has shown that leveraging structural similarities between cities allows for the transfer of knowledge to estimate mobility patterns even when direct local data is limited or unavailable, as might be the case immediately following a disaster. Such transfer learning approaches, grounded in robust similarity metrics like cosine similarity, and refined through dimensionality reduction and clustering techniques, highlight the potential of data-driven frameworks for proactive decision-making in crisis management.

Moreover, this research underscores the broader value of user-level mobile network data beyond traditional network optimization. By integrating advanced analytical methods, it is possible to conduct nuanced studies of urban morphology, socio-economic spatial patterns, and human behavioral responses under stress. The systematic adoption of big data processing techniques ensures that these analyses remain feasible despite the sheer scale of modern network datasets.

Nonetheless, several avenues remain for future investigation. The inclusion of temporal dynamics, adaptive learning algorithms, and multi-source data fusion —incorporating social media, transportation logs, or environmental sensors— could

further enhance the predictive capacity and resilience of mobility models. Additionally, exploring privacy-preserving analytics will be essential to ensure that such powerful insights can be derived responsibly and ethically.

In summary, this chapter establishes that mobile network data, analyzed through scalable and sophisticated methodologies, can substantially advance our ability to understand and predict human mobility. Such capabilities hold significant promise for supporting data-driven urban governance, optimizing infrastructure, and safeguarding populations in the face of both routine and extraordinary challenges.

CHAPTER 4

AI-ASSISTED SECTOR SHUTDOWN FOR MOBILE NETWORK ENERGY EFFICIENCY: ALGORITHMS AND LIVE NETWORK TRIALS

Mobile networks play a critical role in connecting people and businesses to the world around them. However, the rapid growth in mobile data usage has led to an increase in energy consumption by mobile networks, which is a major concern for operators due to the high operating expenses and carbon footprint [64]. Reducing energy consumption in mobile networks is essential, and base stations are the main contributors to this energy consumption, with some studies showing they consume up to 60% of all mobile network power [65]. In order to meet the rising energy demands of mobile networks, operators are turning to energy-saving techniques to reduce their carbon footprint and lower their energy costs [66]. The problem is reducing base stations' energy consumption without negatively affecting the coverage and capacity.

The base station (BS) deployment locations are called sites, each site may have multiple sectors. The sectors have servicing frequency bands called carriers. In the existing radio access network (RAN) configuration of many mobile network operators (MNO) around the world, it is common to see roughly 3 sector – 6 carrier multi-technology configuration for their urban sites.

Base station sleeping is one way to reduce energy consumption in mobile networks by shutting down underutilized base stations. The sleeping strategy is based on carrier shutdown, which means that carriers that are not currently being used are turned off or put into a low-power mode [67]. The main idea of the carrier shutdown strategy is leaving only one carrier up for coverage and turning off the other carriers required for capacity. This strategy is provided as a vendor feature by

equipment manufacturers. The mobile operator may decide the thresholds that define the underutilized base stations or carriers.

The problem with carrier shutdown is that the limited energy savings achieved, as base stations continue to consume significant energy even when carriers are deactivated [68]. This occurs because at least one carrier must remain operational, necessitating continuous power for cooling, maintaining network synchronization, and supporting the baseband unit (BBU) operations. Additionally, understanding and optimizing the energy efficiency of mobile networks are profoundly influenced by power models and energy consumption formulas. Energy usage in base stations is divided into two primary components: load-dependent and non-dependent parts. The load-dependent aspect correlates directly with the network's activity and traffic load, whereas the non-dependent portion accounts for the energy expended even in the absence of active data transmission [69]. A sector that consists of multiple transceivers (TRXs) energy consumption formulated as:

$$P_{\text{in}} = \begin{cases} N_{\text{TRX}} \cdot P_0 + \Delta_p P_{\text{out}} & \text{if } 0 < P_{\text{out}} \leq P_{\text{max}} \\ N_{\text{TRX}} \cdot P_{\text{sleep}} & \text{if } P_{\text{out}} = 0, \end{cases} \tag{4.1}$$

where P_{in} is the total energy consumption of a sector, P_{out} is radio frequency (RF) output, P_0 represents the minimum power consumption when the output power is non-zero, Δ_p denotes the slope of the load-dependent power consumption, N_{TRX} is the number of TRX, and P_{sleep} is a sleep mode energy consumption.

In addition to the carrier shutdown strategy, we propose a sector shutdown algorithm to improve energy saving. A complete sector shutdown is a more comprehensive approach to reduce energy consumption in mobile network base stations. A sector shutdown involves shutting down an entire section of the base station, including the BBU and cooling systems and other associated components. This approach offers more significant energy savings, as the entire sector is powered off and not consuming energy. However, it can also have a more significant impact on network coverage and may require a more complex network design in order to ensure that coverage is not affected. The complete sector shutdown is a very risky operation due to the potential coverage and capacity problems.

To shut down sectors completely, it is key to understand how sites are selected and

deployed. In mobile networks, while deploying sites, there are two main concerns: coverage and capacity [70]. It is essential for mobile operators that users can be connected from anywhere at any time. Therefore, it is crucial to plan the sites to ensure ubiquitous coverage. In this respect, the deployed sites are called coverage sites. These sites are typically located in remote or rural areas, or in areas where there is a lack of existing infrastructure. Coverage sites ensure that users in these areas can connect to the mobile network and have access to the same services as users in more densely populated areas.

On the other hand, capacity sites are aimed to strengthen the mobile network's ability to handle a large number of users in a specific area. These are typically positioned in areas with high population densities or heavy data traffic. To prevent network congestion, these sites are essential to deploy. To ensure meeting the customer requirements, sites are planned and deployed based on peak values of forecasted regional daily traffic demands. Therefore, during low utilization times, some of these sites might not be fully utilized.

It is obvious that in densely populated areas, there are more sites to improve capacity. To meet customers' requirements, mobile operators deploy more sites by considering peak values. Since these sites are densely deployed, there are coverage overlaps of these sites. Therefore, if we determine these overlaps we may determine the redundant sectors that may completely shut down during the rest hours.

One challenge faced in densely deployed areas is that they typically serve as 24/7 living areas. Consequently, performing a complete sector shutdown in such areas poses a considerable risk. However, the nature of this issue necessitates maximizing coverage overlap among sectors in dense urban areas.

To minimize risks and develop a data-driven approach, we utilize live network data, which allows us to ascertain the precise coverages and overlaps of individual sectors. With this comprehensive understanding of the network's operational status, we can effectively forecast the potential impact of shutting down specific sectors. By analyzing user observations from the live network data, we gain valuable insights into which sectors serve particular locations. Consequently, when a sector is scheduled for shutdown, we can identify the sectors capable of com-

pensating for the loss in coverage, ensuring minimal disruption to network performance. This data-driven methodology empowers us to make well-informed decisions, optimizing sector shutdown strategies while maintaining a high-quality customer experience.

In this work, we explored two distinct scenarios to enhance our understanding of sector shutdown algorithms and their impact on mobile networks. The first scenario is conducted offline to rigorously test our algorithms, including the Genetic Algorithm Solution (GAS) and a heuristic-greedy solution known as the Cost-Aware Heuristic Greedy Shutdown (CAHGS). In this scenario, we tested our algorithms by using 30 days historical data from the live network. This initial phase is aimed to identify potential problems and refine our approach by using KPIs of base stations—such as utilization rates, connected user counts, and traffic volumes—alongside geolocated user measurement reports that contains the connected and neighboring base stations and their signal qualities. The intricacies of these developed algorithms are further elaborated in section 4.3.

The second scenario is specifically designed to test the outcomes on a live network, introducing an algorithm that adapts to network dynamics. This algorithm, dubbed the Neighbour Based Expert System Solution (NBESS), represents a heuristic-greedy methodology. Our testing of the NBESS is carried out in a live network environment, specifically in the bustling urban center of Istanbul, Türkiye, a city that operates around the clock. Here, we applied sector shutdown decisions on live network to observe and evaluate the algorithm's efficacy in real-world conditions. This two-pronged approach allowed us to not only pinpoint potential issues through offline simulations but also to validate our findings within a live network, ensuring our solutions are both practical and effective.

The inputs for the proposed sector shutdown algorithm include the given network configuration and daily traffic. The network configuration includes the number of sites, sectors, and carriers. The daily traffic includes information such as user measurements, which are collected by each user device reporting the connected sector and the neighbouring sectors that can be observed by the user. These user measurement reports are geolocated by using machine learning, and this data is

used to determine the coverage areas of the base stations.

The outputs for the proposed algorithm include the number of sectors that can be shut down at each site, as well as the specific times and duration for these shutdowns. The algorithm is able to determine when, how long, and which sectors can be shut down in order to maximize energy savings.

The primary objective of the proposed algorithm is to maximize energy savings by maximizing the number of sectors that can be shut down. While doing that, we need to maintain customer experience. The sub-objective of the algorithm is to determine coverage overlaps in order to identify redundant sectors that may be shut down without causing coverage gaps. It is important to note that we need to shut down only the sectors that are not in use, and not compromise the coverage.

The proposed solutions have several constraints that should be considered in order to avoid negative effects on network coverage and performance. These constraints include maintaining coverage without leading to coverage gaps, and without causing a significant drop in user performance below a threshold. Specifically, each sector that is shut down should be compensated by other sectors, and the drop in RSRP should not exceed the given threshold limits. This ensures that the negative impact of the sector shutdown algorithm is limited and customer satisfaction is maintained.

The rest of the chapter is organized as follows. Section 4.1 presents an overview of related works, exploring prior research and solutions that have been proposed to address similar issues. In Section 4.2, we formally define the problem, precisely outlining the challenges posed by the increasing energy consumption in mobile networks and the need to optimize sector shutdown strategies. In Section 4.3, we delve into the methodologies employed, including data processing techniques, and introduce the three distinct models: GAS, CAHGS, and NBESS. Subsequently, in Section 4.4, we present the results and evaluations obtained from the proposed models and algorithms. Finally, we draw our conclusions and summarize the findings in Section 4.5.

4.1 Related Work

Energy efficiency has become a fundamental design and operational objective in mobile networks, particularly within RANs, which are known to be among the most energy-consuming components of the infrastructure. A significant portion of a mobile network's power budget is consumed by BSs, necessitating innovative approaches to reduce energy usage without compromising network coverage or user experience. Consequently, extensive research has been devoted to developing energy-saving strategies tailored to various components of the RAN. To structure the discussion, we group the relevant literature into key methodological categories, including carrier and cell deactivation strategies, sleep mode orchestration techniques, and sector-level shutdown mechanisms. We also differentiate between simulation-based studies and those evaluated in live network environments, and we explicitly position our proposed work within this landscape. The overview of these representative studies is provided in Table 4.1.

Table 4.1: Summary of representative energy-saving methods in mobile networks, categorized by granularity and evaluation type.

Author/Year	**Main Approach**	**Granularity**	**Test Type**
Dahal et al., 2018	Dynamic Transmitter Shutdown Technique (DTST): selectively switches off idle transceivers based on predicted traffic (Holt-Winters + Erlang B) while maintaining QoS.	Transceiver within BS	**Live Deployment** – tested with Nepal Telecom at a single BS.

Continued on next page

Table 4.1 (continued)

Author/Year	Main Approach	Granularity	Test Type
Yong Tang et al., 2023	Reinforcement learning (RL) based node sleep scheduling using a Confident Information Coverage (CIC) model to minimize active nodes while ensuring adequate sensing coverage and maximizing network lifetime.	Node Level	**Simulation –** Evaluations are conducted using the MATLAB platform.
Masoudi et al., 2020	RL based adaptive sleep mode selection using SARSA algorithm to balance energy savings and service delay based on real-time traffic load.	Cell	**Simulation –** Evaluated using real traffic data from a base station in Stockholm with synthetic fine-grained input.
López-Pérez et al., 2024	Hybrid expert and machine learning-based modeling for energy efficiency and UE throughput prediction under carrier shutdown scenarios, enabling operator-grade modeling using only live network data.	Cell	**Simulation –** Offline modeling and evaluation with an operator's benchmark tool.
Beitelmal et al., 2018	Introduction and evaluation of 26 regular static switch-off patterns for both site-level and sector-level cell switch-off (CSO) strategies.	Cell	**Simulation –** Custom-developed system-level simulator based on standardized UMi and UMa scenarios.

