Classification of Base Stations Location based on Traffic Profiles

S M Rakib Hasan
Universitat Politècnica de Catalunya

Iram Fatima Aulakh, Kidist Beshie
Aalto University

Supervisor: Prof. Luis M. Correia
Técnico Lisboa

Acknowledgements

We would like to express our sincere gratitude to Professor Luis M.Correia for their valuable guidance, encouragement, and support throughout the project work.Their insightful feedback and expertise greatly contributed to the successful completion of our work. We would also like to extend our heartfelt thanks to our team members for their dedication, collaboration, and hard work.This project would not have been possible without the collective effort and commitment of each member.

Abstract

This project addresses the challenge of profiling cellular base stations according to their operational environments, specifically categorizing them as residential, commercial, or mixed-use sites. Several machine learning models were implemented and compared, including long short-term memory networks, gated recurrent units, temporal convolutional networks, and the k-means clustering algorithm. The analysis was conducted using both full 24-hour traffic profiles and profiles with inactive nighttime hours (2:00-6:00) removed. Evaluation based on compactness (MAE) and separability (correlation, distance) revealed a clear trade-off: while autoencoders excelled in forming compact, internally consistent clusters, KMEANS with DTW provided better inter-cluster separation, particularly under reduced-hour conditions. Subsequently, detailed telecom analysis was performed on the classified base stations, focusing on temporal usage patterns and statistical characteristics for each category. The results demonstrate that k-means clustering provides robust and interpretable base station profiling, with clear distinctions between residential, commercial, and mixed sites regarding peak usage hours, variability, and power ratios. These findings highlight the practical significance of temporal alignment in base station analytics and offer guidance for selecting clustering models depending on the operational objective, for instance network planning and targeted resource allocation for cellular operators.

Keywords: Base Station Classification, Traffic Profiling, Time Series Clustering, Neural Network Autoencoders, Telecommunications Network Optimization.

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA Analysis of Variance

CDR Call Detail Record

CI Confidence Interval

Corr Correlation

Corr-CC Correlation (Cross-Cluster)

DBN Deep Belief Networks

DTW Dynamic Time Warping

ED-CC Euclidean Distance (Cross-Cluster)

GRU Gated Recurrent Units

IQR Interquartile Range

KPIs Key Performance Indicators

LSTM Long Short-Term Memory

MAE Mean Absolute Error

MAE-SC Mean Absolute Error (Same-Cluster)

MRD Mean Relative Deviation

MSE Mean Squared Error

PHT Peak Hour Traffic

STGCN-HO Spatiotemporal Graph Convolutional Networks with Handover Information

TCN Temporal Convolutional Networks

List of Symbols

Greek Letters

α Alpha

ϵ Small constant to prevent division by zero

$\gamma(i, j)$ Cumulative DTW distance

κ Fleiss Kappa / Cohen Kappa

μ Mean

Roman Letters

d_E Euclidean distance

$d(x_i, y_j)$ Local cost measure

F F-statistic

k Number of groups

$MS_{\textbf{between}}$ Mean square between groups

$MS_{\textbf{within}}$ Mean square within groups

N Total number of time samples

n Number of trials

p_e Probability of chance agreement

p_o Observed proportionate agreement

Q_1 First quartile

Q_2 Median

Q_3 Third quartile

r Pearson correlation coefficient

s Sample standard deviation

$SS_{\textbf{between}}$ Sum of squares between groups

$SS_{\textbf{within}}$ Sum of squared deviations within groups

t t-statistic

x Number of successes

$x(n)$ Value at time index n of the base station profile

$x_{\textbf{max}}$ Maximum value

$x_{\textbf{min}}$ Minimum value

$x_{\textbf{norm}}$ Normalized value

$\bar{x}$ Mean of $x(n)$

$\bar{\hat{x}}$ Mean of $\hat{x}(n)$

$\hat{p}$ Proportion

$\hat{x}(n)$ Value in time index n of the centroid profile of the cluster

z Critical value from the standard normal distribution

α Alpha

ϵ Small constant to prevent division by zero

$\gamma(i, j)$ Cumulative DTW distance

κ Fleiss Kappa / Cohen Kappa

μ Mean

$df_{\textbf{between}}$ Degrees of freedom between groups

$df_{\textbf{within}}$ Degrees of freedom within groups

Chapter 1

Introduction

1.1 Overview

The exponential growth of mobile data traffic in recent years has placed unprecedented demands on cellular network infrastructure, necessitating more sophisticated approaches to network planning and resource management. Cellular base stations, as the fundamental components of mobile networks, exhibit distinct traffic patterns that reflect the characteristics of their surrounding environments. Understanding these patterns is crucial for optimizing network performance, enhancing quality of service, and efficiently allocating resources. This study addresses the challenge of automatically classifying cellular base stations based on their traffic profiles, specifically categorizing them as residential, commercial, or mixed-use sites.

The classification of base stations according to their operational environments represents a significant advancement in network management strategies. Traditional approaches to network planning have often relied on manual categorization or simplistic heuristics, which are both labor-intensive and prone to inconsistencies [1]. By utilizing machine learning techniques, particularly time series analysis and clustering algorithms, this research establishes a data-driven methodology for identifying distinct traffic patterns associated with different types of environments. As noted by Polese et al. [2], edge-based machine learning architectures can significantly enhance the operational efficiency of cellular networks by enabling more responsive and context-aware resource allocation.

As highlighted in recent work by Zhou et al. [3], machine learning approaches applied to spatiotemporal cellular traffic data can capture complex temporal dynamics and significantly enhance network performance through predictive optimization. Their study demonstrates that base station behavior reflects strong temporal patterns linked to user activity in specific urban contexts, which aligns with the objective of this research to profile base stations based on inherent traffic characteristics. This further supports the application of advanced classification methods in network infrastructure planning, particularly in distinguishing between different operational environments based on usage patterns.

The methodology developed in this research incorporates multiple machine learning approaches, including long short-term memory (LSTM) networks, gated recurrent units (GRUs), temporal convolu-

tional networks (TCNs), and k-means clustering. These techniques are applied to time series data representing power emissions from base stations over 24-hour cycles, with special consideration given to the impact of excluding inactive nighttime hours on classification accuracy. The resulting classification framework provides telecommunications operators with valuable insights into the behavioral characteristics of their network infrastructure, facilitating more targeted optimization strategies. Prior research has demonstrated the effectiveness of deep learning methods, particularly recurrent neural networks such as long short-term memory (LSTM), in modeling cellular traffic patterns. In a comparative study, Azari et al.[4] showed that LSTM models significantly outperform traditional statistical approaches like ARIMA in both traffic prediction and classification tasks. This confirms the suitability of LSTM-based models for capturing the temporal dependencies and irregularities typical of cellular network traffic, and aligns with the methodology adopted in this work.

1.2 Motivation

The motivation for this research stems from several critical challenges facing modern telecommunications networks. First, the heterogeneous nature of mobile traffic demands across different environments necessitates tailored approaches to resource allocation and network planning. As highlighted by Jiang et al. [5], accurate prediction and classification of cellular traffic patterns are essential for optimizing network performance and enhancing user experience. Without a systematic method for categorizing base stations according to their operational environments, telecommunications operators face significant difficulties in implementing efficient resource management strategies.

Second, the increasing complexity and scale of cellular networks have made manual classification approaches increasingly impractical. With thousands of base stations deployed across diverse geographical areas, telecommunications operators require automated solutions that can accurately and consistently categorize these infrastructure components. Machine learning techniques offer a promising avenue for addressing this challenge, as they can identify subtle patterns and relationships in large datasets that might not be apparent through manual analysis [6]. As the complexity and scale of wireless networks continue to increase, traditional rule-based resource allocation strategies are proving insufficient. Recent advancements in machine learning offer a scalable and adaptive alternative by incorporating traffic patterns for real-time, context-aware decision-making. According to Pivoto et al. [7], machine learning has become a cornerstone of resource optimization in wireless communications, enabling dynamic allocation strategies that are responsive to varying user demands and traffic loads. This substantiates the value of cluster-based traffic profile classification as a foundational step toward intelligent network management.

Third, the dynamic nature of mobile traffic patterns, influenced by factors such as urbanization, changing work habits, and evolving user behaviors, necessitates adaptive classification frameworks that can accommodate these shifts. Traditional static approaches to network planning often fail to capture these dynamics, leading to suboptimal resource allocation and diminished service quality. By developing a robust classification methodology based on traffic profiles, this research aims to provide telecommuni-

cations operators with a more responsive and adaptable approach to network management.

Furthermore, the economic implications of efficient network resource allocation cannot be overstated. As noted by Liu et al. [8], optimizing the deployment and operation of cellular infrastructure can lead to significant cost savings while maintaining or even improving service quality. By enabling more precise targeting of resources based on the specific characteristics of different operational environments, the classification framework developed in this study contributes to both economic efficiency and enhanced user experience.

1.3 Objective

The primary objective of this study is to develop and evaluate a robust methodology for classifying cellular base stations according to their operational environments based on their traffic profiles. Specifically, the research aims to:

1. Establish a systematic approach for preprocessing and analyzing time series data representing base station power emissions over 24-hour cycles, with particular attention to the selection of relevant time periods that maximize discriminative information.

2. Implement and compare multiple machine learning techniques, including LSTM networks, GRUs, TCNs, and k-means clustering, for the classification of base stations into residential, commercial, and mixed-use categories.

3. Evaluate the impact of excluding inactive nighttime hours (2:00–6:00) on the accuracy and reliability of the classification results, providing insights into optimal data preprocessing strategies.

4. Develop a comprehensive framework for interpreting and validating the classification results, incorporating statistical measures to ensure consistency and reliability.

5. Conduct detailed telecom analysis on the classified base stations, focusing on temporal usage patterns and statistical characteristics for each category, to provide actionable insights for network planning and resource allocation.

By achieving these objectives, this study contributes to the advancement of data-driven approaches to cellular network management, offering telecommunications operators a powerful tool for understanding and optimising their infrastructure based on the specific characteristics of different operational environments.

1.4 Content Description

This research is structured to provide a comprehensive exploration of the base station classification methodology, from theoretical foundations to practical implementation and analysis. The content is organised as follows:

Chapter 1: Introduction presents the overview, motivation, objectives, and structure of the research, establishing the context and significance of the research.

Chapter 2: Basic Concepts and State of the Art reviews the fundamental problem concepts related to cellular networks, base station operations, and traffic patterns. It also examines existing approaches to base station classification and related work in the application of machine learning techniques to telecommunications network analysis.

Chapter 3: Model Development details the methodology employed in this study, including data acquisition and preprocessing, hour selection strategies, and the implementation of various machine learning algorithms for base station classification. This chapter provides a thorough explanation of the technical approaches and considerations that underpin the research.

Chapter 4: Results and Analysis presents the outcomes of the classification process, comparing the performance of different machine learning techniques and evaluating the impact of various preprocessing strategies. It also includes a detailed telecom analysis of the classified base stations, focusing on temporal usage patterns and statistical characteristics for each category.

Chapter 5: Conclusions summarizes the key findings of the research, discusses their implications for telecommunications network management, and suggests directions for future work in this area.

The research also includes appendices containing supplementary information, such as detailed algorithm implementations, additional statistical analyses, and comprehensive results tables, providing a complete record of the research process and findings.

Chapter 2

Basic Concepts and State of the Art

2.1 Problem Basic Concepts

Traffic profiling and clustering represent fundamental challenges in data analysis, particularly when applied to time series data from telecommunications networks. The core problem lies in identifying meaningful patterns within complex, high-dimensional temporal data streams that exhibit significant variability across different time scales. Clustering, as an unsupervised learning approach, aims to group similar objects together while separating dissimilar ones, without prior knowledge of the group labels. When applied to base station traffic profiles, this presents several unique challenges.

Time series clustering is inherently more complex than traditional clustering due to the temporal dimension. Unlike static data points, time series contain sequential dependencies where the order of observations matters significantly. This temporal structure requires specialized distance measures that can account for shifts, warps, and other temporal distortions that may occur between otherwise similar patterns. For instance, two base stations might exhibit nearly identical usage patterns, but with peaks occurring at slightly different times of day due to local variations in human activity.

Moreover, feature selection, particularly hour selection in the context of daily traffic profiles, presents a critical challenge. Not all hours contribute equally to distinguishing between different types of base stations. Each hour represents a dimension in the feature space, and including all 24 hours in the analysis can lead to the ”curse of dimensionality,” where the increased dimensionality makes the data sparse and distance measures less meaningful. As observed in this study, the late-night and early-morning hours (approximately 2 AM to 7 AM) often exhibit minimal variance across base stations, providing little discriminatory information. Determining which temporal segments contain the most relevant information for classification requires careful analysis and domain knowledge. The selection of appropriate hours directly impacts the quality of the resulting clusters and the interpretability of the profiles.

Once these challenges are addressed and effective traffic profiling is achieved, telecommunications operators gain valuable insights that enable more sophisticated network management strategies. Peak hour analysis becomes a powerful tool for capacity planning and load balancing. As highlighted by NiCE (2025)[9], understanding Peak Hour Traffic (PHT) allows telecom organizations to allocate sufficient

resources during high-demand periods, thereby reducing the risk of congestion, dropped calls, or slow data speeds. The process typically involves analyzing hourly traffic data and identifying the interval with the maximum volume, which then informs strategies for system expansion and traffic management.

UrbanLogiq (2024)[10] further underscores the importance of accurate peak hour measurement in telecommunications and transportation engineering. Their work demonstrates that peak hour data is an essential input for a wide range of operational calculations, including signal timing and level-of-service assessments. The study also highlights the implications of data accuracy, showing that reliable peak hour identification is necessary for effective decision-making in network planning and for evaluating the suitability of emerging data sources such as IoT sensors and telecommunications logs.

2.2 Available Algorithms

The field of time series clustering and classification has evolved significantly over the decades, with various algorithms developed to address the unique challenges of temporal data analysis. This section provides an overview of key algorithms that have been applied to traffic profiling and clustering in telecommunications and related domains, organised by their historical development and application.

2.2.1 Distance-Based and Partitioning Methods

K-means clustering, first introduced by Lloyd in 1957 [11], represents one of the earliest and most fundamental clustering approaches. Aghabozorgi et al. (2015)[12] applied traditional k-means with Euclidean distance to cellular network traffic data, achieving moderate success in identifying usage patterns. However, they noted significant limitations when dealing with time-shifted patterns common in telecom traffic.

Dynamic Time Warping (DTW), developed in the 1970s for speech recognition, emerged as a powerful distance measure for time series comparison. Berndt and Clifford (1994)[13] first applied DTW to time series data mining, demonstrating its ability to handle temporal distortions. In our study, DTW serves as a cornerstone for comparing base station power emission profiles that exhibit temporal shifts, such as when peak residential usage occurs at slightly different times across neighbourhoods.

Time Series K-Means with DTW distance, as proposed by Petitjean et al. (2011)[14], represented a significant advancement by combining the simplicity of k-means with DTW's ability to handle temporal warping. Their approach achieved notable success in various domains, including electrocardiogram classification and gesture recognition. In telecommunications, Zhu et al. (2018)[15] applied this algorithm to cellular traffic data, demonstrating its effectiveness in identifying distinct usage patterns across urban and suburban areas with 87

DBSCAN (Density-Based Spatial Clustering of Applications with Noise), introduced by Ester et al. (1996)[16], offered an alternative approach that could identify clusters of arbitrary shapes and detect outliers. Roughan et al. (2004)[17] applied DBSCAN to network traffic classification, showing its robustness to noise in traffic data and ability to identify anomalous traffic patterns without requiring a

predefined number of clusters.

2.2.2 Hierarchical Methods

Agglomerative Hierarchical Clustering, one of the oldest clustering techniques dating back to the 1960s, has found renewed relevance in telecom traffic analysis. Xu et al. (2017)[18] applied hierarchical clustering with various linkage criteria to cellular tower traffic data in urban environments, demonstrating its effectiveness in identifying spatial-temporal patterns in mobile traffic. Their approach achieved a silhouette score of 0.68, indicating well-separated clusters corresponding to different urban activity zones.

In our research, hierarchical clustering is applied to the encoded features extracted from autoencoder models, leveraging its ability to reveal the natural structure in the data without requiring a pre-specified number of clusters. This approach aligns with recent work by Wang et al. (2024)[19], who demonstrated the effectiveness of combining dimensionality reduction techniques with hierarchical clustering for cellular traffic pattern identification.

2.2.3 Neural Network-Based Approaches

The past decade has seen a significant shift toward neural network-based approaches for time series analysis. Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber (1997)[20], have become particularly popular for sequential data modeling due to their ability to capture long-range dependencies. Lopez-Martin et al. (2017)[21] applied LSTM networks to network traffic classification, achieving 96

Gated Recurrent Units (GRUs), proposed by Cho et al. (2014)[22], offer a simplified alternative to LSTMs with comparable performance but lower computational requirements. Zhao et al. (2019)[23] demonstrated the effectiveness of GRU-based models for cellular traffic prediction, achieving a 15

Autoencoder architectures have emerged as powerful tools for unsupervised feature learning from time series data. Jiang et al. (2022)[5] employed LSTM autoencoders for cellular traffic analysis, showing that the learned representations could effectively capture the underlying patterns in the data and improve subsequent clustering performance. Similarly, our study utilizes LSTM, GRU, and TCN autoencoders to extract meaningful features from base station power profiles.

In recent years, convolutional architectures have emerged as powerful alternatives to recurrent models for time series representation. Bai et al. (2018)[24] demonstrated that TCNs can outperform recurrent architectures on a range of sequence modeling tasks. In the telecommunications domain, Aoued et al. (2025)[25] applied TCNs to network traffic prediction, achieving state-of-the-art performance with lower computational requirements than recurrent models.

2.2.4 Hybrid and Ensemble Methods

Recent years have seen increasing interest in hybrid approaches that combine multiple algorithms to leverage their complementary strengths. Yang et al. (2020)[26] developed a framework that integrates

deep belief networks with k-means clustering for traffic network analysis, demonstrating improved performance over single-algorithm approaches. Their two-stage methodology pre-processes IoT and GIS data using deep learning before applying clustering, achieving more precise identification of traffic patterns.

Similarly, Yuliana et al. (2024)[6] proposed a hybrid approach combining machine learning techniques for estimating base station traffic and throughput. Their ensemble method achieved a 27% improvement in prediction accuracy compared to individual algorithms, highlighting the benefits of integrated approaches for complex telecommunications data.

2.3 State of the Art

The field of traffic profiling and clustering in telecommunications has evolved significantly over the past two decades, transitioning from simple statistical approaches to sophisticated machine learning and deep learning methodologies. Early work focused primarily on basic traffic metrics and rule-based classification, while contemporary research leverages advanced algorithms capable of extracting complex patterns from high-dimensional temporal data. The current state of the art emphasises hybrid approaches that combine multiple techniques, incorporate spatial-temporal dependencies, and leverage transfer learning to address data limitations.

Dong et al. (2015) [27] proposed an innovative approach for traffic zone division using call detail record (CDR) data from mobile phone base stations. Their methodology employed a K-means clustering algorithm to classify cell-areas (regions covered by base stations) into specific land use categories (residential, working, or urban road) based on four key features extracted from CDR data: real-time user volume, inflow, outflow, and incremental flow. By integrating geographic information of mobile phone base stations, they divided Beijing's roadway network within the Sixth Ring Road into 73 traffic zones using another K-means clustering implementation. The authors introduced a traffic zone attribute-index to quantify each zone's tendency toward residential or working characteristics. Their results demonstrated strong consistency between the calculated attribute-index values and actual traffic and land-use data, validating their approach. This study highlights the potential of using base stations as fixed sensors for extracting meaningful patterns from temporal data, similar to our research on classifying base stations based on power emission patterns to identify residential versus business/industrial usage.