Continued on next page

Table 4.1 (continued)

Author/Year	Main Approach	Granularity	Test Type
Alaca et al., 2012	Genetic algorithm-based cell switch-off scheme that iteratively selects the optimal order of cell deactivation while considering user rate constraints and interference.	Cell	**Simulation –** Evaluated using an urban micro-cell scenario with ITU-R International Mobile Telecommunications (IMT) Advanced parameters and interference-aware modeling in MATLAB.
Suárez et al., 2015	Cell breathing-based BS switching-off strategy regulated by traffic thresholds and enhanced with macro-to-femto offloading under Closed Subscriber Groups (CSG) access control.	Sector	**Simulation –** Monte Carlo-based simulation under Code Division Multiple Access (CDMA) and LTE settings, modeling traffic, topology, and femto access restrictions.
Conte et al., 2011	Stepwise transmission power control for base station switch-off (wilting) and switch-on (blossoming), enabling safe handovers and minimizing interference during transitions.	Base Station	**Simulation –** Evaluated using ray-tracing-based simulations (WiSE tool) in a realistic urban scenario (Munich, Germany).

Continued on next page

Table 4.1 (continued)

Author/Year	Main Approach	Granularity	Test Type
Salem et al., 2019	Q-learning-based management of Advanced Sleep Modes (ASMs), optimizing the sleep duration levels of BS components dynamically based on traffic load and operator policy (energy vs. delay trade-off).	Base Station	**Simulation** – An event-based simulator is developed to evaluate performance in terms of energy savings and latency under varying traffic loads.
El-Amine et al., 2019	Q-learning-based multi-level sleep mode management for 5G base stations, using user location and velocity to balance energy savings and service delay.	Base Station	**Simulation** – Evaluated via simulations with mobility and positioning models, assessing energy-delay tradeoffs.
Han et al., 2012	Base station switch-off patterns with cooperative communication and power control to extend coverage and maintain QoS in off-peak hours.	Base Station	**Simulation** – Analytical and numerical evaluation of outage probability and power savings under various switch-off and cooperation patterns.
Oh et al., 2011	Dynamic base station switch-off based on low traffic periods using a greedy algorithm to maintain >95% coverage while deactivating redundant BSs.	Base Station	**Simulation** – Energy savings estimated using real-world traffic traces (anonymized) and BS location data from Manchester.

Continued on next page

Table 4.1 (continued)

Author/Year	Main Approach	Granularity	Test Type
Luís Mata et al., 2025	Introduction of the Energy Sustainability Framework (ESF) combining LightGBM and Shapley Additive Explanations (SHAP) to classify and optimize BSs' energy and spectral efficiency using live network data and root-cause analysis.	Base Station	**Simulation –** Evaluation based on live 4G/5G network data from 38 BSs (233 cells) over 37 days, using performance management/energy management metrics and ML-based classification.
Our proposed method	A hybrid expert system solution combining AI/ML-based spatial analysis with rule-based decision logic to enable sector-level energy savings without compromising QoS.	Sector	**Live Deployment –** Implemented and field-evaluated in a dense urban network with real users and traffic conditions.

A widely studied strategy involves carrier (cell) switch-off mechanisms, supported by industry standards such as those from the 3GPP [71], the ETSI [72], and the O-RAN Alliance [73]. According to the O-RAN Network Energy Savings Technical Report [73], mobile networks commonly deploy multiple carriers to serve the same area. When traffic load falls below a defined threshold, energy savings can be realized by deactivating underutilized carriers or entire cells. UEs connected to these carriers are then offloaded to other active elements by the e/gNodeB. However, making optimal switch-off decisions is complex due to the inherent trade-offs between energy savings, QoS, and coverage preservation. Kundu et al. [74] review such strategies in light of standardization efforts, focusing on architectural enhancements like full-stack acceleration and AI-RAN workload integration. Similarly, Mata et al. [75] introduce a machine learning-based ESF that classifies base

stations according to spectral and energy efficiency using real-world data. While both efforts contribute to energy-performance optimization, our work diverges by utilizing UE geolocation data to perform sector-level switch-off decisions in operational networks, without requiring architectural changes or dependence on composite metrics.

Beyond carrier-level deactivation, additional savings can be achieved through sleep mode mechanisms applied to BS components. These techniques deactivate hardware components during periods of inactivity, targeting short-duration savings on the order of milliseconds to seconds. Masoudi et al. [76] employ reinforcement learning to coordinate group-based deactivation patterns for carriers with similar idle characteristics. Other studies introduce geographical or mobility awareness: El-Amine et al. [77] and Salem et al. [78] incorporate BS positioning and UE movement to formulate multi-level fuzzy sleep modes. An earlier contribution by Conte et al. [79] proposed cell wilting and blossoming at 30-second intervals to balance energy efficiency and service availability. These methods complement longer-interval strategies such as carrier and sector-level shutdowns, which operate over durations ranging from minutes to hours.

Several works have explored longer-term carrier switch-off algorithms aimed at maximizing energy savings while minimizing service degradation. Dahal et al. [80] proposed a transmitter shutdown algorithm that dynamically adjusts power based on channel demand, in 3G deployments. López-Pérez et al. [81] presented SRCON, a hybrid expert- and ML-based approach for predicting throughput and energy use in large-scale networks, demonstrating improvements over traditional shutdown optimization methods. Beitelmal et al. [82] benchmarked 26 Cell Switch-Off (CSO) patterns, calculating SINR-based cost metrics but only validated their method in simulation. In contrast, our approach incorporates real-world geolocated UE distributions via a high-accuracy engine, enabling spatial cost analysis tailored to non-uniform UE patterns.

The complexity of energy-saving decisions increases with user granularity. Alaca et al. [83] observed that per-UE QoS constraints result in NP-hard problem formulations. They applied genetic algorithms to scale down computational require-

ments, focusing on carrier-level switch-offs in simulation environments. Han et al. [84] studied BS switch-off via cooperative communication and deterministic simulations, assuming coordination among active BSs for QoS preservation. Our approach, however, makes no such assumptions and directly uses operational data. Rather than relying on inter-BS communication, we resolve coverage gaps through spatial RSRP-based analysis of non-cosite neighbors. Similarly, Suárez et al. [85] focused on heterogeneous networks with macro/femto deployments, but did not address dynamic sector-level switch-off. Wang et al. [86] proposed RNN-based strategies for switching off micro-BSs in heterogeneous networks while preserving SINR via macro-BSs, differing from our method that dynamically shuts down macro-sector components when spatial UE distributions permit.

Studies such as Oh et al. [87] are particularly close to our work in that they use live network data to evaluate QoS-aware shutdown strategies. However, their method relies on temporal traces of call traffic and does not address geolocated events or spatial coverage analysis at the sector level. Unlike prior studies, our approach is grounded in a cost matrix that integrates power savings, spatial RSRP coverage, and QoS impact using real UE traces in dense urban environments.

In parallel, reinforcement learning has also been explored for energy optimization in next-generation networks. Tang et al. [88] developed a Q-learning-based Confidence Information Coverage (CIC) sleep scheduling algorithm for IoT networks, minimizing active nodes while maintaining required coverage. Although promising for wireless sensor networks, such methods remain limited in their application to RAN-level sector shutdown.

To the best of our knowledge, our study is the first to propose and validate a geolocation-aware, sector-level shutdown algorithm that incorporates live layer-3 trace data for real UE distribution and is evaluated on a live dense-urban LTE network. Our work introduces a hybrid decision-making approach that combines spatial analytics with operator feedback to ensure service continuity while maximizing energy savings.

4.2 Problem Definition

Having identified the pressing concerns surrounding the escalating energy consumption of mobile networks, we now formally define our proposed sector shutdown algorithm. This algorithm aims to reduce the substantial power usage attributable to base stations, which account for a significant portion of network energy consumption. By implementing intelligent shutdown strategies, the objective is to decrease energy expenditure without compromising network coverage or capacity.

In this section, we present a comprehensive mathematical formulation of the problem, taking into account the network configuration, daily traffic patterns, and the goal of maximizing energy savings while ensuring an uninterrupted customer experience. By delineating the inputs, variables, and constraints of the algorithm, we lay the groundwork for an efficient and effective solution.

4.2.1 Network Model and Notation

Let $G = (S, U)$ represent the mobile network graph, where S is the set of sectors and U is the set of users connected to these sectors. We denote

$$S = \{s_i \mid i = 1, 2, \ldots, n\}, \quad U = \{u_j \mid j = 1, 2, \ldots, m\}.$$

The service area is discretized into k grid locations,

$$L = \{l_1, l_2, \ldots, l_k\}.$$

Each user generates measurement reports capturing the RSRP from the sector to which it is connected. We focus on RSRP values because it is a fundamental metric in cellular networks (e.g., LTE/5G) for quantifying signal strength and user experience. Crucially, when a sector is shut down, users previously served by that sector must hand over to alternative sectors, and any degradation in the coverage or signal quality is readily reflected in their RSRP measurements. Hence, this metric provides a direct means to evaluate coverage conditions and service impacts under different shutdown decisions.

We gather these reports in sets denoted by

$$M_{s,l,u} = \{ r \mid r \in R,\, s \in S,\, l \in L,\, u \in U \},$$

where r are rational-valued RSRP measurements. Thus, $M_{s,l,u}$ collects all RSRP measurements reported by user u at location l when served by sector s.

4.2.2 Coverage and Service Quality

Each sector $s \in S$ has a nominal power consumption P_s when it is active. We define a binary vector $\Gamma = [\gamma_1, \gamma_2, \ldots, \gamma_n]$, where each γ_i indicates whether sector s_i is shut down ($\gamma_i = 1$) or not ($\gamma_i = 0$).

Coverage Matrix. A sector s is said to cover a location l if there exists at least one user u located at l with a valid RSRP report for that sector. Formally,

$$A_{s,l} = \begin{cases} 1, & \text{if } \exists\, u \in U : \exists\, r \in M_{s,l,u}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

Hence, $A_{s,l} = 1$ indicates that sector s can provide coverage at location l (according to the measurement reports), while $A_{s,l} = 0$ implies no such measurement exists for that pair.

Service Quality. We define the *service quality* of sector s at location l as the median RSRP value across all user measurement reports for (s, l):

$$Q_{s,l} = \text{median}\Big(\{ r \mid r \in M_{s,l,u},\, u \in U \} \Big). \tag{4.3}$$

Thus, $Q_{s,l}$ reflects the typical (median) RSRP that sector s provides.

To capture the *best* service quality available at location l among all sectors, we define

$$\Phi_l = \max_{s \in S} \{ Q_{s,l} \}. \tag{4.4}$$

Hence, Φ_l is the highest median RSRP observed among all sectors serving location l.

4.2.3 Energy Consumption and Objective

We compute the total energy consumption E of the network by summing the power of the active sectors:

$$E = \sum_{s \in S} P_s \left(1 - \gamma_s\right). \tag{4.5}$$

Here, P_s is included in the sum only if the sector s is on ($\gamma_s = 0$).

To minimize E, one may equivalently maximize the number of shutdown sectors:

$$\max \sum_{s \in S} \gamma_s. \tag{4.6}$$

4.2.4 Constraints and Final Formulation

When a sector s is shut down ($\gamma_s = 1$), coverage and service quality at previously served locations will change. Let A' and Q' denote the coverage matrix and service quality matrix after applying the shutdown decisions. Similarly, let Φ' be the best service quality vector after shutdown.

We enforce two key constraints:

1. **Maintaining coverage**: No location becomes uncovered. Thus,

$$\forall\, l \in L,\ \exists\, s \in S : A'_{s,l} > 0. \tag{4.7}$$

2. **Limiting RSRP drop**: The drop in best service quality at each location must not exceed a threshold τ. Formally,

$$\forall\, l \in L,\ \left(\Phi_l - \Phi'_l\right) \leq \tau. \tag{4.8}$$

Hence, our optimization problem is:

$$\begin{aligned} \underset{\Gamma}{\text{maximize}} \quad & \sum_{s \in S} \gamma_s \\ \text{subject to} \quad & \forall\, l \in L : \exists\, s \in S :\ A'_{s,l} > 0, \quad , \\ & \forall\, l \in L :\ \left(\Phi_l - \Phi'_l\right) \leq \tau, \end{aligned} \tag{4.9}$$

where:

- $\gamma_s = 1$ indicates sector s is shut down,
- A', Q', and Φ' represent the coverage and service quality matrices after shutdown,
- τ is a threshold specifying the acceptable drop in median RSRP for any location.

In the next section, we introduce the dataset collected from a live mobile network and describe the scenarios under which we evaluate our proposed sector shutdown algorithm. These real-world measurements ensure that our model captures practical network behaviors, traffic demands, and coverage patterns.

4.2.5 System Setup

In this study, the Kadikoy region, which is one of the densely populated areas of Istanbul, Türkiye, was chosen as the focus area [89]. The Kadikoy area is known for its high population density, which leads to high demand for mobile services during peak hours. This results in a high number of base stations deployed in the area to meet the increased capacity needs. However, during off-peak hours, this excess capacity becomes redundant and results in wasted energy consumption. This scenario highlights the need for efficient energy management solutions that can adjust to changing network demands. The Kadikoy area, with its high population density and dynamic network usage patterns, serves as an ideal testbed for evaluating the effectiveness of energy-saving techniques such as sector shutdown in real-world environments.

The study uses real-world network data including measurement reports, which the users generate during communication with the core network. These reports contain information about the connected cell and its surrounding neighbour cells. Specifically, each report includes the name of the connected cell, the RSRP of the connected cell, as well as the names and RSRP values of the neighbouring cells. This information is crucial for understanding the quality of service and network coverage in the area and these can be used to make decisions about network optimization.

The geolocated data is identified by using Uber's H3 hierarchical geospatial indexing grid system [62]. It allows the partitioning of a sphere into a hierarchical grid of hexagons, providing a highly efficient and uniform representation of the earth's surface. This makes it an ideal choice for a wide range of applications, including spatial data analysis, location-based services, and mapping applications. The measurement reports are grouped into grids by using the H3-index to analyse which sectors serve which grids. Also, we identify the alternative sectors for each sector by using these data. The selected region (Kadikoy/Istanbul) is presented in Fig. 4.1, and the locations of user measurements are illustrated with blue dots. This region is a dense area where the users are 7/24 active and produce a high amount of network traffic.

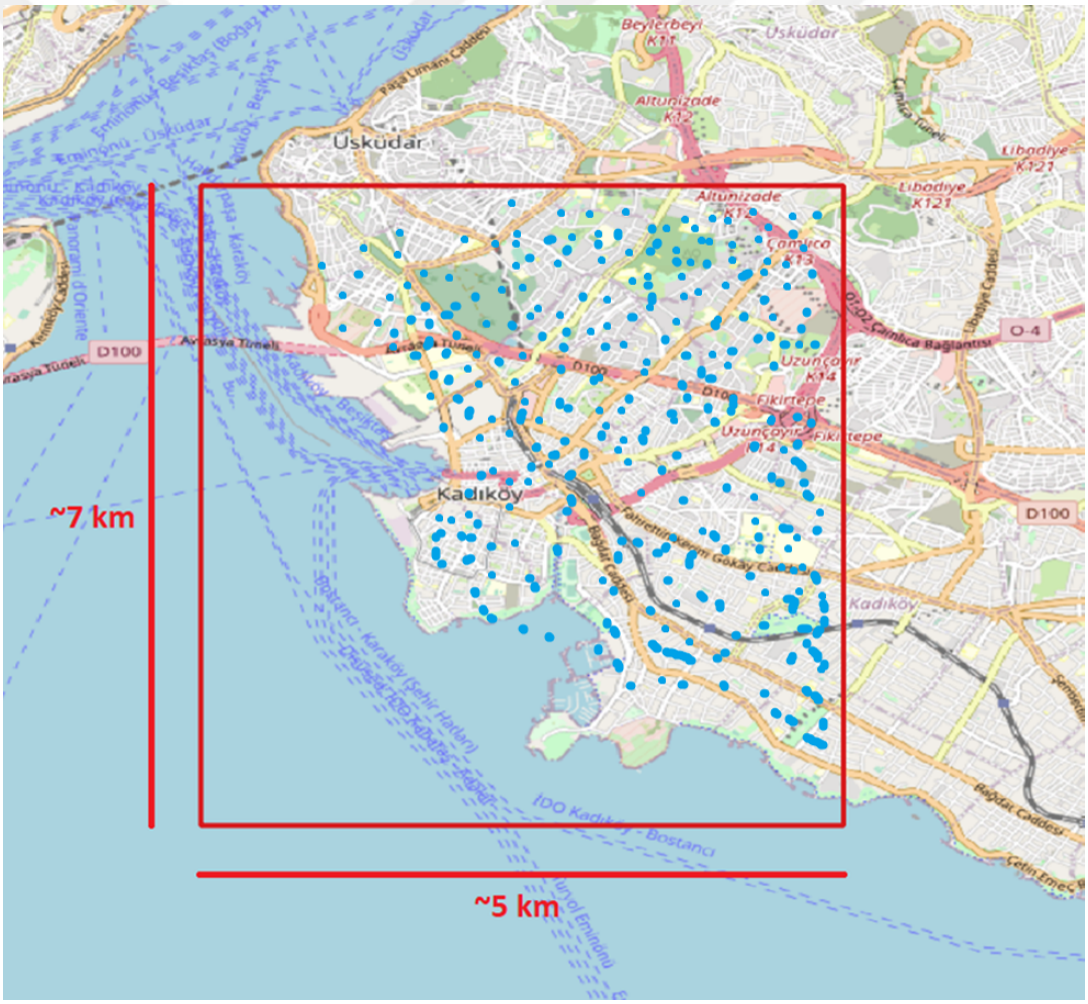


Figure 4.1: Illustration of the selected region. The blue dots represent the locations of measurement reports created by users.

4.2.6 Scenario 1: Offline Network Trial

Our first scenario involves using historical data collected from a live network to analyze the model's performance and identify potential problems through offline testing. We called this scenario "Offline Network Trial with Real Network Data". In this scenario, we compared two solutions: GAS and CAHGS.

We aimed to determine sectors' coverages by using measurement reports. After determining each sector's coverage, we define the compensator sectors which means that alternative sectors that may give the service if the sector is shut off for each sector.

The objective function for this problem is defined in Eq. 4.9. That aims to maximize the number of shutdown sector counts without leading to a coverage gap. Also, maintaining service quality in terms of RSRP value is also ensured in that objective.

We group measurement reports into grids. After this grouping, we form a coverage matrix (Eq. 4.2). Each element of the matrix represents whether a sector (s) servicing a grid (l) or not. By using the coverage matrix, we analyze the coverage areas of the sectors and identify overlaps or gaps in coverage. This information is crucial for determining which sectors can be shut down without affecting network coverage and performance. In addition to that, servicing quality matrix (Eq. 4.3) is formed by using RSRP values of the measurement reports. By using the servicing quality matrix, we forecast the servicing quality after shutting off sectors.

In this scenario, constraints are defined according to grid-based grouped information. The servicing qualities and compensator sectors are defined according to that. However, it should be noted that this method is optimistic in nature. It assumes that if there is a group of users located in the center of a grid and connected to a specific sector, that sector is considered to be servicing the entire grid. However, this assumption does not take into account edge cases where the users are located in the corners of the grid and they may not have the same level of service as those in the center. This is why, this method is considered optimistic and it is important to keep this limitation in mind when interpreting the results and making decisions based on the analysis.

4.2.7 Scenario 2: Live Network Trial

In this scenario, the proposed algorithm is tested on a live mobile network. This scenario is referred to as a Live Network Trial. This type of trial presents a number of unique challenges and constraints that must be taken into consideration. One of the main challenges of a live network trial is that it is difficult to predict how the algorithm will perform in a real-world environment. The algorithm may perform well in a simulation environment, but it may not perform as well on the live network. Due to the network conditions or user behaviour, there may be unexpected variations. Therefore, we define a second scenario that increases the constraints and avoids aggressive actions.

The objective of this scenario is the same as in scenario 1: to maximize the number of shutdown sectors while maintaining network coverage and service quality without leading coverage gaps.

Unlike the previous scenario, in this scenario, compensator sectors are defined according to the neighbour sector information that is stated in the measurement reports. In this scenario, the compensator sectors are determined based on the neighbour sector information included in the measurement reports provided by the users. While users create a measurement report about the sector they are connected to, they specify other sectors that can be connectable from their location. Formally, for a sector $s \in S$ servicing grid $l \in L$, let $C_{s,l} = \{s_i \mid s_i \in S, s_i \neq s\}$ be the set of alternative sectors of s for grid l according to users' measurement reports. The size of $C_{s,l}$ varies based on the number of alternative sectors available for sector s for grid l. Hence, the objective is after turning off a sector s, there exists at least one $s' \in C_{s,l}$ for all l that s is already serving before shutting down.

Based on the user experiences, as we group the measurement reports based on grids, we keep track of which sectors serve which grids and which sectors are alternatives to each other. This information is used to determine the compensator sectors when shutting down sectors. This approach allows us to take into account real-world variations in network conditions and user behaviour, providing a more accurate picture of the network coverage and performance.

Overall, live network trials are a valuable tool for evaluating the performance of an algorithm in a real-world environment but they also present a number of challenges and constraints that must be taken into consideration to ensure the safety and integrity of the network and the satisfaction. The potential negative effects on customer experience should be considered before testing the algorithm on the live network.

4.3 Methodology

In this section, we provide the details of our methodologies for both an offline network scenario and a live network scenario. In the offline network scenario, we utilized historical network data to develop and refine our algorithms. In the live network scenario, we work with real-time network data, which presented new challenges and opportunities. We develop effective algorithms for both offline and live network environments, thereby providing a comprehensive solution for network analysis and management.

4.3.1 Scenario 1: Offline Network Trial

4.3.1.1 Genetic Algorithm Solution (GAS)

A genetic algorithm (GA) is one of the evolutionary algorithms that is based on natural selection [90]. GA is generally used for optimization and search problems that require a heuristic approach. GA is widely used in multi-objective problems to find optimal solutions that satisfy the given conditions [91, 92]. GA is a metaheuristic approach that converges to the solution, although it does not always guarantee the optimum result. The sector turn-off problem is a multi-objective high-dimensional optimization problem. That problem requires maintaining the QoS and maximizing the number of turned-off sectors. The number of solutions exponentially scales with the number of sectors. Sectors can either be running or sleeping, therefore, If there are n sectors in the network, then there are 2^n possible combinations of sector states that need to be evaluated. Hence, the genetic

algorithm, which is a heuristic approach, fits this problem. The GA has mainly 4 phases these are the initialization of a population, fitness function, selection and reproduction. These phases are illustrated in Fig 4.2.

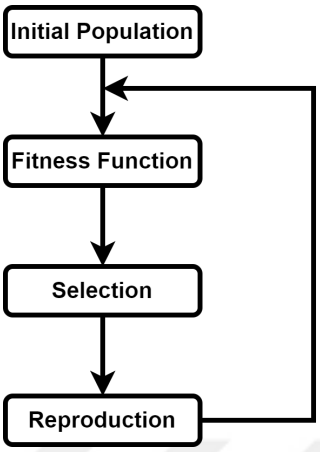


Figure 4.2: Genetic algorithm phases.

The population (set of solutions) consists of individuals (solution). Each solution is represented as a binary vector. Each element of the vector (gen) indicates whether the sector is turned-off or running. Therefore, a solution is a binary decision vector (Γ), and each element of this vector ($\gamma_i \in \Gamma$). represents whether sector (S_i) is running ($\gamma_i = 0$) or not ($\gamma_i = 1$).

The **initial population** is populated with the following methods:

- **Random solutions:** Randomly generated solutions.
- **Untouched solutions:** These solutions are assuming all sectors are running ($\gamma_i = 0$).
- **Randomly single mutated:** For each solution, there is only one sector that is turned off.

The **fitness function** is aimed to measure the suitability of the individuals that are in the population. The sector turn-off problem is a multi-objective problem that requires satisfying the following criteria: QoS, coverage and maximizing the number of turned-off sectors. QoS constraint refers to minimizing or limiting the signal service quality of the users. To satisfy that, we limit the negative effects of the so-

lution on the RSRP of the UE. If the negative effect is higher than the threshold, then we eliminate that solution. Another constraint is coverage. It means that if a solution leads to a coverage gap, then we eliminate that solution. The final objective is to maximize the number of turned-off sectors. If a solution satisfies both constraints (QoS and coverage) the fitness function returns the number of turned-off sectors. The constraints are validated by using the geolocated data. To validate each solution, we remove the turned-off sectors from the original data and report the rest of the sectors servicing quality. The coverage constraint is ensured by comparing the coverage of original data and solution results. If there is a coverage gap after applying a solution, we omit that solution. The QoS constraint is ensured by using the RSRP values of each region. For each region, the difference in the RSRP values between the original data and solution one must be within the acceptable limit.