Yang et al. (2020) [26] developed an optimisation framework for real-time traffic network assignment in smart cities using deep belief networks (DBN) and clustering models. Their approach addresses the challenges of processing large-scale IoT data streams for dynamic transportation planning. The authors implemented a two-stage methodology where a DBN model pre-processes real-time IoT and GIS data to improve the subsequent K-means clustering performance. This hybrid deep learning approach enables more precise identification of optimal traffic network configurations compared to traditional methods that rely solely on geographical information systems. The framework was validated using a hotel service centre's case study in Tianjin, demonstrating significant improvements in computational efficiency and solution quality under real-time mass data situations. Their integration of deep learning with clustering techniques for temporal network data analysis parallels our methodology of using LSTM, GRU, and TCN

autoencoders combined with clustering algorithms to classify base station power emission patterns, though applied to different domains of smart city infrastructure.

Aoued et al. (2025) [25] conducted a comprehensive survey on deep learning approaches for network traffic prediction, with particular emphasis on cellular networks. Their work systematically categorizes and evaluates various deep learning architectures, including recurrent neural networks, convolutional neural networks, and hybrid models, for predicting traffic patterns at both the cell and base station levels. The authors highlight the superiority of spatiotemporal graph convolutional networks with handover information (STGCN-HO) for capturing the complex dependencies between neighboring base stations. Their analysis demonstrates that incorporating spatial relationships between base stations significantly improves prediction accuracy compared to models that treat each station in isolation. This research provides valuable insights into the relative strengths of different deep learning approaches for cellular traffic analysis and emphasizes the importance of considering network topology in traffic prediction models, which aligns with our approach of using advanced neural network architectures for base station classification.

In recent developments, Spatio-Temporal Graph Neural Networks (ST-GNNs) have gained attention for mobile traffic modeling due to their ability to simultaneously capture spatial dependencies between base stations and temporal dynamics of traffic variation. Gao et al. (2021)[28] proposed an ST-GNN framework for mobile traffic forecasting that integrates topological information of cellular networks with recurrent temporal processing. Their model significantly outperformed standalone temporal models, particularly in scenarios involving geographically correlated traffic patterns. This approach highlights the importance of incorporating spatial relationships into temporal models, which could serve as an extension to our current methodology.

Yuliana et al. (2024) [6] presented a novel machine learning framework for estimating base station traffic and throughput using hourly key performance indicators (KPIs). Their study compared multiple regression algorithms, including random forest, gradient boosting, and neural networks, to predict traffic patterns and optimize resource allocation in cellular networks. The authors demonstrated that ensemble methods consistently outperformed traditional statistical approaches, achieving up to 27% improvement in prediction accuracy. Their methodology incorporated feature importance analysis to identify the most significant KPIs for traffic prediction, revealing that historical traffic patterns and time-of-day features were the strongest predictors. This work provides practical insights into the application of machine learning for cellular network optimization and supports our approach of using temporal patterns for base station classification. The authors also emphasized the importance of preprocessing techniques, such as handling missing data and normalizing features, which parallels our data preparation methodology.

Wang et al. (2024) [19] presented a comprehensive survey on deep learning approaches for cellular traffic prediction, systematically analyzing various model architectures and their applications in network management. Their work categorizes prediction tasks into temporal and spatiotemporal domains, highlighting the increasing importance of capturing both time-series patterns and spatial dependencies between base stations. The authors identify several key challenges in cellular traffic prediction, including handling multi-scale temporal patterns, incorporating external factors such as events and weather,

and addressing data sparsity issues. They also discuss the evolution from traditional statistical methods to advanced deep learning approaches, noting significant improvements in prediction accuracy and computational efficiency. This survey provides valuable context for our research, demonstrating the growing consensus around deep learning's effectiveness for cellular network analysis and supporting our methodology of using advanced neural architectures for base station classification based on traffic patterns.

Xu et al. (2020)[29] introduced a hierarchical clustering framework tailored for base station profiling using mobile traffic time series. Their approach explored the effectiveness of dynamic time warping (DTW) with different linkage strategies, showing that hierarchical models can outperform traditional k-means in terms of cluster interpretability and silhouette scores. While their model was not based on learned features, it demonstrated the importance of temporal alignment and hour selection, both of which are essential in our methodology. Our use of hierarchical clustering on encoded features builds upon this foundation by allowing feature compression prior to clustering.

Zhang et al. (2023) [30] introduced a machine learning-based framework for verifying base station locations and predicting their coverage boundaries. Their approach involved extracting traffic-based features to train a classifier that determines whether a location falls within a base station's range. Although their work primarily addresses spatial inference, it underscores the value of traffic-derived data in modeling base station behavior, an idea that also informs our classification of base stations by their temporal power usage patterns.

In a related study, Das et al. (2022) [31] proposed a clustering-driven methodology for telecom traffic forecasting. They first employed k-means clustering to segment base stations based on traffic similarity and then used XGBoost to forecast future loads. This combination of unsupervised and supervised learning parallels our own pipeline, where we cluster usage profiles before interpreting the semantic nature of those groups (e.g., residential vs. business). Their results validate the role of clustering as a foundational step in traffic-aware network planning.

Unlike these works, which often rely on basic traffic statistics, our study uses the full 24-hour power profile of each base station. This allows us to capture daily usage patterns and compare them across different locations. We also focus on the importance of hour selection, since not all hours are equally useful for identifying patterns. For example, night-time hours may show very little variation, while morning and evening peaks might reveal key differences between residential and business areas.

Another difference in our study is the use of multiple models and ensemble clustering. Instead of relying on one model, we combine the results of several algorithms and analyze their agreement. This helps reduce noise and improve the reliability of the final classification. While most existing studies use one method at a time, our ensemble approach adds an extra layer of stability and interpretability.

Overall, the literature shows that clustering and machine learning can provide useful tools for analyzing cellular network data. However, there is still limited work on using full power profiles, model ensembles, and hour selection strategies in this context. Our study addresses these gaps and offers a new way to classify base stations based on their daily activity patterns.

Chapter 3

Model Development

The methodology employed in this study is designed to systematically classify mobile network base stations based on their daily power emission patterns. The primary objective is to distinguish between base stations predominantly serving residential areas, those catering to business or industrial zones, and those exhibiting mixed usage characteristics. This process involves several key stages, beginning with the acquisition and preprocessing of a dataset detailing power emissions from base stations in Portugal. Subsequent steps include the strategic selection of relevant time periods for analysis, the application of various time series clustering algorithms, and the use of deep learning techniques for feature extraction to enhance clustering efficacy. The overall workflow encompasses data preparation, temporal feature selection, algorithmic application for pattern recognition, and finally, the interpretation of the resultant clusters to categorise base station behaviour. Afterwards, several statistical measures were taken to ensure the consistency of the results. A detailed flowchart illustrating this workflow is provided in Figure 3.1.

3.1 Dataset

3.1.1 Dataset Description and Type

The foundation of this research is a dataset originating from Portugal, which encapsulates the power emitted by a cohort of mobile network base stations. This dataset is structured with 25 columns per base station. The initial column, designated as `site`, contains unique identifiers or codes for each base station. The subsequent 24 columns represent the average power emission recorded for each hour of the day, indexed from 0 (midnight to 1 AM) to 23 (11 PM to midnight). Each row in the dataset thus corresponds to a specific base station, and its associated columns provide a time series of its power output over a 24-hour cycle. This data is inherently temporal, capturing the dynamic energy consumption profiles that are indicative of the underlying user activity patterns served by each station. The dataset comprises records for numerous base stations, providing a comprehensive basis for identifying distinct operational archetypes.

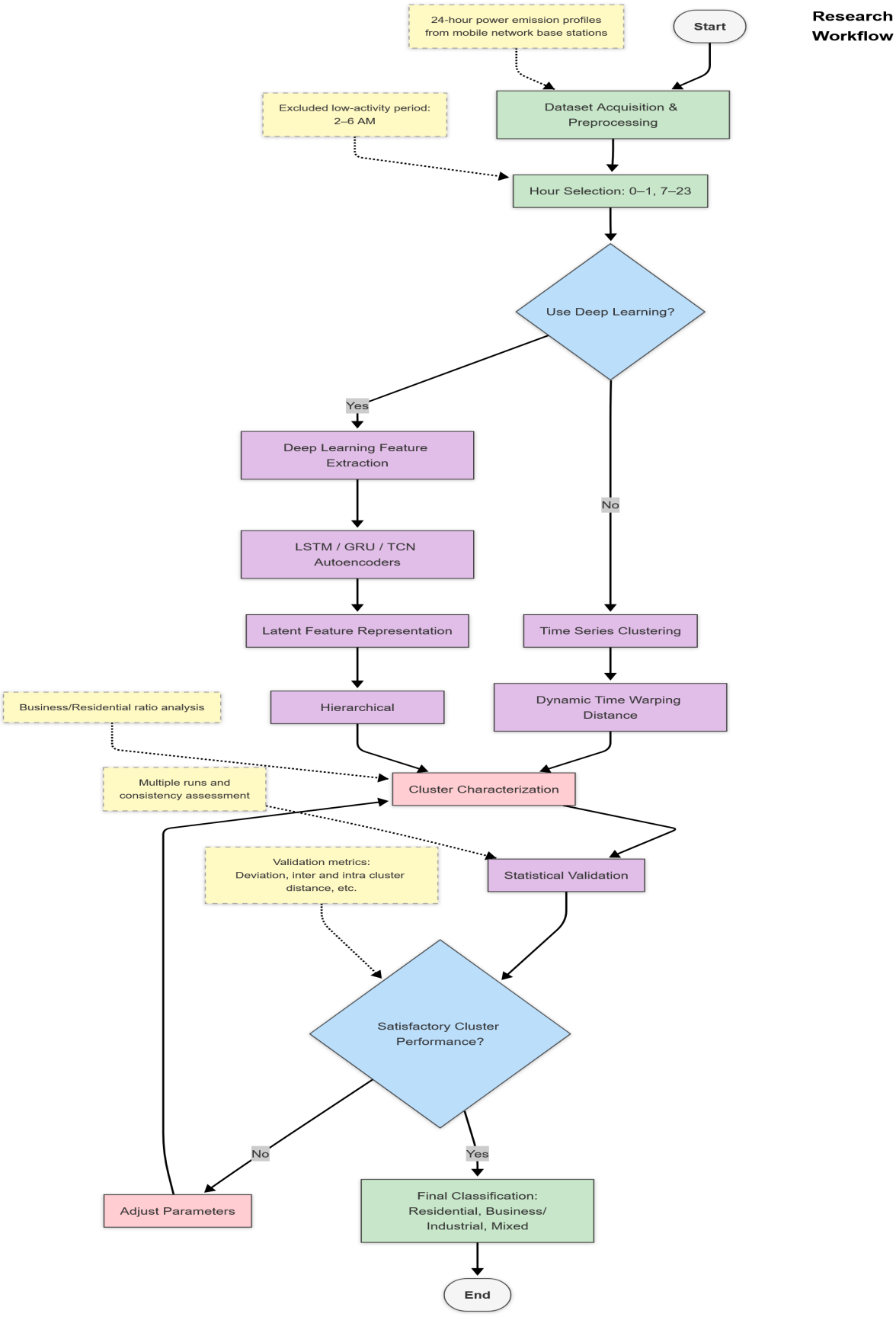

Figure 3.1: High-level view of the complete methodology.

3.1.2 Data Preparation, Manipulation, and Handling

The initial phase of data handling involved loading the dataset, which was provided in an Excel file format (`Data-Traffic.xlsx`). Upon loading, the dataset was partitioned into two primary components: the `site` identifiers, which serve as labels for each base station, and the power emission data, which constitutes the features for analysis. The power data, originally spanning all 24 hours, was then subjected to a filtering process to select specific hours deemed most informative for distinguishing between different types of base station activity, as detailed in the subsequent section. Following hour selection, the power emission values for each base station across the chosen hours were normalized. This normalization was performed using Min-Max scaling, applied independently to each base station's time series. The purpose of normalization is to scale the power values to a common range ([0, 1]), which prevents features with larger magnitudes from disproportionately influencing distance-based algorithms and helps in stabilizing the training process for neural network models. The transformation ensures that the shape of the power emission profile, rather than the absolute power levels, is the primary factor in subsequent clustering and feature learning stages.

3.2 Hour Classification

Rationale for Hour Selection A critical step in the preprocessing pipeline was the strategic selection of specific hours for analysis. The original dataset provided power emission data for all 24 hours of the day. However, for the purpose of classifying base stations into residential, business/industrial, or mixed types, not all hours contribute equally or uniquely to distinguishing these patterns. In this study, the hours selected for analysis were 00:00-01:59 (hours 0 and 1) and 07:00-23:59 (hours 7 through 23 inclusive). Consequently, the period from 02:00 to 06:59 (hours 2 through 6) was excluded from the primary analysis.

The justification for this selection is twofold. Firstly, the power emission patterns during the late-night and early-morning hours (approximately 2 AM to 7 AM) were observed to be relatively low and exhibited minimal variance across a majority of the base stations. During these off-peak times, human activity, and consequently mobile network usage, is generally at its nadir. As such, the power profiles in this window tend to be flat and similar, offering little discriminatory information for distinguishing between residential, business, or mixed usage patterns. Including these hours could potentially introduce noise or dampen the distinct characteristics observable during more active periods.

Secondly, the primary differences in usage patterns between residential and business/industrial areas manifest most strongly during typical waking and working hours, as well as evening residential activity hours. Business and industrial areas are expected to show peak activity during standard working hours (e.g., 9 AM to 5 PM), while residential areas typically exhibit higher usage in the evenings and early mornings. The selected hours (0-1 and 7-23) capture these crucial periods: the early morning hours (0-1) can reflect late-night residential activity or early-morning routines, the daytime hours (7-18) cover the typical business/industrial operational window, and the evening hours (19-23) capture peak resi-

dential usage. By focusing on these more dynamic periods, the aim is to enhance the signal-to-noise ratio, thereby allowing the clustering algorithms to more effectively discern the underlying behavioral typologies of the base stations.

3.3 Algorithms

To classify the base stations, a suite of algorithms was employed, ranging from traditional time series analysis techniques to advanced deep learning models for feature representation. Each algorithm was chosen for its specific strengths in handling temporal data and uncovering latent patterns.

3.3.1 Dynamic Time Warping (DTW)

Working Principle and Equation. Dynamic Time Warping (DTW) is a well-established algorithm used to measure the similarity between two temporal sequences that may vary in time or speed [32]. Unlike Euclidean distance, which compares sequences point-by-point at identical time indices, DTW finds an optimal non-linear alignment between the sequences. This is particularly useful for time series data where patterns might be shifted or stretched along the time axis. Given two time series, $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_m)$, DTW constructs an $n \times m$ distance matrix D, where $D(i, j)$ is the distance (e.g., squared Euclidean distance) between x_i and y_j. The algorithm then finds a warping path $W = (w_1, w_2, \ldots, w_K)$, where $w_k = (i_k, j_k) \in [1..n] \times [1..m]$, that minimizes the total cumulative distance along the path. The path must satisfy boundary conditions (starting at $(1, 1)$ and ending at (n, m)), continuity (steps are to adjacent cells), and monotonicity (path does not go back in time). The DTW distance is the cumulative distance of the optimal warping path, typically computed using dynamic programming with the recurrence relation[32]:

$$\gamma(i, j) = d(x_i, y_j) + \min\{\gamma(i-1, j), \gamma(i, j-1), \gamma(i-1, j-1)\} \tag{3.1}$$

where $d(x_i, y_j)$ is the local cost measure between points x_i and y_j, and $\gamma(i, j)$ is the cumulative DTW distance up to (i, j).

Application in this Study. DTW was a cornerstone in this research, primarily because base station power emission profiles, even within the same category (e.g., residential), can exhibit temporal shifts. For instance, peak residential usage might occur slightly earlier or later in different areas, or business activity might have staggered start and end times. DTW's ability to handle such temporal misalignments makes it a more robust similarity measure than lock-step measures for this type of data. It was specifically utilized as the distance metric within the Time Series K-Means clustering algorithm, enabling a more accurate grouping of base stations based on the shape of their power profiles rather than strict temporal coincidence of peaks and troughs.

3.3.2 Time Series K-Means

Working Principle. Time Series K-Means is an adaptation of the traditional K-Means clustering algorithm specifically designed for time series data. While standard K-Means typically uses Euclidean distance to assign data points to the nearest cluster centroid and to update centroids, Time Series K-Means employs a time series-specific distance measure. In this study, DTW was used as this distance metric as stated in the study by Petitjean et al. (2011)[14]. The algorithm iteratively partitions the dataset of N time series into K clusters. In each iteration, each time series is assigned to the cluster whose centroid is closest (in terms of DTW distance). After all time series are assigned, the centroids are recomputed. For DTW, the centroid (or barycenter) computation is non-trivial and often involves specialized averaging techniques like DBA (DTW Barycenter Averaging).

Training Process. The training process for Time Series K-Means with DTW involves the following steps:

1. Initialization: K initial cluster centroids are selected from the dataset (e.g., randomly or using a k-means++ like strategy for time series).
2. Assignment Step: Each time series in the dataset is assigned to the cluster whose centroid is nearest, according to the DTW distance.
3. Update Step: The centroids of the K clusters are recalculated. This involves finding a representative time series (barycenter) for each cluster that minimizes the sum of DTW distances to all time series within that cluster.
4. Convergence Check: Steps 2 and 3 are repeated until the cluster assignments no longer change significantly, or a maximum number of iterations is reached.

In this study, the `tslearn` library's implementation of `TimeSeriesKMeans` was used, with `n_clusters` set to 3 (aiming for residential, business/industrial, and mixed), DTW as the metric, and a specified number of iterations (e.g., `max_iter=10`).

Application in this Study. Time Series K-Means with DTW was employed as one of the primary clustering methods to directly group the normalized base station power profiles. The goal was to identify distinct operational patterns corresponding to the target categories. The use of DTW was crucial for capturing the inherent temporal flexibility in power usage behaviors. The resulting clusters provided an initial classification of base stations.

3.3.3 LSTM Autoencoder for Feature Extraction

To potentially improve clustering performance by learning more discriminative representations of the time series, a Long Short-Term Memory (LSTM) based autoencoder was developed.

Working Principle. LSTMs are a type of Recurrent Neural Network (RNN) particularly well-suited for learning from sequential data, as they can capture long-range temporal dependencies through their gating mechanisms (input, forget, and output gates) [20]. An autoencoder is a neural network architecture trained to reconstruct its input. It consists of two main parts: an encoder that maps the input data to a lower-dimensional latent representation (encoding), and a decoder that reconstructs the original input from this latent representation. By training the autoencoder to minimize reconstruction error, the encoder learns to extract salient features from the input data.