The objective of the problem, as defined in Eq. 4.9, serves as the foundation for formulating the fitness function to evaluate potential solutions. This fitness function is instrumental in guiding the optimization process towards achieving the desired objectives while adhering to constraints. The fitness function is formulated as follows:

$$\text{Fitness} = \begin{cases} \sum_{i<|\Gamma|} \gamma_i, & \text{if } \forall l \in L, \exists s \in S : A'_{s,l} > 0 \\ & \quad \text{and } \forall l \in L : (\Phi_l - \Phi'_l) \leq \tau \\ 0, & \text{otherwise,} \end{cases} \tag{4.10}$$

where $A^{'}$ and $\Phi^{'}$ represent the situation after sectors shutdown, and τ is the threshold for an acceptable drop in the best RSRP.

Following the calculation of fitness scores for each **solution** within the population, a selection method is applied to filter out individuals that do not satisfy quality of service and coverage constraints. The remaining individuals are then sorted based on the number of turned-off sector counts. Subsequently, the top k individuals are identified as candidates for the optimal solution within the population.

The purpose of **reproduction** is to enrich the solution with new individuals. The reproduction phase is inspired by well-known two biological phenomena that are mutation and crossover, as illustrated in Fig 4.3. The **mutation** is a flipping of one

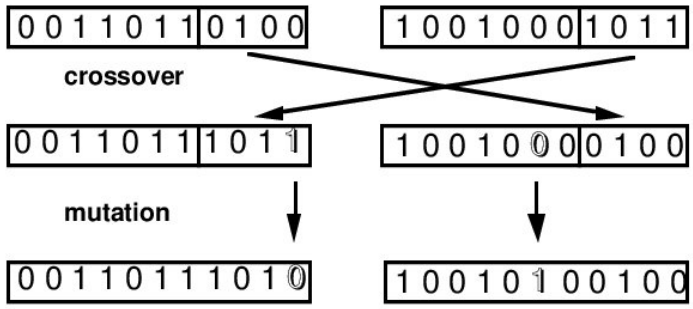


Figure 4.3: Illustration of a mutation and a crossover (adapted from [2]).

or multiple elements of the given solution. In the mutation phase, firstly, we select an individual from the population. After that, we mutate a selected solution to explore the new configurations. The mutated gene is also randomly selected. The mutation may flip a turned-off sector to running or a running sector to turned-off. The **crossover** phase requires two randomly selected individuals. These individuals' genes are crossed over to generate a new instance. The crossing-over point is randomly selected.

The flow of the GA starts with the initialization of a population. The fitness score of each individual in the population is calculated and the top k individuals are selected. After this selection, new individuals are reproduced. That iteration continues with the calculating fitness scores of the new population. In our problem, the termination criterion is the fixed number of iterations. When the iterations are over, the individual with the highest fitness score is the optimum solution found with GA.

4.3.1.2 Cost-Aware Heuristic Greedy Shutdown (CAHGS)

The Cost-Aware Heuristic Greedy Shutdown (CAHGS) algorithm is designed to reduce energy consumption in mobile networks by selectively deactivating underutilized sectors. It relies on predefined, rule-based criteria that evaluate both the cost of shutting down each sector and the availability of alternative (compensator) sectors capable of maintaining coverage. Leveraging geolocated data through the H3 spatial indexing system, CAHGS identifies the grids served by each sector and

systematically assesses whether neighboring sectors can adequately compensate for the coverage, ensuring service continuity throughout the shutdown process.

Algorithm 1 The CAHGS algorithm.

Input : geoLocatedData

Output: safelyTurnOffSectorList

1: $allSectors \leftarrow getAllSectorsFromData()$
2: $allSectorsWithCost \leftarrow calculateCosts(allSectors)$
3: **for** s ∈ sectorsWithCost **do**
4: $isSafeToTurnOff \leftarrow checkSafety(s)$
5: **if** isSafeToTurnOff **then**
6: Add this $sector$ to $safelyTurnOffSectorList$
7: **end if**
8: **end for**

The abstract-level formulation of the CAHGS algorithm is illustrated in Algorithm 1. The process begins by assigning each sector a *turn-off cost*, computed from three key factors: (i) the number of grids for which the sector is the primary server, (ii) its total coverage footprint, and (iii) the presence and capacity of compensator sectors that can maintain coverage in its absence. These metrics are normalized and combined through a weighted summation to yield a single scalar cost value per sector. The sectors are then sorted in ascending order of cost, prioritizing those whose deactivation is expected to have minimal impact on coverage. Following this prioritization, the algorithm evaluates each sector individually by examining its constituent grids. For each grid, it verifies whether adequate compensation is provided by neighboring sectors. If all grids pass this coverage check, the sector's contribution is removed from the coverage matrix, and the sector is marked for shutdown. If any grid is left without sufficient coverage, the sector is excluded from the shutdown list. This evaluation proceeds iteratively for all candidate sectors. While the ordering influences the shutdown sequence, the grid-level coverage validation step guarantees that network integrity is maintained, enabling energy savings without compromising service availability.

The CAHGS algorithm is heuristic in nature, relying on heuristic-greedy safety checks and cost functions rather than optimization through learning. It is also greedy,

as it selects sectors based solely on immediate cost-benefit analysis, without re-optimization once a decision is made. This balance between efficiency and caution enables CAHGS to achieve meaningful energy savings while maintaining robust user coverage.

4.3.2 Scenario 2: Live Network Trial

4.3.2.1 Neighbour-Based Expert System Solution (NBESS)

The Neighbour-Based Expert System solution is a method for optimizing energy consumption in mobile networks by identifying underutilized sectors that can be safely shut down. This approach is based on analyzing geolocated user measurements and considering each cell (or carrier) within a sector individually. By examining each cell's compensating elements and determining whether they can provide adequate coverage in the event of a shutdown, this method is able to identify sectors that can be shut down without negatively impacting network coverage or user experience. The algorithm used in this solution is iterative and evaluates each cell in a sector to determine if it can be safely shut off, ultimately leading to a list of candidate sectors that can be safely shut down for energy savings.

Algorithm 2 outlines the proposed NBESS method. In this context, a *cell* refers to a carrier frequency band (i.e., servicing frequency) associated with a sector. A single sector may contain multiple such cells. When a sector is turned off, all of its constituent cells are simultaneously deactivated. To identify candidate sectors for shutdown, the algorithm evaluates each individual cell to ensure it is fully *compensated*—that is, in every H3 grid it serves, there must exist at least one neighboring cell that can maintain adequate coverage. Only when all cells of a sector are verified to be compensated does the algorithm consider that sector eligible for shutdown. The variables and functions used in this algorithm are summarized in Table 4.2.

In this work, we utilize offline batch data to determine which sectors can be safely turned off during designated rest periods. Specifically, the objective is to identify sectors suitable for shutdown during rest hours without the need for real-time analysis. The algorithm is designed to process the batch data efficiently, ensur-

Algorithm 2 Neighbour-Based Expert Sector Shutdown (NBESS).

Input : D {Geo-located measurement data}

Output: Γ {Set of sectors to turn off}

$\Gamma \leftarrow \emptyset$
while true do
  $C \leftarrow \emptyset$ {Compensated cells}
  $S_{\text{cand}} \leftarrow \emptyset$ {Candidate sectors}
  for each $c \in D$ **do**
    if ISCOMP(c) **then**
      $C \leftarrow C \cup \{c\}$
    end if
  end for
  for each $s \in D$ **do**
    $C_s \leftarrow$ CELLS(s)
    if ISALLCOMP(C, C_s) **then**
      $S_{\text{cand}} \leftarrow S_{\text{cand}} \cup \{s\}$
    end if
  end for
  if ISEMPTY(S_{cand}) **then**
    break
  end if
  $s^* \leftarrow$ MINCOV(S_{cand})
  $\Gamma \leftarrow \Gamma \cup \{s^*\}$
  UPDATE(D, s^*)
end while
return Γ

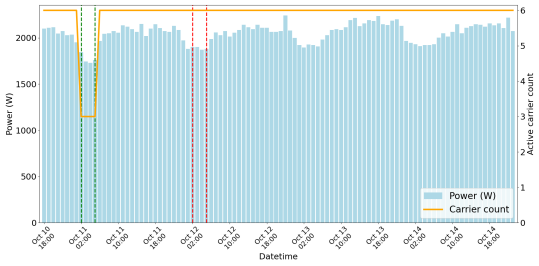
ing timely decision-making while accommodating the size and complexity of the network. After identifying and simulating the shutting down of a single sector, the entire dataset is updated by removing the deactivated sector and its corresponding cells. The algorithm is then re-executed using the updated dataset. This process is repeated iteratively, each time identifying and safely shutting down another sector until all possible energy savings have been achieved.

Table 4.2: Variables and functions used in the NBESS algorithm.

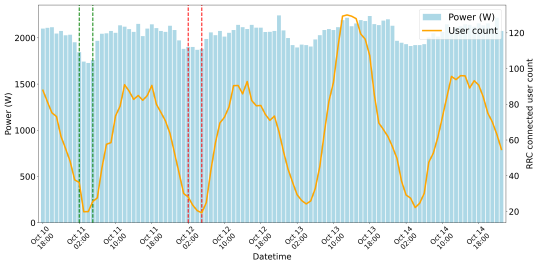
Variable	Meaning
D	Geo-located measurement data, including cell-level user reports.
Γ	Set of sectors that are selected for shutdown.
C	Set of cells that are fully compensated by neighboring cells.
S_{cand}	Set of candidate sectors eligible for shutdown.
c	A single cell element from the dataset.
s	A sector element from the dataset.
C_s	Set of cells that belong to sector s.
s^*	The sector selected for shutdown in the current iteration.
`isComp(`c`)`	Function that checks whether all grids served by cell c are compensated.
`cells(`s`)`	Function that retrieves all cells belonging to sector s.
`isAllComp(`C, C_s`)`	Function that checks whether all cells in C_s are included in C.
`isEmpty(`S_{cand}`)`	Function that checks if the candidate set is empty.
`minCov(`S_{cand}`)`	Function that selects the sector with the smallest coverage area.
`update(`D, s^*`)`	Function that updates dataset D to reflect the shutdown of sector s^*.

4.4 Results and Evaluation

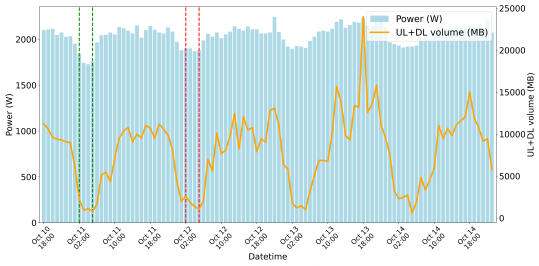
In this section, we present our comprehensive results and evaluations, structured to provide a holistic understanding of our study's impact and insights. Initially, we introduce the base station's energy consumption comparing with active carrier count, connected user count and traffic volume. Following this, we delve into two distinct experimental scenarios designed to evaluate our hypotheses and the practical effectiveness of our proposed solutions. The first scenario, Offline Network



(a) Power vs active carrier count.



(b) Power vs connected user count.



(c) Power vs sum of uplink and downlink volume.

Figure 4.4: Base station enegy consumption model.

Trial with Real Network Data, provides a controlled environment to examine the theoretical and simulation-based aspects of our study, leveraging historical network data to gauge performance and identify potential optimizations.

The second scenario, Live Network Trial, transitions from theoretical exploration to real-world application, presenting the outcomes of deploying our strategies in a live network setting. This scenario offers invaluable insights into the practical challenges and benefits of implementing our proposed methodologies in operational environments, highlighting their scalability, adaptability, and impact on network efficiency and sustainability.

4.4.1 Base Station Energy Consumption Model

To observe the energy consumption of a base station, we gathered data on the power consumption of a single base station. This experiment is centered on a base station located in Kadıköy, İstanbul, which is one of the most densely populated areas. Fig. 4.4 presents the comprehensive power model of the station, showcasing

its hourly energy consumption in watts (W). The power consumption is compared with the active carrier count, the connected user count, and the total traffic in Figures 4.4a, 4.4b, and 4.4c, respectively.

To examine the effect of the number of active carriers on energy consumption, we intentionally deactivated a single carrier. Fig. 4.4a presents the active carrier count vs power consumption of a base station. Between the green lines, the number of active carriers is reduced from five to four. Subsequently, we compared the energy consumption for the following day, demarcated by the red lines, during the same time interval as the previous day. It is evident that although there is a reduction in energy consumption, the impact of the carrier shutdown on the overall energy consumption of the base station is limited.