Model Architecture and Training. The LSTM autoencoder implemented in this study was structured as follows:

- **Encoder**: The encoder part took the time series (normalized power data over selected hours, reshaped to include timesteps and features) as input. It consisted of two LSTM layers (e.g., 32 units and 16 units, with ReLU activation) followed by a Dense layer to produce the final encoding (e.g., 8 dimensions).
- **Decoder**: The decoder took the encoded representation as input. A `RepeatVector` layer was used to replicate the encoding across the time steps of the original sequence. This was followed by two LSTM layers (e.g., 16 units and 32 units, with ReLU activation, returning sequences) and a `TimeDistributed(Dense(features))` layer to reconstruct the time series back to its original shape.

The autoencoder model was compiled using the Adam optimizer and Mean Squared Error (MSE) as the loss function. Training was performed for a specified number of epochs (e.g., 100) with a defined batch size (e.g., 32). An early stopping callback was employed to monitor validation loss and prevent overfitting, restoring the weights from the epoch with the best validation performance.

Application in this Study. The primary application of the LSTM autoencoder was to learn a compressed, yet informative, latent representation (features) of each base station's power emission profile. Once the autoencoder was trained, the encoder part was used to transform the original time series data into these lower-dimensional encoded features. These learned features, which ideally capture the essential dynamics of the power profiles, were then used as input for subsequent clustering algorithms (e.g., Agglomerative Hierarchical Clustering) with the hypothesis that clustering in this learned feature space might yield more distinct and meaningful groupings than clustering directly on the raw or normalized time series data.

3.3.4 GRU-based Autoencoder

In addition to the LSTM autoencoder, a Gated Recurrent Unit (GRU) based autoencoder was implemented to explore alternative recurrent architectures for feature learning from the base station time series data.

Working Principle. GRUs, like LSTMs, are a type of gated RNN designed to overcome the vanishing gradient problem and capture long-range dependencies in sequential data. They achieve this with a simpler architecture than LSTMs, utilizing two main gates: an update gate and a reset gate. The update gate determines how much of the past information to keep and how much new information to add, while the reset gate decides how much of the past information to forget. This streamlined structure often leads to comparable performance with LSTMs but with greater computational efficiency [22].

Model Architecture and Training. The GRU autoencoder employed in this study was constructed with an encoder-decoder architecture analogous to the LSTM model:

- **Encoder**: The input time series (normalized power data with T timesteps and 1 feature) was fed into an encoder consisting of two GRU layers. The first GRU layer had 32 units with ReLU activation and returned sequences. This was followed by a second GRU layer with 16 units and ReLU activation. The output of the second GRU layer was then passed through a Dense layer with 8 units and ReLU activation to produce the 8-dimensional encoded representation.
- **Decoder**: The 8-dimensional encoded vector was the input to the decoder. A `RepeatVector` layer replicated this encoding T times. This sequence was then processed by two GRU layers: the first with 16 units and ReLU activation (returning sequences), and the second with 32 units and ReLU activation (also returning sequences). Finally, a `TimeDistributed(Dense(1))` layer was used to reconstruct the output sequence to match the original input's shape (1 feature per timestep).

The GRU autoencoder model was compiled using the Adam optimizer and Mean Squared Error (MSE) as the loss function. It was trained for 100 epochs with a batch size of 32. The training data was shuffled at each epoch, and 20% of the data was used for validation. An early stopping callback was implemented with a patience of 10 epochs, monitoring the validation loss and restoring the model weights that yielded the best performance on the validation set.

Application in this Study. The GRU autoencoder served the same purpose as the LSTM variant: to learn a compact and informative latent representation of the base station power profiles. After training, the encoder component was utilized to transform the input time series into these 8-dimensional feature vectors. These GRU-derived features were then used as input for Agglomerative Hierarchical Clustering to identify distinct operational patterns among the base stations, providing an alternative set of deep learned features for comparison with those from the LSTM and TCN models.

3.3.5 TCN-based Autoencoder

As a further alternative to recurrent architectures, a Temporal Convolutional Network (TCN) [33] based autoencoder was implemented for feature extraction from the power emission time series.

Working Principle. Temporal Convolutional Networks leverage causal, dilated convolutions to efficiently capture long-range temporal dependencies in sequence data, offering an alternative to RNNs.

Causal convolutions ensure that the prediction for a given timestep only depends on past and current inputs, not future ones. Dilated convolutions enable an exponentially large receptive field with increasing network depth without a proportional increase in parameters or computational cost, by skipping input values with a certain step (dilation rate). TCNs typically stack these convolutional layers, often incorporating residual connections and normalization to facilitate training deeper networks [33][24].

Model Architecture and Training. The TCN autoencoder implemented in this study comprised an encoder and a decoder built with 1D convolutional layers:

- **Encoder**: The input time series (normalized power data with T timesteps and 1 feature) was processed by two `Conv1D` layers. The first `Conv1D` layer had 32 filters, a kernel size of 3, ReLU activation, `same` padding, and a dilation rate of 1. The second `Conv1D` layer had 16 filters, a kernel size of 3, ReLU activation, `same` padding, and a dilation rate of 2. The output of these convolutional layers was then flattened using a `Flatten` layer, followed by a `Dense` layer with 8 units and ReLU activation to produce the 8-dimensional encoded feature vector.
- **Decoder**: The 8-dimensional encoded vector was first passed through a `Dense` layer to expand it to a size suitable for reshaping (e.g., $16 \times T$ units, with ReLU activation). This was then reshaped using a `Reshape` layer to have T timesteps and 16 features. This sequence was processed by two `Conv1D` layers: the first with 16 filters, kernel size 3, ReLU activation, `same` padding, and dilation rate 2; the second with 32 filters, kernel size 3, ReLU activation, `same` padding, and dilation rate 1. Finally, an output `Conv1D` layer with 1 filter (to match the input feature dimension), kernel size 3, sigmoid activation (to ensure output is in [0,1] like the normalized input), and `same` padding was used to reconstruct the time series.

The TCN autoencoder model was compiled with the Adam optimizer and Mean Squared Error (MSE) as the loss function. Training was conducted for 100 epochs using a batch size of 32. Data was shuffled before each epoch, and 20% of the training data was allocated for validation. An early stopping mechanism with a patience of 10 epochs was employed, based on validation loss, to prevent overfitting and retain the best performing model weights.

Application in this Study. Similar to the LSTM and GRU autoencoders, the TCN autoencoder was utilized to learn a condensed feature representation from the base station power profiles. The encoder part of the trained TCN model transformed the input time series into 8-dimensional feature vectors. These TCN-derived features were subsequently used as input for Agglomerative Hierarchical Clustering to identify distinct groups of base stations, providing a convolution-based feature learning perspective for comparison with the recurrent approaches.

3.3.6 Clustering Methods and Types

Following either direct application on time series or feature extraction using autoencoders, clustering techniques were applied to group the base stations.

3.3.7 Hierarchical Clustering (on Encoded Features)

Working Principle. Agglomerative Hierarchical Clustering is a bottom-up clustering approach. It starts by treating each data point (in this case, each base station's encoded feature vector) as a separate cluster [34]. Then, in each step, it merges the two closest clusters until only a single cluster (containing all data points) remains or a predefined number of clusters is reached. The closeness of clusters is determined by a linkage criterion, such as Ward's method, which minimizes the total within-cluster variance. Ward's method aims to find compact, spherical clusters.

Application in this Study. After obtaining the encoded features from the LSTM autoencoder (and potentially GRU/TCN autoencoders), Agglomerative Hierarchical Clustering was applied to these feature vectors. Here `AgglomerativeClustering` from `scikit-learn` was used with a target number of communities/clusters (e.g., `n_clusters=5` initially, though the final goal is 3 types) and Ward linkage. This approach was used to identify groups, termed communities, based on the learned representations. The rationale is that these features might offer a more abstract and potent basis for grouping than the raw time series.

3.4 Cluster Characterization and Interpretation

Regardless of the specific clustering algorithm employed (Time Series K-Means or Hierarchical Clustering on encoded features), a crucial subsequent step was the characterization and interpretation of the resulting clusters. The objective was to assign a meaningful label, namely Residential, Business/Industrial or Mixed to each identified cluster based on its average power emission profile.

To achieve this, the mean power profile for each cluster was calculated by averaging the normalized power values of all member base stations at each selected time point. These mean profiles represent the archetypal power usage pattern for each cluster.

The classification of these mean profiles into the predefined categories was based on a heuristic derived from typical activity patterns. Specifically, a distinction was made between *business hours* (defined, for example, as 08:00 to 18:00) and *night/residential hours* (e.g., 19:00 to 01:00, and potentially early morning hours before the business day starts, while excluding the very low activity period of 02:00-06:00). The average normalized power during business hours was compared to the average normalized power during night/residential hours for each cluster's mean profile.

A ratio was computed:

$$\text{Ratio} = \frac{\text{Average Power during Business Hours}}{\text{Average Power during Night/Residential Hours} + \epsilon}$$

where ϵ is a small constant to prevent division by zero.

Based on this ratio, clusters were categorized as follows:

- **Business/Industrial**: If the ratio was significantly greater than 1 (e.g., Ratio ¿ 1.5), indicating substantially higher power consumption during business hours compared to night hours. This pattern

is characteristic of areas dominated by commercial offices, industrial plants, or other business-related activities that are primarily operational during the day.

- **Residential**: If the ratio was significantly less than 1 (e.g., Ratio ¡ 0.8), indicating higher power consumption during evening/night hours and potentially early mornings, compared to core business hours. This profile aligns with residential areas where mobile usage peaks outside of typical work hours.
- **Mixed**: If the ratio fell within an intermediate range (e.g., $0.8 \leq$ Ratio ≤ 1.5), suggesting that the power consumption was more balanced between business and night/residential hours, or did not show a strong dominance of one over the other. Such profiles could represent areas with a blend of residential and commercial establishments, or base stations whose coverage spans both types of zones.

This characterization provides a qualitative understanding of the types of environments served by the base stations within each cluster, directly addressing the research objective.

Statistical Significance

The clustering of time series data, particularly when employing algorithms with stochastic components (e.g., K-Means initialization, neural network weight initialization and training) or when comparing multiple algorithmic approaches, necessitates a robust evaluation of result stability and significance. A single run of a clustering algorithm might yield a partition that is not representative or is sensitive to initial conditions. Therefore, to ensure the reliability and generalizability of the base station classifications derived in this study, a comprehensive statistical validation framework was adopted. This involved executing the clustering pipelines multiple times and employing statistical tests to assess the consistency of classifications and the agreement between different models. The primary motivation for this rigorous approach is to provide confidence that the identified clusters (Residential, Business/Industrial, Mixed) are not artefacts of random chance but reflect genuine underlying patterns in the base station power emission data. An overview of the statistical measures carried out is outlined in figure 3.2.

3.5 Stability Assessment via Multiple Runs and Probabilistic Assignment

To evaluate the stability of the classifications produced by each of the employed clustering approaches (Time Series K-Means, LSTM-based, GRU-based, and TCN-based clustering), each method was executed 15 independent times. For each run, every base station was assigned to one of the three predefined categories: Residential, Business/Industrial, or Mixed. It is to be mentioned that, for fair comparison among all the algorithms, a fixed seed was maintained in a single run, which is always altered in each subsequent run, providing a fair ground for all the algorithms along with variability in different runs.

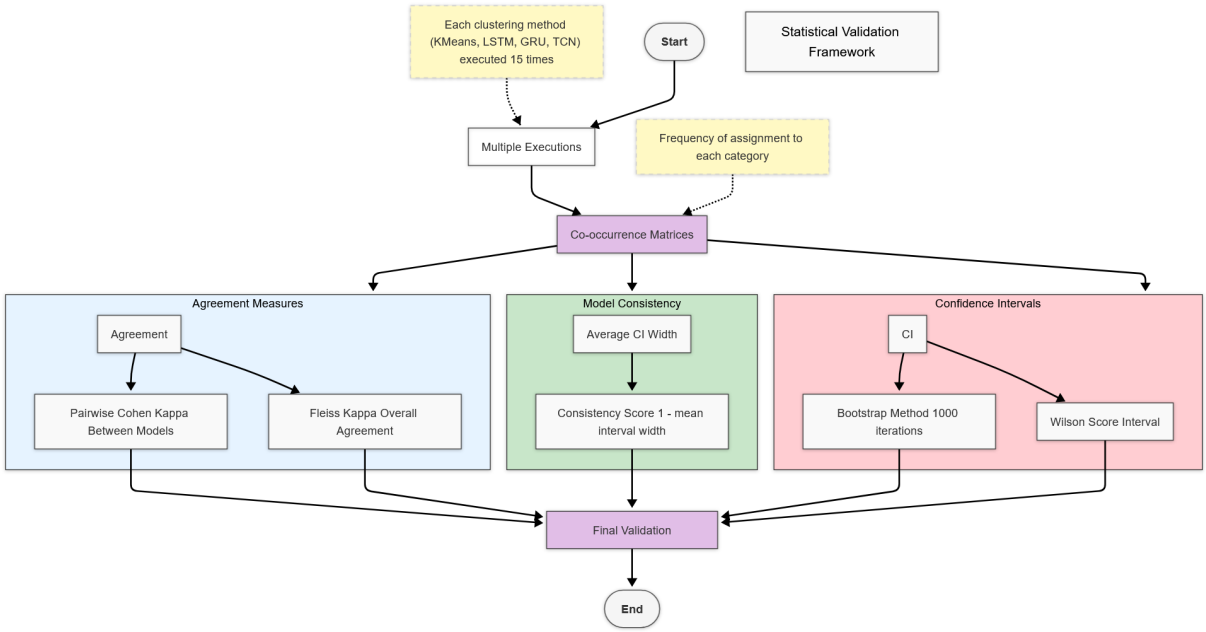

Figure 3.2: Statistical Measures for Ensuring Consistency.

Following these multiple runs, co-occurrence matrices were constructed for each of the four models (KMeans, LSTM, GRU, TCN). These matrices quantified, for each base station, the frequency with which it was assigned to each of the three categories across the 15 runs. These frequencies were then converted into probabilities, representing the empirical likelihood of a given base station belonging to a specific category as determined by a particular model. For a base station BS_i and a model M_j, the probability of being classified into category C_k was calculated as:

$$P(BS_i \in C_k | M_j) = \frac{\text{Number of times } BS_i \text{ classified as } C_k \text{ by } M_j}{\text{Total number of runs (15)}}$$

This probabilistic method of assignment is a commonly used empirical strategy to capture classification consistency and uncertainty across multiple stochastic runs, even though it is not directly adapted from a specific study. This provides a more nuanced view of classification certainty than a single, deterministic assignment.

3.5.1 Confidence Intervals for Classification Proportions

To quantify the uncertainty associated with the classification probabilities derived from the multiple runs, confidence intervals were calculated. Two methods were employed for this purpose: the Wilson score interval and bootstrap resampling.

Wilson Score Interval

The Wilson score interval is a method for calculating a confidence interval for a binomial proportion [35]. It is particularly well-suited for small sample sizes (such as our 15 runs) and for proportions that are close to 0 or 1, where the normal approximation (Wald interval) can be unreliable. For each base station

and each model, a 95% Wilson score confidence interval was computed for the probability of assignment to each of the three categories. The formula for the Wilson score interval for a proportion with number of successes and the number of trials is given by:

$$\text{CI} = \frac{1}{1 + \frac{z^2}{n}} \left(\hat{p} + \frac{z^2}{2n} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}} \right) \tag{3.2}$$

where,

- CI: Wilson score confidence interval for the estimated proportion
- $\hat{p} = \frac{x}{n}$: Sample proportion of successes, with x successes out of n trials
- z: Critical value from the standard normal distribution corresponding to the desired confidence level (e.g., $z \approx 1.96$ for 95% confidence)
- n: Total number of trials

These intervals provide a range within which the true underlying probability of a base station belonging to a category is likely to fall, given the observed variability across runs.

Bootstrap Confidence Intervals

As a complementary approach, bootstrap confidence intervals were also computed. Bootstrapping is a non-parametric resampling technique that can be used to estimate the sampling distribution of a statistic. For each base station and each model, the list of its 15 category assignments was resampled with replacement 1000 times (bootstrap iterations). For each bootstrap sample, the proportion of assignments to each category was calculated. The 95% confidence interval was then determined by taking the 2.5th and 97.5th percentiles of the distribution of these bootstrap proportions. The bootstrap CIs provide an alternative measure of uncertainty and can serve as a validation for the Wilson score intervals, particularly as they make fewer assumptions about the underlying distribution. The comparison between Wilson and bootstrap CIs for example base stations showed largely consistent results, lending further credence to the stability estimates.

3.5.2 Model Consistency Evaluation

To provide a quantitative measure of how consistent each clustering model (KMeans, LSTM, GRU, TCN) was in its classifications across the 15 runs, a model consistency score was derived. This score was based on the average width of the Wilson score confidence intervals associated with the classifications made by each model. For a given model, the widths of the 95% confidence intervals (upper bound - lower bound) were calculated for each base station and each category to which the base station was assigned with non-zero probability. The average of these widths was then computed. A model that consistently assigns base stations to the same categories across runs will have narrower confidence intervals, and thus a smaller average width. The consistency score was defined as $1 -$ mean interval width, making

higher scores indicative of better consistency. This metric allows for a direct comparison of the stability of the different algorithmic approaches.

Inter-Model Agreement

Beyond assessing the internal consistency of each model, it is also important to evaluate the extent to which different clustering models agree on their classifications of the base stations. This was assessed using Fleiss Kappa and pairwise Cohen Kappa statistics, based on the most frequent (modal) classification for each base station by each model across the 15 runs.

Fleiss Kappa

Fleiss Kappa ($kappa$) is a statistical measure for assessing the reliability of agreement between a fixed number of raters when assigning categorical ratings to a number of items or classifying items. In this context, the raters are the four clustering models (KMeans, LSTM, GRU, TCN), and the items are the base stations. Fleiss Kappa corrects for the amount of agreement that would be expected by chance. The value of
$kappa$ ranges from -1 to 1, where 1 indicates perfect agreement, 0 indicates agreement no better than chance, and negative values indicate agreement worse than chance. The calculation involves determining the proportion of all possible pairs of assignments that are in agreement, corrected for the proportion of agreement expected by chance. This provides an overall measure of concordance across all four models.

Pairwise Cohen Kappa

To obtain a more granular understanding of inter-model agreement, Cohen Kappa coefficients [36] were calculated for all pairs of models (e.g., KMeans vs. LSTM, LSTM vs. GRU, etc.). Cohen Kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories. Similar to Fleiss Kappa, it corrects for chance agreement. The formula for Cohen Kappa is:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the observed proportionate agreement between the models, and p_e is the probability of chance agreement. The results of these pairwise comparisons were typically visualized as a heatmap, allowing for easy identification of model pairs that exhibit high or low agreement. This helps in understanding which models produce similar or divergent clustering solutions.

3.5.3 Summary of Assessment

The statistical validation procedures undertaken provide critical insights into the robustness and reliability of the base station classification. The multiple runs and subsequent calculation of classification probabilities with confidence intervals (Wilson score and bootstrap) offer a measure of certainty for each

base station's assignment by each model. The model consistency scores allow for a comparative assessment of the stability of the different clustering algorithms employed. Furthermore, the inter-model agreement analyses, using Fleiss Kappa for overall agreement and pairwise Cohen Kappa for specific model comparisons, shed light on the consensus (or lack thereof) between the different methodological approaches (traditional time series clustering vs. deep learning-based feature extraction followed by clustering). These statistical evaluations are essential for substantiating the findings of the study, ensuring that the proposed classification of base stations into Residential, Business/Industrial, and Mixed categories is well-grounded and that the performance characteristics of the applied algorithms are thoroughly understood. The results from these tests, such as specific Kappa values and consistency scores, are detailed in the results section of the paper and inform the discussion on the most reliable approach for the given task.

3.6 Cluster Profiles Statistical Analysis

3.6.1 Summary Statistics of Hourly Traffic Profiles

To provide insight into the behaviour of base stations across different clusters, this analysis uses statistical visualisation techniques, primarily box plots and line plots with confidence intervals. These tools allow for a structured understanding of the patterns in the data while enabling comparison between categories such as residential, mixed, and commercial base station clusters.

The first type of visualisation employed is the box plot, which offers a compact summary of the distribution of transmit power values.Transmit power data from base stations was first normalised to allow for comparison on a relative scale, removing the influence of absolute magnitude differences between base stations. Each value x was transformed using min-max normalisation:

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{3.3}$$

This ensures that all values fall within the range $[0, 1]$, facilitating an unbiased comparison across different sites.

Box plots represent each cluster's distribution using the five-number summary: minimum, first quartile (Q_1), median (Q_2), third quartile (Q_3), and maximum. The interquartile range (IQR), defined as $Q_3 - Q_1$, captures the central spread of the data, while whiskers extend up to 1.5 times the IQR. Data points beyond this range are plotted individually as outliers. This type of plot is particularly helpful for identifying differences in central tendency and variability among the cluster types. In the case of normalised transmit power, these plots reveal which clusters exhibit higher peak activity, more consistency, or greater fluctuation in usage.

To further characterise the daily patterns of transmit power, time-series line plots were constructed, showing the mean transmit power across each hour of the day for all clusters. For each hourly point, the mean power of all base stations in a given cluster was calculated. In addition to the mean values, the plots include 95% confidence intervals to account for variability and to estimate the reliability of observed

differences. These intervals are computed using the standard formula for the confidence interval of a sample mean:

$$\text{CI} = \bar{x}_i \pm t_{\alpha/2,\, n_i-1} \cdot \frac{s_i}{\sqrt{n_i}} \tag{3.4}$$

where:

- CI: Confidence interval for the mean transmit power at each hour
- $\bar{x}_i$: Sample mean for cluster i at a specific hour
- s_i: Sample standard deviation of the transmit powers in cluster i
- n_i: Number of base stations in cluster i
- $t_{\alpha/2,\, n_i-1}$: Critical value from the Student's t-distribution for a 95% confidence level with $n_i - 1$ degrees of freedom

This approach allows us to assess whether the observed differences in hourly average power between clusters are statistically meaningful or potentially due to sampling variability. Line plots were created for both the full 24-hour day and a reduced window that excludes hours 2-7, which typically represent off-peak or inactive periods in the network. This reduction in hours aims to enhance the visibility of behavioural differences during active times by removing low-variance periods. The confidence bands become narrower in the reduced-hour plots, improving interpretability while preserving the underlying diurnal trends across cluster types.

3.6.2 Peak Hour Analytical Framework

In the context of telecommunications, peak hours are the periods during the day when the majority of internet users in a given area are simultaneously accessing the network. During peak hours, the increased network traffic can strain the infrastructure, leading to slower connection speeds, higher latencies, and potential buffering issues for users. Therefore, internet service providers often take measures to manage network congestion during peak hours, such as implementing traffic shaping techniques or offering plans with higher bandwidth caps. This is why it is necessary to identify the peak hour for a network cell. Peak hour can also indicate the profile of the base station. Based on the time and duration of the peak hour, we can identify base stations into residential, business, and mixed.

Peak Hour Traffic (PHT) is a vital metric for telecommunications systems, helping to ensure that resources are adequately allocated and system performance remains optimal during the busiest times. Peak Hour Traffic (PHT) refers to the highest volume of traffic offered to a telecommunications system during a specific period, often referred to as peak traffic. This metric is crucial for understanding the busiest times of day when the system experiences the most load. Understanding Peak hour traffic is essential for several reasons:

- **Resource Allocation:** Ensures that sufficient resources are available during peak times to handle the load.

- **System Performance:** Helps in maintaining optimal system performance and avoiding congestion.
- **Customer Satisfaction:** Reduces the likelihood of dropped calls, slow data speeds, and improves user experience.

Understanding peak hour patterns for distinct base station clusters enables optimized resource allocation, as clusters exhibit varying peak hour timing and duration. **Peak Hour Traffic (PHT)** is influenced by multiple factors:

- **Temporal Variation:** Traffic patterns correlate with time-of-day trends, such as elevated usage in commercial zones during late morning and residential areas during evenings.
- **Event-Driven Demand:** Transient traffic spikes may arise from localized events, promotions, or emergencies.
- **Behavioral Shifts:** Long-term changes in user behavior, including remote work adoption, can alter PHT dynamics.

To mitigate PHT challenges, cluster-specific strategies are recommended:

- **Load Balancing:** Distributing traffic across servers to prevent congestion.
- **Capacity Planning:** Proactive infrastructure scaling based on historical and predicted demand.
- **Traffic Prioritization:** Implementing quality-of-service policies to prioritize critical applications during peak periods.

Confidence interval

Following the classification of base stations into residential, commercial, and mixed clusters, a detailed peak hour analysis is performed for each cluster. For every base station within a cluster, the specific hour exhibiting the highest transmit power is identified and recorded. This process yields a distribution of peak hours for each cluster type. Subsequently, statistical analysis is conducted by calculating the mean and standard deviation of the peak hours within each cluster, as well as determining the 95% confidence interval to assess the temporal consistency of peak demand. These results are then interpreted in the context of telecommunications network management, providing insights that inform capacity planning and the development of traffic management protocols tailored to the distinct temporal profiles of each cluster.

The mean (average) peak hour for each group is calculated by treating each hour as a numerical value (e.g., 21 for 21:00), and the result is then interpreted by converting it back into hour format.

A 95% confidence interval for this mean is computed using the sample standard deviation and the sample size. This interval provides a range within which the true average peak hour for the group is likely to fall.

The confidence interval is given by:

$$\text{CI} = \bar{x}_i \pm t_{\alpha/2,\, n_i-1} \cdot \frac{s_i}{\sqrt{n_i}} \tag{3.5}$$

where:

- CI: Confidence interval for the average peak hour
- n_i: Number of base stations in group i
- s_i: Sample standard deviation of peak hours in group i
- $\bar{x}_i$: Sample mean peak hour for group i
- $t_{\alpha/2,\, n_i-1}$: t-value from the Student's t-distribution for 95% confidence with $n_i - 1$ degrees of freedom

3.7 Statistical Significance of Inter-Cluster differences

In order to analyze whether the mean peak hour transmit power differs significantly among the three clusters identified by KMeans (Residential, Mixed, and Business/Industrial), we employed statistical hypothesis testing methods, specifically one-way ANOVA and Welch's t-test.

3.7.1 One-Way ANOVA

One-way Analysis of Variance (ANOVA) is a statistical method used to compare the means of three or more independent groups to determine if at least one group mean is significantly different from the others. The null hypothesis H_0 assumes that all group means are equal:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

where k is the number of groups.

The test statistic in ANOVA is the F-statistic, which is calculated as the ratio of the variance between the groups to the variance within the groups:

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} \tag{3.6}$$

Here,

- $MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}}$ is the mean square between groups, representing the variability of the group means around the overall mean.
- $MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}}$ is the mean square within groups, representing the variability of observations within each group.

- SS_{between} is the sum of squares between groups:

$$SS_{\text{between}} = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$$

  where n_i is the number of observations in group i, $\bar{x}_i$ is the group mean, and $\bar{x}$ is the overall mean.

- $SS_{\text{within}} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ is the sum of squared deviations within groups.

- Degrees of freedom: $df_{\text{between}} = k-1$, $df_{\text{within}} = N-k$, where N is the total number of observations.

If the calculated F-statistic is greater than the critical value from the F-distribution (or equivalently, the p-value is below a significance threshold such as 0.05), the null hypothesis is rejected, indicating that at least one cluster's mean peak hour transmit power significantly differs.

In this project, the one-way ANOVA was applied to test for differences in the average peak hour transmit power among the three clusters. The rejection of the null hypothesis justified further pairwise comparisons.

3.7.2 Welch's t-test

While ANOVA assumes equal variances across groups, Welch's t-test is a robust alternative for comparing the means of two groups when the assumption of equal variances may not hold. This test compares two independent samples and adjusts the degrees of freedom accordingly.

The Welch's t-test statistic is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3.7}$$

where

- $\bar{x}_1, \bar{x}_2$ are the sample means of groups 1 and 2
- s_1^2, s_2^2 are the sample variances
- n_1, n_2 are the sample sizes

The degrees of freedom for the test are approximated by the Welch-Satterthwaite equation:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{3.8}$$

A significant t-statistic (with a corresponding low p-value) suggests that the mean peak hour transmit power between the two compared clusters is statistically different, without requiring the assumption of equal variances.

In this project, Welch's t-test was used to perform pairwise comparisons between clusters to complement the ANOVA results and provide a more robust understanding of the differences in temporal usage patterns.

3.8 Clustering Evaluation Metrics and Model Comparison

This section presents a comprehensive evaluation of the clustering models developed in this study. The models; Time Series KMeans (with DTW), LSTM Autoencoder, GRU Autoencoder, and TCN Autoencoder, were rigorously compared using a suite of distance- and similarity-based metrics. These metrics provide internal validation by assessing the compactness of clusters and the separation between operational behaviors of mobile base stations, particularly focusing on Residential and Business patterns, without relying on external ground truth. The analysis was performed on normalized time series data (scaled per profile using Min-Max scaling), ensuring that clustering captured underlying temporal patterns rather than absolute magnitudes.

3.8.1 Evaluation Metrics

The key quantitative metrics used to evaluate both intra-cluster compactness and inter-cluster separation are as follows:

- **Mean Relative Deviation (MRD)**: Quantifies how much a cluster's members deviate from another cluster's mean profile. Lower is better in intra-cluster compactness because it implies tighter grouping.
- **Mean Absolute Error (MAE)**: Measures the average magnitude of error without considering its direction. Lower MAE indicates closer matching between base station profiles and their assigned cluster centroids. A normalized variant of MAE has been used that accounts for the relative difference, ensuring comparability across profiles of different scales.
- **Mean Squared Error (MSE)**: The square of MAE without the final square root. It emphasizes larger deviations by penalizing them more heavily.
- **Pearson Correlation Coefficient**: Measures linear shape similarity between time series, independent of magnitude. A higher value (closer to 1) denotes strong alignment in temporal patterns.
- **Cosine Similarity**: Captures the cosine of the angle between two vectors, focusing on the similarity in shape orientation regardless of magnitude.
- **Euclidean Distance**: The straight-line distance in feature space, offering an absolute measure of profile similarity.

3.8.2 Evaluation Metric Formulas and Interpretations

For each profile $x(n)$ in cluster i, the following metrics were computed against the cluster centroid profile $\hat{x}(n)$ of cluster j. Here, N is the number of time samples in the profile, and ϵ is a small constant to avoid division by zero.

Among these metrics, we propose a novel measure called *Mean Relative Deviation (MRD)*3.9, designed to capture the average relative variation between individual base station profiles and their corresponding cluster centroid. This metric provides an intuitive, scale-independent indicator of how consistently base stations within a cluster follow the typical power usage pattern.

$$\text{Mean Relative Deviation (MRD)}: \quad \text{MRD} = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{x(n)-\hat{x}(n)}{\hat{x}(n)+\epsilon}\right| \tag{3.9}$$

$$\text{Mean Absolute Error (MAE)}: \quad \text{MAE} = \frac{1}{N}\sum_{n=1}^{N}|x(n)-\hat{x}(n)| \tag{3.10}$$

$$\text{Mean Squared Error (MSE)}: \quad \text{MSE} = \frac{1}{N}\sum_{n=1}^{N}(x(n)-\hat{x}(n))^2 \tag{3.11}$$

$$\text{Pearson Correlation Coefficient}: \quad r = \frac{\sum_{n=1}^{N}(x(n)-\bar{x})\left(\hat{x}(n)-\bar{\hat{x}}\right)}{\sqrt{\sum_{n=1}^{N}(x(n)-\bar{x})^2}\sqrt{\sum_{n=1}^{N}\left(\hat{x}(n)-\bar{\hat{x}}\right)^2}} \tag{3.12}$$

$$\text{Cosine Similarity}: \quad \cos\theta = \frac{\sum_{n=1}^{N}x(n)\hat{x}(n)}{\sqrt{\sum_{n=1}^{N}x(n)^2}\sqrt{\sum_{n=1}^{N}\hat{x}(n)^2}} \tag{3.13}$$

$$\text{Euclidean Distance}: \quad d_E = \sqrt{\sum_{n=1}^{N}(x(n)-\hat{x}(n))^2} \tag{3.14}$$

Symbol Definitions:

- $x(n)$: Value at time index n of the base station profile.
- $\hat{x}(n)$: Value in time index n of the centroid profile of the cluster.
- N: Total number of time samples.
- $\bar{x}$, $\bar{\hat{x}}$: Mean of $x(n)$ and $\hat{x}(n)$, respectively.
- ϵ: Small constant to prevent division by zero.
- d_E: Euclidean distance.
- $\cos\theta$: Cosine similarity value.
- r: Pearson correlation coefficient.

These equations comprehensively capture both magnitude-based error measures and shape-based similarities, providing a robust framework for assessing clustering quality.

Chapter 4

Results and Analysis

4.1 Methodology Overview

To classify base stations into *residential*, *business/industrial*, or *mixed* categories, we evaluated four models:

- **Time Series K-Means with DTW (KMEANS)**: Traditional clustering approach using Dynamic Time Warping to align temporal shifts in traffic patterns.
- **LSTM, GRU, and TCN Autoencoders**: Deep learning-based autoencoders were trained to extract latent features for clustering.

Evaluation Framework:

- *Same-Cluster Compactness:* Assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE), Pearson Correlation Coefficient, and Euclidean Distance. Higher correlation and lower error values indicate tighter clusters.
- *Cross-Cluster Separability:* Evaluated by computing MAE, MSE, and correlation between different clusters. Higher error and lower correlation indicate better separation.

Data Variants:

- *Full 24-Hour Profiles:* Includes all hourly traffic features (0-23).
- *Reduced Hour Profiles:* Focuses on hours with higher traffic variability, specifically 0-1 AM and 7-11 PM, excluding the low-variance nighttime period (2-6 AM).

4.2 Robustness Assessment via Multiple Runs

Three Experimental Designs To assess the robustness and consistency of each clustering model, three experimental approaches were conducted:

Single Run

Purpose: This scenario represented a benchmark of best-case model performance without considering variability.

Limitation: Results from a single run may be unrepresentative. For example, the LSTM model achieved an unusually high correlation of **0.936** in one run, which was not consistent across repetitions.

15 Runs Without Seeding

Purpose: This experiment tested the stability of each model under random initialization conditions.

Observation: The autoencoder-based models exhibited high variance. For example, the LSTM model showed a wide Pearson correlation range between **0.728** and **0.886**, depending on the initialization. In contrast, the KMEANS model demonstrated more consistent behavior, with a narrower correlation range of **0.698** to **0.796**.

15 Runs With Seeding

Purpose: Seeding was used to control random initialization, allowing a more direct comparison of algorithmic performance.

Justification for Final Analysis: Seeding enables reproducibility while preserving natural variation in the dataset. For instance, the seeded KMEANS model consistently achieved strong separability, with cross-cluster MAE of **0.113 ± 0.001** compared to **0.112 ± 0.003** without seeding. The full and reduced-hour metric comparisons (e.g., MAE, RMSE, Correlation) for all models can be found in Appendix 5, highlighting consistent trends across all 15 seeded runs.

Observation: Despite the use of seeds, autoencoder models such as TCN still struggled separability, with correlation dropping to **0.084** (vs. **0.078** without seeding), suggesting inherent limitations in feature discrimination.

4.3 Detailed Findings

4.3.1 Model Comparison: Compactness vs. Separation

The evaluation revealed a distinct trade-off between cluster compactness and inter-cluster separability across the applied models. Autoencoder-based approaches namely LSTM, GRU, and TCN, exhibited consistently superior intra-cluster cohesion. Among these, the LSTM model achieved the lowest mean absolute error (MAE) within clusters, with a value of 0.083 under the reduced-hour configuration. This reflects the model's capacity to encode temporal profiles into smooth and compact latent representations, effectively reducing reconstruction error and enhancing internal consistency within each cluster.

However, this compactness came at the cost of reduced separability. Specifically, the LSTM model demonstrated limited ability to distinguish between distinct base station types, particularly in the presence of mixed-use traffic patterns. This limitation was most evident in the reduced-hour evaluation,

where LSTM's cross-cluster correlation dropped to 0.0675, indicating significant overlap between cluster representations. Such degradation in separation performance can be attributed to the over-smoothing of discriminative temporal features during the encoding process, a common drawback of autoencoder architectures when applied to highly variable time series data.

In contrast, the K-means model combined with Dynamic Time Warping (DTW) achieved stronger inter-cluster separation. It yielded a higher cross-cluster MAE of 0.113, suggesting a greater ability to distinguish between base station categories. This performance stems from DTW's robustness in aligning temporal patterns with phase shifts, such as morning versus mid-morning peak traffic across business and mixed-use sites. The incorporation of DTW enabled K-means to preserve meaningful differences in usage patterns that autoencoders tended to smooth out.

Furthermore, an independent deviation analysis reinforced the earlier findings. By computing the mean relative deviation (MRD) between average cluster profiles and individual base station curves, we quantitatively confirmed that LSTM offers the most compact representations within residential and business clusters. For instance, the LSTM model achieved intra-cluster MRDs of 0.295 (Residential) and 0.309 (Business), outperforming all other models. Meanwhile, K-means+DTW exhibited superior separation: its off-diagonal deviations between Residential and Business were as high as 1.351, clearly surpassing LSTM's 1.333. This pattern was consistent across both GRU and TCN as well. Such stark separation margins highlight K-means' effectiveness in distinguishing dissimilar traffic patterns. Ultimately, these findings validate the dual strengths of LSTM in learning smooth latent spaces for compact clusters and K-means in capturing structural diversity across base station categories. The observed metrics from deviation matrices, summarized in Table 4.1, offer robust empirical support for these conclusions. Notably, the adoption of a reduced-hour feature set excluding low-variance nighttime intervals resulted in an 8% improvement in K-means separation performance. This supports the conclusion that removing non-informative temporal dimensions enhances model discriminability. Meanwhile, autoencoder compactness metrics remained relatively stable across full and reduced-hour scenarios, with less than 2% variation in MAE. This suggests that hour reduction selectively amplified class separation for K-means without compromising the internal cohesion achieved by autoencoders.

Overall, the findings underscore a fundamental dichotomy: while autoencoder models are well-suited for ensuring internal profile consistency within clusters, the K-means with DTW approach offers a more effective mechanism for distinguishing operational differences among base station types. The choice between these methods should therefore be guided by the intended application, whether emphasizing intra-cluster uniformity or inter-cluster discrimination. As indicated by figure 4.1, lower MAE indicates tighter cluster cohesion, while higher correlation implies better separation. LSTM demonstrates high compactness, while KMEANS provides better separability.

4.3.2 Summary and Discussion

Table 4.1: Mean Relative Deviation (MRD) values across Residential and Business clusters.