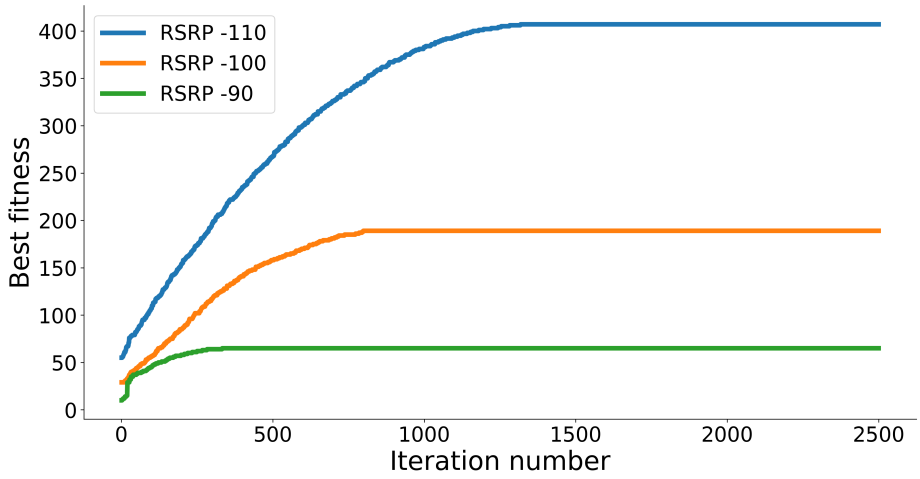
The second observation is gained from Fig. 4.4b and Fig. 4.4c, periods of high connected user count and increased total traffic correspond with higher energy consumption at base stations, indicating a relationship between user engagement and energy usage.

In summary, our investigation into a base station's energy consumption, focusing on a high-density location in Kadıköy, İstanbul, reveals that while reducing active carriers slightly decreases energy usage, this effect is limited. Data indicates a more substantial relationship between energy consumption and periods of high user count and traffic.

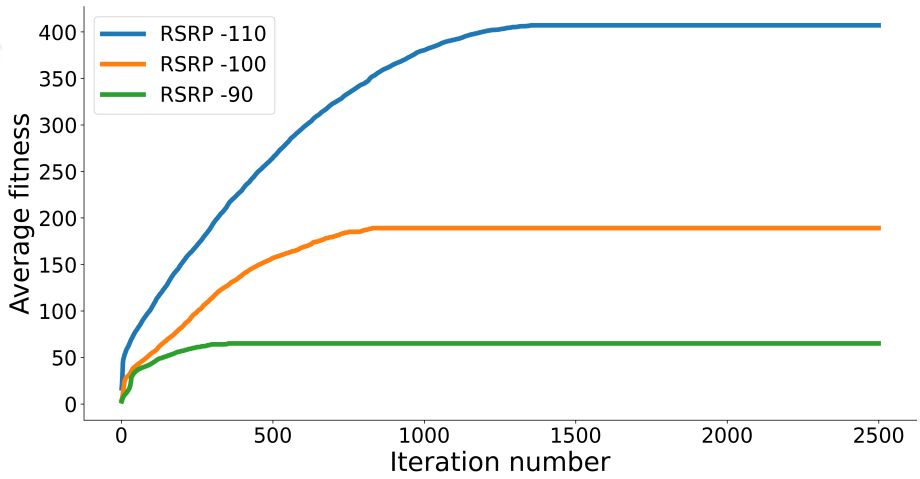
4.4.2 Scenario 1- Offline Network Trial with Real Network Data

In scenario 1, we evaluate our algorithms by using historical real network data. We simulate the results by using real network data to analyse our algorithms. Both the GAS and CAHGS are considered. We evaluate our algorithms to investigate how many sectors can be shut down without violating key coverage/quality constraints.

In Fig. 4.5, the results of testing different RSRP constraints using the GAS are presented. The trade-off between energy savings and QoS is evident in the results. The goal of this analysis is to maximize the number of turned-off sectors while maintaining an acceptable level of QoS. The RSRP constraint is implemented to



(a) Best Fitness.



(b) Average fitness.

Figure 4.5: The GAS results are presented with various RSRP thresholds.

ensure that the solution does not negatively impact the QoS by exceeding a specified threshold. It can be observed that as the RSRP constraint is relaxed, more sectors can be turned off. As expected, the final iteration shows that the highest number of turned-off sectors was achieved with the -110 RSRP limit, followed by -100 and -90. At the final iteration, the scores are 421, 196 and 74 which belong to -110, -100 and -90 RSRP limits, respectively. In parallel, the CAHGS algorithm yields 407, 189, and 65 sectors for the same three RSRP limits, respectively.

Fig. 4.5a presents the best fitness, and Fig. 4.5b shows the average fitness of the population across iterations of the GAS. The best fitness scores converge closely

with the average fitness, indicating that the entire population evolves toward highly similar solutions over time. After reaching this steady state, no further improvements are observed, signifying the GAS's convergence.

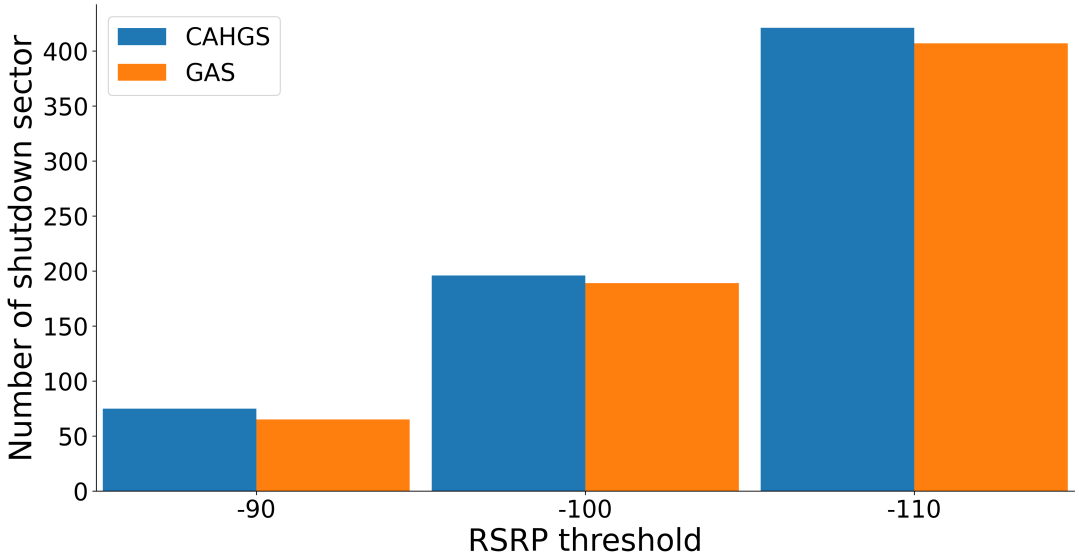


Figure 4.6: Comparision of the GAS and the CAHGS.

Fig. 4.6 compares the GAS and the CAHGS algorithms in terms of how many sectors each method can shut down. Both algorithms achieve comparable outcomes, and they also identify a highly similar set of sectors to turn off. Specifically, there is a 96% intersection between the solutions, suggesting that both approaches consistently prioritize the same sectors for shutdown. These results demonstrate that the GAS can achieve tangible energy reductions while respecting RSRP constraints, and that CAHGS, although less flexible, can also yield effective shutdown strategies.

In summary, our experiments on real historical network data (Scenario 1) demonstrate that by relaxing RSRP thresholds, operators can realize substantial energy savings with only marginal reductions in service quality. Through the GAS and the CAHGS, we observed high agreement in which sectors could be safely shut down, suggesting the robustness of both methods. These offline trials serve as a critical step in validating the feasibility and impact of sector shutdown on energy consumption, while maintaining acceptable QoS. Building on these findings, we move next to a live network trial (Scenario 2) to test a refined version of our approach

under real-time operational conditions, with a focus on preventing any negative impact on customer experience.

Table 4.3: 104 sites energy consumption table for between 2 AM - 6 AM and network-wide estimation.

	Energy Consumption (W)	**Energy Gain** (W)	**Gain Percentage**	**Annual Gain** (GWh)	**Estimated Whole Network Annual Gain (GWh)**
Feature On	932942	58728	5.9%	0.0214	6.18
Feature On + Model	**927305**	**64635**	**6.5%**	**0.0235**	**6.78**
No saving	991671	-	-	-	-

4.4.3 Scenario 2 - Live Network Trial

This scenario is aimed to live network testing. Therefore, we avoid aggressive actions that may negatively affect the QoS or customer experience. The NBESS which is an updated version of the CAHGS is tested on the live network.

In the following parts, the **feature** refers to the vendor sleeping feature, and the **model** refers to the Neighbour-Based Expert System Solution. The energy-saving model decisions were applied on 11 October while the vendors' sleeping feature is enabled.

Our algorithm is designed to complement the vendor energy-saving feature (Ericsson LTE Cell Sleep Feature). Therefore, we compare our algorithm against three scenarios: i) only the vendor feature, ii) our improved energy-saving algorithm, and iii) without any energy-saving feature. To minimize the adverse effects on customer experience, we scheduled the execution of our algorithms from 2 AM to 6 AM. Specifically, on October 11th and 12th, we ran only the vendor's energy-saving features. Our improved algorithm was deployed on October 11th. To assess network energy consumption without any energy-saving measures, we deactivated all energy-saving features from October 13th to October 17th.

Table 4.3 presents a network-wide comparison of three scenarios: vendor feature only, vendor feature combined with our model, and no energy-saving feature. We evaluated the aggregate energy-saving impact across all 104 sites during the same

time window. The vendor feature alone provided a 5.9% energy gain, while the integration of our NBESS model increased this to 6.5%. This improvement corresponds to an estimated additional annual saving of 0.62 GWh, raising the total from 6.18 GWh to 6.78 GWh. These results highlight the scalability and effectiveness of our model in complementing existing vendor solutions at the network level.

Table 4.4: Direct effect of the model on energy consumption by presenting 8 sites where our model takes an action. Energy consumptions between 2 AM - 6 AM.

	Energy Consumption (W)	**Energy Gain (W)**	**Gain Percentage**
Feature On	64529	1104	1.7%
Feature On + Model	**62640**	**2994**	**4.6%**
Feature Off (No saving)	65634	-	-

Table 4.4, presents the comparison of the energy saving results by considering the sites that are action taken. There are 104 sites, but our model only turn-off sectors on the 8 sites. This table shows the direct effect on sites the energy consumption that turn-off action taken. It is obvious that without any energy saving, the network consumes the most energy. In addition to that, our improved model enriches the vendors' energy-saving feature from 1.7% to 4.6% on these sites. During the test, the energy gain of the vendor feature and our improved model is 1104 W and 2994 W, respectively.

The daily energy consumptions of these 8 sites are presented in Fig. 4.7. The vendor feature was activated on 10 October and 12 October. Our improved model is activated on 11 October, and from 13 October to 17 October we deactivate all energy-saving features. There is a significant reduction in energy consumption on 11 October when our improved model ran. On the days when energy saving does not work (13 Oct. - 17 Oct.) due to external factors such as temperature, traffic, etc., there are small fluctuations in energy consumption are seen in the figure, as expected. The average energy consumption on these days is much higher than the days that energy saving is activated.