Model	Res (Self)	Bus (Self)	R→B	B→R
KMEANS+DTW	0.290	0.357	**1.351**	**1.322**
LSTM	**0.295**	**0.309**	1.333	1.258
GRU	0.299	0.337	1.106	1.252
TCN	0.294	0.357	1.057	1.397

In table 4.1, **Self**: Intra-cluster compactness (lower is better). **R→B** and **B→R**: Inter-cluster deviation from one cluster to the other (higher is better).

Table 4.2: Model performance (mean ± std) on reduced-hour data.

Model	MAE-SC	Corr-CC	ED-CC
KMEANS+DTW	0.091 ± 0.002	**0.214 ± 0.015**	**0.507 ± 0.021**
LSTM	**0.083 ± 0.001**	0.067 ± 0.008	0.412 ± 0.018
GRU	0.087 ± 0.002	0.109 ± 0.012	0.438 ± 0.020
TCN	0.089 ± 0.003	0.084 ± 0.010	0.453 ± 0.019

In table 4.2, **MAE-SC**: Mean Absolute Error for same-cluster. **Corr-CC** and **ED-CC**: Correlation and Euclidean Distance for cross-cluster. Best values are bold.

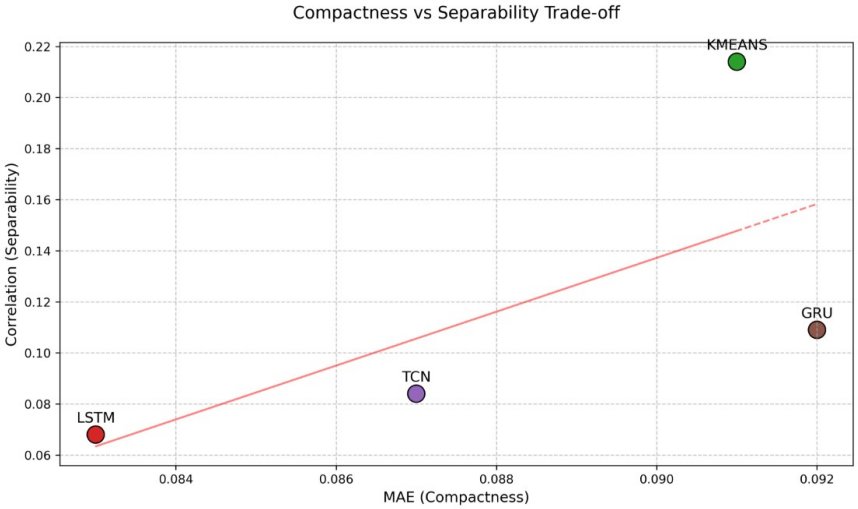

Figure 4.1: Compactness vs Separability trade-off for clustering models.

4.4 Telecom Analysis

The transmit power of the base stations can give insight into the location of the base stations. In turn, classifying base stations helps to improve the service provided, as well as the quality. To understand the temporal usage behavior of base stations, transmit power patterns were analyzed across three clusters identified through K-Means clustering: Residential, Mixed, and Commercial. The plots present the average normalized transmit power across the 24-hour day for each cluster, accompanied by confidence intervals to capture the intra-cluster variability.

In Figure 4.1, which includes the full 24-hour period, a clear distinction emerges between the clusters. The Commercial cluster demonstrates a sharp rise in transmit power during standard working hours, beginning around 8 AM and peaking near midday. This pattern aligns with the expected activity of office spaces, where network usage intensifies during business hours and drops off afterward. In contrast, the Residential cluster shows a dual-peak pattern, with a moderate rise in the early morning and a more prominent peak in the evening, typically between 7 PM and 9 PM. This reflects typical residential usage, where activity increases before and after working hours. The Mixed cluster, as expected, exhibits an intermediate behavior, combining characteristics from both Residential and Commercial zones. During nighttime hours, particularly between 2 AM and 6 AM, all clusters display a marked reduction in transmit power, often approaching zero. This inactive period is accompanied by broader confidence intervals, indicating increased noise and less consistent behavior due to sparse activity.

To better highlight the active usage periods, a second plot, Figure 4.2, which excludes the inactive hours from 2 AM to 7 AM. This reduction enhances the interpretability of the patterns by removing segments of the day where activity is minimal or irregular. In this refined view, the daytime trends become more prominent and the cluster differences more distinct. The Commercial cluster continues to show a strong midday peak, while the Residential cluster's evening peak becomes more pronounced without the visual compression caused by flat nighttime values. Additionally, the confidence intervals become noticeably narrower, particularly during morning and evening periods, suggesting more stable and predictable transmit power behavior when inactive hours are excluded. The Mixed cluster retains its hybrid profile but with reduced variance, further validating the cluster assignments.

Overall, the comparison between the full-day and reduced-hour plots underscores the importance of filtering out inactive periods in telecom analysis. Removing low-activity hours not only improves signal clarity but also reduces the visual and statistical noise that can obscure meaningful patterns. This refinement enables more accurate interpretation of network demand behaviors, which is essential for applications such as energy-efficient network planning, load balancing, and time-aware optimization of base station activity.

In figure 4.3 , the box plot summarises the distribution of the hour of maximum transmit power for every base-station after the K-means partitioning into Residential, Mixed and Commercial groups. Although absolute power levels were first min-max normalised, the timing of each profile's daily maximum remains intact, so differences across clusters can be interpreted as genuine temporal-usage distinctions rather than artefacts of scale.

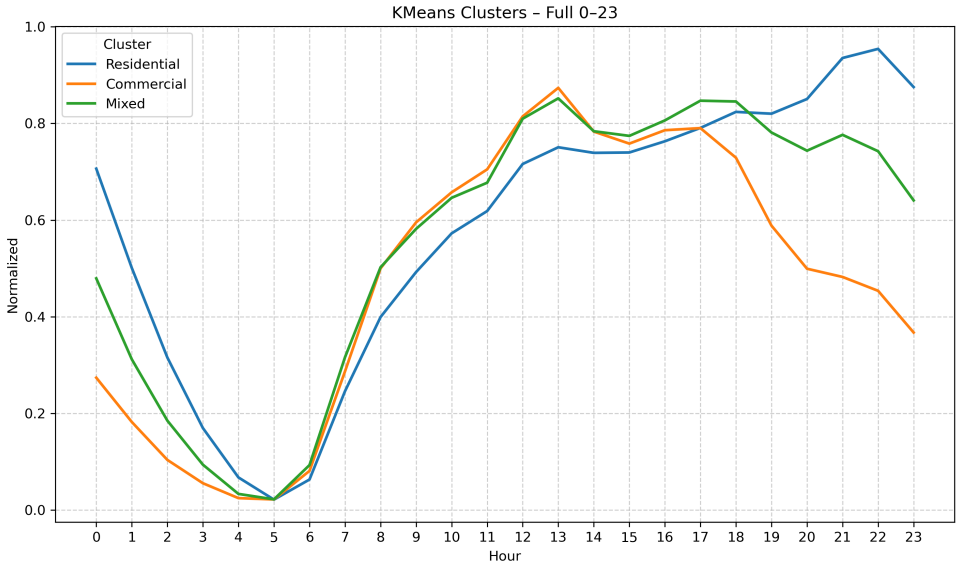

Figure 4.2: Full 24-hour transmit power distribution for base stations.

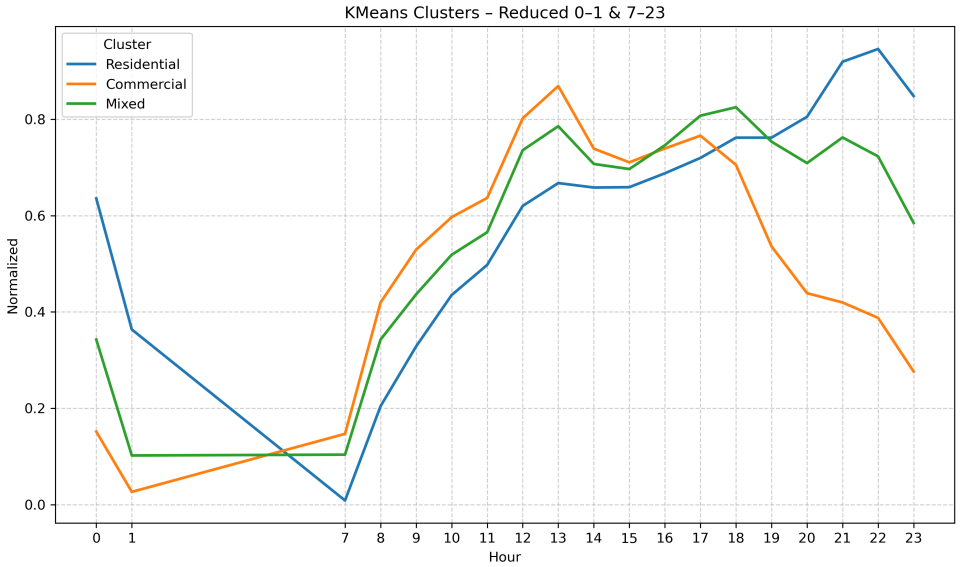

Figure 4.3: Reduced-hour transmit power profile for base stations during peak periods.

Residential sites exhibit their median peak in the early evening, roughly between 18:00 and 20:00. The inter-quartile range spans the late-afternoon to late-evening window, and the whiskers stretch beyond 22:00 with a scattering of after-midnight outliers. This dispersion reflects the heterogeneous lifestyles of household subscribers, while a minority of locations generate overnight spikes, perhaps in student or shift-worker districts. The breadth of the box and length of the upper whisker indicate the largest intra-cluster variability of all three groups.

Commercial stations are markedly different: their median peak occurs just after noon, and the first and third quartiles hug the median tightly, forming the narrowest box in the figure. Short whiskers and the virtual absence of outliers imply that business-district traffic is both highly synchronous and predictable, driven by office hours and retail activity. This cohesiveness validates the cluster label and suggests that

capacity upgrades here can be timed confidently around midday.

The Mixed cluster sits temporally between the other two. Its median peak appears in the late afternoon, around 15:00-16:00, with an inter-quartile range that overlaps both the Residential evening window and the Commercial midday window. The relatively long whiskers on either side, together with several early-morning and late-night outliers, confirm that mixed-use zones experience a prolonged demand rather than a sharp, well-defined surge. Such areas will benefit from more balanced dimensioning across the entire day.

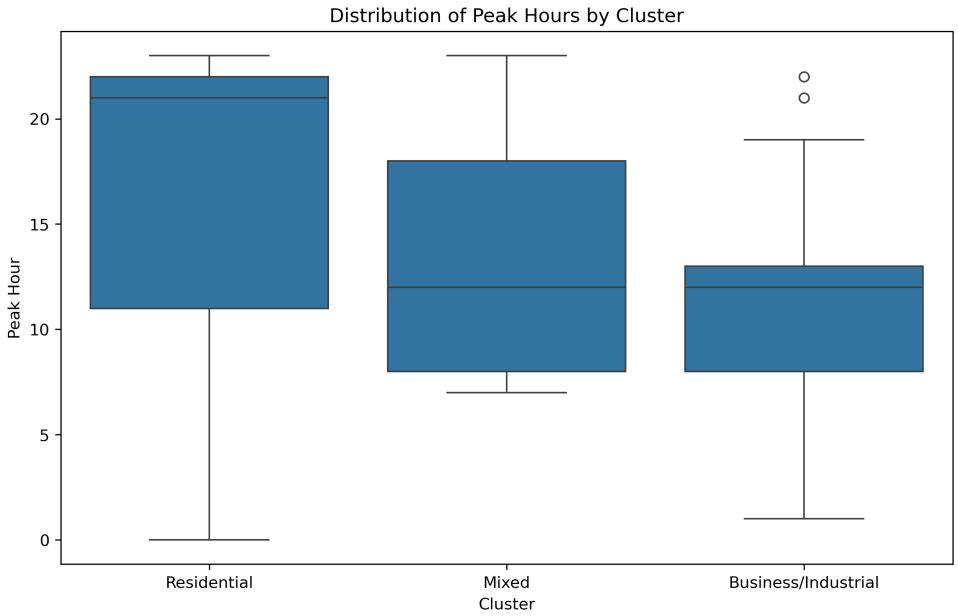

Figure 4.4: Distribution of peak hour values across clusters.

4.4.1 Peak Hour Statistical Analysis

Analysis of the standard deviation and confidence intervals of peak hours provides valuable insights for network resource planning. A cluster exhibiting a low standard deviation and a narrow confidence interval in peak hour distribution enables precise and targeted resource allocation, whereas a cluster with a high standard deviation and a wide confidence interval may require more dynamic or distributed resource management strategies. The results, as illustrated in Table 4.3, reveal that the clusters possess distinct peak hour characteristics. Residential sites display a substantially later mean peak hour (approximately 5:00 PM) compared to Business/Industrial (approximately 11:30 AM) and Mixed clusters (approximately 12:40 PM), indicating that residential areas experience their highest demand in the late afternoon and evening, while business and mixed-use areas peak around midday. The Business/Industrial cluster demonstrates the greatest temporal consistency in peak hour (standard deviation 3.5 hours), whereas the Residential cluster exhibits the highest variability (standard deviation 6.3 hours), suggesting a broader distribution of peak usage times among residential sites.

Furthermore, the confidence intervals for each group show minimal overlap, particularly between the Residential cluster and the other groups, indicating that the observed differences in mean peak hour are

statistically significant. These findings suggest that, for business clusters, resource prioritization should focus on the late morning to noon period; for mixed clusters, optimal resource allocation is achieved in the early afternoon; and for residential clusters, the greater variability in peak times necessitates dynamic resource management and flexible scheduling to accommodate the wider spread of peak demand.

Cluster type	Number of sites	Mean	Std. Deviation	Lower Bound	Upper Bound
Residential	398	17.31	5.98	16.72	17.89
Mixed	225	11.46	4.98	12.07	13.38
Business	120	12.72	3.44	10.84	12.08

Table 4.3: Statistical analysis of peak hours for k-means clusters over a 24-hour period

4.4.2 Statistical Significance

In this analysis, the ANOVA test applied to the three clusters yielded an F-statistic of **7.70** with a p-value of **0.0023**. Since the p-value is below the conventional significance level of 0.05, the null hypothesis is rejected. This result indicates that at least one cluster has a statistically distinct peak hour mean.

To further investigate which clusters differ, both two-sample ANOVA tests (equivalent to independent-sample F-tests) and Welch's t-tests, which do not assume equal variances were conducted between the groups.

The pairwise tests yielded highly significant results for all cluster combinations. When comparing the Residential and Mixed clusters, the ANOVA yielded an F-statistic of 77.46 with a p-value of 1.34×10^{-17}, while the corresponding t-test showed a t-statistic of 9.39 and a p-value of 1.47×10^{-19}. These results indicate a strong and statistically significant difference in the distribution of peak hour usage between these two groups. A similar outcome was observed between the Residential and Business/Industrial clusters, with an ANOVA F-statistic of 85.81 and a p-value of 5.34×10^{-19}, and a t-test t-statistic of 12.49 with an even smaller p-value of 4.57×10^{-30}. This further supports the assertion that residential base stations follow a distinctly different temporal usage pattern compared to business-oriented clusters.

The difference between the Mixed and Business/Industrial clusters, while statistically significant, is less pronounced. The ANOVA returned an F-statistic of 6.23 with a p-value of 0.0131, and the Welch's t-test resulted in a t-statistic of 2.78 with a p-value of 0.0058. Although these values confirm that the distributions are not identical, the lower magnitude of both the test statistics and the p-values relative to the other pairings suggests a more subtle distinction. This may reflect the transitional nature of mixed-use areas, where the traffic pattern partially overlaps with both residential and business profiles.

Overall, these pairwise comparisons validate the clustering approach by confirming that the derived groups differ meaningfully in terms of their peak activity periods. This differentiation is crucial for further network optimization and targeted energy-saving strategies, as it allows operators to model and manage each cluster according to its typical daily load profile.

Chapter 5

Conclusions

The primary objective of this study was to develop a robust methodology for profiling cellular base stations based on their transmit power patterns, with the goal of categorizing them into residential, commercial, or mixed-use environments. By leveraging machine learning techniques, this research aimed to provide a data-driven approach to network planning and resource allocation, addressing the limitations of traditional manual classification methods. The study focused on analyzing 24-hour power emission profiles, with particular attention to the discriminative power of active hours while excluding low-activity nighttime periods (2:00-6:00) to enhance model performance.

To achieve this objective, several machine learning models were implemented and evaluated, including Time Series K-Means with Dynamic Time Warping (DTW), LSTM-based autoencoders, GRU-based autoencoders, and TCN-based autoencoders. These models were selected for their ability to capture temporal dependencies and patterns in the transmit power data. A comprehensive comparison of these models was conducted using clustering evaluation metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), Pearson Correlation, and Cosine Similarity. Autoencoders such as LSTM achieved superior same-cluster compactness (MAE = 0.083), benefiting from their latent representation learning, but they lacked strong separability, particularly in distinguishing mixed-use profiles. KMEANS+DTW, while less compact, provided robust inter-cluster separation (cross-cluster correlation = 0.214), especially for reduced-hour datasets. The results suggest that when interpretability and cluster distinction are critical; for instance, in operational planning traditional methods like KMEANS+DTW may be preferred. However, for profile compression or personalized modeling (for example, for load forecasting within a known tower type), autoencoders remain a strong choice. Future work could explore hybrid models to balance both aspects.

Following the model selection, a detailed telecom analysis was performed using the classified clusters, with peak hour as the primary metric due to its direct correlation with network usage and demand. The analysis revealed distinct temporal patterns for each cluster: commercial base stations exhibited peak usage during midday, residential base stations showed higher activity in the evening, and mixed-use base stations displayed intermediate behavior. Statistical significance tests, including one-way ANOVA and Welch's t-test, confirmed that the differences in peak hour usage between clusters

were statistically significant, further validating the classification framework.

The findings of this study hold significant implications for telecommunications network optimization. By accurately profiling base stations, network operators can implement targeted resource allocation strategies, such as dynamic power adjustment during peak hours, load balancing, and energy-saving measures tailored to the specific usage patterns of each cluster. For instance, commercial base stations may benefit from increased capacity during business hours, while residential base stations could be optimized for evening usage. This granular approach not only enhances service quality but also improves energy efficiency, reducing operational costs and environmental impact.

Despite these contributions, the study has certain limitations. The dataset was geographically limited to base stations in Portugal, which may not fully capture the diversity of usage patterns in other regions. Additionally, the classification framework relies solely on transmit power data, and incorporating additional features such as user density or mobility patterns could further refine the results. Future work could explore the integration of real-time data streams to enable adaptive classification, as well as the application of more advanced deep learning architectures to handle larger and more heterogeneous datasets.

In conclusion, this research provides a scalable and interpretable framework for base station profiling, offering telecommunications operators actionable insights for network planning and optimization. As cellular networks continue to evolve, such data-driven approaches will play an increasingly critical role in ensuring efficient and reliable service delivery.

Appendix A: Detailed Cluster Analysis Report (Seeded Runs)

This appendix presents the full and reduced-hour clustering metric matrices (with C0 = Residential, C1 = Mixed, and C2 = Business/Industrial) for each model across all six metrics (MAE, RMSE, MSE, Correlation, Cosine Similarity, and Euclidean Distance), based on 15 seeded runs.

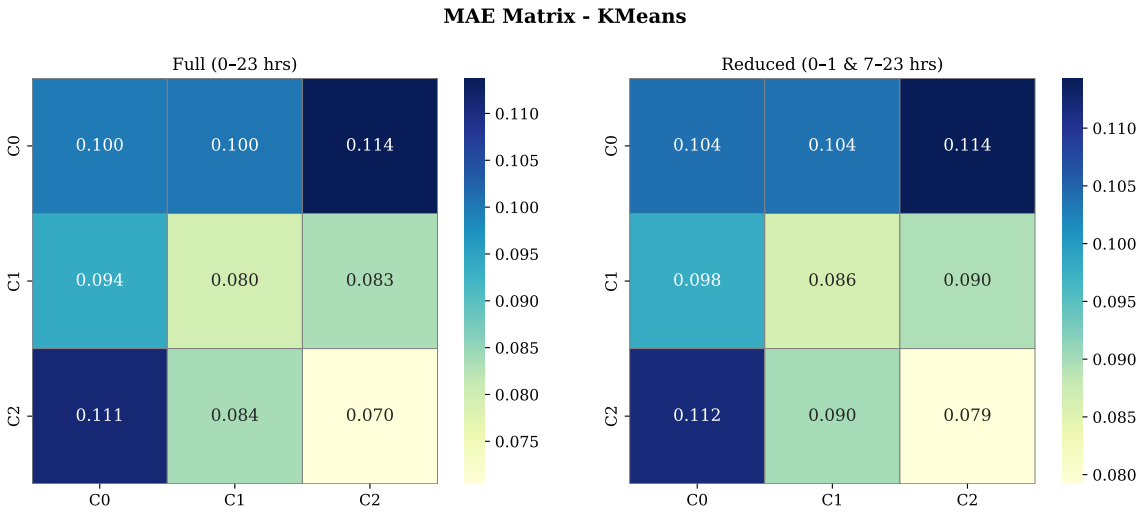

Figure A.1: KMEANS - MAE Matrix (Full vs Reduced Hours)

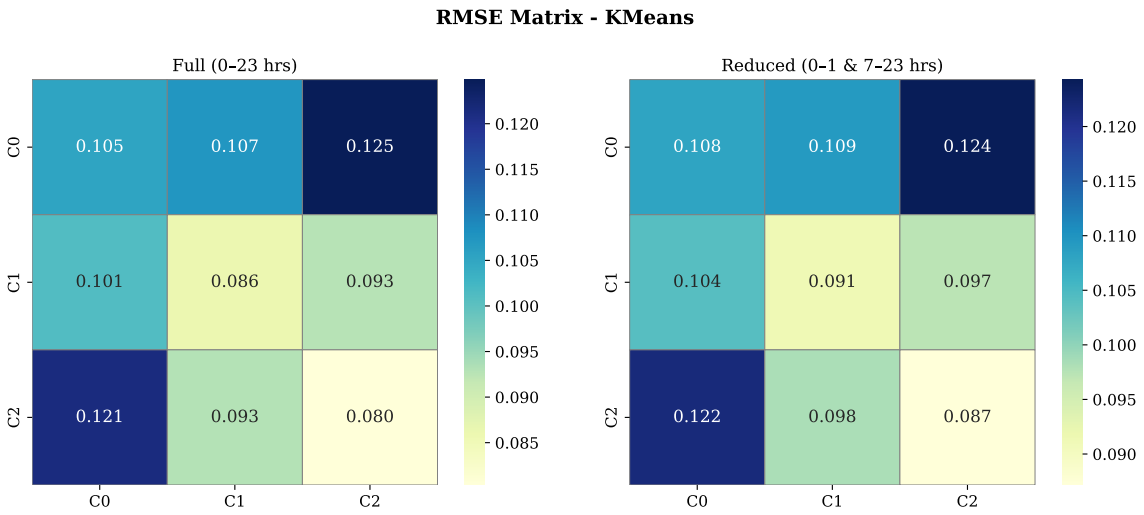

Figure A.2: KMEANS - RMSE Matrix (Full vs Reduced Hours)

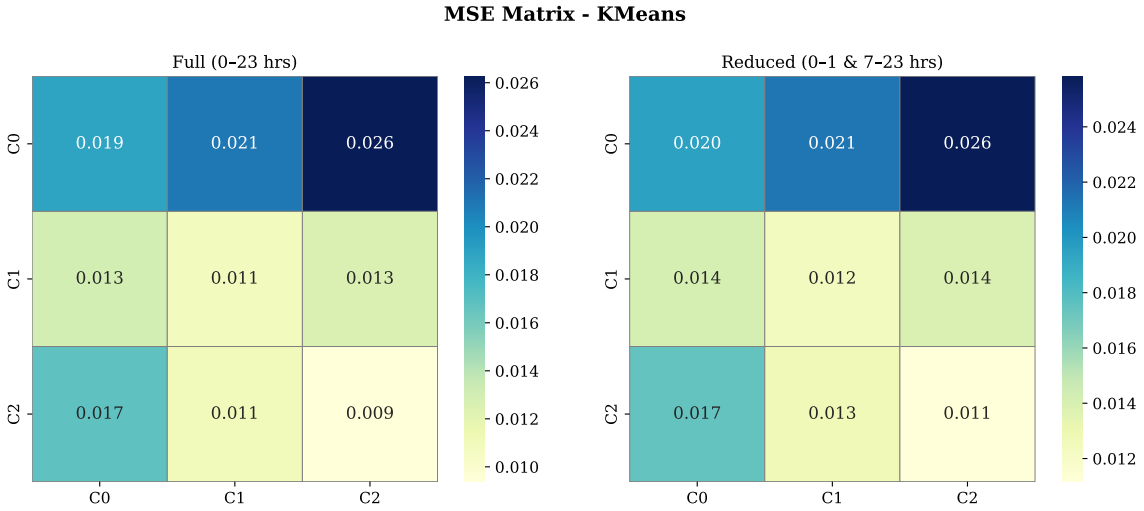

Figure A.3: KMEANS - MSE Matrix (Full vs Reduced Hours)

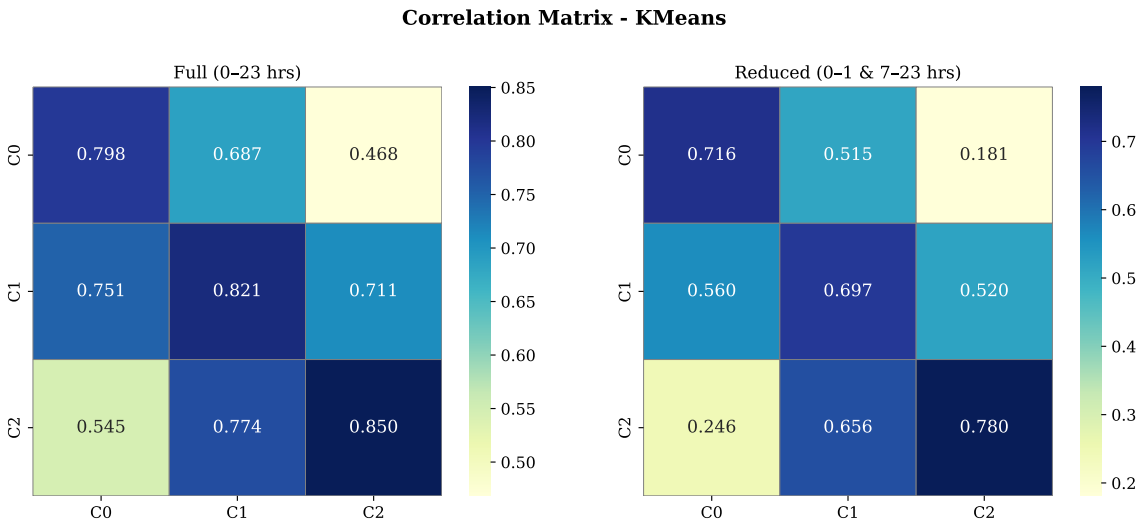

Figure A.4: KMEANS - Correlation Matrix (Full vs Reduced Hours)

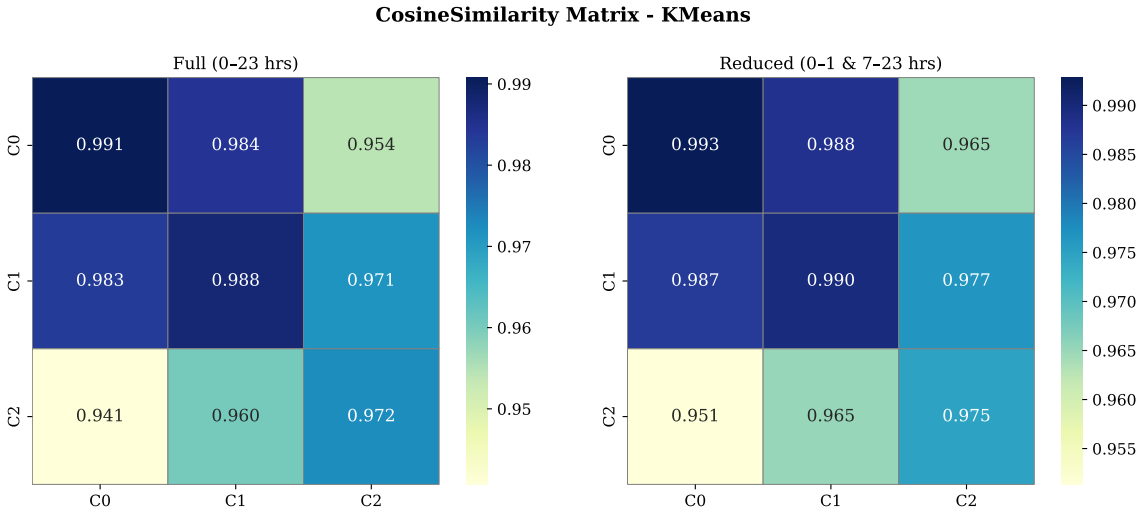

Figure A.5: KMEANS - Cosine Similarity Matrix (Full vs Reduced Hours)

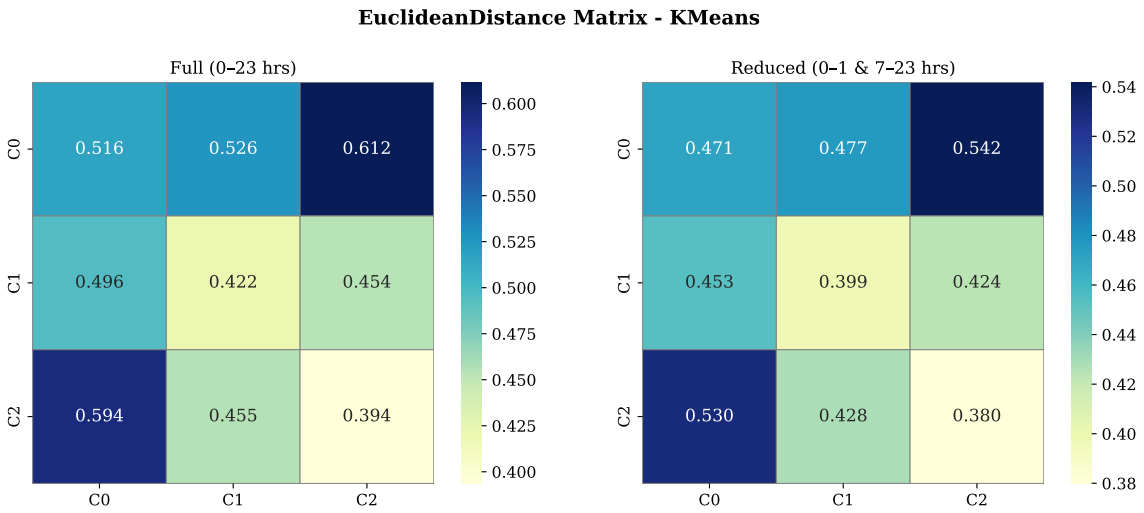

Figure A.6: KMEANS - Euclidean Distance Matrix (Full vs Reduced Hours)

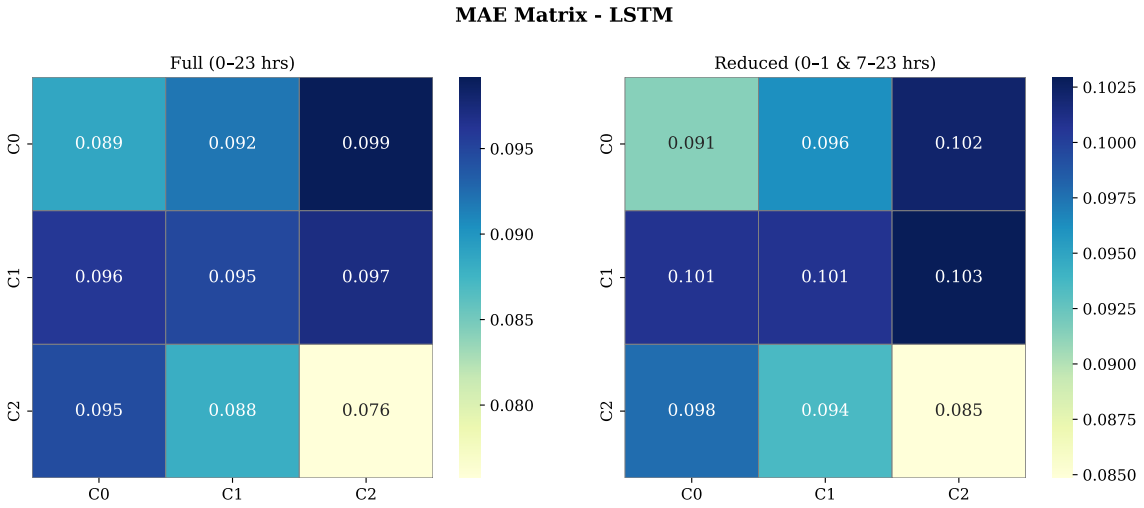

Figure A.7: LSTM - MAE Matrix (Full vs Reduced Hours)

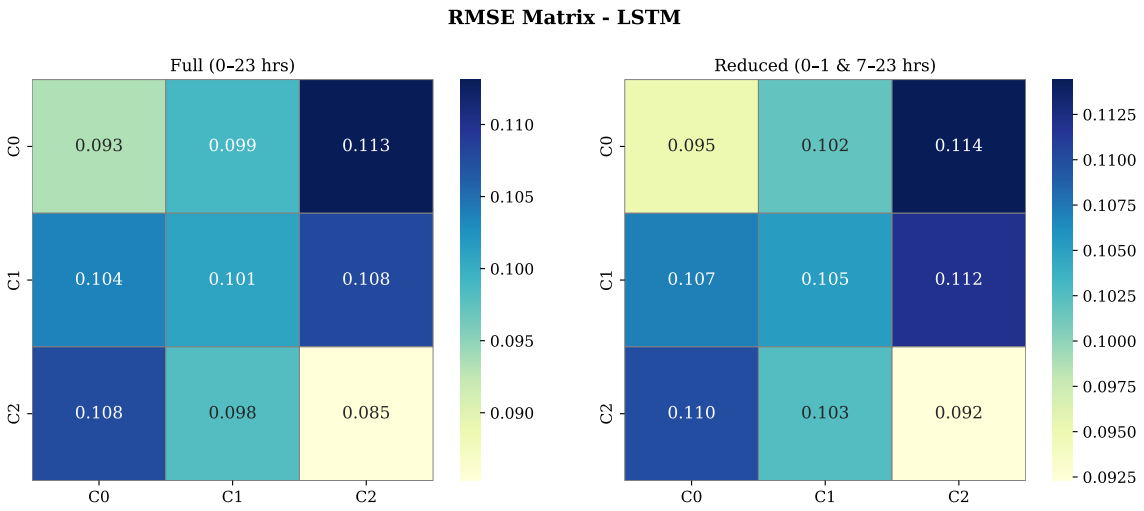

Figure A.8: LSTM - RMSE Matrix (Full vs Reduced Hours)

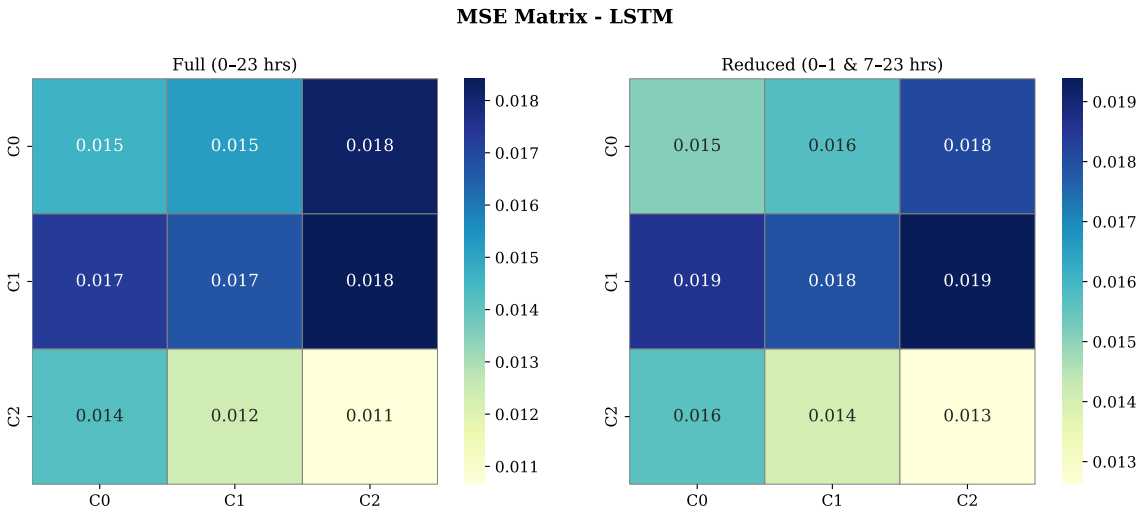

Figure A.9: LSTM - MSE Matrix (Full vs Reduced Hours)

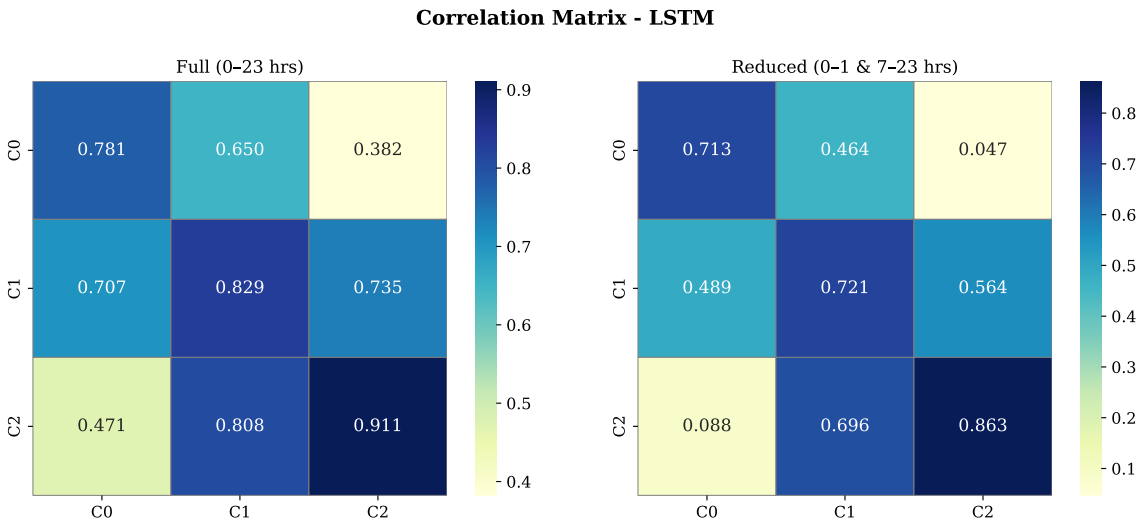

Figure A.10: LSTM - Correlation Matrix (Full vs Reduced Hours)