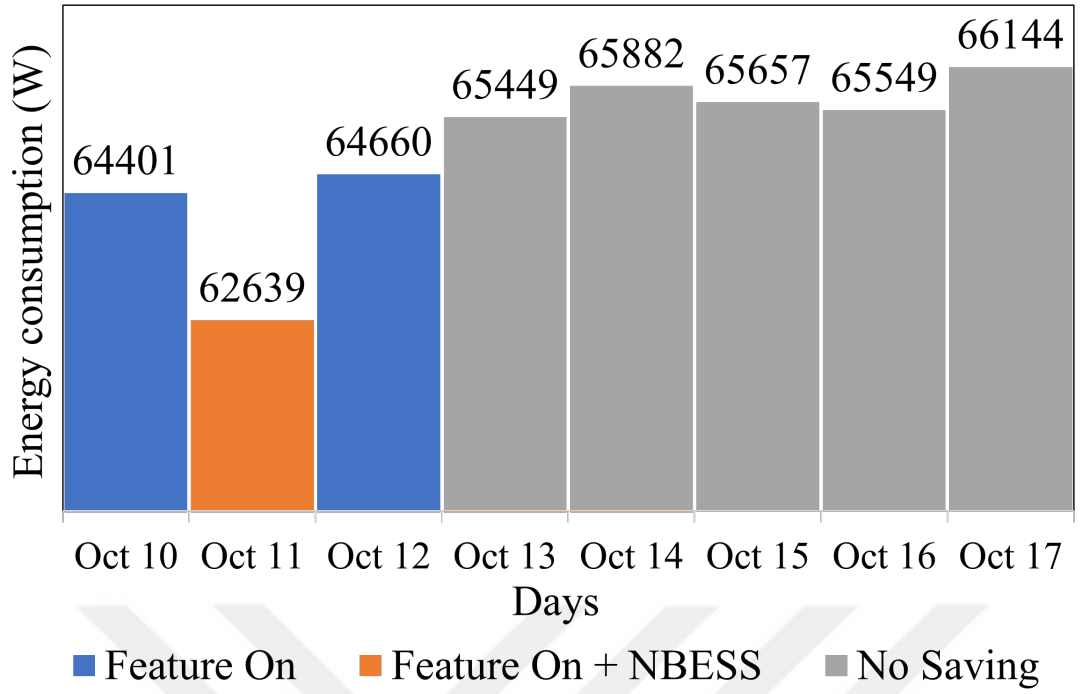


Figure 4.7: Energy consumption between 2 AM - 6 AM of 8 sites that our model takes an action.

Customer Experience Evaluation and Coverage Gap Analysis

In live network trials, the potential impact on customer experience is a critical concern for mobile operators. To validate the safety and effectiveness of our sector shutdown model in this context, we conducted a analysis using geolocated user data. One major risk associated with sector deactivation is the emergence of unintended coverage gaps, particularly in densely populated regions. Such gaps could prevent users from connecting to the network, thereby negatively impacting service continuity.

To evaluate this risk, we analyzed the number of connected users in the test region before, during, and after the sector shutdown period. Our objective was to detect any abnormal activity or significant reduction in user counts that may indicate a coverage issue. The results are presented in Fig. 4.8, which depicts the unique IMSI count for three different time slots across several days. The vendor and model-based sector shutdown operations were executed only on 11 October, and this date is used as the target for comparative analysis.

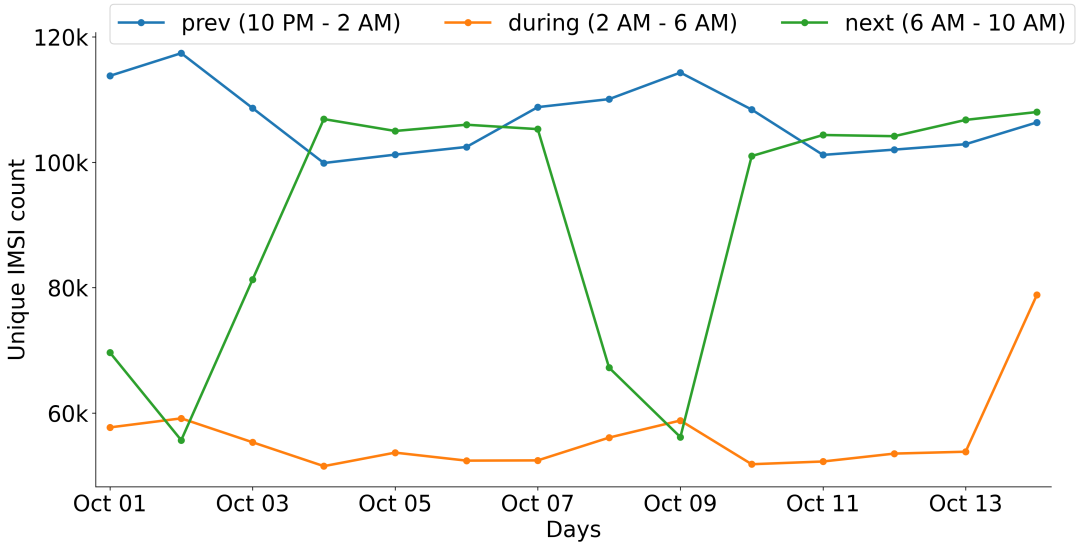


Figure 4.8: Daily unique IMSI count at the work region; blue, red, and green lines represent 10 PM–2 AM, 2 AM–6 AM, and 6 AM–10 AM time slots, respectively.

As shown in Fig. 4.8, during the critical time window (2 AM – 6 AM), the number of connected users on 11 October is consistent with other days, indicating that no noticeable drop occurred due to the sector shutdown. The early night (10 PM – 2 AM) and early morning (6 AM – 10 AM) time frames, which typically exhibit greater variation due to human mobility patterns, also do not show any anomalies. These findings support the conclusion that our algorithm did not induce a coverage gap or affect overall network accessibility.

To further refine our evaluation, we introduced the concept of *residents*—users who regularly connect to the network in the region. We define a resident as an IMSI observed in the area for more than a predefined threshold over the past 14 days. For instance, with a threshold of 0.5, an IMSI must have appeared on at least 7 different days to be classified as a resident.

Fig. 4.9 shows the number of *missing residents*—those who are typically present but were absent on 11 October. The distribution across different thresholds reveals no abnormal increase on the target day. This reinforces our earlier findings that no coverage gap was introduced, as regular users continued to access the network normally.

The results presented in this part demonstrate that the proposed sector shutdown

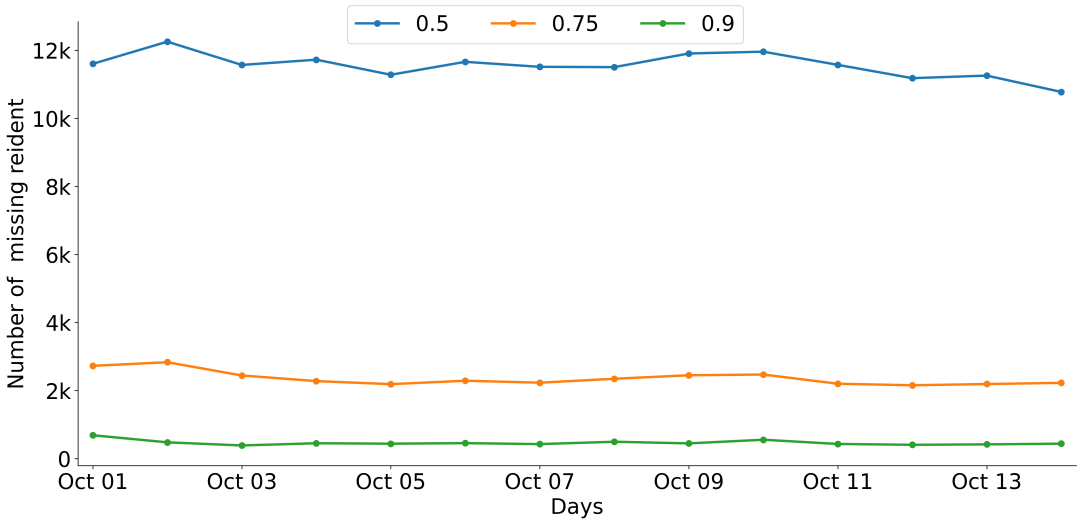


Figure 4.9: The number of missing residents is presented for various thresholds. A missing resident is defined as an IMSI that is normally observed frequently in the region but was not observed on the target date.

algorithm does not introduce any coverage gaps. Both the unique user count analysis and the evaluation of missing residents indicate consistent network accessibility during and after the shutdown operations. These findings confirm that the method can be safely deployed in dense urban areas without compromising service continuity, thus validating its applicability in live network environments as a energy-saving solution.

4.5 Conclusions

Energy saving is an important aspect for mobile operators. While reducing energy consumption, it is crucial for mobile operators to maintain a high level of customer satisfaction. In this study, we propose a sector shutdown algorithm to improve energy gain. The algorithm is designed to shut down entire sections of the sites, including the BBU and cooling systems and other associated components, hence it offers more significant energy savings than traditional carrier shutdown methods. The theoretical approaches using the CAHGS and the GAS showed that approximately 35% of energy could be saved by shutting down 190 out of 500 sectors. However, when applied to a live network, the proposed method was found to be too aggressive and may have a negative impact on customer experience. Therefore,

we proposed a less aggressive approach, the NBESS, which improved the existing solutions by up to 10% without adverse effects on customer experience. The live network trial results showed that the overall network energy saving by using this method was approximately 6.78 GWh. These results demonstrate the potential for significant energy savings in mobile networks while maintaining a high level of customer satisfaction.

The scalability of the proposed sector shutdown algorithm depends on network topology and deployment density. Since the method involves turning off entire sectors, the primary challenge is coverage gaps, especially in sparsely deployed areas where sector redundancy is low. The algorithm is best suited for dense urban regions, where overlapping coverage allows seamless user redistribution without service degradation. In low-density deployments, additional coverage-aware optimizations may be required to prevent isolated coverage holes and maintain network stability.

With the advent of 5G and beyond technologies, mobile networks are expected to become more complex and sophisticated. Data-driven approaches will play a critical role to understand this complex structure of future live networks. Additionally, the use of machine learning techniques could be explored to develop predictive models that adjust energy consumption according to changes in network traffic patterns. Overall, this study provides a promising solution for reducing energy consumption in mobile networks.

CHAPTER 5

CONCLUSION

In the digital era, data has often been described as the "new oil," signifying its immense value as a resource that fuels innovation and economic growth. However, much like crude oil, raw data holds little utility until it is refined into actionable insights and knowledge. Within mobile networks, the vast and continuously growing data streams present an opportunity to transform network management from static, reactive operations into systems imbued with intelligence and foresight. AI stands at the forefront of this revolution, offering tools to elevate network management along the Data–Information–Knowledge–Wisdom hierarchy. The vision for future mobile networks, particularly as we look toward AI-native 6G systems, is not only to deliver unprecedented speed and capacity but to cultivate networks that are context-aware, predictive, and capable of autonomously optimizing themselves in real time. This thesis has been dedicated to advancing this vision by addressing critical challenges in reliability, mobility analysis, and sustainable operations in 5G and beyond mobile networks.

This thesis has advanced the state of AI-driven mobile networking by contributing to three critical areas: reliability, mobility insights, and sustainable operations. The first major focus was on enhancing network reliability through proactive Radio Link Failure (RLF) prediction. As mobile networks evolve to use higher frequency bands, radio links become increasingly vulnerable to disturbances caused by environmental factors such as adverse weather. Traditionally, operators have relied on reactive approaches, intervening only after failures have occurred, resulting in service disruptions and elevated operational costs. In response, this thesis proposed an AI-based RLF prediction framework featuring a novel LSTM+ architecture that

jointly processes time-dependent inputs, including weather measurements and radio link KPIs, alongside static configuration parameters. This innovative model demonstrated superior predictive performance over conventional methods and enables proactive network interventions to minimize disruptions, ensuring more reliable connectivity under adverse conditions. Yet, while effective, the reliance on complex neural architectures introduces challenges around interpretability, as the black-box nature of deep learning can obscure the rationale behind predictions. Future research should therefore explore the integration of explainable AI techniques to clarify the decision-making processes of such models, as well as investigate hybrid approaches that combine deep learning with classical time series analysis for enhanced robustness and insight. In addition, generative models could be employed to suggest optimal link configurations under extreme weather conditions, further empowering proactive network management.

The second contribution of this thesis focused on understanding and modeling human mobility patterns using mobile network data. This capability is essential not only for network optimization but also for broader applications in urban planning, disaster response, and public safety. Capturing and predicting population movements is inherently complex, especially during dynamic or disruptive events where traditional data sources may be unavailable or unreliable. This research developed an end-to-end AI framework that transforms geolocated network measurements into grid-based spatial representations, computes neighbor-based properties, and employs inter-city similarity assessments to enable transfer learning for regions with limited data availability. Through techniques such as grid-based spatial aggregation and scalable data processing, the study demonstrated that anonymized network metrics can be converted into actionable mobility insights. These insights are broadly useful for improving network planning and informing public response strategies, underscoring the value of mobile infrastructure as a powerful sensor for population movements. Future work should focus on integrating multi-modal data sources—such as transportation systems, social media activity, or satellite imagery—to enrich the models' context and resilience. Furthermore, there is a critical need for privacy-preserving techniques to enable the extraction of rich mobility insights while safeguarding user confidentiality and

adhering to evolving data protection regulations.