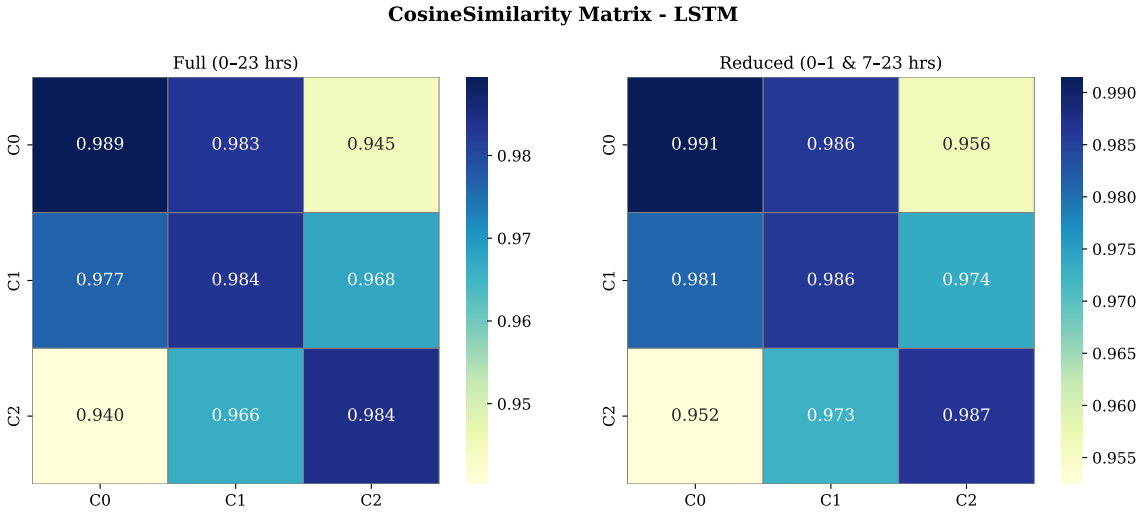

Figure A.11: LSTM - Cosine Similarity Matrix (Full vs Reduced Hours)

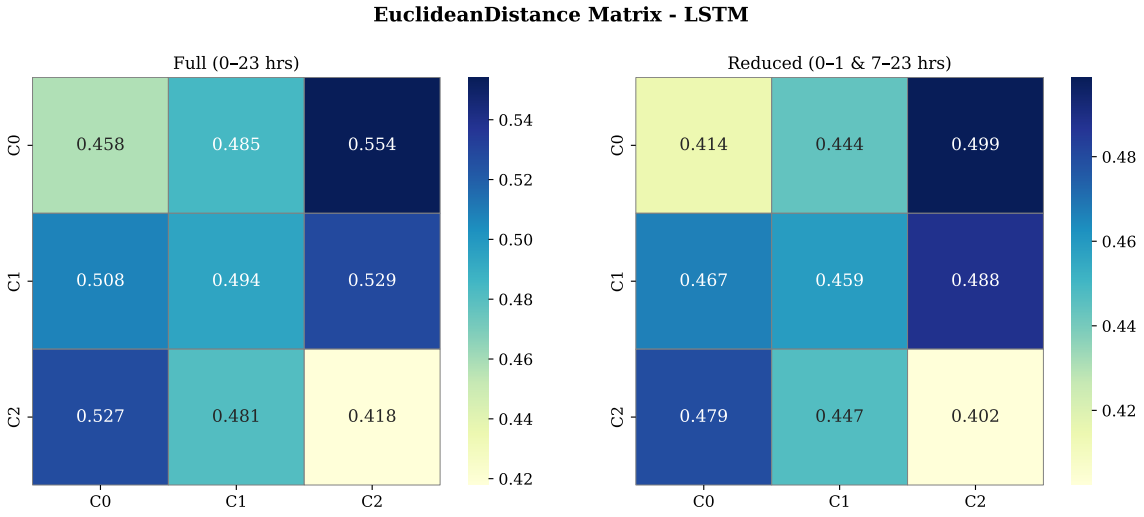

Figure A.12: LSTM - Euclidean Distance Matrix (Full vs Reduced Hours)

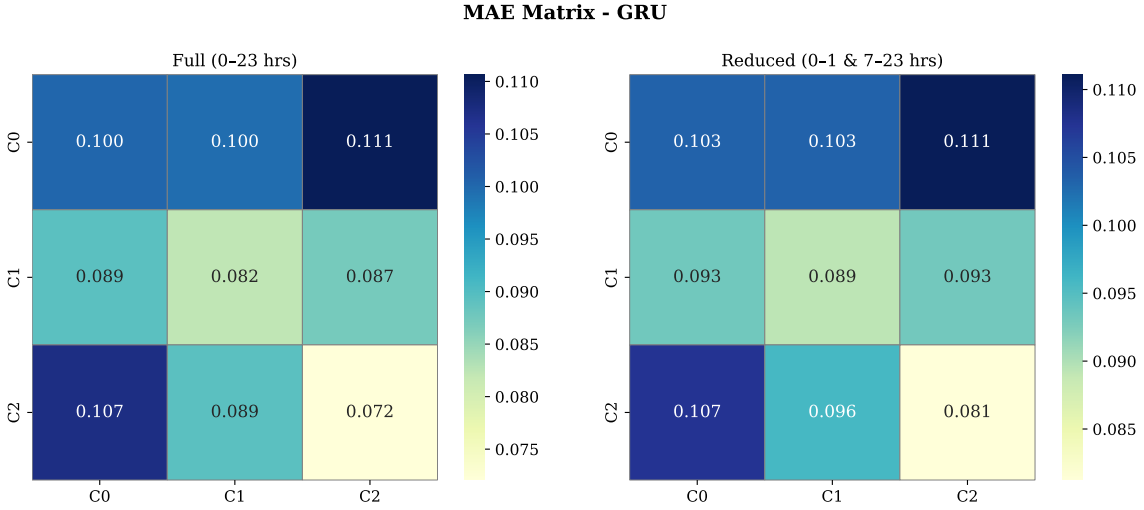

Figure A.13: GRU - MAE Matrix (Full vs Reduced Hours)

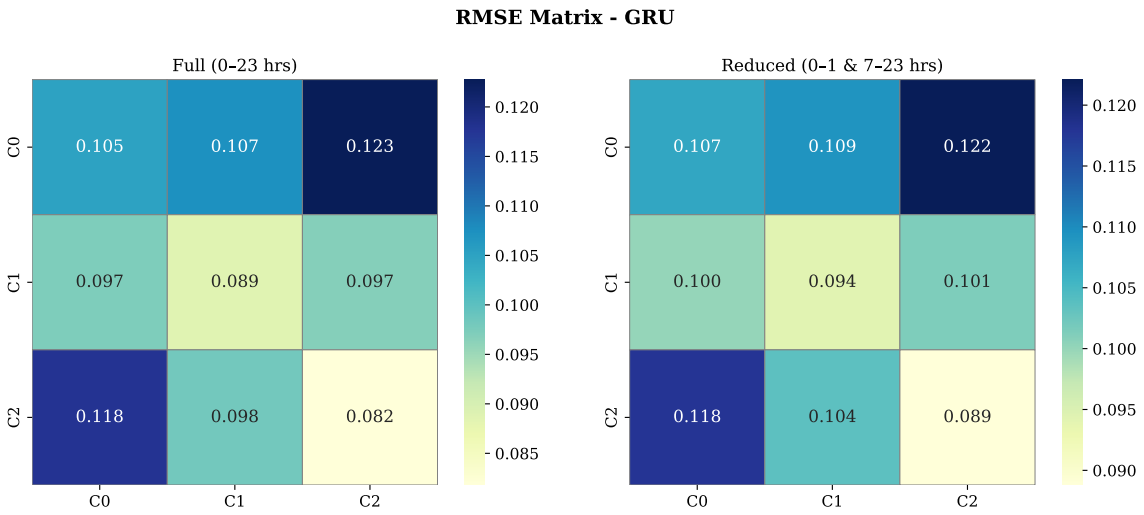

Figure A.14: GRU - RMSE Matrix (Full vs Reduced Hours)

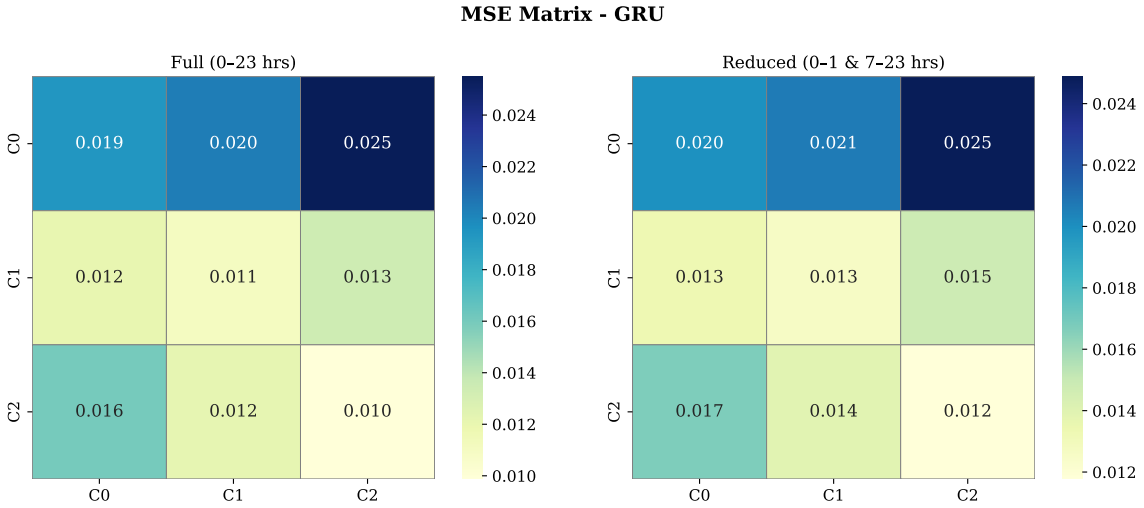

Figure A.15: GRU - MSE Matrix (Full vs Reduced Hours)

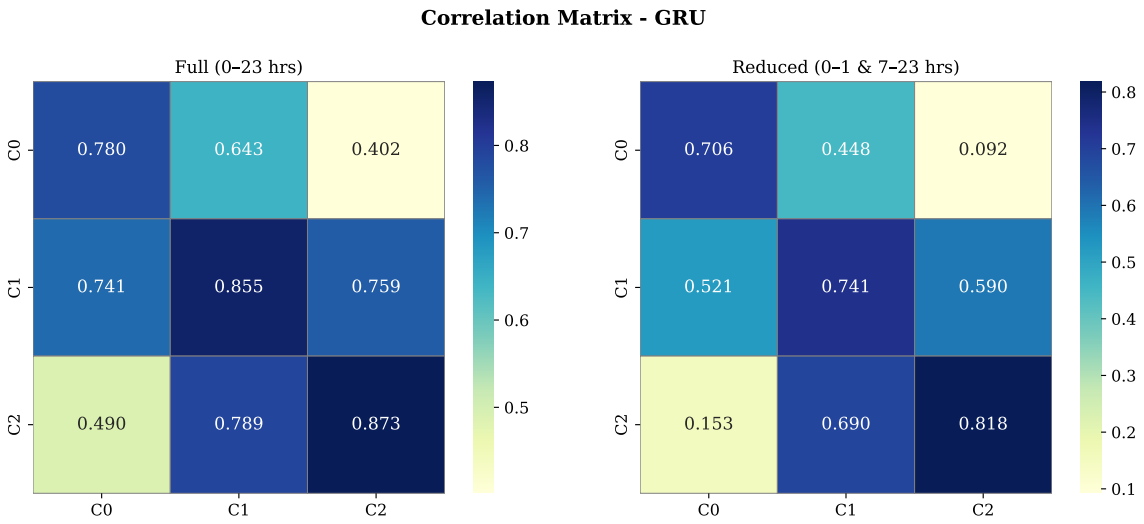

Figure A.16: GRU - Correlation Matrix (Full vs Reduced Hours)

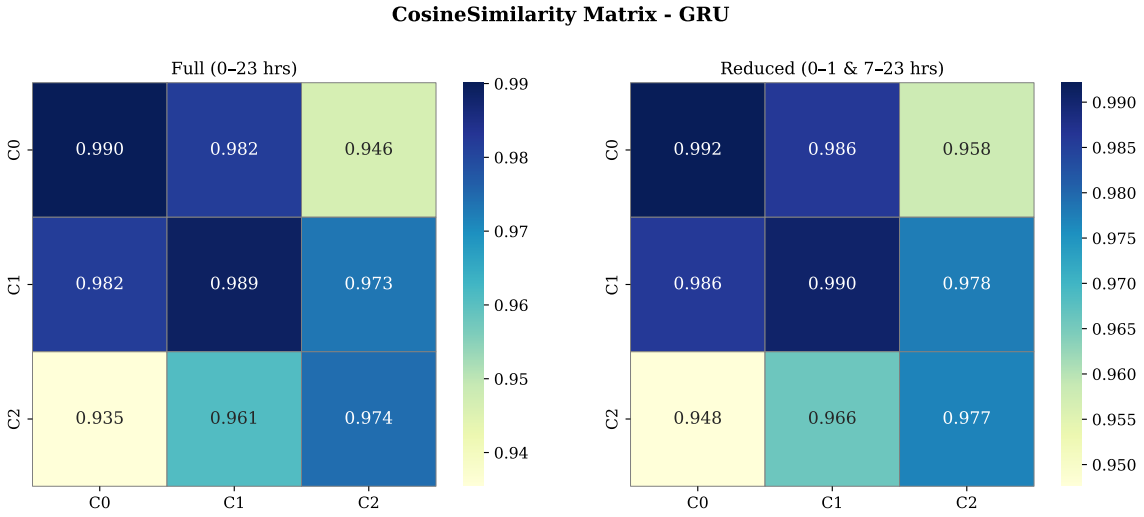

Figure A.17: GRU - Cosine Similarity Matrix (Full vs Reduced Hours)

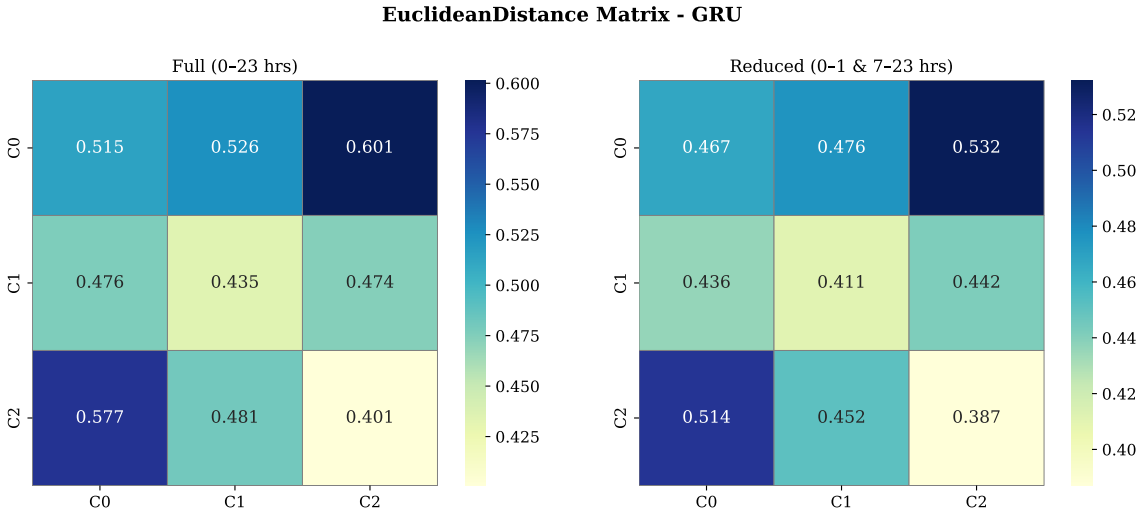

Figure A.18: GRU - Euclidean Distance Matrix (Full vs Reduced Hours)

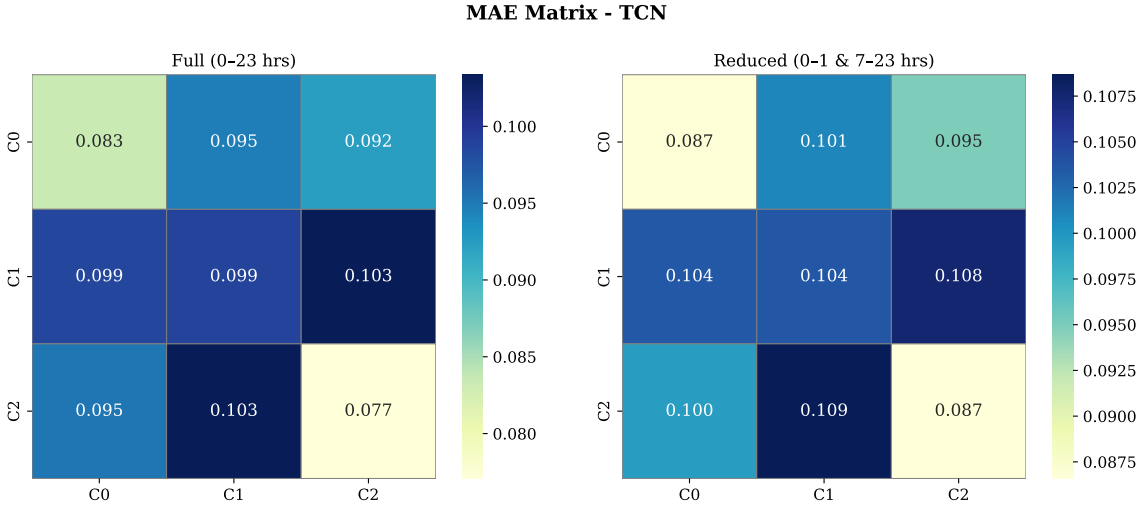

Figure A.19: TCN - MAE Matrix (Full vs Reduced Hours)