The third research focus of this thesis was on sustainability in mobile networks. As networks become denser and more services are introduced, energy consumption rises, creating both environmental and financial challenges for operators. Traditional energy-saving methods can reduce coverage or service quality, making operators cautious about using them too aggressively. This thesis addressed this problem by developing and testing AI-based strategies to save energy. To identify base stations that could be temporarily shut down due to low utilization, the study adopted methods such as genetic algorithms and expert systems. These approaches helped save energy while keeping network services running smoothly. The research was tested both in offline simulations and in live network trials carried out on the Turkcell network, demonstrating improved energy savings without harming user experience. Future research should explore advanced methods like reinforcement learning and multi-agent systems, which can adjust energy-saving actions in real time based on changing traffic and user movement. There is also great potential in using mobility-aware energy management, where knowledge of user movement patterns guides decisions on when and where to power down network parts without harming service quality.

Throughout this work, several important challenges and limitations were encountered. In the RLF prediction part, combining different types of data, like time-based and static information, required careful model design and a lot of tuning, especially because real failure events are rare in the data. In the mobility modeling part, it was challenging to make models work well in different cities and to handle large amounts of data, especially in areas with less data available. In the energy optimization part, it was necessary to find the right balance between saving energy and keeping the network running smoothly, especially during live tests where sudden changes in user activity could cause problems. In all three areas, mobile networks are very complex and change over time, so AI models need to be checked and updated regularly to stay effective.

Looking ahead, this thesis identifies several promising directions for future research. An important avenue is the integration of explainable AI to improve trust

and transparency in critical operational models, especially for failure prediction, where it is essential to provide clear and understandable insights. In mobility analytics, using richer and more diverse datasets, and developing models that can capture time-related changes, could greatly enhance predictive accuracy and adaptability, while maintaining strong privacy protections remains essential. For network sustainability, the application of reinforcement learning and mobility-based optimization offers valuable opportunities to achieve adaptive, real-time energy management without compromising user experience. Furthermore, interdisciplinary research that brings together reliability, mobility, and sustainability goals holds considerable potential. For example, combining proactive failure prediction with mobility-aware energy management could allow networks to dynamically adjust resources while maintaining high performance and resilience.

In conclusion, the research presented in this thesis highlights the role of AI in shaping the future of 5G and beyond mobile networks. By tackling key challenges related to maintaining reliable connections, gaining insights into human mobility, and supporting sustainable network operations, this work offers both new knowledge and practical solutions for next-generation network management. These contributions not only provide immediate benefits, such as fewer service interruptions, better decision making, and reduced energy costs, but also establish a solid foundation for developing intelligent, resilient, and environmentally sustainable communication networks. As the field moves toward AI-native 6G systems, the insights and methods introduced in this thesis will remain valuable for both academic research and practical applications. Ultimately, this work envisions a future where mobile networks function as autonomous, self-optimizing systems that smoothly adapt to changes in their environment, technology, and society, delivering excellent performance while carefully managing resources for the benefit of both operators and society.

REFERENCES

[1] J. Rowley, "The wisdom hierarchy: Representations of the DIKW hierarchy," *Journal of Information Science*, vol. 33, no. 2, pp. 163–180, 2007.

[2] F. Comellas and J. Ozon, "Graph Coloring Algorithms for Assignment Problems in Radio Networks," *Applications of Neural Networks to Telecommunications*, vol. 2, pp. 49–56, 1995.

[3] N. T. M. Lan, "Evolution of Wireless Technology: From 1G to 5G," *Asian Journal of Applied Science and Technology*, vol. 7, no. 04, pp. 68–73, 2023.

[4] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[5] B. C. Jung, "Toward artificial intelligence-native 6G services [Mobile Radio]," *IEEE Vehicular Technology Magazine*, vol. 19, no. 4, pp. 9–14, 2024.

[6] P. Soldati, E. Ghadimi, B. Demirel, Y. Wang, M. Sintorn, and R. Gaigalas, "Approaching AI-native RANs through generalization and scalability of learning," *Ericsson Technology Review*, vol. 2023, no. 3, pp. 2–12, 2023.

[7] Nokia White Paper, "Toward a 6G AI-Native Air Interface." `https://onestore.nokia.com/asset/210299`. Accessed July 2025.

[8] D. D. Hirsch, "The glass house effect: Big Data, the new oil, and the power of analogy," *Me. L. Rev.*, vol. 66, p. 373, 2013.

[9] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE network*, vol. 30, no. 1, pp. 44–51, 2016.

[10] R. L. Ackoff, "From data to wisdom," *Journal of applied systems analysis*, vol. 16, no. 1, pp. 3–9, 1989.

[11] N. Al-Falahy and O. Y. Alani, "Millimetre wave frequency band as a candidate spectrum for 5G network architecture: A survey," *Physical Communication*, vol. 32, pp. 120–144, 2019.

[12] Y.-W. Mal, J.-L. Chen, and H.-K. Lin, "Mobility Robustness Optimization Based on Radio Link Failure Prediction," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 454–457, 2018.

[13] S. Ergüt, "Challenges and opportunities for communication service providers in applying AI/ML," *ITU News Magazine*, no. 5, pp. 38–41, 2020.

[14] J. Moysen and L. Giupponi, "From 4G to 5G: Self-organized network management meets machine learning," *Computer Communications*, vol. 129, pp. 248–268, 2018.

[15] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu, and F. Kojima, "Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-Generation Wireless Networks," *IEEE Access*, vol. 6, pp. 32328–32338, 2018.

[16] N. M. Vural, S. Ergüt, and S. S. Kozat, "An Efficient and Effective Second-Order Training Algorithm for LSTM-Based Adaptive Learning," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2541–2554, 2021.

[17] N. M. Vural, F. Ilhan, S. F. Yilmaz, S. Ergüt, and S. S. Kozat, "Achieving Online Regression Performance of LSTMs With Simple RNNs," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 7632–7643, 2021.

[18] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[19] J. Feng, X. Chen, R. Gao, M. Zeng, and Y. Li, "DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction," *IEEE Network*, vol. 32, no. 6, pp. 108–115, 2018.

[20] L. Chen, D. Yang, D. Zhang, C. Wang, J. Li, and T.-M.-T. Nguyen, "Deep mobile traffic forecast and complementary base station clustering for C-RAN optimization," *Journal of Network and Computer Applications*, vol. 121, pp. 59–69, 2018.

[21] Y. Wei, Z. Zhang, F. R. Yu, and Z. Han, "Joint User Scheduling and Content Caching Strategy for Mobile Edge Networks Using Deep Reinforcement Learning," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2018.

[22] M. T. Vega, C. Liaskos, S. Abadal, E. Papapetrou, A. Jain, B. Mouhouche, G. Kalem, S. Ergüt, M. Mach, T. Sabol, *et al.*, "Immersive interconnected virtual and augmented reality: a 5G and IoT perspective," *Journal of Network and Systems Management*, vol. 28, no. 4, pp. 796–826, 2020.

[23] S. K. Mohanty and S. K. Udgata, "SATPAS: SINR-based adaptive transmission power assignment with scheduling in wireless sensor network," *Engineering Applications of Artificial Intelligence*, vol. 103, p. 104313, 2021.

[24] J. Luomala and I. Hakala, "Effects of Temperature and Humidity on Radio Signal Strength in Outdoor Wireless Sensor Networks," in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 1247–1255, 2015.

[25] S.-H. Fang and Y.-H. S. Yang, "The Impact of Weather Condition on Radio-Based Distance Estimation: A Case Study in GSM Networks With Mobile Measurements," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6444–6453, 2016.

[26] N. Javed, E. Lyons, M. Zink, and T. Wolf, "Adaptive Wireless Mesh Networks: Surviving Weather without Sensing It," *Computer Communications*, vol. 54, pp. 120–130, 2014.

[27] [Dataset] Turkcell, " ITU-AI ML in 5G Challenge." `https://github.com/Turkcell/ITU-AIMLin5GChallenge-2021`, 2021.

[28] X.-Y. Wang, H.-Y. Yang, Y.-W. Li, W.-Y. Li, and J.-W. Chen, "A new SVM-based active feedback scheme for image retrieval," *Engineering Applications of Artificial Intelligence*, vol. 37, pp. 43–53, 2015.

[29] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[30] H. D. Trinh, L. Giupponi, and P. Dini, "Mobile Traffic Prediction from Raw Data Using LSTM Networks," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 2018-September, pp. 1827–1832, 2018.

[31] A. Dal Pozzolo, O. Caelen, and G. Bontempi, "When is Undersampling Effective in Unbalanced Classification Tasks?," in *Machine Learning and Knowledge Discovery in Databases* (A. Appice, P. P. Rodrigues, V. Santos Costa, C. Soares, J. Gama, and A. Jorge, eds.), pp. 200–215, Springer International Publishing, 2015.

[32] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, p. 27, Mar 2019.

[33] P. Branco, L. Torgo, and R. P. Ribeiro, "A Survey of Predictive Modeling on Imbalanced Domains," *Computing Surveys (CSUR)*, vol. 49, no. 2, pp. 1–50, 2016.

[34] Z. Yang, T. Zhang, J. Lu, D. Zhang, and D. Kalui, "Optimizing area under the ROC curve via extreme learning machines," *Knowledge-Based Systems*, vol. 130, pp. 74–89, 2017.

[35] Turkish Republic Presidency, Strategy and Budget Office, "Kahramanmaras and Hatay earthquakes report (in Turkish)," 2023. Last accessed 14 July. 2025.

[36] Turkish Informatics Society, "Earthquake and Informatics Workshop report (in Turkish)," 2023. Last accessed 20 Aug. 2024.

[37] M. Başaran, S. Aktaş, and B. Bilgin, "6G Network Slicing Vision for Post-Disaster: AI-Enabled User Prioritization and Energy Management," in *2024 IEEE International Conference on Advanced Telecommunication and Networking Technologies (ATNT)*, vol. 1, pp. 1–4, 2024.

[38] Turkish Aerospace Industries Inc., "Aksungur İHA, üzerinde yer alan baz istasyonu ile deprem bölgesinde iletişime destek veriyor (in Turkish)," 2023. Last accessed 20 Aug. 2024.

[39] X. Song, R. Shibasaki, N. J. Yuan, X. Xie, T. Li, and R. Adachi, "DeepMob: Learning Deep Knowledge of Human Emergency Behavior and Mobility from Big and Heterogeneous Data," *ACM Trans. Inf. Syst.*, vol. 35, June 2017.

[40] F. Calabrese, L. Ferrari, and V. D. Blondel, "Urban Sensing Using Mobile Phone Network Data: A Survey of Research," *ACM Comput. Surv.*, vol. 47, Nov. 2014.

[41] 3GPP, "TS 28.550: Management Data Analytics Services (MDAS)," tech. rep., 3rd Generation Partnership Project (3GPP), 2022. Release 17.

[42] 3GPP, "TS 23.288: Architecture Enhancements for 5G System (5GS) to support Network Data Analytics Services (NWDAF)," tech. rep., 3rd Generation Partnership Project (3GPP), 2022. Release 17.

[43] O-RAN Alliance, "O-RAN Architecture Description v3.0." `https://www.o-ran.org/specifications`, 2021.

[44] M. Volk and J. Sterle, "5G Experimentation for Public Safety: Technologies, Facilities and Use Cases," *IEEE Access*, vol. 9, pp. 41184–41217, 2021.

[45] T. Salam and U. Dieter, "5G as Enabler Technology for Smart City Use Cases," in *2022 9th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*, pp. 1–7, 2022.

[46] T. Yabe, N. K. Jones, P. S. C. Rao, M. C. Gonzalez, and S. V. Ukkusuri, "Mobile phone location data for disasters: A review from natural hazards and epidemics," *Computers, Environment and Urban Systems*, vol. 94, p. 101777, 2022.

[47] N. Shiwakoti, M. Sarvi, and G. Rose, "Modelling pedestrian behaviour under emergency conditions - State-of-the-art and future directions," in *31st Australasian Transport Research Forum, ATRF 2008*, 2008.

[48] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating Origin-Destination Flows Using Mobile Phone Location Data," *IEEE Pervasive Computing*, vol. 10, p. 36–44, Oct. 2011.

[49] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, no. 1, p. 1376, 2013.

[50] R. Wang, X. Zhang, and N. Li, "Zooming into mobility to understand cities: A review of mobility-driven urban studies," *Cities*, vol. 130, p. 103939, 2022.

[51] L. Alessandretti, U. Aslak, and S. Lehmann, "The scales of human mobility," *Nature*, vol. 587, no. 7834, pp. 402–407, 2020.

[52] A. Hadachi, A. Lind, J. Lõmps, and P. Piksarv, "From Mobility Analysis to Mobility Hubs Discovery: A Concept Based on Using CDR Data of the Mobile Networks," in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 1–6, 2018.

[53] Y. Jo, J. Lim, and D. Hong, "Mobility Management Based on Beam-Level Measurement Report in 5G Massive MIMO Cellular Networks," *Electronics*, vol. 9, no. 5, 2020.

[54] S. Jiang, Y. Yang, S. Gupta, D. Veneziano, S. Athavale, and M. C. González, "The TimeGeo modeling framework for urban mobility without travel surveys," *Proceedings of the National Academy of Sciences*, vol. 113, no. 37, pp. E5370–E5378, 2016.

[55] T. Yabe, K. Tsubouchi, N. Fujiwara, T. Wada, Y. Sekimoto, and S. V. Ukkusuri, "Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic," *Scientific reports*, vol. 10, no. 1, p. 18053, 2020.

[56] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, p. 1655–1661, AAAI Press, 2017.

[57] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 162–177, 2015. Big Data in Transportation and Traffic Engineering.

[58] S. Wang, H. Miao, J. Li, and J. Cao, "Spatio-Temporal Knowledge Transfer for Urban Crowd Flow Prediction via Deep Attentive Adaptation Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4695–4705, 2022.

[59] L. Wang, X. Geng, X. Ma, F. Liu, and Q. Yang, "Cross-city transfer learning for deep spatio-temporal prediction," in *Proceedings of the 28th International*

Joint Conference on Artificial Intelligence, IJCAI'19, p. 1893–1899, AAAI Press, 2019.

[60] M. Amirijoo, P. Frenger, F. Gunnarsson, H. Kallin, J. Moe, and K. Zetterberg, "Neighbor cell relation list and measured cell identity management in LTE," in *NOMS 2008 - 2008 IEEE Network Operations and Management Symposium*, pp. 152–159, 2008.

[61] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, p. 107–113, 2008.

[62] Uber Technologies, "Uber's Hexagonal Hierarchical Spatial Index." `https://h3geo.org/`, accessed on 2023-03-25.

[63] P. Riehmann, M. Hanfler, and B. Froehlich, "Interactive sankey diagrams," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 233–240, IEEE, 2005.

[64] H. Pihkola, M. Hongisto, O. Apilo, and M. Lasanen, "Evaluating the energy consumption of mobile data transfer—from technology development to consumer behaviour and life cycle thinking," *Sustainability*, vol. 10, no. 7, 2018.

[65] S. Euttamarajah, Y. H. Ng, and C. K. Tan, "Energy-efficient joint base station switching and power allocation for smart grid based hybrid-powered comp-enabled hetnet," *Future Internet*, vol. 13, no. 8, 2021.

[66] P. Gandotra, R. K. Jha, and S. Jain, "Green Communication in Next Generation Cellular Networks: A Survey," *IEEE Access*, vol. 5, pp. 11727–11758, 2017.

[67] F. Salahdine, J. Opadere, Q. Liu, T. Han, N. Zhang, and S. Wu, "A survey on sleep mode techniques for ultra-dense networks in 5G and beyond," *Computer Networks*, vol. 201, p. 108567, 2021.

[68] E. Chavarria-Reyes, I. F. Akyildiz, and E. Fadel, "Energy-Efficient Multi-Stream Carrier Aggregation for Heterogeneous Networks in 5G Wireless Systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7432–7443, 2016.

[69] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, 2011.

[70] P. Rahmawati, M. I. Nashiruddin, and M. A. Nugraha, "Capacity and Coverage Analysis of 5G NR Mobile Network Deployment for Indonesia's Urban Market," in *2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pp. 90–96, 2021.

[71] 3GPP, "3GPP TS 28.310: Management and orchestration; Energy efficiency of 5G." https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3550/, accessed on 2023-04-20.

[72] ETSI, "ETSI ES 203.228 v1.3.1: Environmental Engineering (EE); Assessment of mobile network energy efficiency." https://www.etsi.org/deliver/etsi_es/203200_203299/203228/01.03.01_60/es_203228v010301p.pdf, accessed on 2023-04-20.

[73] O-RAN Alliance, "O-RAN.WG1.Network-Energy-Savings-Technical-Report-R003-v01.00: Network Energy Saving Use Cases Technical Report." https://orandownloadsweb.azurewebsites.net/specifications, accessed on 2023-04-20.

[74] L. Kundu, X. Lin, and R. Gadiyar, "Toward energy efficient ran: From industry standards to trending practice," *IEEE Wireless Communications*, vol. 32, no. 1, pp. 36–43, 2025.

[75] L. Mata, M. Sousa, P. Vieira, M. P. Queluz, and A. Rodrigues, "Optimizing energy and spectral efficiency in mobile networks: A comprehensive energy sustainability framework for network operators," *IEEE Access*, vol. 13, pp. 22342–22364, 2025.

[76] M. Masoudi, M. G. Khafagy, E. Soroush, D. Giacomelli, S. Morosi, and C. Cavdar, "Reinforcement learning for traffic-adaptive sleep mode management in

5g networks," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–6, 2020.

[77] A. El-Amine, H. A. Haj Hassan, M. Iturralde, and L. Nuaymi, "Location-Aware Sleep Strategy for Energy-Delay Tradeoffs in 5G with Reinforcement Learning," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2019.

[78] F. E. Salem, T. Chahed, Z. Altman, and A. Gati, "Traffic-aware Advanced Sleep Modes management in 5G networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2019.

[79] A. Conte, A. Feki, L. Chiaraviglio, D. Ciullo, M. Meo, and M. A. Marsan, "Cell wilting and blossoming for energy efficiency," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 50–57, 2011.

[80] M. S. Dahal, J. N. Shrestha, and S. R. Shakya, "Energy saving technique and measurement in green wireless communication," *Energy*, vol. 159, pp. 21–31, 2018.

[81] D. López-Pérez, A. De Domenico, N. Piovesan, and M. Debbah, "Data-driven energy efficiency modeling in large-scale networks: An expert knowledge and ml-based approach," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 780–804, 2024.

[82] T. Beitelmal, S. S. Szyszkowicz, D. G. González, and H. Yanikomeroglu, "Sector and Site Switch-Off Regular Patterns for Energy Saving in Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2932–2945, 2018.

[83] F. Alaca, A. B. Sediq, and H. Yanikomeroglu, "A genetic algorithm based cell switch-off scheme for energy saving in dense cell deployments," *2012 IEEE Globecom Workshops, GC Wkshps 2012*, pp. 63–68, 2012.

[84] F. Han, Z. Safar, W. S. Lin, Y. Chen, and K. J. R. Liu, "Energy-efficient cellular network operation via base station cooperation," in *2012 IEEE International Conference on Communications (ICC)*, pp. 4374–4378, 2012.

[85] L. Suárez, L. Nuaymi, and J.-M. Bonnin, “Energy-efficient bs switching-off and cell topology management for macro/femto environments,” *Computer Networks*, vol. 78, pp. 182–201, 2015. Special Issue: Green Communications.

[86] X. Wang, B. Lyu, C. Guo, J. Xu, and M. Zukerman, “A Base Station Sleeping Strategy in Heterogeneous Cellular Networks Based on User Traffic Prediction,” *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 1, pp. 134–149, 2024.

[87] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, “Toward dynamic energy-efficient operation of cellular network infrastructure,” *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56–61, 2011.

[88] Y. Tang, X. Deng, L. Yi, Y. Xia, L. T. Yang, and X. Tang, “Collaborative Intelligent Confident Information Coverage Node Sleep Scheduling for 6G-Empowered Green IoT,” *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 2, pp. 1066–1077, 2023.

[89] Endeksa, “Population, income level and education level information.” Accessed on 31 Jan, 2023.

[90] S. Mirjalili, *"Genetic Algorithm"*, pp. "43–55". "Cham": "Springer International Publishing", 2019.

[91] H. Afshari, W. Hare, and S. Tesfamariam, “Constrained multi-objective optimization algorithms: Review and comparison with application in reinforced concrete structures,” *Applied Soft Computing*, vol. 83, p. 105631, 2019.

[92] I. Costa-Carrapiço, R. Raslan, and J. N. González, “A systematic review of genetic algorithm-based multi-objective optimisation for building retrofitting strategies towards energy efficiency,” *Energy and Buildings*, vol. 210, p. 109690, 2020.

CURRICULUM VITAE

PERSONAL

NAME : Semih Aktaş

WORK EXPERIENCE

2023-	**6Gen Researcher** Turkcell, İstanbul, Turkiye
2022-2024	**Researcher** Gazi University, Ankara, Turkiye
2018-2023	**5G VATS R&D** Turkcell, İstanbul, Turkiye
2016-2017	**Software Engineer** DİA Software, Ankara, Türkiye
2014-2015	**Student Teaching Assistant** Department of Computer Enginering, METU, Ankara, Türkiye

EDUCATION

2025 | Ph.D. in COMPUTER ENGINEERING
Graduate School of Natural and Applied Sciences
Middle East Technical University (METU), Türkiye
Thesis: "AI-Driven 5G and Beyond Mobile Networks: Reliability, Mobility Insights, and Sustainable Operations"
Advisor: Hande ALEMDAR

2019 | MSc. Degree in COMPUTER ENGINEERING
Graduate School of Natural and Applied Sciences
Middle East Technical University (METU), Türkiye
Thesis: "Base Station Power Optimization for Green Networks Using Reinforcement Learning"
Advisor: Hande ALEMDAR

2016 | Bs Degree in COMPUTER ENGINEERING
Faculty of Engineering
Middle East Technical University (METU), Türkiye

PUBLICATIONS

Journal and conference publications are annotated with the "**J**" and "**C**" prefixes.

C B. Bilgin, **S. Aktaş** and S. Bas, "Intent-based End-to-End Explainability Orchestration Framework for AI-Native Networks" 2025 11st IEEE International Conference on Network Softwarization (NETSOFT), Budapest, Hungary, June 2025, pp. 1-6.

C M. Başaran, **S. Aktaş** and B. Bilgin, "6G Network Slicing Vision for Post-Disaster: AI-Enabled User Prioritization and Energy Management," 2024 IEEE International Conference on Advanced Telecommunication and Networking Technologies (ATNT), Johor Bahru, Malaysia, 2024, pp. 1-4.

J Çodur, S., Erkayman, B., Alp, S. S., Özenir, O., Pamucar, D., Yıldız, G., ... & **Aktaş, S.** (2024). "Application of the full consistency method (FUCOM)-Cosine similarity framework in 5G infrastructure investment planning: An approach for telecommunication quality improvements." Heliyon, 10(9).

C **S. Aktaş**, S. Çodur, B. Erkayman, G. Yıldız, A. Gemalmaz and A. Dikel, "Evaluation of 5G Infrastructure Investment Options Using the AHP-TOPSIS Method," 2023 31st Signal Processing and Communications Applications Conference (SIU), Istanbul, Turkiye, 2023, pp. 1-4.

J **Aktaş, S.**, Alemdar, H., & Ergüt, S. (2022). Towards 5G and beyond radio link diagnosis: Radio link failure prediction by using historical weather, link parameters. Computers and Electrical Engineering, 99, 107742.

J **Aktaş, S.**, & Alemdar, H. Base Station Power Optimization for Green Networks Using Reinforcement Learning. SAUCIS. August 2021;4(2):244-265.

NATIONAL AND INTERNATIONAL PROJECTS

1. **TUBİTAK 1515, 6Gen. Lab.:** 6G ve Yapay Zeka Laboratuvarı, (2023-...)
2. **SNS-JU, VERGE**: AI-powered eVolution towards opEn and secuRe edGe architEctures (2024-2025)
3. **H2020-EU, StorAIge:** Embedded storage elements on next MCU generation ready for AI on the edge (2023-2024)
4. **TUBİTAK 1001**, Deprem Anı ve Sonrasındaki Olası Kentsel Nüfus Hareketliliğinin Tahmini ve Modellenmesi (2022-2024)
5. **Celtic, AI4GREEN:** Artificial Intelligence for Green networks (2019-2022)