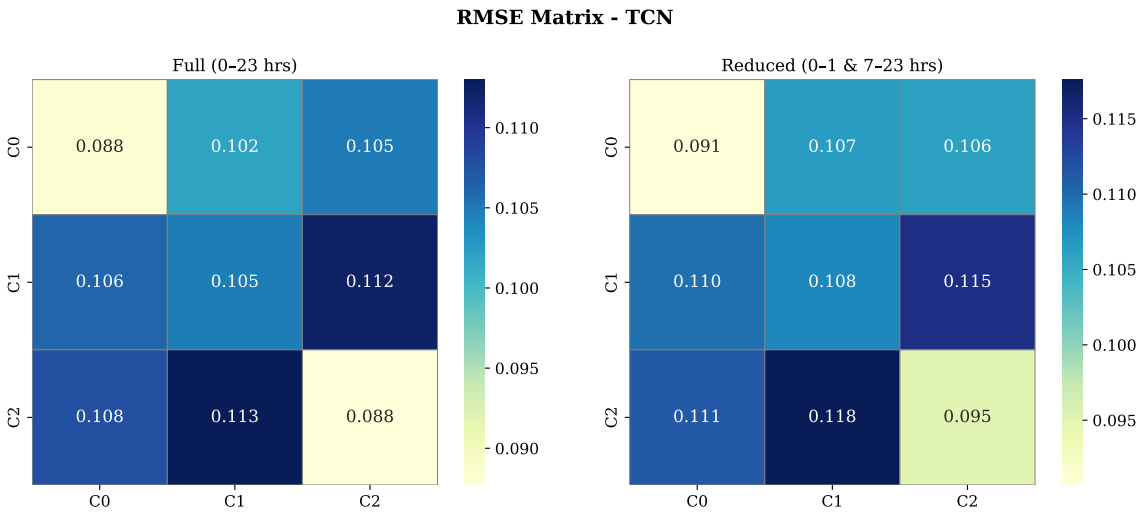

Figure A.20: TCN - RMSE Matrix (Full vs Reduced Hours)

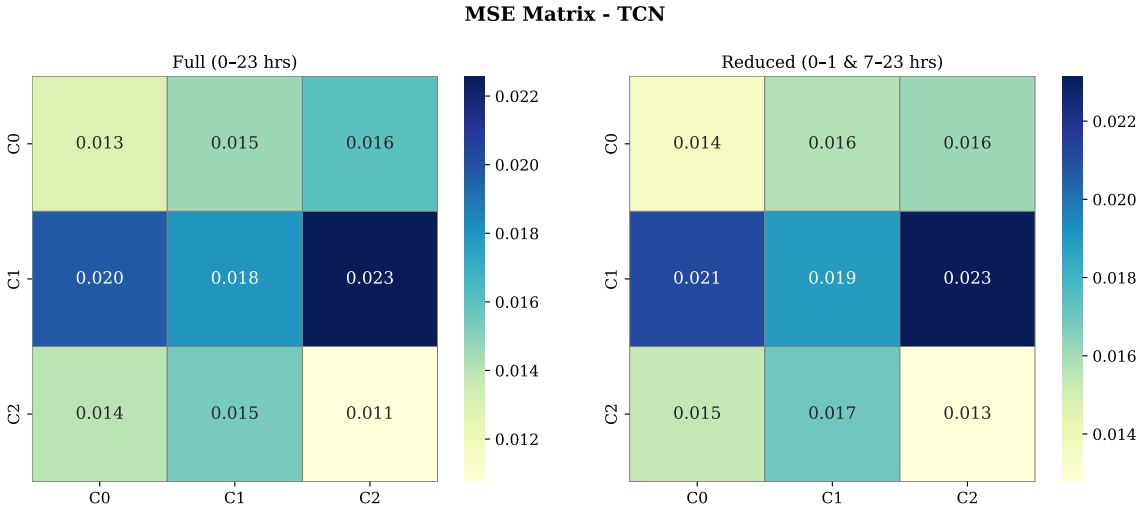

Figure A.21: TCN - MSE Matrix (Full vs Reduced Hours)

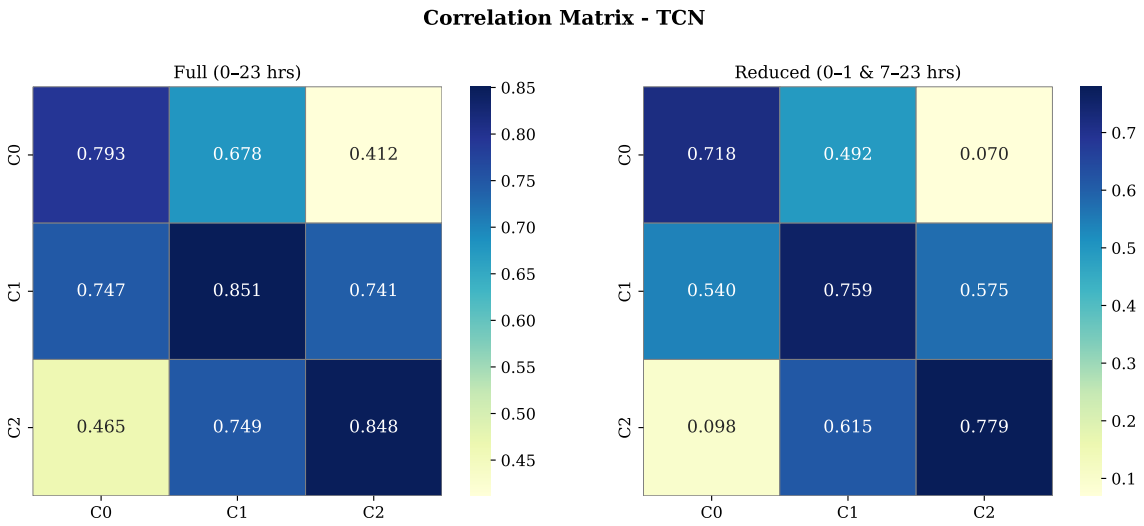

Figure A.22: TCN - Correlation Matrix (Full vs Reduced Hours)

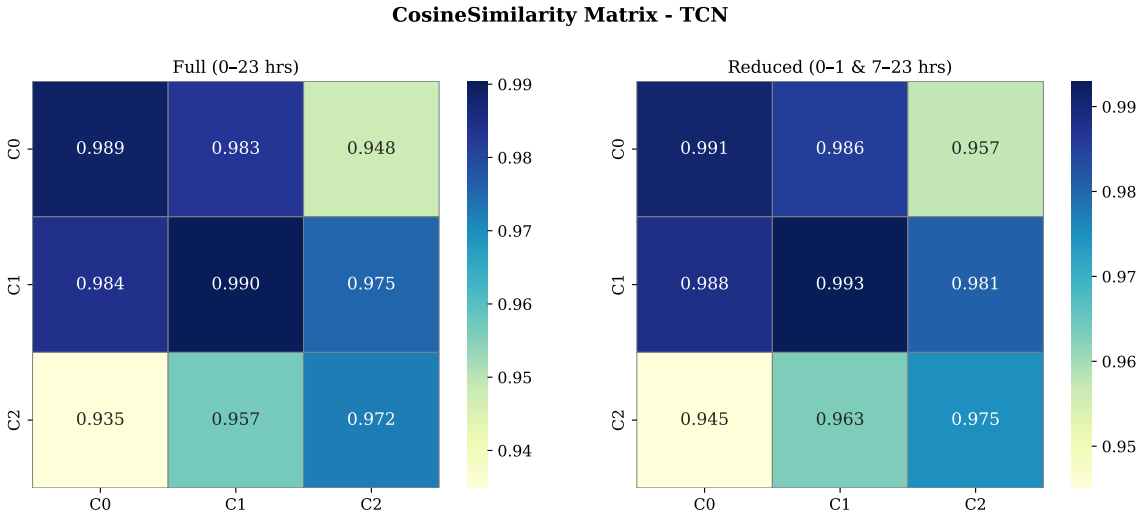

Figure A.23: TCN - Cosine Similarity Matrix (Full vs Reduced Hours)

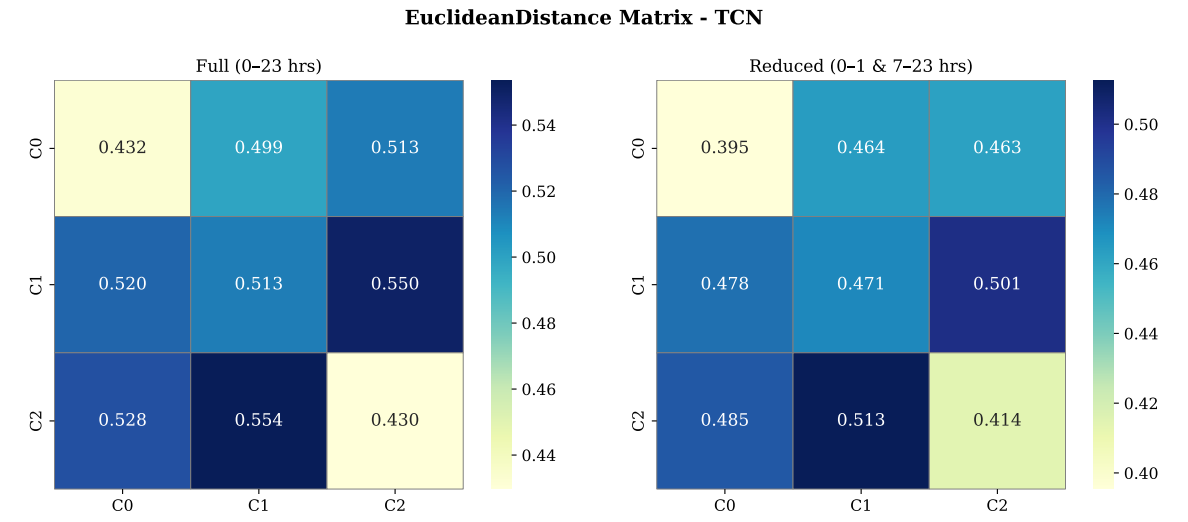

Figure A.24: TCN - Euclidean Distance Matrix (Full vs Reduced Hours)

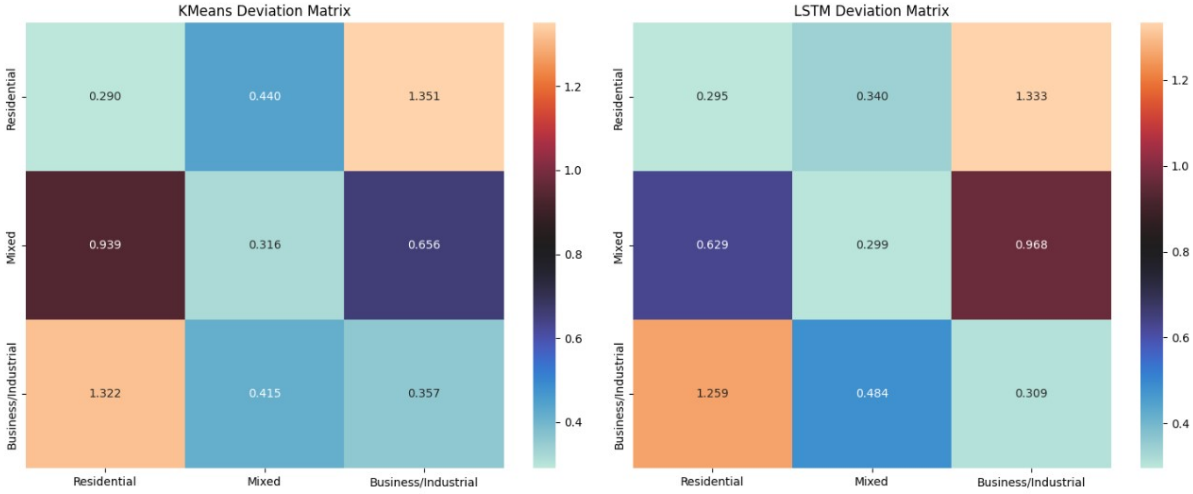

Figure A.25: KMEANS & LSTM - Mean Relative Deviation

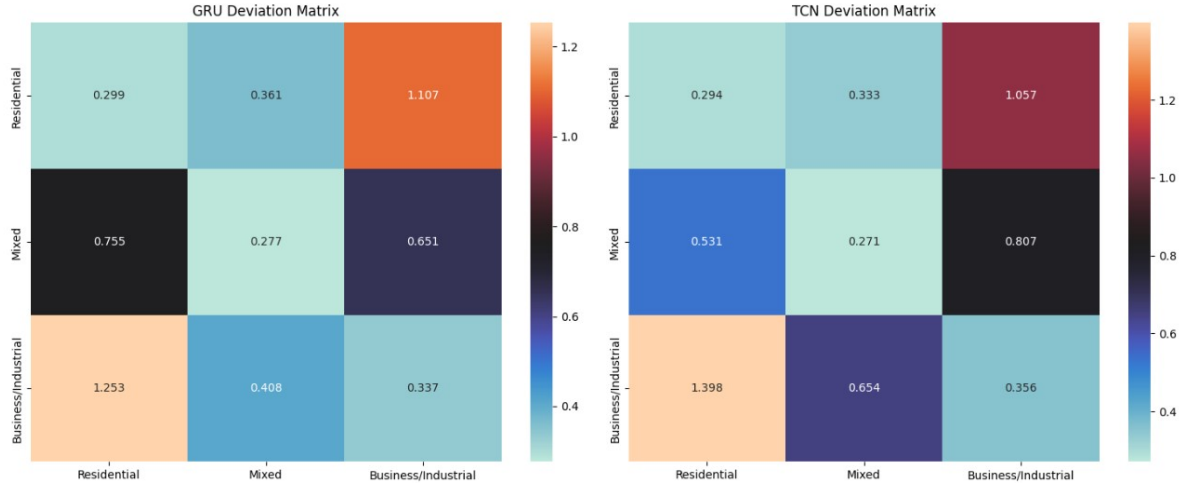

Figure A.26: GRU & TCN - Mean Relative Deviation

Bibliography

[1] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *Journal of Internet Services and Applications*, vol. 9, no. 1, p. 16, 2018.

[2] M. Polese, E. Pateromichelakis, T. Guan, G. Imbarlina, K. Ramantas, M. G. Sarret, C. Deng, I. Karls, K. Abdoli, S. K. Sagari, S. Saur, P. Kulkarni, and A. Dutta, "Machine learning at the edge: A data-driven architecture for next-generation communications systems," *IEEE Transactions on Mobile Computing*, vol. 20, no. 10, pp. 2991–3004, 2021.

[3] Z. Zhou, L. Mao, T. Wang, and W. Ma, "Spatio-temporal analysis and prediction of cellular traffic using machine learning," *IEEE Access*, vol. 8, pp. 72904–72917, 2020.

[4] A. Azari, P. Papapetrou, S. Denic, and G. Peters, "Cellular traffic prediction and classification: a comparative evaluation of lstm and arima," *arXiv preprint arXiv:1906.00939*, 2019.

[5] W. Jiang, H. Luo, H. Xu, W. Song, and X. Xu, "Cellular traffic prediction with machine learning: A survey," *Expert Systems with Applications*, vol. 210, p. 117163, 2022.

[6] H. Yuliana, A. S. Hidayat, A. Suhartomo, and N. Mufti, "Estimating base station traffic and throughput using machine learning techniques," *IEEE Access*, vol. 12, pp. 57089–57102, 2024.

[7] D. G. S. Pivoto, F. A. P. de Figueiredo, C. Cavdar, G. R. de Lima Tejerina, and L. L. Mendes, "A comprehensive survey of machine learning applied to resource allocation in wireless communications," *IEEE Communications Surveys Tutorials*, 2024. Early Access.

[8] Z. Liu, X. Yin, Z. Xie, Z. Zhao, Y. Jiang, Y. Zhao, V. Aggarwal, and D. Vasisht, "Exploring practical vulnerabilities of machine learning-based wireless systems," *Proceedings of the 20th USENIX Symposium on Networked Systems Design and Implementation*, pp. 1–17, 2023.

[9] NiCE, "Peak hour traffic (pht)," *NiCE Glossary*, 2025. Accessed: 2025-06-29.

[10] UrbanLogiq, "Peak hour measurements: How accurate do data sources need to be?," *UrbanLogiq Resource Library*, 2024. Accessed: 2025-06-29.

[11] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, pp. 129–137, Mar. 1982.

[12] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, “Time-series clustering – a decade review,” *Information Systems*, vol. 53, pp. 16–38, 2015.

[13] D. J. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series,” pp. 359–370, 1994.

[14] F. Petitjean, A. Ketterlin, and P. Gançarski, “A global averaging method for dynamic time warping, with applications to clustering,” *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

[15] G. Zhu, S. Wang, L. Sun, Y. Ge, and Z. Zhang, “A clustering-based spatial-temporal propagation model for metro passenger flow prediction,” *IEEE Access*, vol. 6, pp. 43222–43232, 2018.

[16] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” pp. 226–231, 1996.

[17] M. Roughan, S. Sen, O. Spatscheck, and N. Duffield, “Class-of-service mapping for qos: a statistical signature-based approach to ip traffic classification,” pp. 135–148, 2004.

[18] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, J. Song, and Y. Li, “Big data driven mobile traffic understanding and forecasting: A time series approach,” *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 796–805, 2017.

[19] X. Wang, J. Guo, S. Chen, H. Feng, and H. Chen, “A survey on deep learning for cellular traffic prediction,” *Intelligence Computing*, vol. 3, pp. 1–25, January 2024.

[20] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[21] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, “Network traffic classifier with convolutional and recurrent neural networks for internet of things,” *IEEE Access*, vol. 5, pp. 18042–18050, 2017.

[22] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (A. Moschitti, B. Pang, and W. Daelemans, eds.), (Doha, Qatar), pp. 1724–1734, Association for Computational Linguistics, Oct. 2014.

[23] H. Zhao, J. Liu, X. Dong, X. He, and J. Sun, “A traffic prediction model for self-adapting routing overlay network in publish/subscribe system,” *IEEE Systems Journal*, vol. 13, no. 2, pp. 1726–1737, 2019.

[24] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, “Temporal convolutional networks for action segmentation and detection,” in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 156–165, 2017.

[25] O. Aoued, V. A. Le, K. Piamrat, and Y. Ji, “Deep learning on network traffic prediction: Recent advances, analysis, and future directions,” *ACM Computing Surveys*, vol. 57, pp. 1–37, February 2025.

[26] J. Yang, Y. Han, Y. Wang, B. Jiang, Z. Lv, and H. Song, “Optimization of real-time traffic network assignment based on iot data using dbn and clustering model in smart city,” *Future Generation Computer Systems*, vol. 108, pp. 976–986, 2020.

[27] H. Dong, M. Wu, X. Ding, L. Chu, L. Jia, Y. Qin, and X. Zhou, “Traffic zone division based on big data from mobile phone base stations,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 278–291, 2015. Big Data in Transportation and Traffic Engineering.

[28] S. Gao, J. Chen, L. Ma, Y. Zhang, and X. Zhang, “Spatio-temporal graph neural network for mobile traffic forecasting,” *IEEE Transactions on Mobile Computing*, vol. 20, no. 10, pp. 2874–2886, 2021.

[29] L. Xu, S. Yu, and H. Zheng, “Hierarchical time-series clustering approach to mobile traffic data for base station profiling,” *Expert Systems with Applications*, vol. 160, p. 113709, 2020.

[30] X. Zhang, Y. Gao, W. Ding, Z. Li, and S. Xu, “Cell boundary prediction and base station location verification based on machine learning,” *IEEE Transactions on Mobile Computing*, vol. XX, no. X, pp. 1–13, 2023. Early Access.

[31] R. Das, A. Ray, D. Datta, B. Saha, and P. P. Dutta, “Clustering driven approach to predict traffic load of base stations using real network data,” *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4450–4462, 2022.

[32] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[33] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *CoRR*, vol. abs/1803.01271, 2018.

[34] J. H. W. J. and, “Hierarchical grouping to optimize an objective function,” *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[35] E. B. Wilson, “Probable inference, the law of succession, and statistical inference,” *Journal of the American Statistical Association*, vol. 22, no. 158, pp. 209–212, 1927.

[36] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.